JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

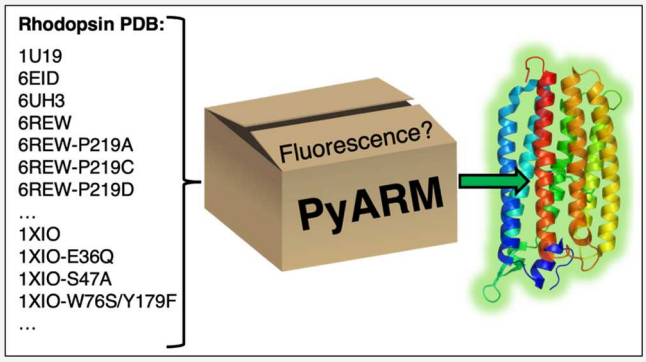
Automated QM/MM Screening of Rhodopsin Variants with Enhanced Fluorescence

Laura Pedraza-González, Leonardo Barneschi, Michał Marszałek, Daniele Padula, Luca De Vico, and Massimo Olivucci*

Cite This: https://doi.org/10.1021/acs.jctc.2c00928 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We present a computational protocol for the fast and automated screening of excited-state hybrid quantum mechanics/molecular mechanics (QM/MM) models of rhodopsins to be used as fluorescent probes based on the automatic rhodopsin modeling protocol (a-ARM). Such "a-ARM fluorescence screening protocol" is implemented through a general Python-based driver, PyARM, that is also proposed here. The implementation and performance of the protocol are benchmarked using different sets of rhodopsin variants whose absorption and, more relevantly, emission spectra have been experimentally assessed. We show that, despite important limitations that make unsafe to use it as a black-box tool, the protocol reproduces the observed trends in fluorescence and it is capable of selecting novel potentially fluorescent rhodopsins. We also show that the protocol can be used in mechanistic investigations to discern fluorescence enhancement effects associated with a near degeneracy of the S_1/S_2 states or, alternatively, with a barrier generated via coupling of the S_0/S_1 wave functions.



1. INTRODUCTION

Microbial rhodopsins are trans-membrane proteins formed by an opsin apoprotein hosting an all-*trans* protonated retinal Schiff base (rPSB) chromophore (see Figure 1A). They are found in extremely diverse microorganisms (*e.g.*, archaea, bacteria, algae, and fungi) and even in giant viruses[1−3] and carry out different light-dependent biological functions initiated by the photoisomerization of rPSB from the all-*trans* to the 13-*cis* configuration.[3−15]

Specific microbial rhodopsins, exhibiting ion-transporting functions (*e.g.*, light-gated ion channels or light-driven ion pumps), have been found to be instrumental to the development of optogenetics tools.[3,10,12,15−21] For this reason, they have been engineered to function as light-driven actuators, silencers, or fluorescent reporters of neuronal action potentials.[10] Regarding the latter, specific microbial rhodopsins have been used to construct genetically encodable voltage indicators (GEVIs).[22−26] In such applications, a change in membrane voltage induces a variation in rhodopsin fluorescence intensity, which is directly used as a voltage indicator.[12,25,27]

It is evident that rhodopsins used as GEVIs must be fluorescent. A prototype fluorescent reporter is Archaerhodopsin-3 (Arch3), an archaeal rhodopsin from *Halorubrum sodomense*, with light-driven outward proton pumping activity.[26−29] However, since the fluorescence of Arch3 is extremely dim with a fluorescence quantum yield (ϕ^f) of ca. 1.1×10^{-4}, extensive efforts have been made to search variants with enhanced fluorescence (*i.e.*, via directed evolutionary approaches and random mutagenesis).[22−24,26,27] Among others, variants such as QuasArs,[30] Archons,[31,32] Archers,[33,34] Arch5, and Arch7[34] (see Table S1 in the Supporting Information) have been reported to feature a brighter fluorescence, enabling applications not only in the imaging of thin brain slices but also in living mammals and invertebrates.[35−37] Nevertheless, as shown in Table S2 in the Supporting Information, their ϕ^f value remains in the range of 10^{-3}−10^{-2} and, therefore, not yet as bright as desirable.

The enhanced fluorescence exhibited by the Arch3 variants is triggered by the direct one-photon irradiation of the dark adapted (DA) state, leading to the direct formation of a fluorescent excited state (FS) upon relaxation from the Franck−Condon (FC) point of Figure 1B,[30,31] through a mechanism herein called "mechanism A". This is very different and more effective than the fluorescence of Arch3 that

Received: September 9, 2022

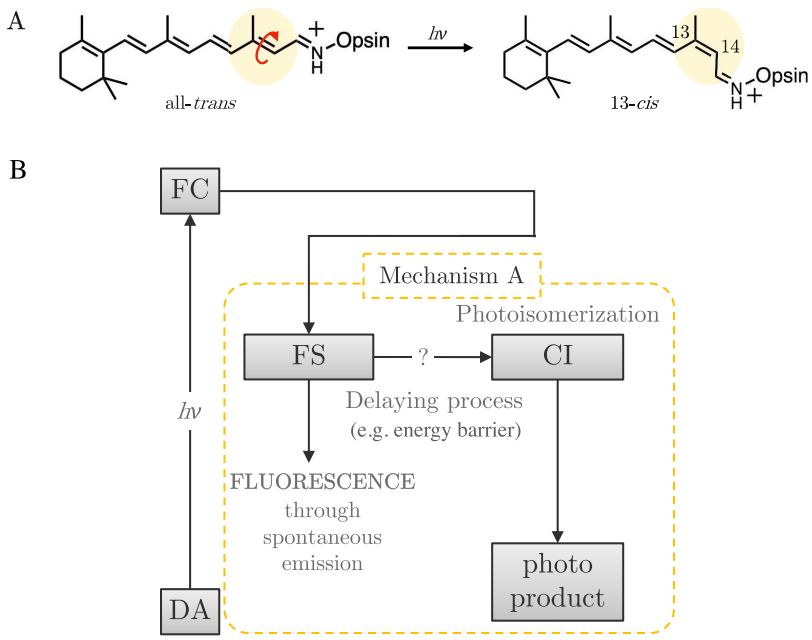


Figure 1. Fluorescence in microbial rhodopsins. (A) Structure of the microbial rPSB chromophore in its dark adapted (DA) state (left) and of its photo product after photoisomerization (right). (B) Fluorescence, through spontaneous emission, after photoexcitation of the DA state to its Franck−Condon (FC) and first-excited (S_1) state relaxation. In mechanism A, the photoisomerization process proceeds somewhat slowly, giving rise to a sufficiently stable fluorescent state (FS). In this case, fluorescence occurs from the reactive S_1 state and requires the absorption of just one photon. ? indicates a process capable of slowing down the isomerization.

originates from a photocycle intermediate and requires three photons to generate the spontaneously emitting state.[13,27,38,39] In mechanism A (see Figure 1B), the spontaneous emission of FS competes with the rPSB isomerization, ultimately leading to excited-state decay through a conical intersection (CI). Such a competition is controlled by the processes capable of slowing down the progression along the isomerization coordinate. For example, it has been reported that in a blue-shifted mutant of *Anabaena* sensory rhodopsin (ASR),[40] mixing of the S_1 and S_2 states can be responsible for longer FS lifetime and, therefore, enhanced emission.[41−43] As we will see below, such mixing may, however, not be the most common fluorescence enhancement process.

Here, we report on the design, implementation, and benchmarking of an automated protocol for selecting rhodopsin variants that display enhanced fluorescence with respect to a reference. The protocol, hereinafter referred to as "a-ARM fluorescence screening", leverages the framework of the advanced-automatic rhodopsin modeling protocol (a-ARM)[42,44−46] that uses QM/MM models for predicting the properties of the DA state of the protein. The a-ARM fluorescence screening protocol is tailored to "search" for fluorescent rhodopsin variants exploiting mechanism A (see Figure 1B) and, thus, assumes that the ϕ^f value is determined by the competition between the rate of spontaneous emission (k_{fl}) of FS and excited-state isomerization (k_{iso}). This implies that the isomerization dominates the FS lifetime (τ_{ESL}), as quantified in eq 1, where k_{other} indicates other radiationless deactivation processes

$$\phi^f = k_{fl} \cdot \tau_{ESL} = \frac{k_{fl}}{k_{fl} + k_{iso} + k_{other}} \quad (1)$$

Through different benchmarking studies, we show that the generated models can be used to simulate and assist predicting trends in the DA spectroscopic and photochemical properties. We also show that the analysis of the selected fluorescent variant models supports the existence of two different processes for the generation of fluorescence that we call mechanism A1 and mechanism A2.

Beyond the development of the protocol itself, we present a Python-based software package, called `PyARM`, comprising automated building, analyzing, and screening parts.[46] The a-ARM S_0 QM/MM model building protocol (incorporated in `PyARM` as `a_arm_protocol` driver; Figure 2A)[45] is used to build models characterized by the electrostatic embedding of the QM subsystem (the rPSB chromophore) and a hydrogen-link frontier between the QM subsystem and MM subsystem (the opsin apoprotein). In such basic models, rhodopsins are monomeric, gas-phase (*i.e.*, without membrane and solvent environment), and globally uncharged molecules (Figure 2B). `PyARM` also incorporates the a-ARM fluorescence screening protocol as `a_arm_fluorescence_searcher` driver (Figure 2C), representing a one-click user-friendly command-line architecture capable of generating candidate fluorescent rhodopsins starting from a set of ground-state (*i.e.*, S_0) a-ARM models.

2. METHODS

2.1. a-ARM Fluorescence Screening Protocol. The workflow of the a-ARM fluorescence screening protocol was partially inspired by a recent report[41] on the fluorescence mechanism operating in the W76S/Y179F$^{ASR}_{AT}$ mutant of ASR (see also ref 42). The authors found that an energy barrier (E^f_{S1}) located along the S_1 potential energy surface (PES) driving the rPSB isomerization (see Figure S1 in the Supporting Information) justifies the fluorescent enhancement observed when comparing the mutant with the wild type. While the workflow automates the steps taken to determine if

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

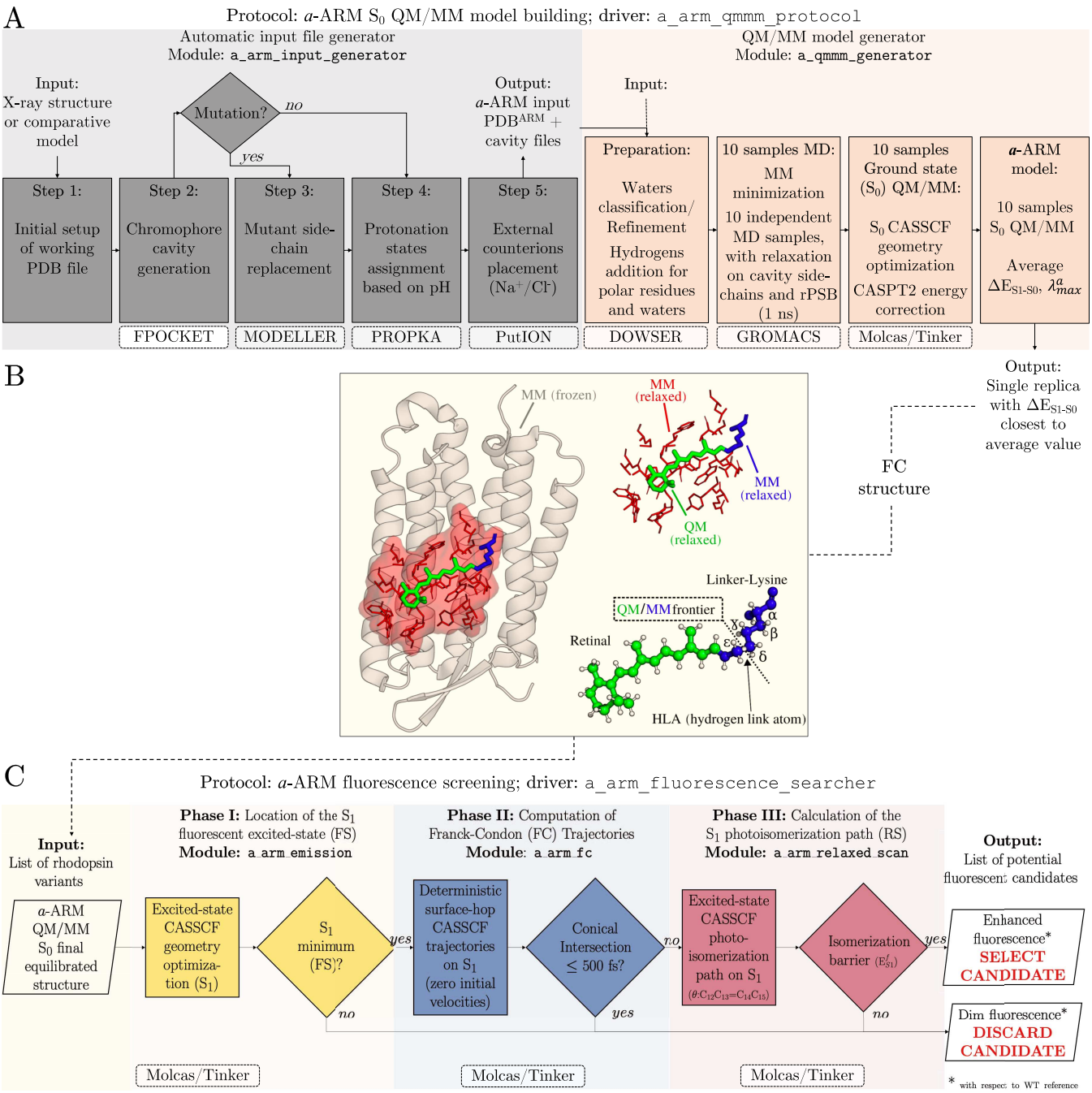


Figure 2. General workflow for the selection of rhodopsin fluorescent variants controlled by `PyARM`. (A) Flow-chart displaying the two-phase structure of the a-ARM S_0 QM/MM model building protocol, controlling the construction of ground-state a-ARM models and detailed in refs 42, 44−46. (B) Schematic illustration of the QM and MM subsystems and of the mobile part of a prototypical a-ARM model. These are the MM frozen atoms (in gray), MM relaxed atoms (in red and blue), and QM atoms (in green). A frontier is defined along the Cδ and Cε carbons of the chromophore linker-lysine. The QM valence is saturated by a dummy atom called hydrogen-link atom (HLA). (C) Flow-chart displaying the selection steps used in the a-ARM fluorescence screening protocol, controlling the automatic search for enhanced fluorescent rhodopsin variants. These are collected in three sequential phases, each imposing a distinct criterion for selecting/discarding possible fluorescent candidates.

an isomerization barrier exists in potentially fluorescent rhodopsins, the automation reduces human errors (*i.e.*, model preparation, data manipulation, and data interpretation) as well as makes possible the screening of entire arrays of fluorescent candidates.

As anticipated above, and illustrated in Figure 2, the workflow is based on two sequential protocols. The first one (see Figure 2A) prepares the input for the second one, in the form of one single S_0 a-ARM model (see Figure 2B) for each investigated variant. The second one is the actual screening protocol (see Figure 2C), and it is based on three phases that will now be detailed (see Section S3 of the Supporting Information).

2.1.1. Phase I: Location of the Fluorescent State. In phase I (see Figure 2C), the protocol answers the question of whether or not a rhodopsin variant has, close to the FC point, an S_1 energy minimum corresponding to FS. Such information provides a first, relatively fast, filter to determine if a variant has

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

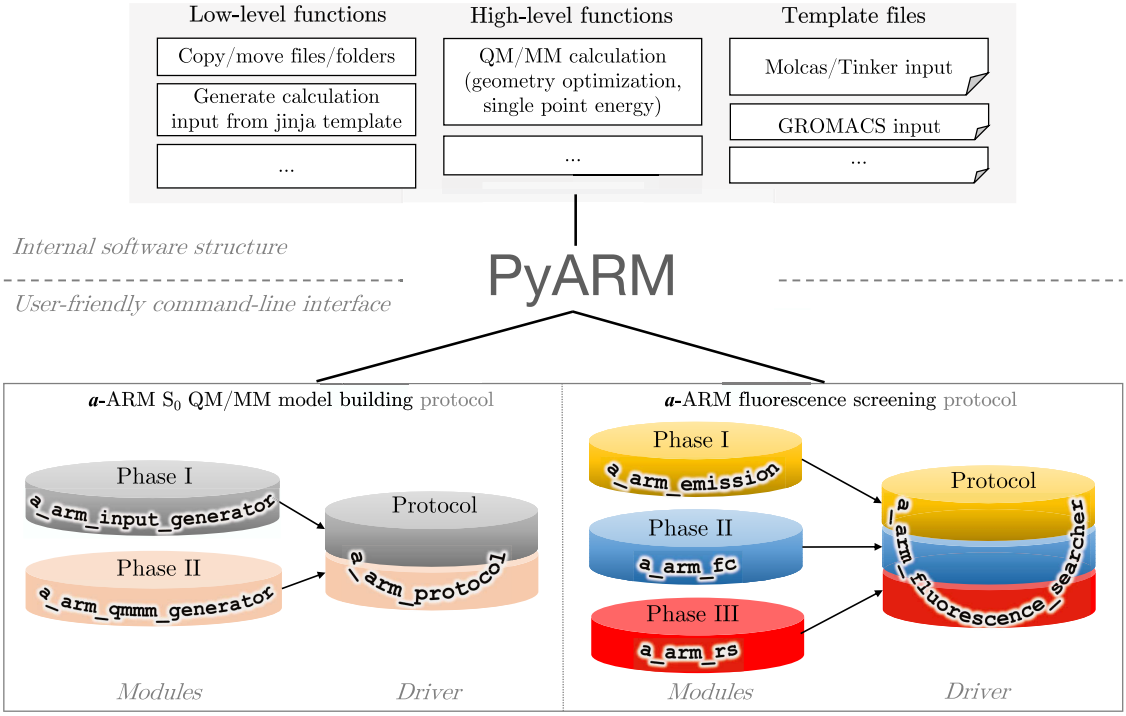


Figure 3. Structure of PyARM, a Python-based package. PyARM contains independent modules whose connection is defined by a specific driver such as a_arm_protocol or a_arm_fluorescence_searcher (names of modules and driver names are given in a typewriter font). Modules make use of high- and low-level functions, as well as template files that are integral parts of the package. Adapted with permission from,[46] open access under a CC BY license (Creative Commons Attribution 4.0 International License). Notice that colors marking the different modules (see the bottom part) create a link to the flow charts of Figure 2C marked with the same colors.

a fluorescent DA state or has to be immediately discarded. The only required input is a list of rhodopsin variants along with their FC geometries (*i.e.*, the ones corresponding, for each variant, to the replica featuring the maximum absorption wavelength (λ^{a}_{max}) value closest to the average of the canonical $N = 10$ replicas of a-ARM models (see Figure 2A)). Consistently with the a-ARM model structure[42,44−46] described in Figure 2B, the rPSB, water molecules, and side-chain atoms of the residues forming the chromophore cavity are allowed to relax from the FC point along the S_1 PES. Instead, the backbone atoms for all of the residues, as well as the noncavity residues (*e.g.*, side chain of amino acids, water molecules, and external counterions), are fixed at the original input coordinates.

The S_1 geometry optimization is driven by the potential energy gradient obtained at the state-averaged SA*n*-CASSCF/6-31G(d)/AMBER level with $n = 2$ (*i.e.*, two-roots) with the full rPSB π-system as the active space. The energy values of the optimized FS geometry are then recomputed with $n = 3$ (*i.e.*, three-roots, S_0, S_1, and S_2) using the CASPT2 rather than the CASSCF level. This choice is based on refs 41, 47 as the best compromise between computational time and quality of the electronic structure description.

The output is a list of potentially fluorescent candidates along with their FS geometries and estimated λ^{f}_{max} computed in terms of the vertical emission energy (δE^{f}_{S1-S0}) values (see Section S3.1 of the Supporting Information for an illustrative example).

2.1.2. Phase II: Calculation of Franck−Condon Trajectories. In phase II, the S_1 dynamics of the selected variants is probed using a single quantum-classical trajectory using the same S_0 models (*e.g.*, geometrical constraints and level of theory) used during phase I. We use the so-called FC trajectory, namely, a surface-hop trajectory released from the FC point with zero initial velocities (see the top right panel of Figure S1 in the Supporting Information), that allows us to see if a modest amount of kinetic energy would lead to relaxation of the identified FS. If a sizable energy barrier exists between the FS and a CI point, the trajectory is trapped in the FS region (long decay time), thus indicating a possible fluorescent candidate. However, if the barrier is shallow, the trajectory would overcome the barrier and rapidly reach the CI (short decay time), thus pointing to a weakly fluorescent species. Based on previous studies,[41,48−50] we define a threshold of 500 fs for the decay time to categorize rhodopsins as having dim or enhanced fluorescence. More specifically, the target of phase II is to qualitatively determine if the excited-state potential energy surface features a barrier high enough to block rapid progression to the ground state. In this phase and in phase III, we also look at the change in electronic structure along the trajectory (or path) by reporting the C14−C15=NH^+ charge variation. At the FC point, such a charge is large in the ground state and S_2 state (these have covalent/diradical characters) and it is small in the S_1 states that carries a charge-transfer character.[49] Coupled (i.e., mirror-image) variations along an FC trajectory or an RC of such a charge in different states are taken as evidence of electronic coupling between such states.[41]

The output is an updated list of potentially fluorescent candidates, along with their λ^{f}_{max} values, this time calculated as the average ΔE^{f}_{S1-S0} along the FC trajectory.[41] An illustrative example is provided in Section S3.2 of the Supporting Information.

2.1.3. Phase III: Calculation of Photoisomerization Paths. In phase III, the protocol computes the S_1 isomerization path

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

of rPSB to get more quantitative information on the E^{f}_{S1} and, consequently, on the fluorescence intensity of the variants selected in phase II. This is achieved, in an approximate fashion, via a relaxed scan (RS) along the C13=C14 (see Figure 1A) twisting coordinate starting from the FS minimum. The calculation locates the presence (or absence) of an energy maximum along the S_1 energy profile. The main output is a list of potentially fluorescent variants along with their corresponding calculated E^{f}_{S1} in terms of the energy difference between the maximum and FS. An illustrative example is provided in Section S3.3 of the Supporting Information.

2.2. Software Design and Implementation: The `PyARM` Package. `PyARM` has been mainly developed to automate the two protocols described in Figure 2. The `PyARM` functions, template files, modules, and drivers have been recently introduced in ref 46 and are detailed in Section S4 of the Supporting Information. A snapshot of the `PyARM` code containing the drivers reported in this work is provided as a supplementary file, while access to the developer's repository can be obtained by contacting the corresponding author.

2.2.1. Structure and Flow. As illustrated in Figure 3, `PyARM` implements low-level and high-level functions that define specialized protocols for handling the building and analysis of a-ARM models (see Section 2.2.2). `PyARM` uses external software, such as GROMACS,[51] for the MD engine, and the [Open]Molcas[52]/Tinker[53] interface,[54] for describing and linking the model QM and MM subsystems (see ref 46). As depicted in Figure 3, all components of `PyARM` are modular. A user can simply use a module to perform, automatically, a given type of application (*e.g.*, perform a geometry optimization) or use a driver to launch a protocol defined by connected modules and executing more complex computations (*e.g.*, generate an S_0 or S_1 QM/MM model, locate a fluorescent rhodopsin, etc.). An experienced user can create new drivers capable of executing new protocols.[42,44,45,55] We now describe the two drivers used in the present work, while others are reported in ref 46

2.2.2. `a_arm_protocol` Driver. As illustrated in Figures 2A and 3, this driver executes the a-ARM S_0 QM/MM model building protocol for the automated (*i.e.*, default) or semi-automated (*i.e.*, customized) construction of the S_0 a-ARM models necessary for the subsequent fluorescent rhodopsin search. For each rhodopsin variant, 10 different replicas are generated and the corresponding λ^{a}_{max} is computed as the average value.[42,45,54,56,57] The `a_arm_protocol` driver executes two modules incorporated into `PyARM` in a stand-alone fashion: the `a_arm_input_generator` and `a_arm_qmmm_generator`, respectively. The modules ultimately produce the necessary S_0 a-ARM models in two phases. Details on the a-ARM model description and building are given in Figure 2A, and a tutorial for each module is provided in Section S4.1 of the Supporting Information. The properties/limitations of the a-ARM technology have been reviewed in refs 42, 46.

2.2.3. `a_arm_fluorescence_searcher` Driver. As illustrated in Figures 2B and 3, each of the three phases of the a-ARM fluorescence screening protocol is implemented into `PyARM` as an independent module: phase I is implemented by the `a_arm_emission` module (see Section S4.2.4 in the Supporting Information), phase II by the `a_arm_fc` module (see Section S4.2.5 in the Supporting Information), and phase III by the `a_arm_relaxed_scan` module (see Section S4.2.6 in the Supporting Information). The `a_arm_fluorescence_searcher` driver links the three modules and automates the pipeline connecting the list of target variants (and the corresponding a-ARM models) to the list of candidates predicted to feature an enhanced fluorescence, their S_1 a-ARM models along with the trends in λ^{a}_{max}, λ^{f}_{max}, and E^{f}_{S1} values. As discussed below, the output also provides insight into the mechanism of enhanced fluorescence. The methodological details as well as some technical aspects of the implementation and usability are given in Sections S4.2 and S5 of the Supporting Information.

3. RESULTS AND DISCUSSION

To benchmark the a-ARM S_0 QM/MM model building and a-ARM fluorescence screening protocols, we employed three sets of rhodopsins. In Section 3.1, we discuss the first set, which from now on is called the benchmark set. As detailed in Table S1 of the Supporting Information, this is composed of 84 rhodopsin proteins whose measured λ^{a}_{max} ranges from 470 to 628 nm. It includes vertebrate, invertebrate, and microbial rhodopsins that feature either all-*trans*, 11-*cis*, or 9-*cis* rPSBs and significantly expands the previously reported benchmark of a-ARM models.[45,54,55] In addition, we present a few customized, rather than default, models that expose the present limitations of the input used to search for fluorescence variants.

The second and third sets are subsets of the benchmark set. In Section 3.2, we discuss the application set (see Table S1 of the Supporting Information) defined by Arch3 and 9 of its variants. While the X-ray structure of Arch3 is available in the Protein Data Bank as the 6GUX entry,[58] the variants' three-dimensional structures were obtained via comparative (homology) modeling carried out in our lab. Table S2 shows measured quantities such as λ^{f}_{max}, excited-state lifetime (ESL), and ϕ^{f} for each member of the set. Notice that, due to a general lack of fluorescence measurements, the application set is a relatively small set used for testing trends in fluorescence intensity. As shown below, this is done by correlating the computed S_1 isomerization barriers E^{f}_{S1} with the observed ϕ^{f}. We will also discuss the transition oscillator strength (f_{Osc}) computed at FS.

Finally, in Section 3.3, we discuss the search set. As reported in Table S1 of the Supporting Information, this set includes 15 entries whose fluorescent behavior has not yet been experimentally characterized and 3 entries that have been previously studied in ref 41, for a total of 18 variants. More specifically, the search set includes 17 mutants with known λ^{a}_{max} of the all-*trans* form of wild-type ASR (ASR_{AT}) and, thus, expands the set of just three mutants studied in ref 41. While the X-ray structure of ASR is available in the Protein Data Bank as the 1XIO entry,[40] the three-dimensional structure of its variants was obtained through a side-chain replacement procedure recently implemented in the a-ARM S_0 QM/MM model building protocol[46,57] (step 3 in Figure 2A). This set is employed for assessing the ability of the protocol to predict the behavior of variants that have shown to be fluorescent in the lab as well as for ranking variants that have never been tested experimentally.

3.1. Benchmark Set. To test the `a_arm_protocol` driver, we computed the DA state vertical excitation energy (δE^{a}_{S1-S0}) for the full benchmark set. Consistently with earlier reports, we set a 4.0 kcal mol^{-1} computed-observed difference threshold ($\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$) for acceptable models. The results are

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

shown in Figure 4, and the underlying data is given in the Supporting Information.

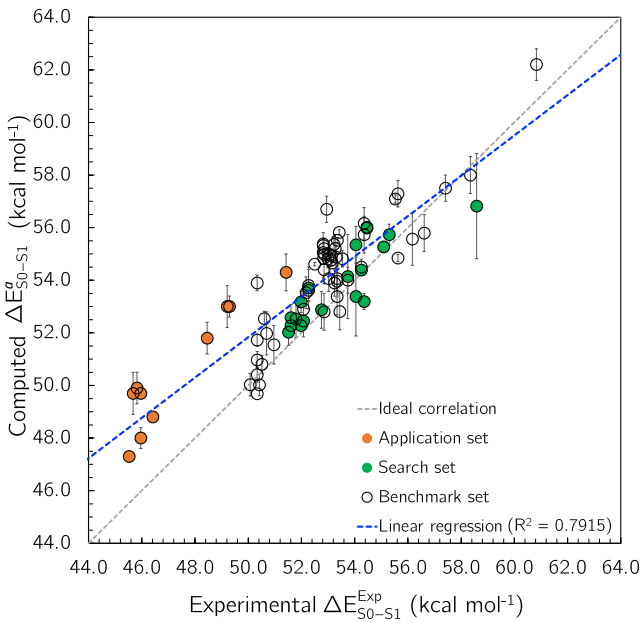


Figure 4. Vertical excitation energies. Extended benchmark of S_0 a-ARM model protocol in terms of trends of ΔE^{a}_{S1-S0}. The computed models and excitation energy values were obtained using the `a_arm_protocol` driver, where the average value from the 10 replicas is plotted for the benchmark set (empty circles), application set (orange circles), and search set (green circles), along with the corresponding standard deviation for the computed data (vertical error bars). Linear regression corresponds to the whole benchmark set. Experimental and computed data for KR2 WT and mutants are taken from ref 57. The row data are reported in the Supporting Information.

When the models are generated with the automated default approach (a-ARM$_{default}$),[42,45,46] 73 out of the 84 models (87%) have a $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ value below the threshold. The 11 outliers were modeled using a customized approach (a-ARM$_{customized}$)[42,45,46] to investigate the origin of such limitation. Briefly, when the default model fails to yield a $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ value below the threshold, it is modified by following a systematic (*i.e.*, nonarbitrary) procedure (see ref 46). Such a procedure probes the effects of varying the protonation state of ionizable residues (usually those surrounding the chromophore), which is a frequent cause of inconsistency between computed and observed ΔE^{a}_{S1-S0} values. Furthermore, the effect of the variation of selected side-chain rotamers is explored as the second cause of inconsistency (see refs 46, 57).

The a-ARM$_{default}$ and a-ARM$_{customized}$ approaches/models are detailed in refs 42, 45, 46. When considering both model types, the final parallelism between the computed and experimental trends (see Figure 4) is quantified by the trend deviation with respect to bovine rhodopsin (Rh_{11C}), calculated as $\|\text{Trend-Dev.}\| = |\Delta^{Rh\text{-}Exp,X\text{-}Exp}_{max,X} - \Delta^{Rh\text{-}Calc,X\text{-}Calc}_{max,X}|$, where $\Delta^{Rh\text{-}Exp,X\text{-}Exp}_{max,X}$ is the difference between the experimental λ^{a}_{max} of each rhodopsin in the set with respect to the experimental value of Rh_{11C}, while $\Delta^{Rh\text{-}Calc,X\text{-}Calc}_{max,X}$ is the difference between the corresponding computed a-ARM values. The obtained value is 0.6 ± 0.8 kcal mol^{-1} (0.03 ± 0.03 eV), which agrees with the 0.7 ± 0.5 kcal mol^{-1} (0.03 ± 0.02 eV) value reported in ref 45

As described in Sections 3.2 and 3.3, the obtained S_0 a-ARM models can be used for excited-state calculations. However, if a fully automated procedure were to be used for the input generation, only 87% of the models would match the selected threshold and represents a serious decrease in the model accuracy that has to be considered. Methods for improving the prediction of the residue protonation states during the construction of a-ARM models are being investigated in our laboratory (see for instance ref 59).

3.2. Application Set. We now discuss the a-ARM fluorescence screening protocol, using the application set (see Figure 3).

3.2.1. Phase I. The existence of FS categorizes a variant as potentially fluorescent. Accordingly, in phase I, the protocol tries to locate, for each entry, an S_1 energy minimum starting from one S_0 a-ARM model representing the FC point (from now on called the FC model). Such a model is defined as the replica whose λ^{a}_{max} is closest to the average value of the 10 a-ARM replicas generated by the `a_arm_protocol` driver. The use of a single model is necessary given the large number of calculations: phase I requires seven QM/MM calculations per model (see Figure S7 in the Supporting Information). Thus, a total of 700 QM/MM calculations would be required to generate one FS model for each of the 10 replicas of the 10 variants of the application set. The usage of a single a-ARM model reduces such a number to just 70 (*i.e.*, a 90% reduction).

The above strategy appears to be viable after looking at how a single model, rather than 10 replicas, affects the computed (i) λ^{a}_{max}, (ii) λ^{f}_{max}, and (iii) experimental-computed differences relative to Arch3. Figure 5A presents a view of the trend in ΔE^{a}_{S1-S0} (turquoise triangles), and the corresponding $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ values are given in Figure 5B. The individual ΔE^{a}_{S1-S0} computed for the single selected replica are also reported (dark-blue circles). It is shown (see error bars in Figure 5A and Table S3 in the Supporting Information) that the standard deviation in ΔE^{a}_{S1-S0} is relatively small, ranging between 0.0 and 0.8 kcal mol^{-1}, in line with what was previously reported.[44] Accordingly, one hypothesizes that, in general, the FC model of a specific variant would lead to an FS structure (or no FS) very close to those achieved by taking each of the corresponding 10 replicas as input for the S_1 relaxation.

To corroborate the above hypothesis, S_1 geometry optimizations were carried out for the 10 replicas of six variants (Arch3, Arch5, Arch7, Archon2, QuasAr1, and QuasAr2). The use of a subset allowed us to decrease the number of QM/MM calculations from ca. 700 to ca. 420. The produced trend in λ^{f}_{max} and standard deviation are reported in Figure 5C (green triangles). With the exception of Arch3, all tested variants achieved a stable FS and show no apparent differences between the results of the 10 replicas and the single selected model. However, in Arch3, 8 out of the 10 initial S_0 replicas provided an FS structure. Furthermore, in Figure 5D, one observes a difference between the $\Delta^{Exp}_{calc}\Delta E^{f}_{S1-S0}$ values from the 10 replicas and the single replica for the case of Archon2. This indicates that the results of phase I may not be conclusive.

Despite the above uncertainty on Arch3, the comparison of the computed average ΔE^{f}_{S1-S0} value with the value of the single replica suggests that the latter is an acceptable estimate for the FS structure and ΔE^{f}_{S1-S0}, justifying its use in phase I. Therefore, the rest of the application set (D95E/T99C$^{Arch3}_{AT}$, D95E/T99C/V59A$^{Arch3}_{AT}$, D95E/T99C/P296S$^{Arch3}_{AT}$, and D95E/

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

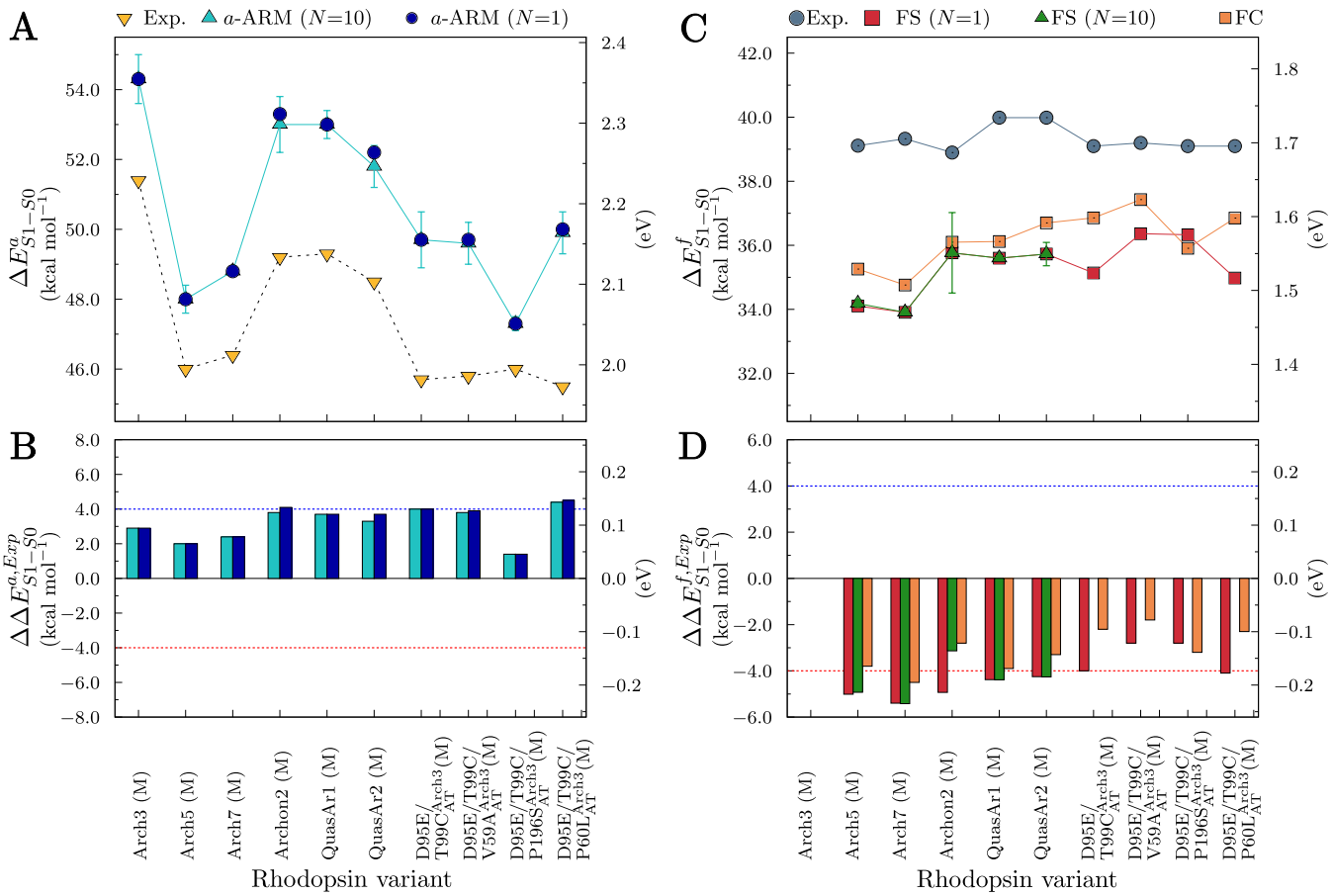


Figure 5. Trends in vertical absorption (ΔE^{a}_{S1-S0}) and emission (ΔE^{f}_{S1-S0}) excitation energies for the application set. (A) Average ΔE^{a}_{S1-S0} values computed for N = 10 replicas of an a-ARM S_0 model (green up-triangles), along with their corresponding error bars. The values relative to the replica used for the FS calculations are included as dark-blue circles. Experimental data is also provided (yellow triangles). (B) Difference between computed and experimental absorption data ($\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$). (C) Computed ΔE^{f}_{S1-S0} using the representative S_0 replica (red squares), along with experimental data (indigo circles). The average values for N = 10 replicas are provided for a subset of the application set (green triangles), with their corresponding error bars. Values corrected with kinetics energy (phase II) are shown in orange squares. (D) Difference between calculated and experimental data ($\Delta^{Exp}_{calc}\Delta\ E^{f}_{S1-S0}$). Bars of panels B and D are colored according to the computed data reported in panels A and C, respectively. Data on emission for Arch3 is not presented since its reported fluorescence comes from a photointermediate. Row data is reported in Table S3 in the Supporting Information.

T99C/P60L$^{Arch3}_{AT}$) was treated using only the representative replica.

Since the nine variant models were generated from the Arch3 X-ray crystallographic structure via comparative modeling, one expects that all computed λ^{a}_{max} and λ^{f}_{max} display differences from the experimental value (*e.g.*, $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$) consistent with those displayed by the WT template. However, inspection of Figure 5B reveals that this is not the case as one sees variations (irrespective of the number of replicas) up to 2 kcal mol^{-1} in absolute value (*e.g.*, for the triple mutant D95E/T99C/P196S$^{Arch3}_{AT}$). Similar deviations are seen in the emission values of Figure 5D. Considering that all variant S_0 models were generated using the same setup for the input file generator of Figure 2A (*i.e.*, chromophore cavity, protonation states, counterion placement), the details of the comparative modeling protocol (*e.g.*, the possible rearrangement of the chromophore cavity and water location) may be responsible for the deviation. This reveals an additional limitation of the a-ARM protocol discussed in refs 42, 57.

We now discuss the differences in computed and experimental trends, for λ^{f}_{max}, when using the FS model central to our automated protocol. As seen in Figure 5C, the trend is only loosely reproduced, with an error bar of about 2.2−5.4 kcal mol^{-1} red-shifted (Figure 5D). This indicates that the quality of the FS models appears to be decreased with respect to the corresponding FC models. We conclude that, presently, Phase I is only capable of categorizing rhodopsins with relatively large differences in λ^{f}_{max} (*e.g.*, discriminate between blue and red fluorescence).

Finally, we looked at the fluorescence mechanism. Accordingly, we evaluated the total energies for the S_1 and S_2 wave functions, computed at the CASPT2 level. As reported in Table S4 in the Supporting Information, no degeneracy is detected between S_1 and S_2, suggesting that the mechanism does not involve the mixing of S_1/S_2 wave functions proposed by Marín et al. Below, we will see that phase II and phase III corroborate such a hypothesis.

3.2.2. Phase II. A 500 fs FC trajectory was computed for each candidate selected in phase I to deal with the following issue: even if a stable FS is located, a variant may still have a too short ESL for emission. Figure 6 displays the relevant progression along the FC trajectories for the limiting cases of Arch3 and Arch7 (the progression of the other variants of the

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

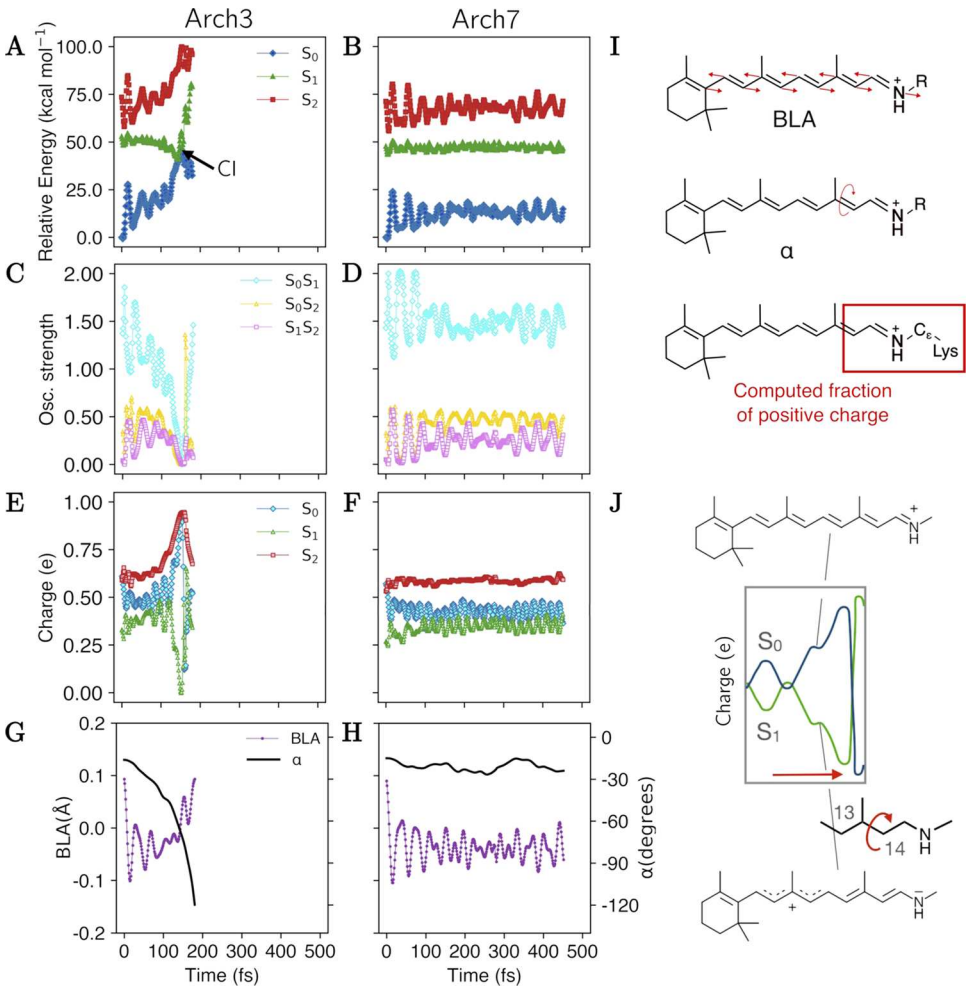


Figure 6. Franck−Condon (FC) trajectory computations on S_1 for Arch3 (A, C, E, G) and Arch7 (B, D, F, H). (A, B) a-ARM QM/MM FC trajectories computed at a two-root state-averaged SA2-CASSCF/AMBER/6-31G(d) level of theory and corrected at the 3-root CASPT2 level, using the `a_arm_fc` module. Blue and green dashed lines represent the CASSCF S_0 and S_1 profiles, respectively. The corrected S_0, S_1, and S_2 energies are represented as blue diamonds, green squares, and red triangles, respectively. (C, D) Transition oscillator strength. (E, F) Mulliken charge variations of the ═CH−CH═CH−CH═NH_2 moiety of the chromophore. (G, H) Torsional angle (α) of the C13═C14 isomerization coordinate and evolution of the total bond length alternation (BLA) computed along the FC trajectory. (I) Pictorial representation of the computed BLA and α dihedral for panels G and H, and chromophore moiety for panels E and F. (J) Representation of the mirror-like behavior of S_0 and S_1 charges in Arch3-based variants. Similar data, corresponding to the rest of mutants of the application set, is given in Section S6 in the Supporting Information.

application set is reported in Section S6 in the Supporting Information).

Since phase I did not produce a conclusive outcome on Arch3, we first analyze its FC trajectory. As shown in Figure 6A, the trajectory energy profile of Arch3 is consistent with the presence of a decay channel located in the vicinity of a CI, which is reached in ca. 200 fs. Consistently, the f_{Osc} (Figure 6C) and geometrical progressions (Figure 6G) show that S_1 is the spectroscopic allowed state and that promotion to such state leads to a ca. 90° twisted C13═C14 bond typical of the CI region. These results confirm that Arch3 must be discarded as its FS structure is unstable. The analogous procedure for the nine rhodopsins selected in phase I (see the case of Arch7 shown in Figure 6B,D,H) shows that no candidate reaches the CI in the same time scale and, actually, no significant double bond twisting progression is seen in the 500 fs simulation (see the Supporting Information).

We also looked at the kinetic energy effect on the quality of the λ^{f}_{max} calculation. To do so, λ^{f}_{max} was recomputed in terms of the average ΔE^{a}_{S1-S0} along the 200 fs FC trajectory. The results shown in Figure 5C,D (orange squares and bars, respectively) are slightly less red-shifted with respect to the experimental data. This indicates that the effect of the kinetic energy is not a major contributor to the $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ value. Thus, the discrepancy must be due to the limitations of the a-ARM model.

To learn more about the mechanism of fluorescence enhancement, we analyzed the electronic character changes along the S_1 trajectories of Arch3 and Arch7.

Accordingly, we computed the S_0, S_1, and S_2 charges (as derived from the CASSCF Mulliken population analysis) of a suitable moiety (C14−C15-NH^+) of rPSB and tracked its variation along the trajectory, as an indication of electronic state coupling. Figure 6E shows mixing between S_1 and S_0 states, revealed by mirror-image oscillations along the Arch3 trajectory, while there is no mixing between S_2 and S_1. This is in contrast with what was previously reported for mutants ASR in ref 41. Figure 6G shows that these oscillations are related to

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

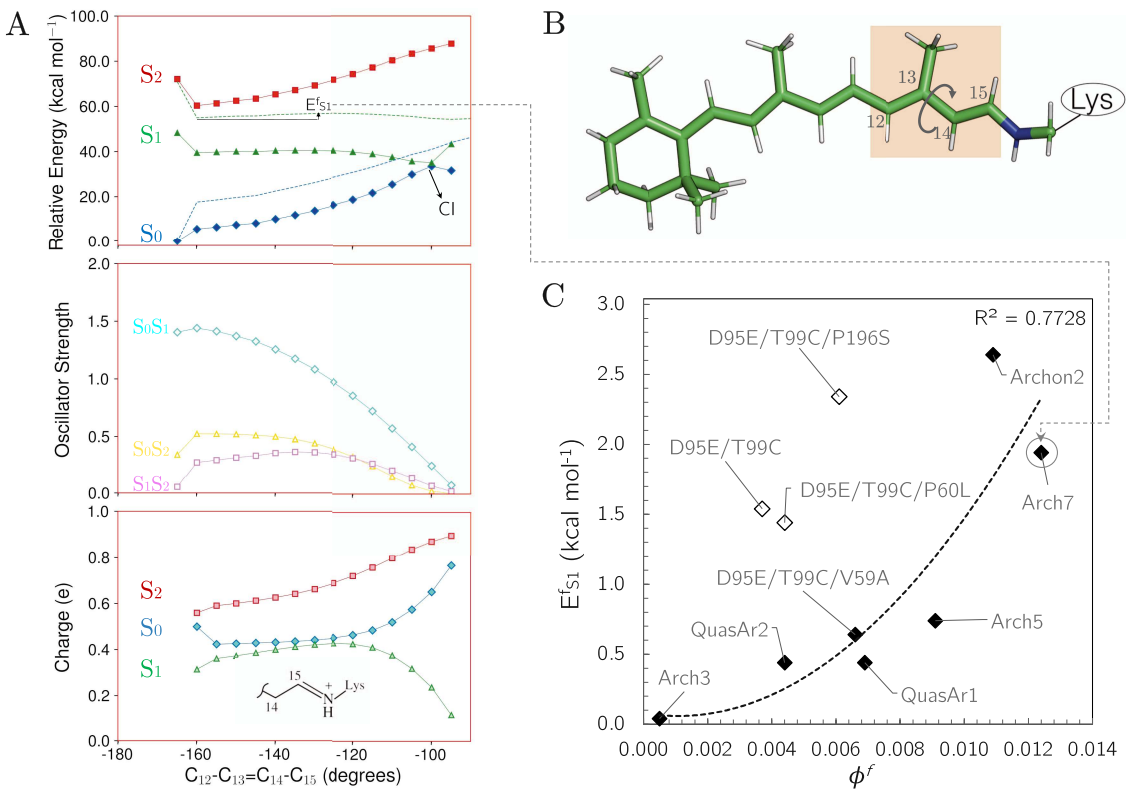


Figure 7. (A) Relaxed scan (RS) computation on S_1 for Arch7 of the application set. (top) a-ARM QM/MM RS computed at 2-roots from FS structure. Constrained geometry optimizations performed at the state-averaged (SA) 2-roots CASSCF/AMBER/6-31G(d) and energies corrected at the 3-roots CASPT2/6-31G(d) level. Blue and green dashed lines represent the CASSCF S_0 and S_1 profiles, respectively. The corrected S_0, S_1, and S_2 energies are represented as blue diamonds, green squares, and red triangles, respectively. (Middle) Transition oscillator strength (f_{Osc}). (Bottom) Mulliken charge variations of the ═CH−CH ═ CH−CH═NH_2 moiety of the chromophore. (B) Retinal protonated Schiff base (rPSB) all-*trans* representation; the arrow illustrates the coordinate of photoisomerization plotted in panel A. (C) Correlation between the computed energy isomerization barrier and experimental fluorescence quantum yield (ϕ^f) for the application set. The E^f_{S1} values were obtained, with the `a_arm_relaxed_scan` module, as the difference between the S_1 energy value at the energy maximum located along the S_1 path and the S_1 energy at FS, calculated at the 2-roots SA-CASSCF/AMBER/6-31G(d). The seven variants that present a second-order polynomial correlation between the E^f_{S1} and the ϕ^f (R^2 = 0.7728) are presented as filled diamonds, whereas the other three variants without apparent correlation are presented as empty diamonds. Similar data, corresponding to the rest of mutants of the application set, is given in Section S8 in the Supporting Information.

a change in bond order along a bond length alternation (BLA) stretching mode. Another observation is that along the progression, involving a 90° twisting of the C13═C14 bond, the charge distribution along the two coupled states not only oscillates but ultimately inverts. This suggests that the S_1 and S_0 (adiabatic) energy profiles are naturally described by the mixing of two (diabatic) states represented by the resonance formulas of Figure 6J. Specifically, S_1 corresponds to a charge-transfer state where the charge is spread along the rPSB backbone, which is far from the −C14═NH− location it had in the S_0 state. The fact that the S_1 state tends to change the "electronic character" along the isomerization path indicates that the S_1 energy profile, including the possible existence of an energy barrier, can be understood in terms of the electronic coupling between these two electronic configurations.

Figure 6F displays the same charge analysis performed with Arch7 and confirms that a similar S_1−S_0 coupling is also seen in variants that do not readily react (*i.e.*, that are prone to fluoresce). Thus, the above information, along with the conclusions of phase I about the nondegeneracy of S_1 and S_2 states in the FS structure, allows us to hypothesize a fluorescence mechanism where an E^f_{S1} barrier emerges along the isomerization coordinate due to S_1−S_0 coupling. This hypothesis is investigated in phase III.

Thus, with phase II, we conclude that the nine candidates of the application set selected by phase I must be transferred to phase III. This result is in line with the experimental fact that most entries of the application set have been observed to emit light (see Table S2) and some have already been employed as GEVIs.

3.2.3. Phase III. As anticipated above, phase III relies on the assumption that there is a relationship between the E^f_{S1} barrier and ESL and that, in turn, there is a proportionality between ϕ^f and ESL.[41] For Archaea rhodopsins, the barrier is located along the S_1 C13═C14 isomerization coordinate that is mapped via an RS calculation (*i.e.*, from FS to the CI featuring a 90° twisted C12−C13−C14−C15 rPSB dihedral angle; see Section S4.2.6 of the Supporting Information). The E^f_{S1} value is then computed as the CASPT2 energy difference between the RS energy maximum (*i.e.*, assumed to approximate the position of the transition structure) and the energy at FS. We take "as a conservative threshold" for classifying a member of the application set that has passed phases i and ii as non-fluorescent, a barrier value of <0.1 kcal mol^{-1}. Thus, members

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Summary of the Analysis Performed in Phases I−III for the Rhodopsins of the Application Set

rhodopsin variant	phase I[a]				phase II[b]		phase III[c]		fluorescent candidate?	mechanism[d]
	FS	α	ΔE^{f}_{S1-S0}	$\Delta\Delta E^{f,Exp}_{S1-S0}$	CI			E^{f}_{S1}		
Arch3	N.L.				R.	<200			no	
Arch5	L.	−153.3	34.1	−5.0	N.R.		E.B.	0.7	yes	A2
Arch7	L.	−160.1	33.9	−5.4	N.R.		E.B.	1.9	yes	A2
Archon2	L.	−162.4	32.4	−4.9	N.R.		E.B.	2.6	yes	A2
QuasAr1	L.	−157.4	35.5	−4.5	N.R.		E.B.	0.4	yes	A2
QuasAr2	L.	−156.1	35.5	−4.5	N.R.		E.B.	0.4	yes	A2
D95E/T99C$^{Arch3}_{AT}$	L.	−158.2	35.1	−4.0	N.R.		E.B.	1.5	yes	A2
D95E/T99C/P196S$^{Arch3}_{AT}$	L.	−156.9	36.3	−2.8	N.R.		E.B.	2.3	yes	A2
D95E/T99C/V59A$^{Arch3}_{AT}$	L.	−155.6	37.0	−2.2	N.R.		E.B.	0.6	yes	A2
D95E/T99C/P60L$^{Arch3}_{AT}$	L.	−157.2	35.0	−4.1	N.R.		E.B.	1.3	yes	A2

[a]FS structure located? N.L.: not located; L.: located, with an α dihedral angle for the C_{13}=C_{14} bond and computed ΔE^{f}_{S1-S0} (kcal mol^{-1}) SA2-CASSCF(12,12)/AMBER//3-CASPT2(12,12)/6-31G(d). [b]CI reached within 200 fs? N.R.: not reactive; R.: reactive, with the estimated photoisomerization time (fs); phase II not performed if no corresponding FS was located. [c]Energy barrier along 2-roots SA-CASSCF S_1 RS profile? B.L.: barrier-less; E.B.: energy barrier, 2-roots SA-CASSCF E^{f}_{S1} (kcal mol^{-1}); phase III not performed if no corresponding FS was located. [d]See Section 4 for the definition of mechanisms A1 and A2.

Table 2. Maximum Emission Wavelength (λ^{f}_{max}) Computed with the `a_arm_emission` Module for the Rhodopsins in the Search Set[a]

rhodopsin[b]	ΔE^{f}_{S1-S0}	λ^{f}_{max}	$f_{Osc}(S_1-S_0)$	ΔE^{f}_{S2-S0}	λ^{f}_{max}	$f_{Osc}(S_2-S_0)$
			$f_{Osc}(S_1-S_0) < f_{Osc}(S_2-S_0)$			
W76F$^{ASR}_{AT}$ (BS)	42.3 (1.56)	675	0.29	43.6 (1.60)	656	1.45
L83Q$^{ASR}_{AT}$ (BS)	44.5 (1.92)	643	0.17	43.6 (1.89)	656	1.54
W76S/Y179F$^{ASR}_{AT}$ (BS)	42.2 (1.55)	677	0.27	43.0 (1.58)	665	1.50
P206H$^{ASR}_{AT}$ (BS)	41.9 (1.81)	681	0.30	43.6 (1.89)	655	1.48
P206Q$^{ASR}_{AT}$ (BS)	44.7 (1.91)	640	0.14	42.9 (1.86)	665	1.58
P206Y$^{ASR}_{AT}$ (BS)	38.7 (1.67)	739	0.55	44.2 (1.92)	647	1.26
			$f_{Osc}(S_1-S_0) > f_{Osc}(S_2-S_0)$			
WT$^{ASR}_{AT}$	35.7 (1.31)	800	1.39	54.5 (2.00)	525	0.58
S214D$^{ASR}_{AT}$	33.3 (1.22)	858	1.24	54.5 (2.00)	525	0.60
Y73Q$^{ASR}_{AT}$ (RS)	34.0 (1.25)	841	1.30	54.8 (2.01)	522	0.60
D217E$^{ASR}_{AT}$ (RS)	34.3 (1.49)	834	1.33	54.9 (2.38)	521	0.59
D217N$^{ASR}_{AT}$ (RS)	34.4 (1.49)	831	1.37	54.9 (2.38)	520	0.58
E36Q$^{ASR}_{AT}$ (RS)	35.4 (1.53)	808	1.36	54.5 (2.36)	524	0.58
S86D$^{ASR}_{AT}$ (BS)	34.7 (1.28)	824	1.34	54.6 (2.01)	523	0.60
P206C$^{ASR}_{AT}$ (BS)	37.0 (1.36)	774	1.53	55.2 (2.03)	518	0.53
P206K$^{ASR}_{AT}$ (BS)	32.7 (1.41)	874	1.19	54.5 (2.36)	524	0.61
V112N$^{ASR}_{AT}$ (BS)	34.6 (1.27)	826	1.24	53.2 (1.95)	538	0.61
D75E$^{ASR}_{AT}$ (BS)	36.6 (1.59)	782	1.39	53.8 (2.33)	531	0.59
S214D-D217E$^{ASR}_{AT}$ (BS)	33.0 (1.43)	865	1.20	54.6 (2.36)	523	0.59

[a]S_1 geometry optimization using a two-state-averaged roots (n = 2) CASSCF wave function; FS CASPT2 energy re-evaluation based on a three-roots wave function, *i.e.*, SA2-CASSCF(12,12)/AMBER//3-CASPT2(12,12)/6-31G(d). f_{Osc} calculated at the CASSCF level. ΔE^{f}_{S1-S0} and ΔE^{f}_{S2-S0} in kcal mol^{-1} and eV (in parentheses); λ^{f}_{max} in nm; f_{Osc} unitless. [b]BS and RS stand for blue- and red-shifted, respectively, with respect to λ^{a}_{max} of ASR_{AT}.

with E^{f}_{S1} value equal to or larger than this threshold are selected as potentially fluorescent.

As an example, Figure 7A shows the computed isomerization path for Arch7. The reaction barrier is found between a −140 and −130° value of the dihedral angle (see Figure 7B). Section S9 in the Supporting Information reports the RS of the other variants of the application set. In all cases, the energy profiles have the structure expected for an only slightly fluorescent system (see Section 2.1.3), that is, with a barrier separating an FS and CI of just a few kcal mol^{-1}. The predicted trend in the E^{f}_{S1} value is given in Table S5, while Figure 7C shows a possible correlation with the experimental ϕ^{f} data.

Nevertheless, as seen for λ^{a}_{max} and λ^{f}_{max}, the observed trend is better reproduced when considering only a subset of the computed barriers. In fact, the data in Figure 7C, shows that it is possible to establish a reasonable polynomial fitting between ϕ^{f} and E^{f}_{S1} only for Arch3, Arch5, Arch7, Archon2, QuasAr1, QuasAr2, and D95E/T99C/V59A (olive diamonds). A more appropriate way of verifying the quality of the predicted S_1 barriers would be to compare the computed data with ESL measurements, as reported by Marín et al.[41] However, ESL measurements have not been performed for several application set members.

Table 1 collates the final results. All variants of the application set analyzed with phase III show an E^{f}_{S1} value above the defined threshold and thus can be considered potentially fluorescent. This is consistent with the observation of an enhanced DA fluorescence with respect to the WT. We therefore conclude that the proposed protocol has a certain degree of reliability.

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

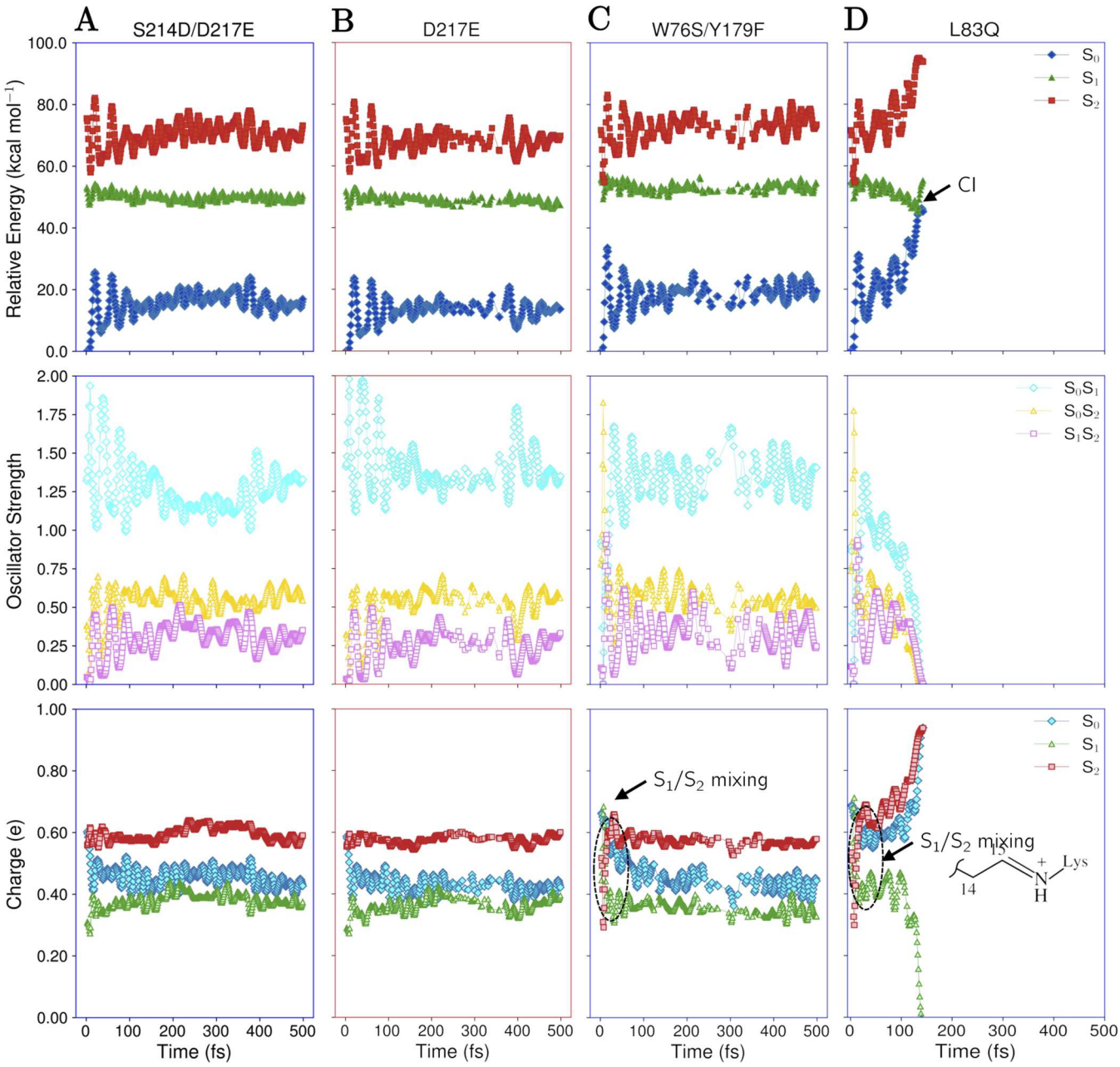


Figure 8. Franck−Condon (FC) trajectory computation on S_1 for four rhodopsins of the search set. S214D/D217E (A) and D217E (B) do not exhibit S_1−S_2 mixing, while W76S/Y179F (C) and L83Q (D) exhibit S_1−S_2 mixing. Methods, level of theory, colors, and symbols as in Figure 6A−F. Blue and red frames correspond to blue- and red-shifted variants, respectively. Similar data, corresponding to the rest of mutants of the search set, is given in Section S7 in the Supporting Information.

Finally, the RS was analyzed to gain insights into the fluorescence enhancement mechanism by looking at the change in charge distribution along the paths. As observed in Figure 7A for Arch7, and for all of the other variants of the application set in Section S9 of the Supporting Information, there is no coupling between the S_1 and S_2 energy profiles consistently with the fact that the two states are well separated in energy. On the other hand, the charge displacement occurring along the FC trajectory of Arch7 indicates that the S_0 and S_1 PESs are electronically coupled since they display a mirror-image-like behavior (see, for instance, also Figure 6F). This is consistent with the close S_1 and S_0 charge distributions seen in the mid-panel of Figure 7A. These results validate the hypothesis that the mechanism of enhanced fluorescence in the investigated red-shifted Arch3 variants is not the same as that reported for the ASR blue-shifted mutants but involves an S_1−S_0 coupling. As we will discuss below and in Section 4, it is such a coupling that is responsible for the generation of a barrier along the isomerization coordinate.

3.3. Search Set. This set includes 18 ASR-based rhodopsin variants (*i.e.*, 4 red-shifted and 13 blue-shifted), of which only two mutants ($L83Q^{ASR}_{AT}$ and $W76S/Y179F^{ASR}_{AT}$) have been experimentally investigated (see ref 41).

3.3.1. Phase I. An FS was located in all variants of the search set. As reported in Table 2, the computed S_1−S_0 λ^{f}_{max} values span a 643−874 nm range and, thus, considerably expanded with respect to that of the application set. The table shows that, in a number of cases (*i.e.*, 6 out of 18), the S_1 and S_2 states

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. Summary of the Analysis Performed in Phases I–III for the Rhodopsins of the Search Set

	phase I[a]			phase II[b]		phase III[c]			
rhodopsin variant	FS	α	ΔE^{f}_{S1-S0}	CI	time (fs)		E^{f}_{S1}	fluorescent candidate?	mechanism
WT^{ASR}_{AT}	L.	−158.0	35.7	N.R.		E.B.	0.3 (0.0)	yes	A2
$S214D^{ASR}_{AT}$	L.	−151.4	33.3	N.R.		E.B.	0.3 (0.0)	yes	A2
$Y73Q^{ASR}_{AT}$	L.	−153.2	34.0	N.R.		E.B.	0.2 (0.0)	yes	A2
$P206C^{ASR}_{AT}$	L.	−161.5	37.0	N.R.		E.B.	0.4 (0.1)	yes	A2
$P206K^{ASR}_{AT}$	L.	−153.5	32.7	N.R.		E.B.	0.2 (0.0)	yes	A2
$D75E^{ASR}_{AT}$	L.	−168.8	36.6	N.R.		E.B.	1.5 (0.4)	yes	A2
$D217E^{ASR}_{AT}$	L.	−156.0	34.3	N.R.		E.B.	0.5 (0.1)	yes	A2
$D217N^{ASR}_{AT}$	L.	−156.0	34.4	N.R.		E.B.	0.4 (0.0)	yes	A2
$E36Q^{ASR}_{AT}$	L.	−154.8	35.4	N.R.		E.B.	0.2 (0.0)	yes	A2
$S214D\text{-}D217E^{ASR}_{AT}$	L.	−150.2	33.0	N.R.		E.B.	0.1 (0.0)	yes	A2
$W76S/Y179F^{ASR}_{AT}$	L.	−171.0	42.2	N.R.		E.B.	N.A.	yes	A1
$P206H^{ASR}_{AT}$	L.	−178.0	41.9	N.R.		E.B.	N.A.	yes	A1
$P206Y^{ASR}_{AT}$	L.	-175.8	38.7	N.R.		E.B.	N.A.	yes	A1
$P206Q^{ASR}_{AT}$	L.	−157.6	44.7	R.	<200			nO	
$S86D^{ASR}_{AT}$	L.	−160.6	34.7	N.R.		B.L.	0.0 (0.0)	nO	
$W76F^{ASR}_{AT}$	L.	−155.5	42.3	R.	>200			nO	
$L83Q^{ASR}_{AT}$	L.	−153.8	44.5	R.	<200			nO	
$V112N^{ASR}_{AT}$	L.	−157.0	34.6	R.	≈320			nO	

[a]FS structure located? N.L.: not located; L.: located, with α dihedral angle for the $C_{13} = C_{14}$ bond and computed ΔE^{f}_{S1-S0} (kcal mol^{-1}) SA2-CASSCF(12,12)/AMBER//3-CASPT2(12,12)/6-31G(d). [b]CI reached within 200 fs? N.R.: not reactive; R.: reactive, with the estimated photoisomerization time (fs); phase II not performed if no corresponding FS was located. [c]Energy barrier along the S_1 RS profile? B.L.: barrier-less; E.B.: energy barrier, E^{f}_{S1} at the SA2-CASSCF(12,12)/AMBER level, values corrected at the 3-CASPT2(12,12)/6-31G(d) level are shown in parentheses (kcal mol^{-1}), N.A.: not applicable for RS with discontinuities; phase III not performed if no corresponding FS was located.

are close in energy and mixed (see Table S4 in the Supporting Information), pointing to the fluorescence mechanism reported in ref 41. All six mutants displaying S_2 and S_1 mixing are blue-shifted and have f_{Osc} values consistently with $S_2 \rightarrow S_0$ allowed transitions. However, since their S_1 and S_2 states are nearly degenerate, transitions to S_1 or S_2 are indistinguishable and both feature a switching electronic character. The other 12 members (including WT) of the search set show a behavior similar to that of the members of the application set. They do not show S_2–S_1 mixing, while the $S_1 \rightarrow S_0$ transition is allowed. Notice also that four of these variants are red-shifted.

In conclusion, our results indicate that all 18 variants are potentially fluorescent and have to be transferred to phase II. However, it is suggested that 6 (blue-shifted) variants have a fluorescence mechanism different from that of 12 (5 red-shifted, 6 blue-shifted, and WT) variants. Such conclusions will be further explored and corroborated in phase II and phase III.

3.3.2. Phase II. We start by inspecting the FC trajectories of the 12 variants not showing S_1–S_2 mixing or degeneracy. Figure 8A,B compares the blue-shifted variant S214D/$D217E^{ASR}_{AT}$ and the red-shifted variant $D217E^{ASR}_{AT}$. The energy profiles (top panels) show that both variants propagate exclusively along the S_1 PES and that the f_{Osc} value shows that S_1 is the spectroscopic state. The mirror-image-like C14–C15=NH^+ charge variation of the S_1 and S_0 states (bottom panels) indicates that these states are electronically coupled.

The results above are consistent with FC trajectories similar to those belonging to the fluorescent candidates of the application set and, therefore, characterized by a barrier origin related to S_0 and S_1 mixing. It is also apparent that no CI or twisted rPSB conformations are reached within 500 fs, with the only exception of the $V112N^{ASR}_{AT}$ variant that reaches the CI in about 320 fs (see Figure S16H in the Supporting Information) and is thus discarded. Notice that, within the selected subset showing $f_{Osc}(S_1–S_0) > f_{Osc}(S_2–S_0)$ according to Table 2, there are no significant differences between the FC trajectories computed for either blue- or red-shifted mutants (see also Figure S16 in Section S7 in the Supporting Information).

In contrast with the above variants, Figure 8C,D reports examples ($W76S/Y179F^{ASR}_{AT}$ and $L83Q^{ASR}_{AT}$) of FC trajectories characterized by mixing between S_1 and S_2 and $f_{Osc}(S_1–S_0) < f_{Osc}(S_2–S_0)$. During the very initial trajectory propagation, both energies and charges display a behavior consistent with near degeneracy in the FS region. However, in three blue-shifted cases ($P206Q^{ASR}_{AT}$, $W76F^{ASR}_{AT}$, and $L83Q^{ASR}_{AT}$), the degeneracy is abandoned and the variants reach the photochemically relevant CI and decay to S_0 within 500 fs (see Table 3, Figure 8D, and Figure S17 in Section S7 of the Supporting Information). Such decays imply that the corresponding rhodopsins must be discarded.

In conclusion, phase II reduces the number of fluorescent candidates from 18 to 14, as 4 rhodopsins, *i.e.*, $V112N^{ASR}_{AT}$, $W76F^{ASR}_{AT}$, $P206Q^{ASR}_{AT}$, and $L83Q^{ASR}_{AT}$, are predicted to decay rapidly.

3.3.3. Phase III. The trend in E^{f}_{S1} values for the search set is computed using the same strategy employed for the application set. The RS for the blue-shifted S214D/$D217E^{ASR}_{AT}$ and red-shifted $D217E^{ASR}_{AT}$ candidates that do not show S_1/S_2 mixing are plotted in Figure 9A,B, respectively, whereas the $W76S/Y179F^{ASR}_{AT}$ variant displaying S_1/S_2 mixing is presented in Figure 9C.

The energy profiles along RS may present discontinuities. These are attributed to the too fast relaxation of certain internal coordinates relative to the scanned coordinate and/or mixing with excited states not included in the state-averaging. One case is shown in Figure 9C (top panel) for a candidate exhibiting S_1/S_2 mixing. In the middle and bottom panels, there is an evident crossing of both the f_{Osc} and charges immediately after the energy discontinuity. All variants with S_1/S_2 mixing (see Table 2) present a similar behavior (as

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

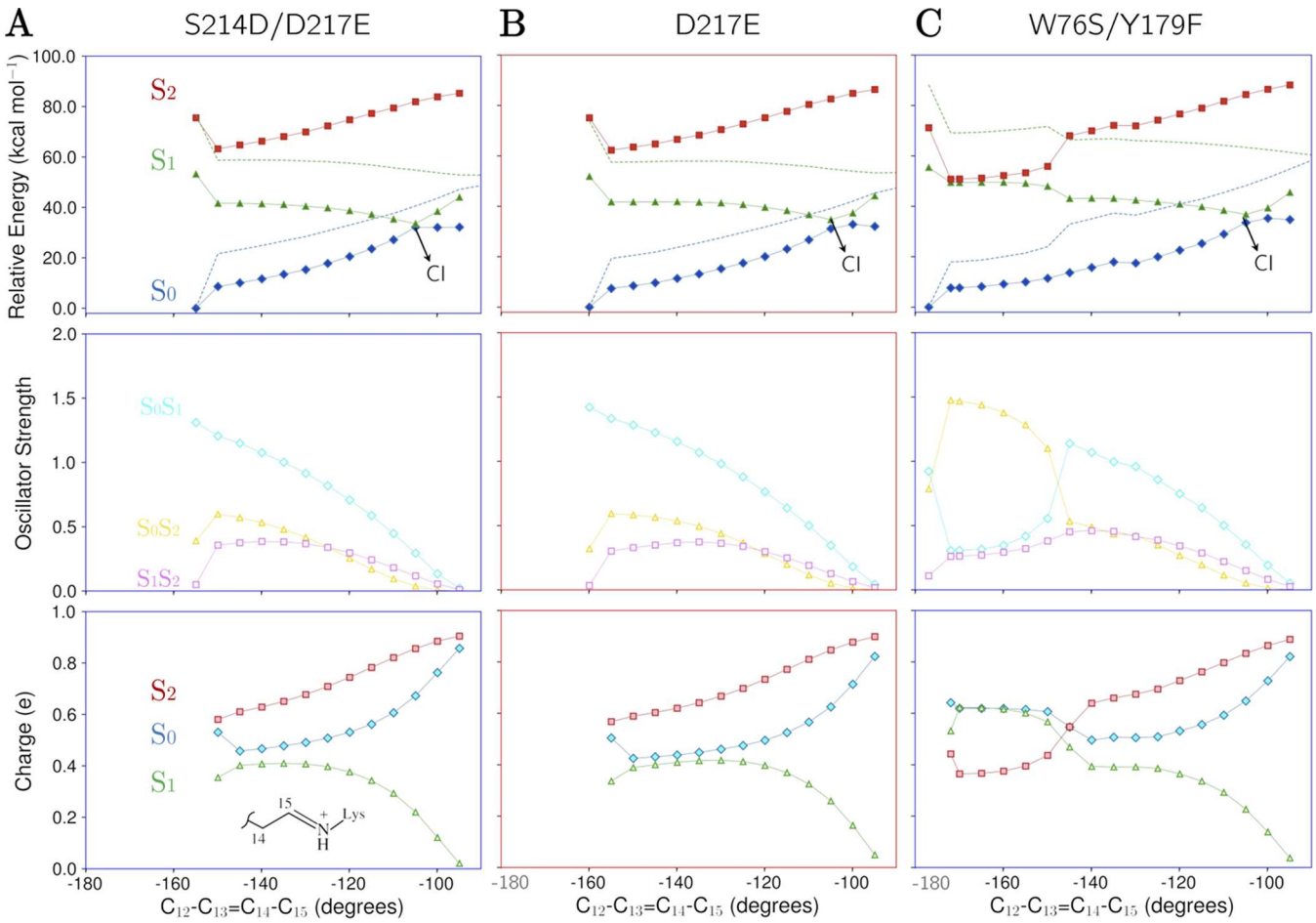


Figure 9. Relaxed scan (RS) for three members of the search set. (A) S214D/D217E$^{ASR}_{AT}$ and (B) D217E$^{ASR}_{AT}$ do not present S_1/S_2 mixing, while (C) W76S/Y179F$^{ASR}_{AT}$ does. Methods, level of theory, colors, and symbols as in Figure 7A. Blue and red frames correspond to blue- and red-shifted variants, respectively. Similar data, corresponding to the rest of mutants of the search set, is given in Section S9 in the Supporting Information.

shown in Figure S20 in Section S9 of the Supporting Information).

Due to the discontinuities along the S_1 energy profile, it is not possible to compute E^f_{S1} for the W76S/Y179F$^{ASR}_{AT}$ variant. However, it is possible to qualitatively identify whether a sizable barrier is present or not. For instance, it is evident from inspection of the W76S/Y179F$^{ASR}_{AT}$ energy profile that, after the FS region, the S_1 energy changes monotonically in a slightly downhill direction but remains almost degenerate with S_2 that instead changes in an uphill direction, suggesting that the S_1/S_2 mixing may generate a barrier (see also below). This result is consistent with the outcome of phase II where FC trajectory plotted in Figure 8C exhibits a nonreactive behavior. This variant should thus be considered fluorescent in agreement with experimental and theoretical data (see ref 41).

As illustrated in Figure 9A for S214D/D217E, no discontinuity is detected when S_1/S_2 mixing does not occur (see also Figure S19 in Section S9 of the Supporting Information). When comparing the RS energy profiles for the blue-shifted S214D/D217E and red-shifted D217E variants, the S_1 profile starting at FS is slightly uphill. The fact that E^f_{S1} of the red-shifted variant is almost 5 times higher than that of the blue-shifted suggests a correlation between computed E^f_{S1} and red-shifted or blue-shifted character of the variant. However, after looking at Table 3, which reports the E^f_{S1} for the whole set, we find that such correlation is not supported. For instance, the variant exhibiting the largest E^f_{S1} value of the set is the blue-shifted D75E. Whereas the analysis of the charges for the variants without S_1/S_2 mixing shows an evident electronic coupling between S_1 and S_0, as previously identified for mechanism A2, the variants with S_1/S_2 mixing show a mirror-like behavior between S_1 and S_2 characteristic of mechanism A1. In conclusion, the analyses of the energy profiles, charges, and oscillator strength, performed for both phase II and phase III, suggest that six variants (L83Q$^{ASR}_{AT}$, P206H$^{ASR}_{AT}$, P206Q$^{ASR}_{AT}$, P206Y$^{ASR}_{AT}$, W76F$^{ASR}_{AT}$, and W76S/Y179F$^{ASR}_{AT}$) feature a mechanism A1, whereas the other 12 variants present mechanism A2.

The above results suggest the following screening protocol based on the S_1 energy change: (i) if the RS energy profile is changing monotonically in a downhill direction (even in the presence of an S_1/S_2 degeneracy), the candidate is discarded; (ii) if the RS energy profile is flat and does not show a barrier the candidate is considered weakly fluorescent and discarded; and (iii) if the RS energy profile is initially uphill, the rhodopsin variant is selected as potentially fluorescent (this would incorporate also all cases where a >0.1 kcal mol^{-1} barrier has been located). We stress that these criteria are enforced by the need to automate the protocol (see Section S4.2) based on an RC computed at 2-root state CASSCF level and whose energy profile is then recomputed using three roots.

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

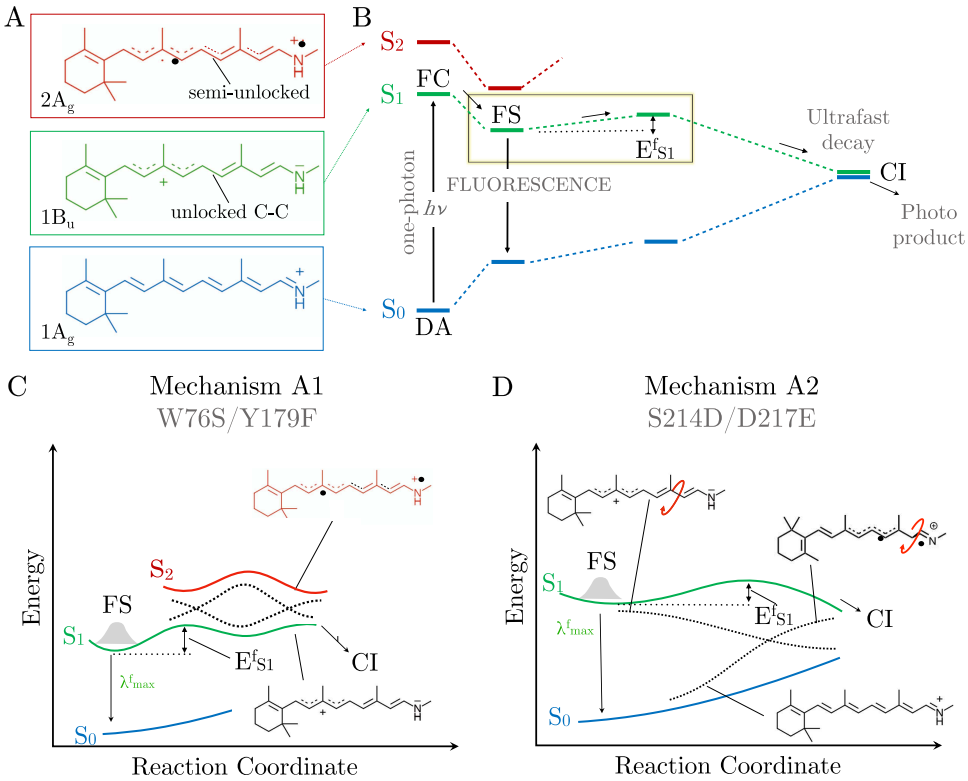


Figure 10. Fluorescence mechanisms and S_1 progression. (A) Resonance formulas representing the electronic character of the S_0, S_1, and S_2 states. (B) General structure of the S_1 isomerization path of a fluorescent rhodopsin. (C) Rugged surface generated via S_2/S_1 mixing. (D) Barrier generated via coupling between the S_1 and S_0. The dotted curves represent diabatic energies associated with the different electronic characters (see the main text).

Further calculations with, for instance, an increased number of roots have not been pursued.

After applying the above procedure to the rhodopsins selected in phase II, only $S86D_{AT}^{ASR}$ is discarded (see Table 3), while 13 out of 18 variants are classified as potentially fluorescent. Of these, 11 have not been experimentally investigated. Notice that, in contrast to the application set, there are no available observed ϕ^f data for the search set.

4. LIMITATIONS, FLUORESCENCE MECHANISMS, AND CONCLUSIONS

Above, we have presented an automated procedure that allows us to build the entire sets of QM/MM models of potentially fluorescent rhodopsins starting from X-ray crystallographic structures or comparative models as the only input. The procedure leverages the a-ARM protocol for the construction of FC models (including models of mutants) that are then screened via the a-ARM fluorescence screening protocol. Technically, this is performed by the presented `PyARM` python package, which automatically and sequentially executes the `a_arm_protocol` and `a_arm_fluorescence_searcher` drivers.

The quality of the FC models has been assessed via λ^a_{max} computations using a set of 84 rhodopsins (*i.e.*, the benchmark set). We have shown that, if a fully automated procedure had to be used, only 87% of the models would reproduce the observed trend. This represents a first "uncertainty issue" of the presented procedure. A second level of uncertainty has been documented for the FS models using the application set. In this case, the fact that a single replica is used for producing the FC model, as well as the related possible inaccuracies in the computed FS model, may have led to non-negligible differences in computed and observed λ^f_{max}. A consequence of this fact is probably reflected by the lack of perfect correlation between computed and observed values within the application set (see Figure 7C). Due to these limitations, the procedure must be used with care, for instance while testing or attempting predictions using models with (likely) different ionization states. Clearly, these issues prompt further research in the area of automated building of QM/MM models (*e.g.*, in comparative modeling, mutant modeling, and ionization state prediction methods, in the context of construction of a-ARM models), part of which is being carried out in our lab (see for instance ref 59).

Despite the above limitations, it was possible to reproduce the trend between computed and observed λ^f_{max} and document/demonstrate a correlation between the computed E^f_{S1} values and the measured ϕ^f. Indeed, for the GEVI system Arch3, we found that the computationally selected fluorescent rhodopsins match those experimentally observed to show fluorescence. For instance, the trend of ϕ^f is correctly predicted for 6 out of 10 members of the set. On the other hand, for ASR, we proposed 13 out of 17 possible fluorescent candidates, only 2 of which have been successfully expressed and observed to be fluorescent. These results justify further research efforts, with the final objective of providing a tool useful in optogenetics.

The investigation of the pool of Arch3 (application set) and ASR (search set) variant models allowed us to detect two distinct fluorescence enhancement mechanisms: mechanism A1 and mechanism A2. These mechanisms are characterized on the basis of the electronic characters of the first three singlet

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

electronic states of rPSB. In Figure 10A, we show the resonance formulas representing such characters, while the scheme of Figure 10B describes the isomerization path and energy profile used to interpret the computational results of phase I−phase III. In the same figure, we also frame (in yellow) the path region responsible for fluorescence modulation. As we will now discuss, the resonance formula and reaction path are used to propose a diabatic model for the origin of the E^{f}_{S1} located along the path.

Mechanism A1 (see Figure 10C), first reported by Marín et al.,[41] is characterized by the mixing of the reactive S_1 state with the partially bound S_2 state. This creates small S_1 energy barriers or, better, a "rugged" region of the PES whose origin is attributed to the increase of the 2Ag diradical character characterizing the S_2 wave function. Since the diradical character leads to a partially locked C13═C14 bond (see Figure 10A, red panel), the motion toward the CI becomes restrained when such character dominates. Accordingly, the existence of a rugged S_1 path can be explained via crossing and recrossing of 2Ag and 1Bu diabatic state energies along a nearly degenerate S_1/S_2 region. Such a behavior has also been recently described for the canonical microbial proton−pump bacteriorhodopsin.[60] The A1 mechanism was assigned on the basis of phase I−phase III calculations by detecting (i) a near-S_1/S_2 energy degeneracy at the FS structure leading to an S_1 state with a switching character along the FC trajectory, (ii) a mirror-image variation of the rPSB positive charge distribution of the S_1 and S_2 states along the same trajectory, and (iii) an S_1/S_2 degeneracy along the computed RS isomerization coordinate also showing mirror-image variations of the charge distribution of the two states. In this context, the mirror-image variation is associated with the coupled variation of the weights of the two diabatic states along the computed adiabatic states.

Mechanism A2 (see Figure 10D), originally reported in this work, is instead due to the coupling between the reactive S_1 state with the bounded S_0 state and on the destabilization of the CI structure with respect to the FS region (see refs 61, 62 for a more detailed description of this mechanism). Notice that, in contrast to mechanism A1, this mechanism involves a double bond twisting motion/deformation. As a result, the barrier is located along a highly distorted (nonplanar) geometry of the chromophore. It must also be noticed that, again due to the twisting, the S_1 electronic character becomes gradually that of a covalent diradical diabatic state that we associate with the 1Ag resonance formula of Figure 10A in the CI region. In this case, we hypothesize that the diabatic states associated with the 1Ag and 1Bu cross one time in the region corresponding to the S_1 energy maxima. Therefore, the barrier would originate from a large coupling between the two diabatic states in the diabatic crossing region. Regarding phase I−phase III calculations, the mechanism is characterized by (i) a large gap between the S_1 and S_2 energies at FS with an S_1 state displaying a charge-transfer character, (ii) a large gap between the S_1 and S_2 together with a mirror-image variation of the rPSB positive charge distribution of the S_0 and S_1 states, and (iii) an RS that does not exhibit any S_1/S_2 degeneracy but S_0 and S_1 charge distributions displaying mirror-image variations. Notice, however, that, due to the complexity of the rhodopsin systems, a precise assignment of the structural factors selecting the exact (*e.g.*, A1 vs A2) fluorescence mechanism would be premature and more in-depth studies are required.

As already mentioned above, we hope that, in the near future, the modular framework of PyARM will provide a platform "hosting" additional modules, therefore making complex workflows/protocols available to users with minimal programming knowledge. This may also include modules that use TD-DFT as a less costly QM methods in applications that allow the use of such a level of theory (*i.e.*, when investigating absorption properties). Another possibility is to include a module for the screening of or screening specific emitted colors after that a mutant or variant has been shown to be emissive. More specifically, we envision that, using such a tool, new protocols (also based on new modules and drivers) can be easily achieved, with only a minimal modification of the code (see also ref 46). In conclusion, we believe that the work here presented represents an original contribution toward the automation of multiscale computational photochemistry and photobiology that, in our opinion, constitutes one of the most needed and awaited advancements by the scientific community. This research need is easily demonstrated by the recent activity of other computational photochemistry laboratories working in the automation area.[63]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00928.

Row-data (ZIP)

Section S1 in the provided Supporting Information document includes details on all rhodopsins used in the benchmark set, application set, and search set; Section S3 provides further details on the a-ARM fluorescence screening protocol, including operational flow charts of its three phases; Section S4 reports technical details of PyARM, its drivers, modules, and templates, including screenshots, input and output files, usage, commands, software prerequisites, and requirements; Section S5 provides implementation and computational details specifically about the a-ARM fluorescence screening protocol; Sections S2 and S6−S9 contain extra Figures S1 and S15−S20; Section S10 reports Tables S3−S5 of raw data (PDF)

(ZIP)

(PDF)

■ AUTHOR INFORMATION

Corresponding Author

Massimo Olivucci − *Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, I-53100 Siena, Italy; Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States;* Email: massimo.olivucci@unisi.it, molivuc@bgsu.edu

Authors

Laura Pedraza-González − *Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, I-53100 Siena, Italy;* Present Address: Department of Chemistry and Industrial Chemistry, Università di Pisa, Via Moruzzi 13, 56124 Pisa, Italy

Leonardo Barneschi − *Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, I-53100 Siena, Italy*

Michał Marszałek − *Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, I-53100*

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Siena, Italy; Faculty of Chemistry, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

Daniele Padula − *Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, I-53100 Siena, Italy*

Luca De Vico − *Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, I-53100 Siena, Italy*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00928

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

L.P.G., L.B., D.P., L.D.V., and M.O. acknowledge partial support provided by MUR (Ministero dell'Università e della Ricerca) grant "Dipartimento di Eccellenza 2018−2022". M.M. acknowledges for ERASMUS+ Internship 2019−2020 from the Wrocław University of Science and Technology. D.P. acknowledges the Italian Ministry of University and Research (MUR) for a Rita Levi Montalcini grant (PGR18PJMBW). M.O. acknowledges partial support provided by grants NSF CHE-CLP-1710191, NIH 1R15GM126627 01, and Banca d'Italia. The authors acknowledge Dr. Alessio Valentini for fruitful discussions during the writing of the code.

■ REFERENCES

(1) Bratanov, D.; Kovalev, K.; Machtens, J.-P.; Astashkin, R.; Chizhov, I.; Soloviov, D.; Volkov, D.; Polovinkin, V.; Zabelskii, D.; Mager, T.; Gushchin, I.; Rokitskaya, T.; Antonenko, Y.; Alekseev, A.; Shevchenko, V.; Yutin, N.; Rosselli, R.; Baeken, C.; Borshchevskiy, V.; Bourenkov, G.; Popov, A.; Balandin, T.; Büldt, G.; Manstein, D. J.; Rodriguez-Valera, F.; Fahlke, C.; Bamberg, E.; Koonin, E.; Gordeliy, V. Unique structure and function of viral rhodopsins. *Nat. Commun* **2019**, *10*, No. 4939.

(2) Needham, D. M.; Yoshizawa, S.; Hosaka, T.; Poirier, C.; Choi, C. J.; Hehenberger, E.; Irwin, N. A.; Wilken, S.; Yung, C.-M.; Bachy, C.; Kurihara, R.; Nakajima, Y.; Kojima, K.; Kimura-Someya, T.; Leonard, G.; Malmstrom, R. R.; Mende, D. R.; Olson, D. K.; Sudo, Y.; Sudek, S.; Richards, T. A.; DeLong, E. F.; Keeling, P. J.; Santoro, A. E.; Shirouzu, M.; Iwasaki, W.; Worden, A. Z. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 20574−20583.

(3) Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev.* **2014**, *114*, 126−163.

(4) Skulachev, V. P. *Membrane Bioenergetics*; Springer-Verlag: Berlin, Heidelberg, 1988.

(5) Mendes, H. F.; Van Der Spuy, J.; Chapple, J. P.; Cheetham, M. E. Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. *Trends Mol. Med.* **2005**, *11*, 177−185.

(6) Mendes, H. F.; Cheetham, M. E. Pharmacological manipulation of gain-of-function and dominant-negative mechanisms in rhodopsin retinitis pigmentosa. *Hum. Mol. Genet.* **2008**, *17*, 3043−3054.

(7) Kurihara, M.; Sudo, Y. Microbial rhodopsins: wide distribution, rich diversity and great potential. *Biophys. Physicobiol.* **2015**, *12*, 121−129.

(8) Govorunova, E. G.; Sineshchekov, O. A.; Li, H.; Spudich, J. L. Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications. *Annu. Rev. Biochem.* **2017**, *86*, 845−872.

(9) Athanasiou, D.; Aguila, M.; Bellingham, J.; Li, W.; McCulley, C.; Reeves, P. J.; Cheetham, M. E. The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy. *Prog. Retinal Eye Res.* **2018**, *62*, 1−23.

(10) Kandori, H. Retinal Proteins: Photochemistry and Optogenetics. *Bull. Chem. Soc. Jpn.* **2020**, *93*, 76−85.

(11) Lenahan, C.; Sanghavi, R.; Huang, L.; Zhang, J. H. Rhodopsin: a potential biomarker for neurodegenerative diseases. *Front. Neurosci.* **2020**, *14*, No. 326.

(12) Kojima, K.; Shibukawa, A.; Sudo, Y. The Unlimited Potential of Microbial Rhodopsins as Optical Tools. *Biochemistry* **2020**, *59*, 218−229.

(13) Kojima, K.; Kurihara, R.; Sakamoto, M.; Takanashi, T.; Kuramochi, H.; Zhang, X. M.; Bito, H.; Tahara, T.; Sudo, Y. Comparative studies of the fluorescence properties of microbial rhodopsins: spontaneous emission versus photointermediate fluorescence. *J. Phys. Chem. B* **2020**, *124*, 7361−7367.

(14) Rozenberg, A.; Inoue, K.; Kandori, H.; Béjà, O. Microbial rhodopsins: the last two decades. *Annu. Rev. Microbiol.* **2021**, *75*, 427−447.

(15) Willem, J.; Ganapathy, S. Rhodopsins: an excitingly versatile protein species for research, development and creative engineering. *Front. Chem.* **2022**, *10*, No. 879609.

(16) Mattis, J.; Tye, K. M.; Ferenczi, E. A.; Ramakrishnan, C.; O'Shea, D. J.; Prakash, R.; Gunaydin, L. A.; Hyun, M.; Fenno, L. E.; Gradinaru, V.; Yizhar, O.; Deisseroth, K. Principles for Applying Optogenetic Tools Derived from Direct Comparative Analysis of Microbial Opsins. *Nat. Methods* **2012**, *9*, 159−172.

(17) Deisseroth, K. Optogenetics. *Nat. Methods.* **2011**, *8*, 26−29.

(18) Rein, M. L.; Deussing, J. M. The Optogenetic (R)evolution. *Mol. Genet. Genomics* **2012**, *287*, 95−109.

(19) Nakao, S.; Kojima, K.; Sudo, Y. Microbial Rhodopsins as Multifunctional Photoreactive Membrane Proteins for Optogenetics. *Biol. Pharm. Bull.* **2021**, *44*, 1357−1363.

(20) Broser, M. Far-red absorbing rhodopsins, insights from heterodimeric Rhodopsin-cyclases. *Front. Mol. Biosci.* **2022**, *8*, No. 1384.

(21) Alekseev, A.; Gordeliy, V.; Bamberg, E. *Rhodopsin. Methods in Molecular Biology*; Gordeliy, V., Ed.; Springer, Humana: New York, NY, 2022; pp 71−100.

(22) Fan, L. Z.; Kheifets, S.; Böhm, U. L.; Wu, H.; Piatkevich, K. D.; Xie, M. E.; Parot, V.; Ha, Y.; Evans, K. E.; Boyden, E. S.; Takesian, A.; Cohen, A. All-optical electrophysiology reveals the role of lateral inhibition in sensory processing in cortical layer 1. *Cell* **2020**, *180*, 521−535.

(23) Adam, Y.; Kim, J. J.; Lou, S.; Zhao, Y.; Xie, M. E.; Brinks, D.; Wu, H.; Mostajo-Radji, M. A.; Kheifets, S.; Parot, V.; Chettih, S.; Williams, K. J.; Gmeiner, B.; Farhi, S. L.; Madisen, L.; Buchanan, E. K.; Kinsella, I.; Zhou, D.; Paninski, L.; Harvey, C. D.; Zeng, H.; Arlotta, P.; Campbell, R. E.; Cohen, A. E. Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. *Nature* **2019**, *569*, 413−417.

(24) Piatkevich, K. D.; Bensussen, S.; Tseng, H.-a.; Shroff, S. N.; Lopez-Huerta, V. G.; Park, D.; Jung, E. E.; Shemesh, O. A.; Straub, C.; Gritton, H. J.; Romano, M. F.; Costa, E.; Sabatini, B. L.; Fu, Z.; Boyden, E. S.; Han, X. Population imaging of neural activity in awake behaving mice. *Nature* **2019**, *574*, 413−417.

(25) Panzera, L. C.; Hoppa, M. B. Genetically encoded voltage indicators are illuminating subcellular physiology of the axon. *Front. Cell. Neurosci.* **2019**, *13*, No. 52.

(26) Chien, M.-P.; Brinks, D.; Testa-Silva, G.; Tian, H.; Brooks, F. P.; Adam, Y.; Bloxham, W.; Gmeiner, B.; Kheifets, S.; Cohen, A. E. Photoactivated voltage imaging in tissue with an archaerhodopsin-derived reporter. *Sci. Adv.* **2021**, *7*, No. eabe3216.

(27) Zhang, X. M.; Yokoyama, T.; Sakamoto, M. Imaging Voltage with Microbial Rhodopsins. *Front. Mol. Biosci.* **2021**, *8*, No. 811.

(28) Chow, B. Y.; Han, X.; Dobry, A. S.; Qian, X.; Chuong, A. S.; Li, M.; Henninger, M. A.; Belfort, G. M.; Lin, Y.; Monahan, P. E.; Boyden, E. S. High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **2010**, *463*, 98−102.

(29) Kralj, J. M.; Douglass, A. D.; Hochbaum, D. R.; Maclaurin, D.; Cohen, A. E. Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. *Nat. Methods* **2012**, *9*, 90−95.

(30) Hochbaum, D. R.; Zhao, Y.; Farhi, S. L.; Klapoetke, N.; Werley, C. A.; Kapoor, V.; Zou, P.; Kralj, J. M.; Maclaurin, D.; Smedemark-Margulies, N.; Saulnier, J. L.; Boulting, G. L.; Straub, C.; Cho, Y. K.; Melkonian, M.; Wong, G. K.-S.; Harrison, D. J.; Murthy, V. N.; Sabatini, B. L.; Boyden, E. S.; Campbell, R. E.; Cohen, A. E. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **2014**, *11*, 825−833.

(31) Piatkevich, K. D.; Jung, E.; Straub, C.; Linghu, C.; Park, D.; Suk, H.; Hochbaum, D.; Goodwin, D.; Pnevmatikakis, E.; Pak, N.; Kawashima, T.; Yang, C.; Rhoades, J.; Shemesh, O.; Asano, S.; Yoon, Y.; Freifeld, L.; Saulnier, J.; Riegler, C.; Engert, F.; Hughes, T.; Drobizhev, M.; Szabo, B.; Ahrens, M.; Flavell, S.; Sabatini, B.; Boyden, E. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol* **2018**, *14*, 352−360.

(32) Penzkofer, A.; Silapetere, A.; Hegemann, P. Absorption and Emission Spectroscopic Investigation of the Thermal Dynamics of the Archaerhodopsin 3 Based Fluorescent Voltage Sensor Archon2. *Int. J. Mol. Sci.* **2020**, *21*, No. 6576.

(33) Flytzanis, N. C.; Bedbrook, C. N.; Chiu, H.; Engqvist, M. K.; Xiao, C.; Chan, K. Y.; Sternberg, P. W.; Arnold, F. H.; Gradinaru, V. Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons. *Nat. Commun* **2014**, *5*, No. 4894.

(34) McIsaac, R. S.; Engqvist, M. K.; Wannier, T.; Rosenthal, A. Z.; Herwig, L.; Flytzanis, N. C.; Imasheva, E. S.; Lanyi, J. K.; Balashov, S. P.; Gradinaru, V.; Arnold, F. H. Directed evolution of a far-red fluorescent rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 13034−13039.

(35) Lou, S.; Adam, Y.; Weinstein, E. N.; Williams, E.; Williams, K.; Parot, V.; Kavokine, N.; Liberles, S.; Madisen, L.; Zeng, H.; Cohen, A. Genetically targeted all-optical electrophysiology with a transgenic Cre-dependent optopatch mouse. *J. Neurosci.* **2016**, *36*, 11059−11073.

(36) Dempsey, G. T.; Chaudhary, K. W.; Atwater, N.; Nguyen, C.; Brown, B. S.; McNeish, J. D.; Cohen, A. E.; Kralj, J. M. Cardiotoxicity screening with simultaneous optogenetic pacing, voltage imaging and calcium imaging. *J. Pharmacol. Toxicol. Methods* **2016**, *81*, 240−250.

(37) Azimi Hashemi, N.; Bergs, A. C.; Schüler, C.; Scheiwe, A. R.; Costa, W. S.; Bach, M.; Liewald, J. F.; Gottschalk, A. Rhodopsin-based voltage imaging tools for use in muscles and neurons of Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 17051−17060.

(38) Maclaurin, D.; Venkatachalam, V.; Lee, H.; Cohen, A. E. Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5939−5944.

(39) Ohtani, H.; Itoh, H.; Shinmura, T. Time-resolved fluorometry of purple membrane of Halobacterium halobium O640 and an O-like red-shifted intermediate Q. *FEBS Lett.* **1992**, *305*, 6−8.

(40) Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H. Anabaena sensory rhodopsin: a photochromic color sensor at 2.0 Å. *Science* **2004**, *306*, 1390−1393.

(41) Marín, M. d. C.; Agathangelou, D.; Orozco-Gonzalez, Y.; Valentini, A.; Kato, Y.; Abe-Yoshizumi, R.; Kandori, H.; Choi, A.; Jung, K.-H.; Haacke, S.; Olivucci, M. Fluorescence Enhancement of a Microbial Rhodopsin via Electronic Reprogramming. *J. Am. Chem. Soc.* **2019**, *141*, 262−271.

(42) Pedraza-González, L.; Marín, M. d. C.; De Vico, L.; Yang, X.; Olivucci, M. *QM/MM Studies of Light-responsive Biological Systems*; Andruniów, T.; Olivucci, M., Eds.; Springer Nature: Switzerland AG, 2021; Vol. *31*, pp 1−75.

(43) Agathangelou, D.; Roy, P. P.; del Carmen Marín, M.; Ferre, N.; Olivucci, M.; Buckup, T.; Léonard, J.; Haacke, S. Sub-picosecond C=C bond photo-isomerization: evidence for the role of excited state mixing. *C. R. Phys.* **2021**, *22*, 111−138.

(44) Melaccio, F.; Marín, M. dC.; Valentini, A.; Montisci, F.; Rinaldi, S.; Cherubini, M.; Yang, X.; Kato, Y.; Stenrup, M.; Orozco-González, Y.; Ferré, N.; Luk, H. L.; Kandori, H.; Olivucci, M. Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *J. Chem. Theory Comput.* **2016**, *12*, 6020−6034.

(45) Pedraza-González, L.; De Vico, L.; Marín, M. dC.; Fanelli, F.; Olivucci, M. *a*-ARM: Automatic Rhodopsin Modeling with Chromophore Cavity Generation, Ionization State Selection and External Counter-ion Placement. *J. Chem. Theory Comput.* **2019**, *15*, 3134−3152.

(46) Pedraza-González, L.; Barneschi, L.; Padula, D.; De Vico, L.; Olivucci, M. Evolution of the Automatic Rhodopsin Modeling (ARM) Protocol. *Top. Curr. Chem.* **2022**, *380*, 1−48.

(47) Melaccio, F.; Olivucci, M.; Lindh, R.; Ferré, N. Unique QM/MM potential energy surface exploration using microiterations. *Int. J. Quantum Chem.* **2011**, *111*, 3339−3346.

(48) Luk, H. L.; Melaccio, F.; Rinaldi, S.; Gozem, S.; Olivucci, M. Molecular bases for the selection of the chromophore of animal rhodopsins. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 15297−15302.

(49) Gozem, S.; Luk, H. L.; Schapiro, I.; Olivucci, M. Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores. *Chem. Rev.* **2017**, *117*, 13502−13565.

(50) Gholami, S.; Pedraza-González, L.; Yang, X.; Granovsky, A. A.; Ioffe, I. N.; Olivucci, M. Multistate Multiconfiguration Quantum Chemical Computation of the Two-Photon Absorption Spectra of Bovine Rhodopsin. *J. Phys. Chem. Lett.* **2019**, *10*, 6293−6300.

(51) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(52) Aquilante, F.; Autschbach, J.; Carlson, R. K.; Chibotaru, L. F.; Delcey, M. G.; De Vico, L.; Fdez Galván, I.; Ferré, N.; Frutos, L. M.; Gagliardi, L.; Garavelli, M.; Giussani, A.; Hoyer, C. E.; Li Manni, G.; Lischka, H.; Ma, D.; Malmqvist, PÅ.; Müller, T.; Nenov, A.; Olivucci, M.; Pedersen, T. B.; Peng, D.; Plasser, F.; Pritchard, B.; Reiher, M.; Rivalta, I.; Schapiro, I.; Segarra-Martí, J.; Stenrup, M.; Truhlar, D. G.; Ungur, L.; Valentini, A.; Vancoillie, S.; Veryazov, V.; Vysotskiy, V. P.; Weingart, O.; Zapata, F.; Lindh, R. Molcas 8: New Capabilities for Multiconfigurational Quantum Chemical Calculations Across the Periodic Table. *J. Comput. Chem.* **2016**, *37*, 506−541.

(53) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardére, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273−5289.

(54) Aquilante, F.; Autschbach, J.; Baiardi, A.; Battaglia, S.; Borin, V. A.; Chibotaru, L. F.; 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. Modern quantum chemistry with [Open] Molcas. *J. Chem. Phys.* **2020**, *152*, No. 214117.

(55) Pedraza-González, L.; Marín, M. dC.; Jorge, A. N.; Ruck, T. D.; Yang, X.; Valentini, A.; Olivucci, M.; De Vico, L. Web-ARM: a Web-Based Interface for the Automatic Construction of QM/MM Models of Rhodopsins. *J. Chem. Inf. Model.* **2020**, *60*, 1481−1493.

(56) Mroginski, M.-A.; Adam, S.; Amoyal, G. S.; Barnoy, A.; Bondar, A.-N.; Borin, V. A.; Church, J. R.; Domratcheva, T.; Ensing, B.; Fanelli, F.; Ferré, N.; Filiba, O.; Pedraza-González, L.; González, R.; González-Espinoza, C. E.; Kar, R. K.; Kemmler, L.; Kim, S. S.; Kongsted, J.; Krylov, A. I.; Lahav, Y.; Lazaratos, M.; NasserEddin, Q.; Navizet, I.; Nemukhin, A.; Olivucci, M.; Olsen, J. M. H.; Pérez de Alba Ortíz, A.; Pieri, E.; Rao, A. G.; Rhee, Y. M.; Ricardi, N.; Sen, S.; Solov'yov, I. A.; De Vico, L.; Wesolowski, T. A.; Wiebeler, C.; Yang, X.; Schapiro, I. Frontiers in Multiscale Modeling of Photoreceptor Proteins. *Photochem. Photobiol.* **2021**, *97*, 243−269.

(57) Nakajima, Y.; Pedraza-González, L.; Barneschi, L.; Inoue, K.; Olivucci, M.; Kandori, H. Pro219 is an Electrostatic Color Determinant in the Light-driven Sodium Pump KR2. *Commun. Biol.* **2021**, *4*, No. 1185.

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(58) Juarez, J. F. B.; Judge, P. J.; Adam, S.; Axford, D.; Vinals, J.; Birch, J.; Kwan, T. O.; Hoi, K. K.; Yen, H.-Y.; Vial, A.; Milhiet, P.-E.; Robinson, C. V.; Schapiro, I.; Moraes, I.; Watts, A. Structures of the archaerhodopsin-3 transporter reveal that disordering of internal water networks underpins receptor sensitization. *Nat. Commun.* **2021**, *12*, No. 629.

(59) Pieri, E.; Ledentu, V.; Sahlin, M.; Dehez, F.; Olivucci, M.; Ferré, N. CpHMD-Then-QM/MM Identification of the Amino Acids Responsible for the Anabaena Sensory Rhodopsin pH-Dependent Electronic Absorption Spectrum. *J. Chem. Theory Comput.* **2019**, *15*, 4535−4546.

(60) Gozem, S.; Johnson, P. J. M.; Halpin, A.; Luk, H. L.; Morizumi, T.; Prokhorenko, V. I.; Ernst, O. P.; Olivucci, M.; Miller, R. J. D. Excited-State Vibronic Dynamics of Bacteriorhodopsin from Two-Dimensional Electronic Photon Echo Spectroscopy and Multi-configurational Quantum Chemistry. *J. Phys. Chem. Lett.* **2020**, *11*, 3889−3896.

(61) Barneschi, L.; Marsili, E.; Pedraza-González, L.; Padula, D.; De Vico, L.; Kaliakin, D.; Blanco-González, A.; Ferré, N.; Huix-Rotllant, M.; Filatov, M.; Olivucci, M. On the fluorescence enhancement of arch neuronal optogenetic reporters. *Nat. Commun.* **2022**, *13*, No. 6432.

(62) Palombo, R.; Barneschi, L.; Pedraza-González, L.; Padula, D.; Schapiro, I.; Olivucci, M. Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin. *Nat. Commun.* **2022**, *13*, No. 6652.

(63) Pieri, E.; Lahana, D.; Chang, A. M.; Aldaz, C. R.; Thompson, K. C.; Martínez, T. J. The non-adiabatic nanoreactor: towards the automated discovery of photochemistry. *Chem. Sci* **2021**, *12*, 7294−7307.

https://doi.org/10.1021/acs.jctc.2c00928
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX