



Published in final edited form as:

J Chem Inf Model. 2020 March 23; 60(3): 1481–1493. doi:10.1021/acs.jcim.9b00615.

Web-ARM: a Web-Based Interface for the Automatic Construction of QM/MM Models of Rhodopsins

Laura Pedraza-González[†], María Del Carmen Marín[†], Alejandro N. Jorge[‡], Tyler D. Ruck[‡], Xuchun Yang[‡], Alessio Valentini[¶], Massimo Olivucci^{†,‡}, Luca De Vico[†]

[†]Department of Biotechnology, Chemistry and Pharmacy, Università degli Studi di Siena, Via A. Moro 2, I-53100 Siena, Italy

[‡]Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States of America

[¶]Theoretical Physical Chemistry, Research Unit MolSys, Université de Liège, Allée du 6 Août, 11, 4000 Liège, Belgium

Abstract

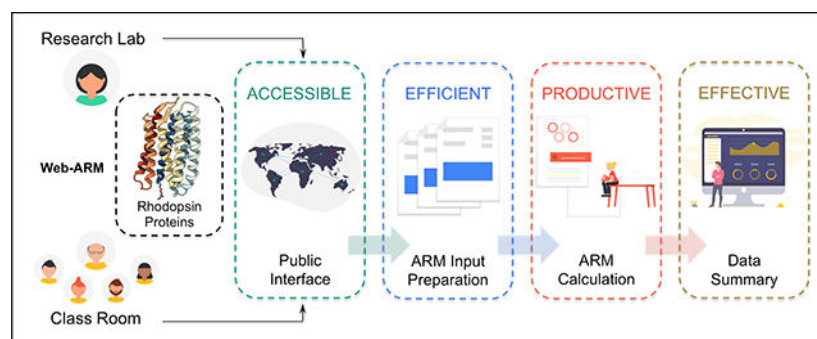
This article introduces Web-ARM, a specialized, on-line available, tool designed to build quantum mechanical/molecular mechanical models of rhodopsins, a widely spread family of light-responsive proteins. Web-ARM allows to rapidly build models of rhodopsins with a documented quality and to predict trends in UV-Vis absorption maximum wavelengths, based on their excitation energies computed at the CASPT2//CASSCF/Amber level of theory. Web-ARM builds upon the recently reported, python-based *a*-ARM protocol [*J. Chem. Theory Comput.*, **2019**, *15*, 3134–3152], and as such necessitates only a crystallographic structure or a comparative model in PDB format and a very basic knowledge of the studied rhodopsin system. The user-friendly web interface uses such input to generate congruous, gas-phase models of rhodopsins and, if requested, their mutants. We present two possible applications of Web-ARM, which showcase how the interface can be employed to assist both research and educational activities in fields at the interface between chemistry and biology. The first application shows how, through Web-ARM, research projects (*e.g.*, rhodopsin and rhodopsin mutant screening) can be carried out in significantly less time with respect to using the required computational photochemistry tools via a command line. The second application documents the use of Web-ARM in a real-life educational/training activity, through a hands-on experience illustrating the concepts of rhodopsin color tuning.

Graphical Abstract

molivuc@bgsu.edu; Luca.DeVico@unisi.it.

Supporting Information Available

Details on the features of the *a*-ARM QM/MM Models (section S1) and on the accepted .pdb files formats (section S2); default parameters for the molecular dynamics employed by Web-ARM (section S3, Table S1); details on the bR_{AT} and ASR_{AT} models building (section S4, Table S2); effect of doubling the simulated MD timings in bR_{AT}-Y185F and bR_{AT}-M145F (section S3, Table S3); complete energetics data for bR_{AT} and ASR_{AT} with N=1 (section S6.1, Table S4) and N=10 (section S6.2, Table S5); and a Tutorial on the use of the Web-ARM interface are given as Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org/>



1 Introduction

Rhodopsin proteins (from now on simply called rhodopsins) constitute a group of transmembrane, light-responsive proteins widely distributed among animals and microorganisms.^{1,2} The photoexcitation of the retinal protonated Schiff base chromophore of rhodopsins triggers various biological functions, such as vision in vertebrates and invertebrates and ion-pumping, ion-gating and light-sensing in unicellular algae, archaea and eubacteria.³ In recent years, it has been shown that basic hybrid quantum mechanical/molecular mechanical (QM/MM) models based on an *ab initio* multi-configurational QM treatment, are capable of reproducing both spectroscopic and photochemical properties of rhodopsins^{4,5} and their mutants⁶ with semi-quantitative accuracy (*e.g.*, with a few kcal mol⁻¹ / 1 – 2 tenths of eV error in vertical excitation energy prediction). Such body of results, provided the motivation for the development of an Automatic Rhodopsin Modeling (ARM) protocol,⁷ namely, an automated protocol for the fast and congruous construction of standardized gas-phase QM/MM models of rhodopsin-like proteins. The sought after aim was to facilitate the systematic studies of large sets of rhodopsins as demanded, for instance, by chemists and biologists involved in evolutionary and protein engineering studies. Indeed, ARM models have already been successfully applied to the study of spectroscopic properties of various rhodopsins.^{7–10}

The first version of ARM⁷ was, essentially, a Bash shell script, interfacing a set of publicly available programs, which took as input either a crystallographic structure or a previously generated comparative model in PDB format. The quality of ARM models (*i.e.*, the QM/MM models generated with ARM) was validated by computing, semi-automatically (as few user-dependent manipulation steps were required) the absorption vertical excitation energy (E_{S1-S0}) for a set of more than 30 phylogenetically distant rhodopsins whose “experimental” E_{S1-S0} , assumed to equal the energy associated to the wavelength (λ_{max}^a) corresponding to the maximum of the absorption band, were reproduced with a maximum error of 3.0 kcal mol⁻¹ (0.13 eV).^{7–10}

Recently, some of the authors have reported on an updated, python-based version of the ARM protocol, called *a*-ARM,¹¹ capable to automatically, rather than semi-automatically, generate rhodopsin QM/MM models. In fact, *a*-ARM incorporates the original ARM protocol but automates a number of its, otherwise user-dependent, construction steps such as the definition of the retinal chromophore cavity, the assignment of the protonation states of

the ionizable residues, the neutralization of the rhodopsin protein with external counterions, and finally the generation of single or multiple mutations. The *a*-ARM protocol, employing default parameters, proved to be capable of reproducing the E_{S1-S0} values for 77% (30/39) of the rhodopsins in the benchmark set, which included both wild type and mutant rhodopsins, within the previously mentioned error. However, *a*-ARM still remains a command line code directed to users familiar with the Linux environment since (i) it has no user-friendly interface and (ii) its set-up requires the installation of several computational packages in one's own local computational facility. These points can significantly slow down or make cumbersome the construction and application of *a*-ARM models, except perhaps for researchers familiar with file manipulation and access to an information technology service and with a computational chemist background. Thus, a user friendly interface to *a*-ARM would provide access to a fast and automated construction and analysis of rhodopsin QM/MM models to an interdisciplinary community.

Following the general interest in accessing complex computational chemistry tools through the web,^{12–25} here we report on the design, implementation and testing of Web-ARM, a web-accessible interface (web-arm.org) built upon *a*-ARM, and intended as a user-friendly and fast (with respect to the command-line version) builder of congruous and reproducible gas-phase QM/MM models (*i.e.*, *a*-ARM models, see also section S1 in the Supporting Information for a description of the term gas-phase) of rhodopsins. Briefly, such gas-phase model refer to a globally uncharged trans-membrane protein where external Cl^- and/or Na^+ counterions are properly placed in both the intracellular (IS) and extracellular (OS) protein surfaces to approximately mimic the effect of the water-solvation on the internal chromophore. We recommend Web-ARM users to familiarize with previous literature^{7,11} on *a*-ARM, and QM/MM-based models generation in general, prior to the use of the interface.

Web-ARM avoids portability issues, since the whole software resides on a remote machine, and thus there is no need for the user to install anything other than a standard, up-to-date, web browser. In practice, a Web-ARM user will be able to generate (with limitations, see later), a set of *a*-ARM models whose building steps are taken care of on the server side by a cron-daemon and leveraging a dedicated computational facility. Finally, Web-ARM is capable to perform “job-farming” of *a*-ARM model construction (*i.e.*, the parallel and independent constructions of many ARM models) ultimately predicting trends in structural and spectroscopic properties of rhodopsins in a, substantially, high-throughput manner.

In sections 2.1.1–2.1.4 we provide a detailed description of the Web-ARM availability and “architecture” including the input preparation for the QM/MM model building, sequence mutation, the actual QM/MM model building and, finally, monitoring. In sections 2.2.1 and 2.2.2 we give an introduction to two rhodopsins employed in our own benchmark studies.

In section 3.1 we benchmark Web-ARM at the research level, by simulating the search for microbial rhodopsin mutants displaying a blue-shifted absorption with respect to the wild-type (WT). In section 3.2 we, instead, present an application of Web-ARM in an undergraduate educational research activity, actually carried out in our laboratory by two of the authors. Here the students focused on a different microbial rhodopsin to investigate the

color-tuning effect of a larger set of single mutations and, through the inspection of the results, learn about some of the basic principles of color modulation in rhodopsins.

2 Methods

2.1 Design and Implementation of Web-ARM

The Web-ARM interface is available at the address web-arm.org (Figure 1). Web-ARM is written in Python 3, and makes use of the following software packages: PROPKA 3.1,²⁶ SCWRL4,²⁷ Modeller,²⁸ fpocket,²⁹ vmd,³⁰ and pdb2pqr.^{12,31} At the home page level (Figure 1), Web-ARM allows to perform 4 actions, each corresponding to the 4 phases described in subsections 2.1.1 – 2.1.4. Figure 2 depicts the parts of each phase and how such parts are interconnected. Registered users are allowed to build as many concurrent models as wished (default 10, see later). However, given the current limited resources, guest users are presently allowed to build only one *a*-ASR model at a time on the developer's dedicated resources. More specifically, as described in the following sections, one single QM/MM model is enough to perform basic analysis useful, for instance, to complement an educational project. Moreover, a dedicated guest user can still construct the same model multiple times providing a set of different (*i.e.*, uncorrelated) models which can then be employed to achieve statistically more robust results in terms, for instance, of structure and E_{S1-S0} values. A Tutorial, along with working examples, is available as Supporting Material and can also be downloaded from the interface itself. The Tutorial also reports on possible errors or issues in the execution of the interface, and how to solve them.

2.1.1 Preparation—Phase 1 (Figure 2, left column) is based on the *a*-ARM protocol,¹¹ and includes the sequential actions/steps a-h. (a) The interface requests a file containing a rhodopsin structure, which has to be uploaded by the user in pdb format, or downloaded from the RCSB database³² by issuing the corresponding pdb ID (see section S2 in the Supporting Information for further details on the capabilities of the interface in handling pdb files). (b) If the pdb file contains more than one chain, the interface asks the user to choose which chain to use. (c) If the provided crystal structure contains various rotamers for one or more residues, the user is requested to choose the rotamer to be used (the relative occupancy factor of each rotamer is displayed). (d) While Web-ARM recognizes automatically the lysine residue linked to the chromophore, the interface asks the user to select the main counterion (MC) among the two negative residues closest to the linker, or any residue. (e) If available, Web-ARM suggests the user to employ the crystallographic pH as default value to compute residues charges. The physiological pH of 7.4 is suggested if no crystallographic value is present, but the users can impose a different pH value. (f) The interface uses PROPKA to retrieve the pKa and buried percentage values of all ionizable residues, and based on such values and the selected pH, determine the protonation state for each residue. Again, the user has the possibility to overwrite such choice (*e.g.* when experimental data points to a different choice). On the one hand, as explained in Ref. 11, the *a*-ARM method, being based on the PropKa method, features a certain percentage of failure in predicting the protonation state of ionizable residues. We refer the reader to section 2.2.3 of Ref. 11 for a complete discussion on how to, in case, choose a different protonation state. On the other hand, notice that *a*-ARM generated QM/MM models have already been used as initial guess

for subsequent sophisticated modelling, exploring rhodopsin protein protonation states.³³ Such application demonstrates the potential use of our interface for preparing, for instance, pH-constant dynamics calculations. (g) The interface decides number, relative position (extracellular or intracellular) and coordinates (executing PUTION)⁷ of the counterions necessary to neutralize the model.¹¹ (h) The interface employs the program fpocket to locate the residues forming the cavity around the retinal chromophore. This step is skipped in case the structure has been created as part of a mutants set (Phase 2d). (i) The Phase 1 output is an input file for the generation of an *a*-ASR model (in pdb format) and a cavity file, which lists the cavity residues. The user can save these files, together with the log file of all Phase 1 activities and a structure file in PyMol³⁴ format useful for visually inspecting the output. Figure 3 was produced using such PyMol-formatted file. The input and cavity files can be transferred directly to Phase 2b or Phase 3b, to set up mutations or submit the calculation that generates the corresponding *a*-ASR model, respectively.

2.1.2 Mutation—Phase 2 (Figure 2, right column, top) is optional, but we expect to be frequently used. It includes the sequential steps a-d. (a) If the user has previously prepared input and cavity files (see above), it is possible to upload them here. (b) The user has the possibility to insert single, double or triple mutations repeatedly, which will, ultimately, allow for the preparation of an entire set of mutants (arguably, the major strengths of the interface). The mutant structures are generated using SCRWL4, which requires a sequence file (.pir file, created with Modeller) for specifying the new sequence, as described in section 2.2.5 of Pedraza-González et al.¹¹ (c) The original input and cavity files can be automatically transferred to Phase 3b, and associated to the *a*-ASR model of the WT protein. (d) Each generated mutant can be transferred to Phase 1d (and later skipping Phase 1h). Transferring from either Phase 2c or Phase 2d opens a new browser page, where the user can perform the necessary tasks on each model, WT or mutant.

As mentioned in Phase 2b, the Web-ARM interface manages mutations through the *a*-ARM protocol.¹¹ Accordingly, the interface assumes to start from a WT rhodopsin, selects the residues of the cavity, and performs mutations exclusively on the chromophore cavity residues (the user should ensure that the residues to be replaced are part of the computed cavity). Otherwise, it is also possible to simply download the pdb file generated by Phase 2, and upload it directly in Phase 1a.

Model consistency between the WT and mutants is unavoidable when comparing the members of a set of mutants. Therefore, the workflow must ensure that the WT and mutant structures share the same rotamers (if present in the original pdb file. See Phase 1c) and the same number and type of cavity residues. However, due to the possibility that different mutants include, due to a change in pKa value determined by *a*-ARM, the same ionizable residue in different ionization states, Phase 1f and Phase 1g have to be repeated for each mutant, so as to re-compute the appropriate number of counter ions, and their location. Finally, while the interface is capable of retrieving the crystallization pH from a pdb file from the RCSB data bank, such information is lost for subsequent calculations. The user should ensure the consistent use of the same pH value for WT and all mutants.

Notice that the SCWRL4 software is used just to provide, in the preparation phase, an initial geometry for the side-chain of the mutated residue. As presented below in Section 2.1.3, during the calculations phase, the geometry of the new modelled side-chain as well as the side-chain of its neighbor amino acids (belonging to the chromophore cavity), are readjusted during the molecular dynamics phase. Therefore, possible conformational changes in the residues of the chromophore cavity as well as in the rPSB, due to steric/electrostatic effects induced by their interaction with the mutated residue, are considered. For these reasons, we stress that under the philosophy of *a*-ARM only mutations of residues that belong to the chromophore cavity sub-system are allowed.

2.1.3 Calculation—Phase 3 (Figure 2, right column, middle). After input preparation (Phase 1), the Web-ARM interface performs automatically all steps necessary to construct a relaxed QM/MM model (*i.e.*, the final *a*-ARM model) by using the *a*-ARM protocol. As mentioned above and detailed in Melaccio et al.,⁷ such protocol is made of Bash shell scripts performing a series of file format manipulations, followed by calculations with different software packages. Web-ARM is able to run all the scripts autonomously through steps a-f, thus decreasing the user's workload and increase productivity. (a) If the user has previously prepared a *a*-ARM input and cavity files in Phase 1, it is possible to upload them in this phase. (b) The user is asked to provide a valid email address, which will be used to recognize registered users (see Section 2.1). Once the calculation is completed, the user will be alerted through email (please note, this feature is still not active at the time of writing). (c) Following the *a*-ARM protocol, the users chooses the type of retinal (all-*trans* with charge +1, 11-*cis* with charge +1, 13-*cis* with charge +1, all-*trans* with charge 0) embedded in the prepared rhodopsin guess structure. (d) The interface then asks the user to accept the default *a*-ARM parameters (Table S1) and go to Phase 3f or (e) provide a set of personalized parameters. (f) Finally, the interface presents one or more available queues where it is possible to submit the calculation. Then, it confirms the job submission and provides a unique job ID, to be used in Phase 4a.

Web-ARM takes care of all *a*-ARM manipulations/computations through a Python 3 based driver residing on the offered computational resource. To achieve that, the interface creates, after the parameters set in Phase 3d or Phase 3e, a blueprint of commands to feed the protocol. Briefly, each QM/MM model building comprises 1 or more (default 10) MM-based molecular dynamics (MD) calculations performed using Gromacs 4.5.4,³⁵ followed by a QM/MM ground state geometry optimization at the CASSCF(12,12)/6-31G*/AMBER level of theory³⁶⁻³⁸ ultimately yielding 10 different QM/MM models of the same protein. Then finally, the 10 corresponding E_{S1-S0} values are computed at the 3-states state-average CASPT2(12,12) level of theory.^{39,40} Suitable level shiftings are employed during the CASSCF and CASPT2 calculations, so as to minimize the possibility of convergence failure via state mixing and intruder state problems.⁷ If required, a single representative QM/MM equilibrium structure is also provided as the structure whose excitation energy is closest to the average of the 10 computed excitation energies. All QM/MM calculations are performed using coupled Molcas 8.1⁴¹ and Tinker 6.3 packages.⁴²

2.1.4 Status Check—Phase 4 (Figure 2, right column, bottom) includes the checking for completion of the *a*-ARM calculation and the subsequent retrieval of the results. (a) In this phase the user should provide the job ID obtained in Phase 3f, which will prompt the interface to show the calculation status. (b) In case a job is still running, the user is informed on the advancement of each calculations associated to each MD run. (c) If instead the job is done, a final report is displayed and a .tgz file, containing the relevant results and the output files, is made available for downloading.

Calculations submitted through Web-ARM are expected to be finished in 1 – 1.5 days, depending on the number of jobs already present in the queue, and the number of jobs submitted through the interface itself. Users are encouraged to transfer, at the earliest possible, results of interest in a personal storage unit, for further study. No Web-ARM produced data is indefinitely retained on our facility, other than the results of all calculations here presented, which will be kept always available. The corresponding job IDs can be found in Tables S3, S4 and S5 in the Supporting Information, and assessed through Phase 4a at any time.

2.2 Model Systems

To illustrate the capabilities of Web-ARM, we present two applications involving bacteriorhodopsin (bR_{AT}; an *Archea* proton-pumping rhodopsin) and *Anabaena* sensory rhodopsin (ASR_{AT}; an eubacterial photochromic rhodopsin). The first, which is research oriented, focuses on the use of the interface as a tool for the prediction of absorption data (sections 2.2.1 and 3.1), whereas the second uses the interface in a training activity focusing on the relationship between structure and spectroscopy in rhodopsins and in proteins in general (sections 2.2.2 and 3.2). In the present section we describe the building of the bR_{AT} and ASR_{AT} WT structures (bR_{AT}-WT and ASR_{AT}-WT respectively) also performed with Web-ARM.

2.2.1 Bacteriorhodopsin Model—As also detailed in the Tutorial found in the Supporting Information (the Tutorial is based on the calculations revised in the present section), the *a*-ARM model for bR_{AT}-WT was generated with the interface, employing the most recent crystallographic structure (PDB ID 6G7H)⁴³ as a template. The parameters employed for the generation of the model were: (i) an all-*trans* retinal chromophore, (ii) the chain A for modeling the protein environment, which was oriented by aligning the principal axis of helix VI in the direction of the *z*-axis, (iii) the default 23 chromophore cavity residues defined based on Voronoi tessellation and alpha spheres theory, and including the Lys-216 residue covalently linked to the chromophore, plus the Asp-85 main (MC) and Asp-212 secondary (SC) counter-ion residues, (iv) rotamers AAsp-104, ALeu-109 and ALeu-15 with occupancy number of 0.80, 0.54 and 0.57, respectively, (v) protonation states predicted at crystallographic pH 5.6 as neutral, for the ionizable residues Asp(96,115,212) and Glu(194) and, finally, (vi) with the inclusion of 1 Cl⁻ inner (IS) and 1 Cl⁻ outer (OS) counter-ions (see Table S2), with positions optimized with respect to an electrostatic potential grid constructed around each charged target residue (see Figure 3 for an exemplification of such subsystems division).¹¹

The obtained wild type *a*-ARM input for bR_{AT}-WT was then transferred to Phase 2b (see section 2.1.2) to prepare 5 single-point mutants (as reported in Table 1) of residues belonging to the chromophore cavity, following the same procedure as for the wild type.

2.2.2 Anabaena Sensory Rhodopsin Model—ASR_{AT}-WT was constructed by two undergraduate students (A.N.J. and T.D.R.) under the supervision of X.Y. and then used in the generation and analysis of the protein mutants. The building followed the same procedure described for bR_{AT}; but using as template the 1XIO⁴⁴ crystallographic structure. In addition, in this case was necessary to generate a customized cavity in order to include 5 residues of potential interest to the study of color tuning. Accordingly, the parameters employed for the generation of this model were (i) an all-*trans* retinal chromophore found as B conformer, and previously indicated as ASR_{AT}-1 in Ref. 11, (ii) the unique chain A for the environment protein, (iii) a chromophore cavity customized by adding the residues Tyr-11, Cys-134, Cys-137, Cys-203 and Pro-206 to the list of 23 default residues found by fpocket and including the Lys-210 residue covalently linked to the chromophore plus the Asp-75 MC residue, (iv) conformer ALys-310 with occupancy number 0.50 (see Section 3.1 in Ref. 11), (v) protonation states predicted at crystallographic pH 5.6 as neutral for the ionizable residues Asp(198, 217), Glu(36) and His(8); and, finally, (vi) the inclusion of 7 Cl⁻ inner (IS) and 1 Na⁺ outer (OS) counter-ions (see Table S2).

Based on such wild type model and through Web-ARM (Phase 2, section 2.1.2), A.N.J. and T.D.R. generated and analyzed two sets of 10 ASR mutants each (ANJ Set and TDR Set) respectively, for a total of 20 mutants as reported in Table 2.

3 Results and Discussion

As anticipated above we demonstrate the Web-ARM capabilities by reporting on two different applications. Firstly (Section 3.1), we show how Web-ARM can assist a researcher in the computational screening of rhodopsin mutants. Accordingly, Web-ARM is used to generate a set of 5 mutant models of bR_{AT} and compute the corresponding E_{S1-S0} trend which is then validated by comparison with the experimental trend. The “target” is to discover mutations capable to shift the absorption maximum to the blue. Secondly (Section 3.2), we test Web-ARM as a tool for education/training of non-expert computational chemists. Accordingly, we asked two students to build and analyze 20 QM/MM mutant models and describe and analyze the relationship between the type of replacement (location with respect to the retinal chromophore and change in the side chain charges) and the E_{S1-S0} variation in ASR_{AT}. In this way, the trainee shall begin to comprehend the mechanism of color-tuning at the molecular level. This could eventually be integrated by briefings providing information on different color-tuning mechanisms, which can then be verified through Web-ARM. We encourage the curious reader to employ Web-ARM to explore possible color-tuning mechanisms, by, *e.g.*, introducing mutations, as shown here.

3.1 Web-ARM as a Research Tool

As detailed below we use the interface to predict the maximum absorption wavelengths (λ_{max}^a) of wild type bR_{AT} (bR_{AT}-WT), along with that of 5 mutants. See Table 1. Through

this small benchmark set we intend to illustrate the capabilities of the interface itself to provide absorption data which result in reasonable agreement with experimental trends. The $\text{bR}_{AT}\text{-WT}$ model is produced following the description of section 3.2.6 of Pedraza-González et al.,¹¹ with regard to the choice of main counter-ion and, correspondingly, protonation of ionizable residues, as also detailed above in section 2.2.1. The $\text{bR}_{AT}\text{-WT}$ model was then transferred to Phase 2b (section 2.1.2), to introduce the selected single point mutations, each then transferred to Phase 1b (section 2.1.1). Finally, the generated models were transferred to Phase 3b (section 2.1.3), and the obtained results are shown in the right columns of Table 1 and in Figure 4. We now analyze the obtained results. First, we evaluated the quality of the $\text{bR}_{AT}\text{-WT}$ QM/MM model in terms of E_{S1-S0} . As shown in Table 1, the result is in line with that previously reported^{7,11} when considering the documented blue-shift error of $+3.0 \text{ kcal mol}^{-1}$ (0.13 eV). It is also in line with experimental data (the observed absorption wavelengths of bR_{AT} and five of its mutants are given in the left columns of Table 1).

The small blue shift deviation of the calculated E_{S1-S0} with respect to the experimental value suggests that the generated QM/MM models are suitable for the screening of bR_{AT} -mutants. Therefore, from now on, we consider the calculated E_{S1-S0} of $50.7 \text{ kcal mol}^{-1}$ (2.20 eV) as a baseline for analyzing the hypsochromic effect induced for each of the 5 selected point mutations. For each mutation, the differences between the calculated E_{S1-S0} and its corresponding experimental value is reported in the left most column of Table 1. All mutants are blue shifted with respect to $\text{bR}_{AT}\text{-WT}$ and feature a difference with respect to experimental data in the $0.1\text{--}1.9 \text{ kcal mol}^{-1}$ (0.01–0.08 eV) range. Considering the linear regression line, we can conclude that both WT and mutants are blue-shifted of no more than $3.0 \text{ kcal mol}^{-1}$ (0.13 eV) with respect to the experimental data, as previously reported (less than $3.0 \text{ kcal mol}^{-1}$ error was found for ca. 77% of cases using large sets of rhodopsins and their variants).^{7,11}

Among the set, the two mutants with larger deviation between computed and experimental absorption energies are $\text{bR}_{AT}\text{-Y185F}$ and $\text{bR}_{AT}\text{-M145F}$, that are two mutants where the mutated residue has a large side chain. We could attribute the larger deviation to a non-optimal conformation of the mutated residue, with respect to the original residue $\text{bR}_{AT}\text{-WT}$, possibly due to the basic manipulation used for the construction of the mutant model. In fact, every mutation is generated based on a conformational search algorithm (as implemented in SCWRL4),²⁷ followed by the choice of the conformer with highest score. There is no certainty that such conformer is the most stable for the given protein. In order, to test the robustness of Web-ARM mutant modeling, without using any manual intervention and/or visual inspection, we re-computed the two aforementioned mutants, doubling the time allocated for the MD simulations in Phase 3e, with respect to the default values of Table S1. Table S3 shows that, indeed, longer MD timings have an effect on the obtained results. However, only a small improvement on the excitation energy of $\text{bR}_{AT}\text{-M145F}$, and a reduction of the standard deviation for $\text{bR}_{AT}\text{-Y185F}$ were found. While these results confirm the possibility that the obtained mutant structures could be improved, the MD default timings (Table S1) still represent a more than valid compromise, as previously noticed.⁷ Thus, the results reveal that a method more effective than the α -ARM default one is necessary to improve the prediction of the residue conformation. One strategy has been

recently proposed as the ARM+ASEC algorithm.⁵⁰ Future versions of the interface will include the possibility to perform also such kind of calculations, or in general improve on the mutants generation algorithm.

3.2 Web-ARM as a Training/Educational Tool

Here we present the work performed by two undergraduate co-authors who employed Web-ARM to generate two separate sets of ASR_{AT} mutants. The mutants were generated starting from the ASR_{AT}-WT model (also produced with the interface, see section 2.2.2) that reproduces the available experimental absorption data (550 nm, 52.0 kcal mol⁻¹, 2.25 eV),⁴⁴ with a $\Delta\Delta E_{S1-S0}^{Exp.}$ of just +0.4 kcal mol⁻¹ (0.02 eV). Each set was assigned to a student who had only a basic knowledge of the rhodopsin structure and of the QM/MM technology. In the following the two sets, made by 10 mutants each, are named ANJ and TDR sets after the student names. The scope was to simulate a search for ASR variants showing red or blue-shifted absorption maxima. The two students followed the procedure reported in sections 2.1.1 – 2.1.3 to produce the results reported in the left columns of Table 2 and as blue bars in Figure 5 (N=1 data).

The difference, with respect to the research results presented in the previous section, is that for each mutant only 1 seed, rather than 10, was requested (Phase 3e) for the molecular dynamics calculation, resulting in a single QM/MM model and a single excitation energy evaluation. While this much faster procedure produces excitation energies displaying larger discrepancies from those obtained with the default protocol, the educational/training target, *i.e.*, learning how specific amino acid replacements can shift the absorption maximum, remains valid. This can be further appreciated in Figure 4, comparing the results obtained with only 1 molecular dynamics (red dots) as opposed to those with the default 10 (blue dots). Nevertheless, using just 1 repetition represents a useful research tool to preliminary screen large numbers of rhodopsin variants. The so-obtained excitation energies trends hold only a qualitative value and thus, representative models should be re-evaluated with the standard, 10 repetitions, protocol. For this reason, the E_{S1-S0} change of the computed most blue- and red-shifted mutants of each set with respect to the computed E_{S1-S0} of ASR_{AT}-WT were re-evaluated, still using the Web-ARM interface, employing the default 10 repetitions. The corresponding results are reported in the right columns of Table 2 (N=10 data) and as yellow bars in Figure 5. As observed in Figure 5, in both the ANJ and TDR sets, the mutants expected to be blue- or red-shifted on the basis of the N=1 data, are confirmed as having computed blue- or red-shifted absorption maxima, with respect to ASR_{AT}-WT.

From both a research and/or training/educational perspective, the automatic building and spectroscopic characterization of mutants of rhodopsins offered by Web-ARM are rather obvious, when focusing on the way in which the cavity amino acids tune the color corresponding to the absorption maxima.^{8–10} In fact, as also mentioned above, such spectroscopic quantity is determined by the E_{S1-S0} value (see Figure 6B, green E_{S1-S0} gap). In turn, such value depends on both steric and electrostatic interaction between retinal chromophore and protein environment. Using the models produced by Web-ARM, researchers can learn about the effect controlling the E_{S1-S0} variations based on the amino acid side-chain replacement (*e.g.*, of a polar with an apolar residues) and the specific

position where the substitution occurs, close to the proton Schiff base (see Figure 6A, bottom, orange circle) or close to the β -ionone ring (see Figure 6A, top, violet circle).

For example, in the case of ASR_{AT}-L83N mutation (TDR set), the amino acid substitution occurs in the proton Schiff base region (Figure 7A, orange circle). This replacement involved the substitution of Leucine (L) (apolar residue) by a Asparagine (N) (a polar residue carrying a (-)O-CR=NH2(+) dipole). As a consequence, the energy of S₀ was stabilized by the negative pole (the (-)O-CR) which is located closer to the proton Schiff base than the positive pole (the CR-NH2(+)), causing an increase of the energy gap, translating in a blue-shift effect on the E_{S₁-S₀} with respect the WT model (see Figure 6B, blue E_{S₁-S₀} gap). On the other hand, and always in the same set, ASR_{AT}-C137S mutation presented the opposite effect. The substitution of Cysteine (C) by a Serine (S) (both polar residues) in the region close to the β -ionone ring (Figure 7C, violet circle), is showing a stabilization of the S₁ energy. Thus, the trainee notices how the stabilization is due to shorter distance between the electronegative atom (O(-)) in the mutated residue and the retinal chromophore (Figure 7, left) with respect to the same electronegative atom (S(-)) in the wild type (Figure 7B), and therefore a red-shift effect with respect the WT model is observed (see Figure 6B, red E_{S₁-S₀} gap).

The exercise described above shows that trainees are able to perform scientifically significant calculations using Web-ARM. The two undergraduates were able to search for variants of ASR with different E_{S₁-S₀}, and make a meaningful suggestion as to what to choose for further experimental testing. In order to substantiate their findings, ANJ and TDR produced 3D models of the WT and most significant mutants of ASR, as obtained from their work. The 3D printed sculptures in Figure 8 were produced starting from the pdb files produced by Web-ARM as those which computed excitation energy is closest to the N=10 average.

4 Conclusions

Web-ARM represents a new and effective tool for the automated production of QM/MM models of rhodopsins. As detailed above, its most remarkable features are:

- i. A promptly accessible and user-friendly interface to the *a*-ARM protocol for the automated building of sizable sets of *a*-ARM models. It is therefore likely that Web-ARM will be used in studies aimed to predict or rationalize trends in spectroscopic properties of different rhodopsins and, most importantly rhodopsin mutants.
- ii. Web-ARM represents, to the best of our knowledge, the first web-based tools for producing congruous (*i.e.*, standardized) QM/MM models of proteins employing a link-atom frontier strategy. In fact, the *a*-ARM code¹¹ ensure consistency of the generated *a*-ARM-models. It must be stressed that such models are suitable for the study of both ground and excited state properties.
- iii. The Web-ARM generated models can be employed as rapidly building “guess models” of documented quality for the generation of more realistic and complex QM/MM models including, for instance, other monomers (e.g. in the modeling

of multimeric rhodopsins), the solvent/membrane environment, an higher level of QM theory or a different force-field (*e.g.*, a polarizable force-fields).

Features i-iii can be exploited in research projects requiring the generation and study of entire sets of *a*-ARM-models. In fact, the underlying *a*-ARM protocol is exclusively geared towards speed of input preparation, with the extra option of performing such task with completely standardized choices, while the execution of the calculations necessary to build the models is left to the user, which leads to difficulties when many models have to be built and characterized, possibly simultaneously (*e.g.*, for phylogenetically related rhodopsins and for rhodopsin mutants). We have shown how the Web-ARM interface facilitates the executions of such tasks, also by removing the waiting times between completion of each stage of an *a*-ARM-model calculation and submission of the next, and actually speeding up the whole process. Accordingly, even a less experienced user can use the Web-ARM interface to prepare *a*-ARM inputs in batches, and submit the corresponding *a*-ARM calculations on our server, without any further concern. Also experienced users looking for a more detailed description of a rhodopsin and not interested in large sets of rhodopsins, can use Web-ARM to obtain, in a standardized manner, a starting QM/MM model, which can be later refined using their preferred methods, or simply used as guess structures for specific calculations, such as pH-constant dynamics (see Ref. 33) or for the construction of fully relaxed QM/MM models embedded in a biological membrane.

Above we have also shown how Web-ARM can be used in training activities at the border between chemistry and biology. High-quality ground or excited state QM/MM calculations of spectroscopic and photochemical properties of rhodopsins are not widely accessible. With Web-ARM, researchers and even students non-expert in computational chemistry, can perform QM/MM calculations without the need to install specialized software. Both junior researchers and trainees will thus be able to perform meaningful calculations focusing on the underlying research targets, methodological concepts and data analysis, while remaining confident that the calculations are internally consistent.

One further appealing feature of Web-ARM is that its standardized output allows a quick retrieval of different data (*e.g.*, detailed protein structure, average excitation energy, standard deviation of the computed results, etc.), which are then ready to be used in any kind of analysis. The Tutorial incorporated in the Supporting Information, which is also accessible from the web page, provides a further detailed example on the possible interface uses.

It is our hope that, Web-ARM will allow scientists from different backgrounds to efficiently study rhodopsins and their characteristics. Although our current computational resources allow 10 replicas only for registered users, it is our hope that, through future development, we will be able to extend our resources pool or develop technologies allowing users to use their own resources. For example, we envision the possibility of incorporating a collaborator resources into those available for Web-ARM, as well as acquiring *ad hoc* cloud-based resources.

As an outlook, future versions of Web-ARM will include the possibility of employing the QM/MM model obtained with the current implementation, for a subsequent complete population analysis and absorption spectrum simulation.⁵¹ Finally, current efforts towards

the improvement of the mutations routine are directed to replace the SCWRL4 (*i.e.*, a rotamer library) by a software based on comparative modelling. Future versions of the protocol will include this feature.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

L.P.-G., M.D.C.M., M.O. and L.D.V. acknowledge a MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) grant "Dipartimento di Eccellenza 2018 – 2022". M.O., A.N.J., T.D.R. and X.Y are grateful for grants NSF CHE-CLP-1710191 and NIH GM126627 01. M.O. and M.D.C.M. are grateful to the Università degli Studi di Siena for an "Assegno Premiale 2017". A.V. thanks the research project #T.0132.16 from Fonds National de la Recherche Scientifique (Belgium) for a post-doctoral fellowship.

References

- (1). Ernst OP; Lodowski DT; Elstner M; Hegemann P; Brown LS; Kandori H Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev* 2013, 114, 126–163. [PubMed: 24364740]
- (2). Spudich JL; Yang C-S; Jung K-H; Spudich EN Retinylidene Proteins: Structures and Functions from Archaea to Humans. *Annu. Rev. Cell Dev. Biol* 2000, 16, 365–392. [PubMed: 11031241]
- (3). Tahara S; Takeuchi S; Abe-Yoshizumi R; Inoue K; Ohtani H; Kandori H; Tahara T Origin of the Reactive and Nonreactive Excited States in the Primary Reaction of Rhodopsins: pH Dependence of Femtosecond Absorption of Light-Driven Sodium Ion Pump Rhodopsin KR2. *J. Phys. Chem. B* 2018, 122, 4784–4792. [PubMed: 29708342]
- (4). Gozem S; Schapiro I; Ferré N; Olivucci M The Molecular Mechanism of Thermal Noise in Rod Photoreceptors. *Science* 2012, 337, 1225–1228. [PubMed: 22955833]
- (5). Rinaldi S; Melaccio F; Gozem S; Fanelli F; Olivucci M Comparison of the Isomerization Mechanisms of Human Melanopsin and Invertebrate and Vertebrate Rhodopsins. *Proc. Natl. Acad. Sci. U.S.A* 2014, 111, 1714–1719. [PubMed: 24449866]
- (6). Melaccio F; Ferré N; Olivucci M Quantum Chemical Modeling of Rhodopsin Mutants Displaying Switchable Colors. *Phys. Chem. Chem. Phys* 2012, 14, 12485–12495. [PubMed: 22699180]
- (7). Melaccio F; Marín M. d. C.; Valentini A; Montisci F; Rinaldi S; Cherubini M; Yang X; Kato Y; Stenrup M; Orozco-González Y; Ferré N; Luk HL; Kandori H; Olivucci M Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *J. Chem. Theory Comput* 2016, 12, 6020–6034. [PubMed: 27779842]
- (8). Agathangelou D; Orozco-Gonzalez Y; Marín M. d. C.; Roy P; Brazard J; Kandori H; Jung K-H; Léonard J; Buckup T; Ferré N; Olivucci M; Haacke S Effect of Point Mutations on the Ultrafast Photo-isomerization of *Anabaena* Sensory Rhodopsin. *Faraday Discuss.* 2018, 207, 55–75. [PubMed: 29388996]
- (9). 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. [PubMed: 30532962]
- (10). Inoue K; Marín M. d. C.; Tomida S; Nakamura R; Nakajima Y; Olivucci M; Kandori H Red-shifting Mutation of Light-driven Sodium-pump Rhodopsin. *Nat. Commun* 2019, 10, 1993. [PubMed: 31040285]
- (11). Pedraza-González L; De Vico L; Marín M. d. C.; Fanelli F; Olivucci M a-ARM: Automatic Rhodopsin Modeling with Chromophore Cavity Generation, Ionization State Selection and External Counter-ion Placement. *J. Chem. Theory Comput* 2019, 15, 3134–3152. [PubMed: 30916955]

- (12). Dolinsky TJ; Czodrowski P; Li H; Nielsen JE; Jensen JH; Klebe G; Baker NA PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* 2007, 35, W522–W525. [PubMed: 17488841]
- (13). Hediger MR; Jensen JH; De Vico L BioFET-SIM Web Interface: Implementation and Two Applications. *PLoS ONE* 2012, 7, e45379. [PubMed: 23056201]
- (14). Van Dijk M; Wassenaar TA; Bonvin AM A Flexible, Grid-enabled Web Portal for GROMACS Molecular Dynamics Simulations. *J. Chem. Theory Comput* 2012, 8, 3463–3472. [PubMed: 26592996]
- (15). Khoury GA; Thompson JP; Smadbeck J; Kieslich CA; Floudas CA Force-field_PTM: Ab initio Charge and AMBER Forcefield Parameters for Frequently Occurring Post-translational Modifications. *J. Chem. Theory Comput* 2013, 9, 5653–5674. [PubMed: 24489522]
- (16). Sandal M; Duy TP; Cona M; Zung H; Carloni P; Musiani F; Giorgetti A GOMoDo: A GPCRs Online Modeling and Docking Webserver. *PLoS One* 2013, 8, e74092. [PubMed: 24058518]
- (17). Krüger J; Grunzke R; Gesing S; Breuers S; Brinkmann A; de la Garza L; Kohlbacher O; Kruse M; Nagel WE; Packschies L; Müller-Pfefferkorn R; Schäfer P; Schärfe C; Steinke T; Schlemmer T; Warzecha KD; Zink A; Herres-Pawlis S The MoSGrid Science Gateway – A Complete Solution for Molecular Simulations. *J. Chem. Theory Comput* 2014, 10, 2232–2245. [PubMed: 26580747]
- (18). Weber J; Achenbach J; Moser D; Proschak E VAMMPIRE-LORD: A Web Server for Straightforward Lead Optimization Using Matched Molecular Pairs. *J. Chem. Inf. Model* 2015, 55, 207–213. [PubMed: 25629725]
- (19). Chomenidis C; Drakakis G; Tsiliki G; Anagnostopoulou E; Valsamis A; Doganis P; Sopasakis P; Sarimveis H Jaqpot Quattro: A Novel Computational Web Platform for Modeling and Analysis in Nanoinformatics. *J. Chem. Inf. Model* 2017, 57, 2161–2172. [PubMed: 28812890]
- (20). Martínez-Rosell G; Giorgino T; De Fabritiis G PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. *J. Chem. Inf. Model* 2017, 57, 1511–1516. [PubMed: 28594549]
- (21). Doerr S; Giorgino T; Martínez-Rosell G; Damas JM; De Fabritiis G High-throughput Automated Preparation and Simulation of Membrane Proteins with HTMD. *J. Chem. Theory Comput* 2017, 13, 4003–4011. [PubMed: 28723224]
- (22). Zhou H; Cao H; Skolnick J FINDSITEcomb2.0: A New Approach for Virtual Ligand Screening of Proteins and Virtual Target Screening of Biomolecules. *J. Chem. Inf. Model* 2018, 58, 2343–2354. [PubMed: 30278128]
- (23). Lee J; Patel DS; Stähle J; Park S-J; Kern NR; Kim S; Lee J; Cheng X; Valvano MA; Holst O; Knirel YA; Qi Y; Jo S; Klauda JB; Widmalm G; Im W CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput* 2018, 15, 775–786. [PubMed: 30525595]
- (24). Schmit JD; Kariyawasam NL; Needham V; Smith PE SLTCAP: A simple method for calculating the number of ions needed for MD simulation. *J. Chem. Theory Comput* 2018, 14, 1823–1827. [PubMed: 29506385]
- (25). Franco-Ulloa S; Riccardi L; Rimembrana F; Pini M; De Vivo M NanoModeler: A Webserver for Molecular Simulations and Engineering of Nanoparticles. *J. Chem. Theory Comput* 2019, 15, 2022–2032. [PubMed: 30758952]
- (26). Olsson MH; Søndergaard CR; Rostkowski M; Jensen JH PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput* 2011, 7, 525–537. [PubMed: 26596171]
- (27). Krivov GG; Shapovalov MV; Dunbrack RL Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins: Struct., Funct., Bioinf* 2009, 77, 778–795.
- (28). Eswar N; Webb B; Marti-Renom MA; Madhusudhan M; Eramian D; Shen M.-y.; Pieper U; Sali A Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinf* 15, 5.6.1–5.6.30.
- (29). Le Guilloux V; Schmidtke P; Tuffery P Fpocket: An Open Source Platform for Ligand Pocket Detection. *Bioinformatics* 2009, 10, 168–179. [PubMed: 19486540]

- (30). Humphrey W; Dalke A; Schulten K VMD: Visual Molecular Dynamics. *J. Mol. Graph* 1996, 14, 33–38. [PubMed: 8744570]
- (31). Dolinsky TJ; Nielsen JE; McCammon JA; Baker NA PDB2PQR: An Automated Pipeline for the Setup of Poisson–Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* 2004, 32, W66–W67.
- (32). Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE The Protein Data Bank. *Nucleic Acids Res.* 2000, 28, 235–242. [PubMed: 10592235]
- (33). 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. [PubMed: 31264415]
- (34). Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.2r3pre. 2010.
- (35). Pronk S; Páll S; Schulz R; Larsson P; Bjelkmar P; Apostolov R; Shirts MR; Smith JC; Kasson PM; Van Der Spoel D; Linddahl E GROMACS 4.5: a High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* 2013, 29, 845–854. [PubMed: 23407358]
- (36). Roos BO; Taylor PR; Siegbahn PE A Complete Active Space SCF Method (CASSCF) using a Density Matrix Formulated Super-CI Approach. *Chem. Phys* 1980, 48, 157–173.
- (37). Hariharan PC; Pople JA The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* 1973, 28, 213–222.
- (38). Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc* 1995, 117, 5179–5197.
- (39). Andersson K; Malmqvist P-Å; Roos BO; Sadlej AJ; Wolinski K Second-order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem* 1990, 94, 5483–5488.
- (40). Andersson K; Malmqvist P-Å; Roos BO Second-order Perturbation Theory with a Complete Active Space Self-consistent Field Function. *J. Chem. Phys* 1992, 96, 1218–1226.
- (41). Aquilante F; Autschbach J; Carlson RK; Chibotaru LF; Delcey MG; De Vico L; Fdez. Galván I; Ferré N; Frutos LM; Gagliardi L; Garavelli M; Giussani A; Hoyer CE; Li Manni G; Lischka H; Ma D; Malmqvist PÅ; Müller T; Nenov A; Olivucci M; Bondo Pedersen T; Peng D; Plasser F; Pritchard B; Reiher M; Rivalta I; Schapiro I; Segarra-Martí J; Stenrup M; Truhlar DG; Ungur L; Valentini A; Vancoillie S; Veryazov V; Vysotskiy VP; Weingart O; Zapata F; Lindh R Molcas 8: New Capabilities for Multiconfigurational Quantum Chemical Calculations Across the Periodic Table. *J. Comput. Chem.* 2016, 37, 506–541. [PubMed: 26561362]
- (42). Coutinho K; Georg HC; Fonseca TL; Ludwig V; Canuto S An Efficient Statistically Converged Average Configuration for Solvent Effects. *Chem. Phys. Lett* 2007, 437, 148–152.
- (43). Nogly P; Weinert T; James D; Carbajo S; Ozerov D; Furrer A; Gashi D; Borin V; Skopintsev P; Jaeger K; Nass K; Bath P; Bosman R; Koglin J; Seaberg M; Lane T; Kekilli D; Brünle S; Tanaka T; Wu W; Milne C; White T; Barty A; Weierstall U; Panneels V; Nango E; Iwata S; Hunter M; Schapiro I; Schertler G; Neutze R; Standfuss J Retinal Isomerization in Bacteriorhodopsin Captured by a Femtosecond X-ray Laser. *Science* 2018, 361, eaat0094. [PubMed: 29903883]
- (44). Vogeley L; Sineshchekov OA; Trivedi VD; Sasaki J; Spudich JL; Luecke H Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science* 2004, 306, 1390–1393. [PubMed: 15459346]
- (45). Richter H-T; Needleman R; Lanyi JK Perturbed Interaction Between Residues 85 and 204 in Tyr-185-Phe and Asp-85-Glu Bacteriorhodopsins. *Biophys. J* 1996, 71, 3392–3398. [PubMed: 8968608]
- (46). Perálvarez-Marín A; Bourdelande J-L; Querol E; Padrós E The Role of Proline Residues in the Dynamics of Transmembrane Helices: the Case of Bacteriorhodopsin. *Mol. Membr. Biol.* 2006, 23, 127–135. [PubMed: 16754356]
- (47). Ihara K; Amemiya T; Miyashita Y; Mukohata Y Met-145 is a Key Residue in the Dark Adaptation of Bacteriorhodopsin Homologs. *Biophys. J* 1994, 67, 1187–1191. [PubMed: 7811932]

- (48). Perálvarez-Marín A; Márquez M; Bourdelande J-L; Querol E; Padrós E Thr-90 Plays a Vital Role in the Structure and Function of Bacteriorhodopsin. *J. Biol. Chem* 2004, 279, 16403–16409. [PubMed: 14757760]
- (49). Yoshitsugu M; Yamada J; Kandori H Color-Changing Mutation in the E-F Loop of Proteorhodopsin. *Biochemistry* 2009, 48, 4324–4330. [PubMed: 19334675]
- (50). Orozco-Gonzalez Y; Manathunga M; Marín M. d. C.; Agathangelou D; Jung K-H; Melaccio F; Ferré N; Haacke S; Coutinho K; Canuto S; Olivucci M An Average Solvent Electrostatic Configuration Protocol for QM/MM Free Energy Optimization: Implementation and Application to Rhodopsin Systems. *J. Chem. Theory Comput* 2017, 13, 6391–6404. [PubMed: 29112449]
- (51). Gholami S; Pedraza-González L; Yang X; Granovsky AA; Ioffe IN; Olivucci M Multi-State Multi-Configuration Quantum Chemical Computation of the Two-Photon Absorption Spectra of Bovine Rhodopsin. *J. Phys. Chem. Lett* 2019, 10, 6293–6300. [PubMed: 31545053]

The screenshot shows the web-ARM interface in a browser window. The page title is "The Web-ARM Interface ver. 1.0". It includes a header with author information and a "Welcome to the ARM Web Interface! Please, make your choice." section. This section contains four main options, each with a button and a corresponding phase label on the right:

- Phase 1:** "Click here if you wish to prepare a pdb crystal structure for a subsequent ARM calculation." with a "Prepare" button.
- Phase 2:** "Click here if you already have pdb and cavity files and wish to create mutants." with a "Mutate" button.
- Phase 3:** "Click here if you have an already assembled pdb structure and cavity file and wish to start a new ARM calculation." with a "Compute" button.
- Phase 4:** "Click here to check the status of a previous ARM calculation (requires a Unique ID)." with a text input field and a "Check" button.

Below the main options, there is a link to a tutorial and a disclaimer about academic use. At the bottom, it lists the software programs used: PROPKA 3.1, SCWRL4, Modeller, fpocket, vmd, and pdb2pqr, each with a link and citation.

Figure 1:
Web-ARM Interface Home Page.

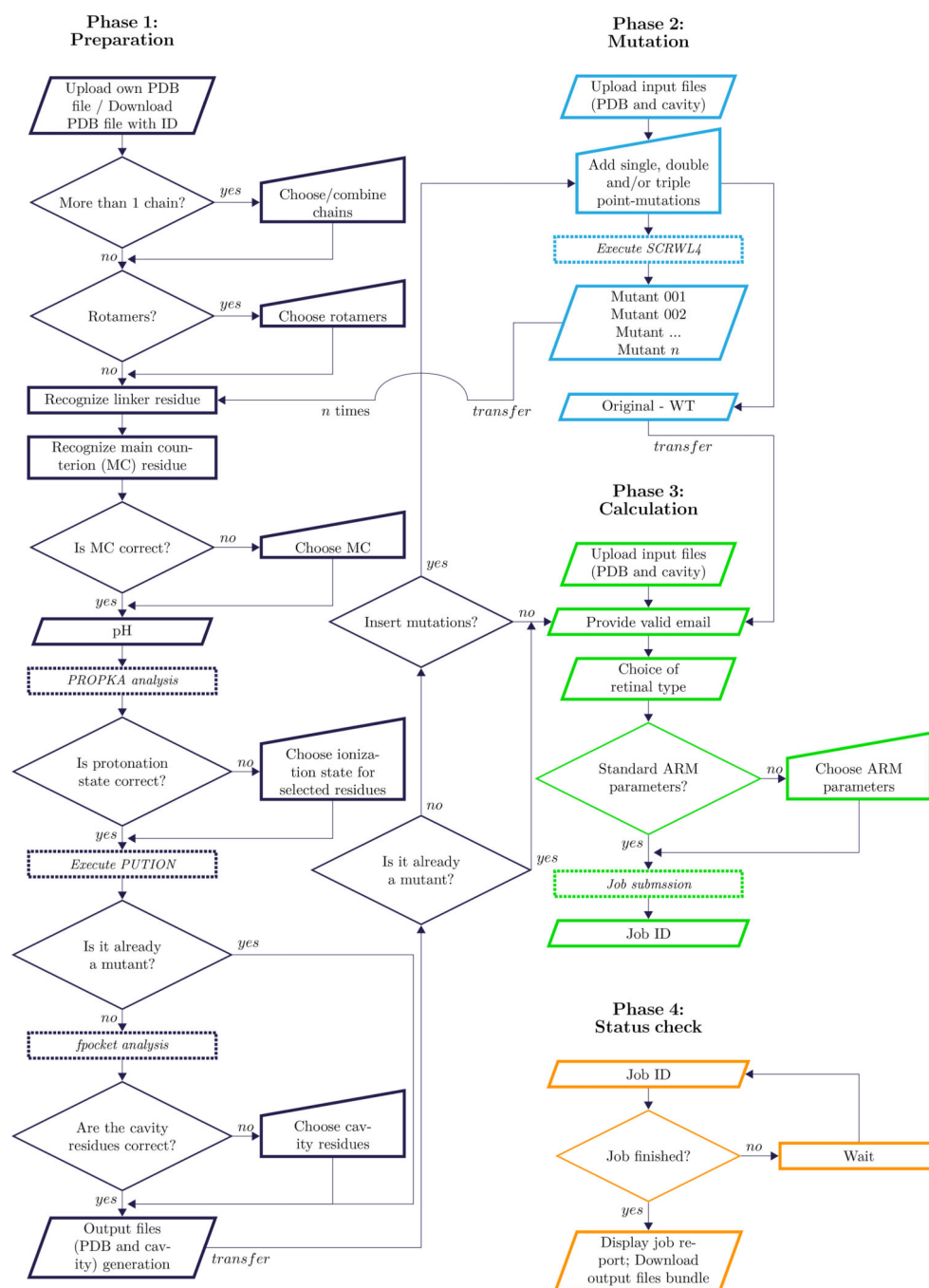


Figure 2: Overview of the web interface. The four different phases are highlighted in different colours.

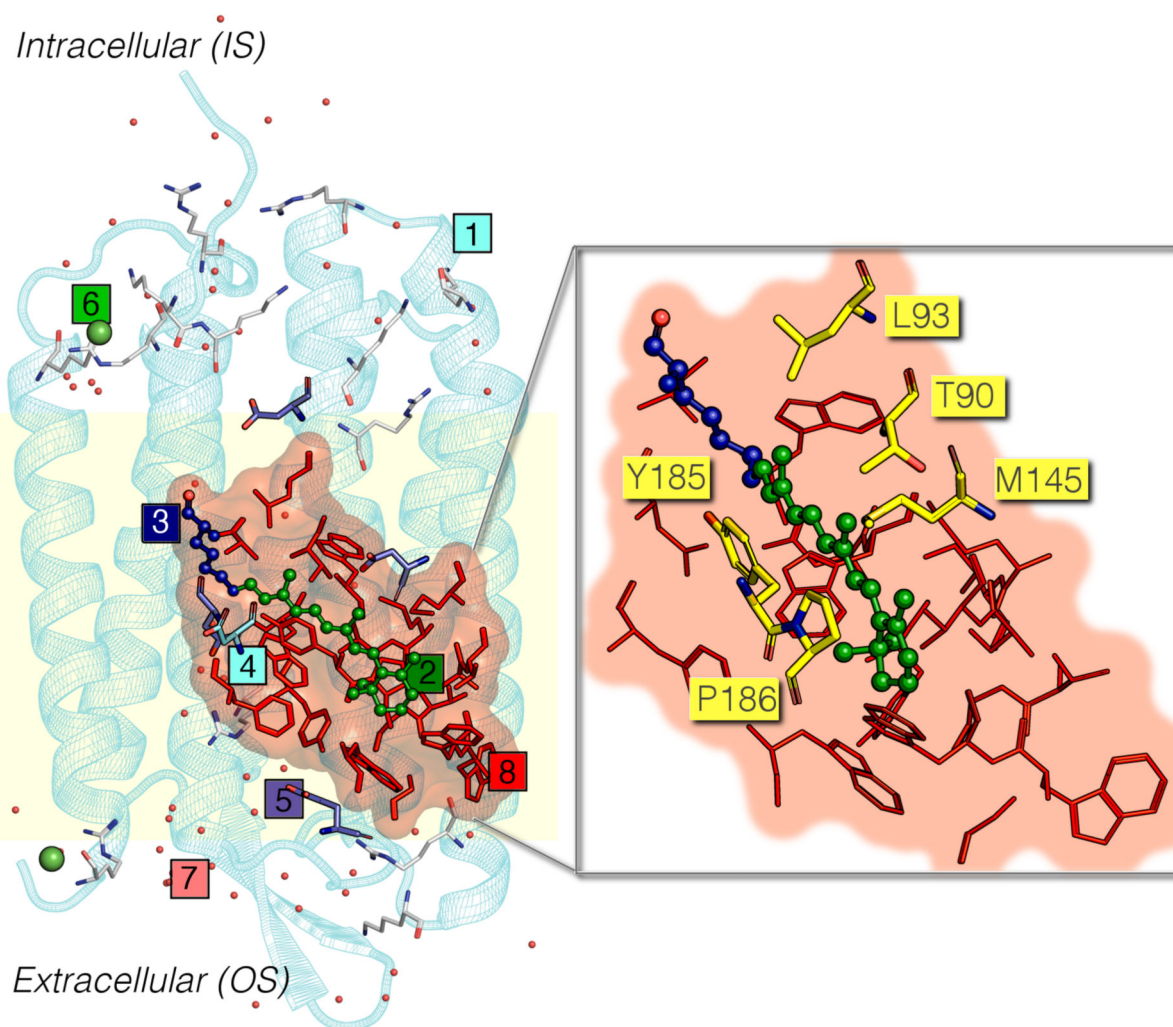


Figure 3: General scheme of an *a*-ARM input generated by Web-ARM for bacteriorhodopsin. The structure is composed of: (1) main chain (cyan cartoon), (2) chromophore rPSB (green ball-and-sticks), (3) Lys side-chain covalently linked to the chromophore (blue ball-and-sticks), (4) main counter-ion MC (cyan tubes), (5) protonated residues (violet tubes), (6) external Cl counter-ions (green balls), (7) water molecules (red balls) and the (8) residues of the chromophore cavity subsystem (red frames). Parts 1 and 6 form the environment subsystem. Parts 2 and 3 form the Lys-QM subsystem which includes the H-link atom located along the only bond connecting blue and green atoms. Parts 4 and 8 form the cavity subsystem. The water molecules (Part 7) may be part of the environment or cavity subsystems. The external charged residues are shown in frame representation. A brief description of the importance and use of each subsystem can be found in section S1 of the Supporting Information. Inset: Highlight of the cavity residues, with the target residues around the chromophore chosen for single mutants (Table 1) presented as yellow tubes. This figure was prepared using the PyMol-formatted³⁴ output file of the interface (Phase 1i, section 2.1.1).

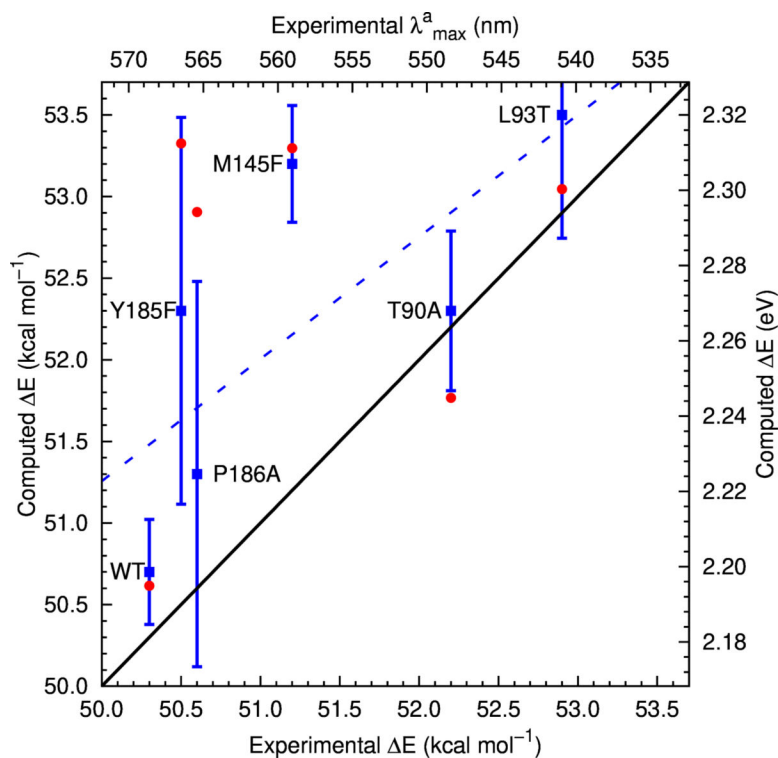


Figure 4: Bacteriorhodopsin wild type (WT) and mutants computed first vertical excitation energies (y axis) versus experimental data (x axis), in kcal mol⁻¹ (blue squares). Experimental absorption maxima λ_{max}^a in nm and computed excitation energies in eV are also shown on secondary axes. The error bars show the standard deviation of the computed data over 10 dynamic repetitions. The dashed blue line represents the linear regression of the computed data, while the black solid line represents the perfect correlation line between computed and experimental data. Red dots show the computed excitation energies using just 1 molecular dynamics run (namely, the first of each of the default 10).

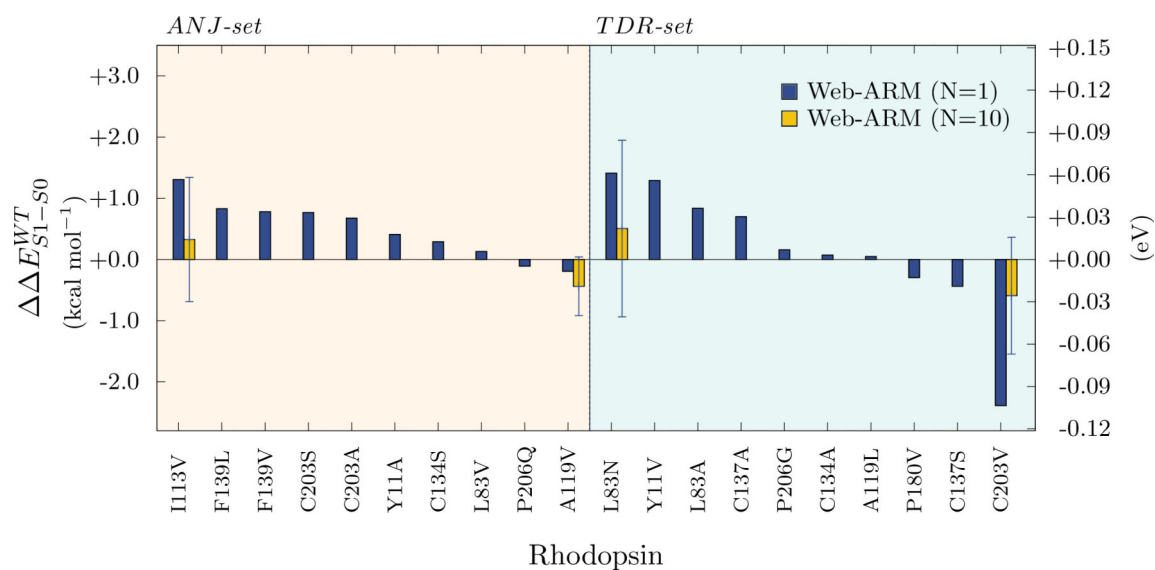


Figure 5: Differences between calculated vertical excitation energies (E_{S1-S0}) of ASR_{AT} WT and each of the mutants studied in both ANJ-set and TDR-set ($\Delta\Delta E_{S1-S0}^{WT}$). Blue bars represent values computed with one single seed (N=1), whereas yellow bars represent values computed with the standard 10 seeds (N=10). Data is presented in kcal mol⁻¹ (left vertical axis) and eV (right vertical axis).

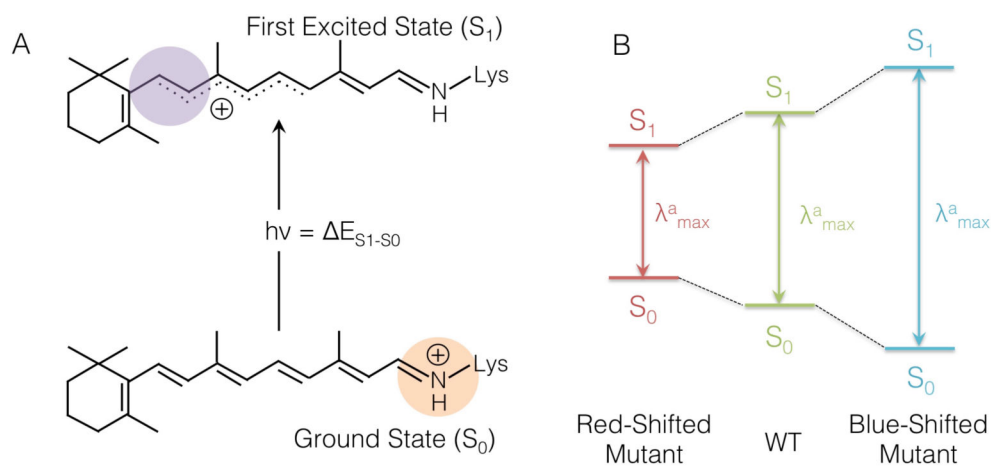


Figure 6: (A) Schematic illustration of the vertical excitation energy ($E_{S_1-S_0}$) of the all-*trans* rPSB chromophore, as embedded in ASR. The orange circle indicates the proton Schiff base region, while the violet circle indicates the β -ionone ring region. Upon photo-excitation, the main location of the positive charge of rPSB shifts from the orange to the violet region. (B) Schematic representation of the red- and blue-shift effects on the absorption maximum wavelength (λ^a_{max}). A red-shifted mutant can be obtained by destabilizing the ground and/or stabilizing the excited state of rPSB, with respect to the WT conditions. Alternatively, a blue-shifted mutant can be obtained by stabilizing the ground and/or destabilizing the excited state of rPSB.

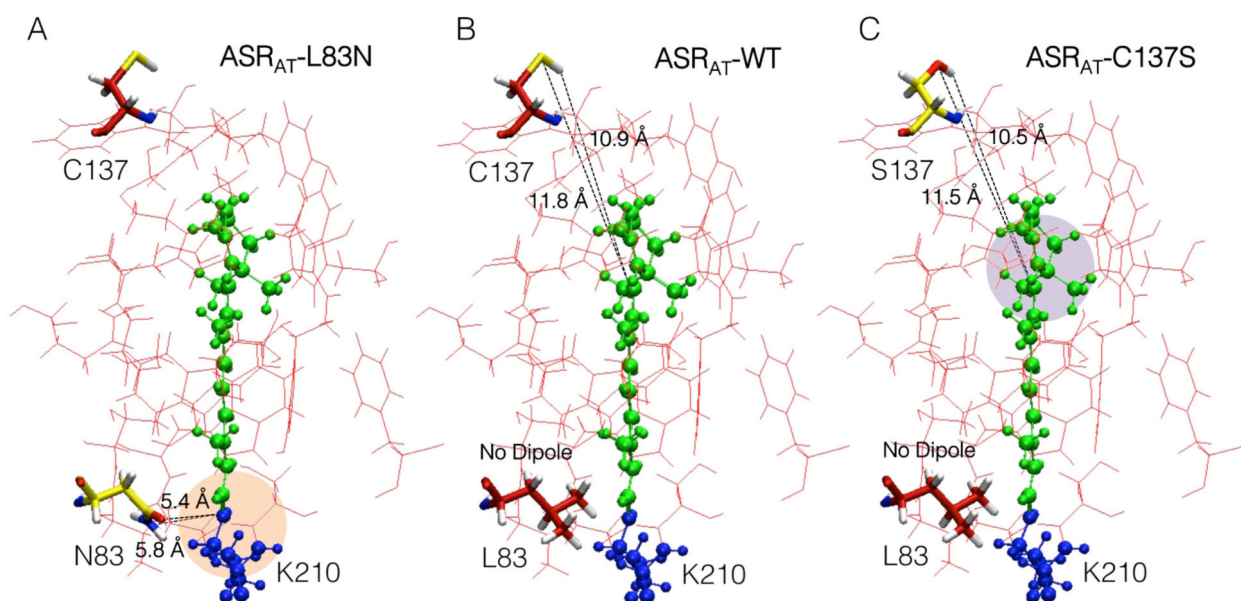


Figure 7:

Cavity residues of ASR_{AT}-WT and mutants Web-ARM models. (a) ASR_{AT}-L83N mutant model, found to have a blue-shifted absorption with respect to WT. (b) ASR_{AT}-WT model. (c) ASR_{AT}-C137S mutant model, found to be red-shifted with respect to WT. The cavity amino acids are shown in red lines. The variable cavity residues 83 and 137 are shown in tube representation in red for WT and in yellow for the mutants. Retinal chromophore is shown in green and lysine linker (K210) is shown in blue, both in ball-and-sticks representation. Violet and orange circles indicate the β -ionone ring and the proton Schiff base regions, respectively. Black dashed lines indicate the distances between point mutation residues and retinal chromophore.

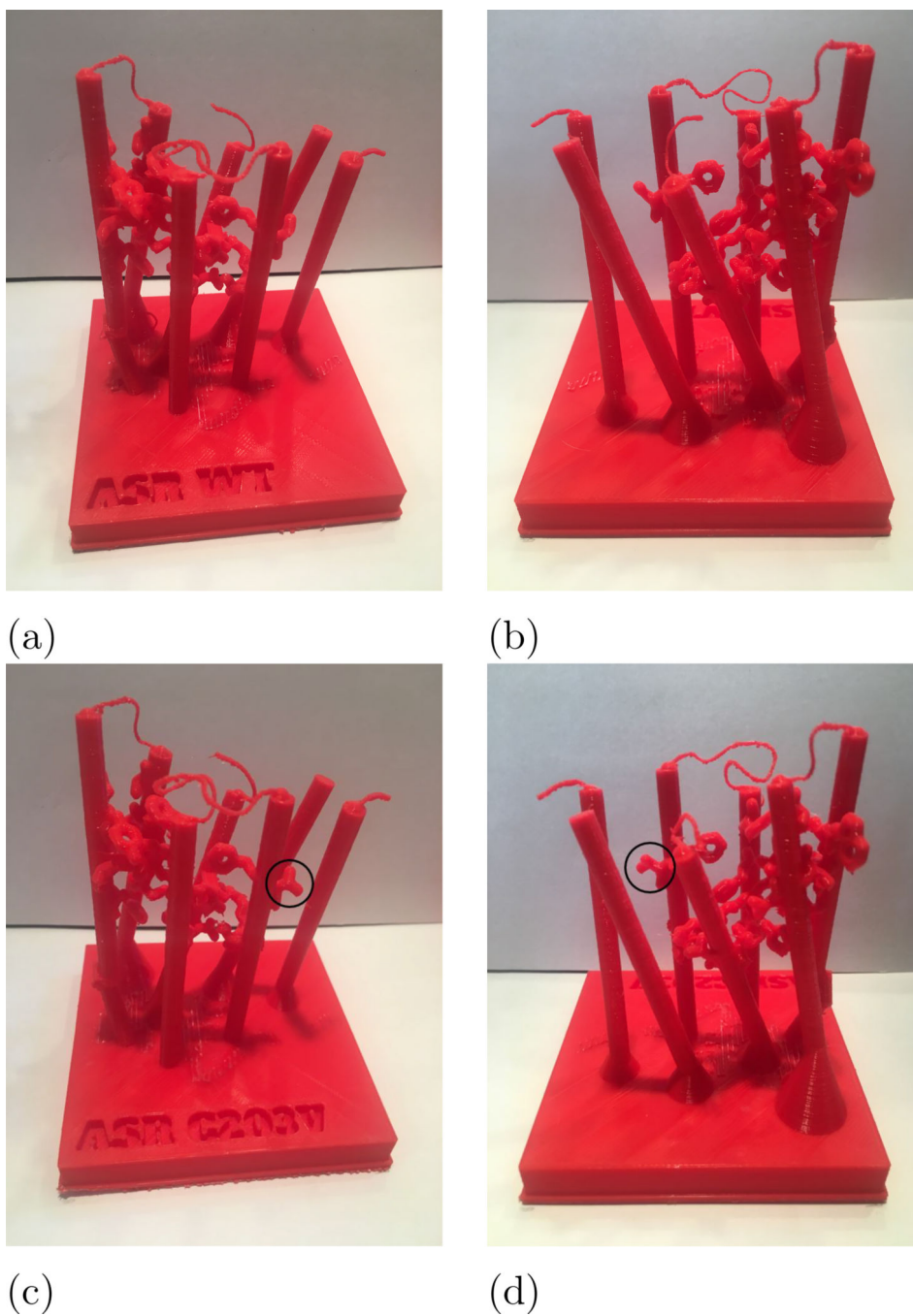


Figure 8: 3D printed sculptures of the ASR_{AT}-WT model (a and b, front and back view, respectively), as well as the ASR_{AT}-C203V model (c and d, front and back view, respectively), highlighting the mutated residue. The 3D sculptures were produced at facilities in BGSU starting from the pdb files of the closest to the average among the N=10 obtained structures, as provided by Web-ARM. Photos courtesy of TDR.

Table 1:

Experimental and computed first vertical excitation energies for studied Bacteriorhodopsin wild type (bR_{AT}-WT) and mutant models.

Code	Experimental			N=10 WARM ^a			f_{osc}	$\Delta\Delta E_{S1-S0}^{Exp}$ ^d (kcal mol ⁻¹)
	(nm)	E_{S0-S1} ^b (kcal mol ⁻¹)	(eV)	(nm) ^e	E_{S0-S1} ^c (kcal mol ⁻¹)	(eV)		
bR _{AT} -WT	568 ¹	50.3	2.18	564	50.7 (0.32)	2.20	1.43	+0.4
<i>bR Mutants</i>								
bR _{AT} -Y185F	566 ⁴⁵	50.5	2.19	547	52.3 (1.19)	2.27	1.28	+1.8
bR _{AT} -P186A	565 ⁴⁶	50.6	2.19	557	51.3 (1.18)	2.23	1.41	+0.7
bR _{AT} -M145F	558 ⁴⁷	51.2	2.22	537	53.2 (0.36)	2.31	1.29	+1.9
bR _{AT} -T90A	548 ⁴⁸	52.2	2.26	546	52.3 (0.49)	2.27	1.34	+0.1
bR _{AT} -L93T	540 ⁴⁹	52.9	2.29	534	53.5 (0.76)	2.32	1.28	+0.6

^aNumber of parallel dynamics repetitions.

^bExperimental excitation energy.

^cAverage computed excitation energy. Standard deviation in parenthesis.

^dTo be compared with the experimental value.

Table 2:

Computed first vertical excitation energies for *Anabaena* sensory rhodopsin wild type (ASR_{AT}-WT) and mutants models. ^a

Code	N=1 WARM ^b			N=10 WARM ^b			f_{osc}	$\Delta\Delta E_{S1-S0}^{WT}$ ^e (kcal mol ⁻¹)
	(nm) ^f	E _{S0-S1} ^c (kcal mol ⁻¹)	(eV)	(nm) ^f	E _{S0-S1} ^d (kcal mol ⁻¹)	(eV)		
ASR _{AT} -WT				547	52.3 (0.3)	2.27	1.30	–
<i>ANJ Set</i>								
ASR _{AT} -I113V	533	53.6	2.33	543	52.6 (1.2)	2.28	1.3	+0.3
ASR _{AT} -F139L	537	53.2	2.31					
ASR _{AT} -F139V	538	53.1	2.30					
ASR _{AT} -C203S	538	53.1	2.30					
ASR _{AT} -C203A	539	53.0	2.30					
ASR _{AT} -Y11A	542	52.7	2.29					
ASR _{AT} -C134S	543	52.6	2.28					
ASR _{AT} -L83V	545	52.5	2.27					
ASR _{AT} -P206Q	547	52.2	2.26					
ASR _{AT} -A119V	548	52.1	2.26	551	51.9 (0.6)	2.25	1.4	–0.4
<i>TDR Set</i>								
ASR _{AT} -L83N	532	53.7	2.33	542	52.8 (1.8)	2.29	1.28	+0.5
ASR _{AT} -Y11V	533	53.6	2.33					
ASR _{AT} -L83A	537	53.2	2.31					
ASR _{AT} -C137A	544	52.5	2.30					
ASR _{AT} -P206G	544	52.5	2.28					
ASR _{AT} -C134A	545	52.4	2.27					
ASR _{AT} -A119L	545	52.4	2.27					
ASR _{AT} -P180V	549	52.0	2.26					
ASR _{AT} -C137S	550	51.9	2.25					
ASR _{AT} -C203V	572	49.9	2.17	553	51.7 (1.2)	2.24	1.32	–0.6

^aThe mutants are ordered, inside each set, from shortest to longest computed absorption λ_{max}^a .

^bNumber of parallel molecular dynamics repetitions.

^cComputed excitation energy.

^dAverage computed excitation energy. Standard deviation in parenthesis.

^eComputed excitation energy difference mutant minus WT. Only values for N=10 are reported.

^fTo be compared with the experimental absorption maxima λ_{max}^a .