

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

Yuta Nakajima<sup>1,6</sup>, Laura Pedraza-González<sup>2,6</sup>, Leonardo Barneschi<sup>2</sup>, Keiichi Inoue<sup>3</sup>, Massimo Olivucci<sup>2,4</sup>✉ & Hideki Kandori<sup>1,5</sup>✉

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

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

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

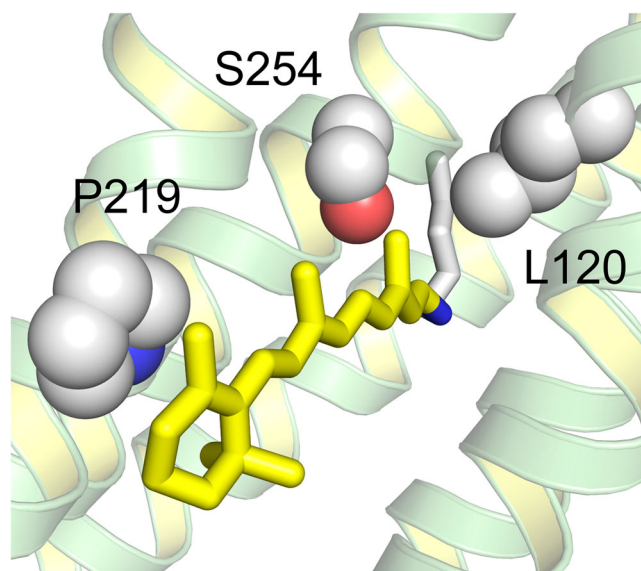
In most cases, the all-*trans* retinal chromophore of microbial rhodopsins features a protonated, Schiff base linkage (rPSBAT) and such  $-\text{CH}=\text{NH}(+)-$  state is stabilized by a negatively charged  $-\text{COO}(-)$  counterion, a glutamate or aspartate, placed in its vicinity. While the electrostatic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

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

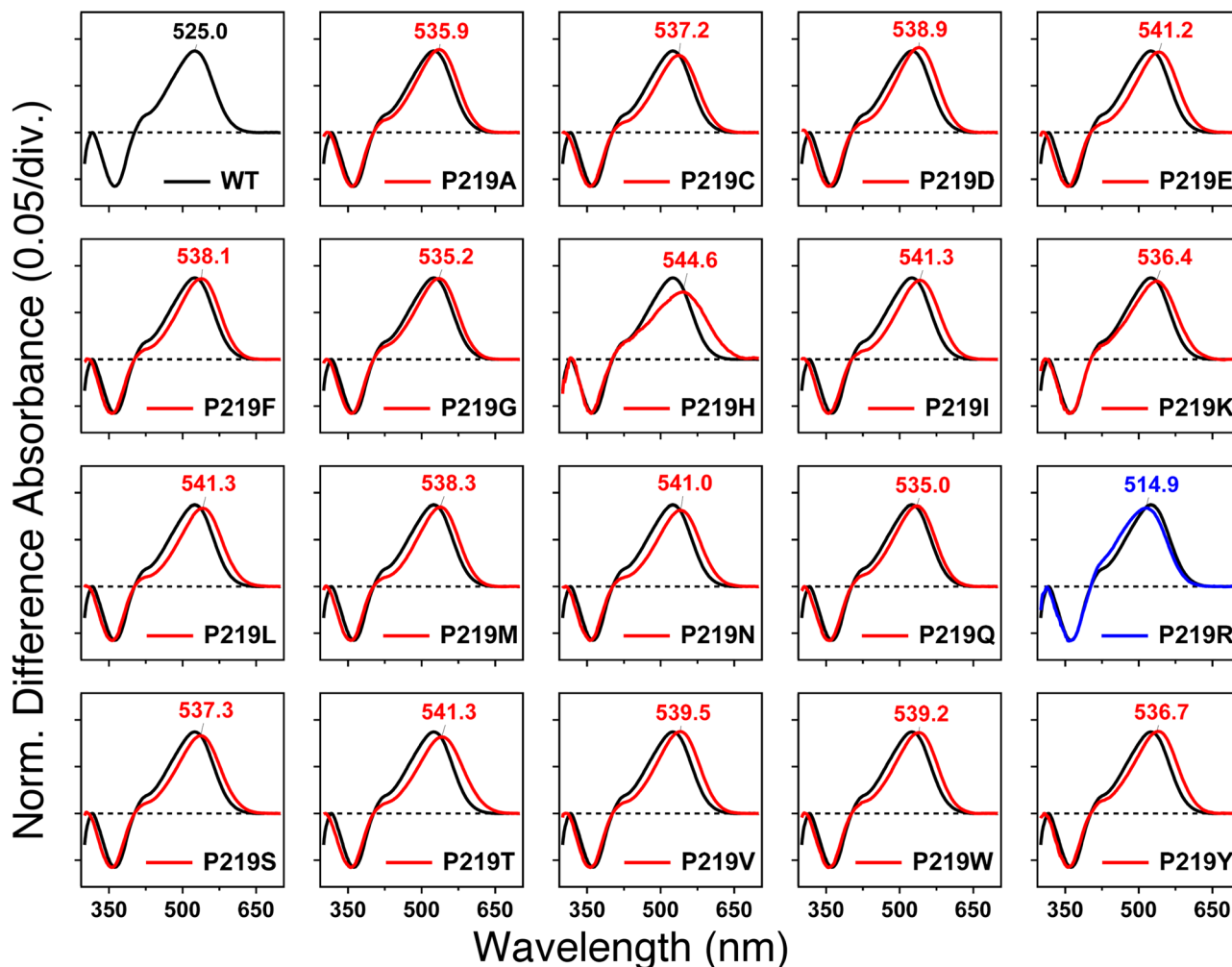
would stabilize  $S_0$  with respect to  $S_1$  leading to a blue-shift of the  $\lambda_{\max}$  value. In contrast, if the negative atoms are located in the vicinity of the  $\beta$ -ionone ring they would stabilize  $S_1$  with respect to  $S_0$  leading to a red-shift. Of course, positively charged atoms will have an opposite effect<sup>7</sup>.

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

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



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



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

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

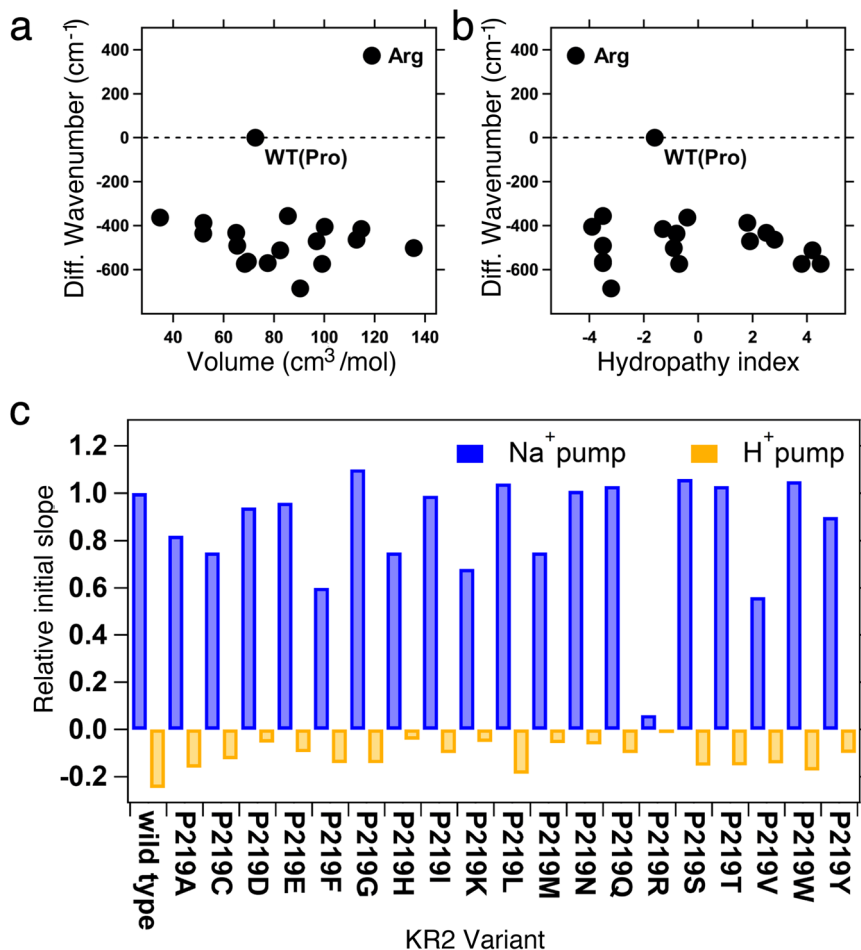
## Results and discussion

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

To quantify the absorption properties of each mutant, the sample was illuminated in the presence of hydroxylamine. This

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

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



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

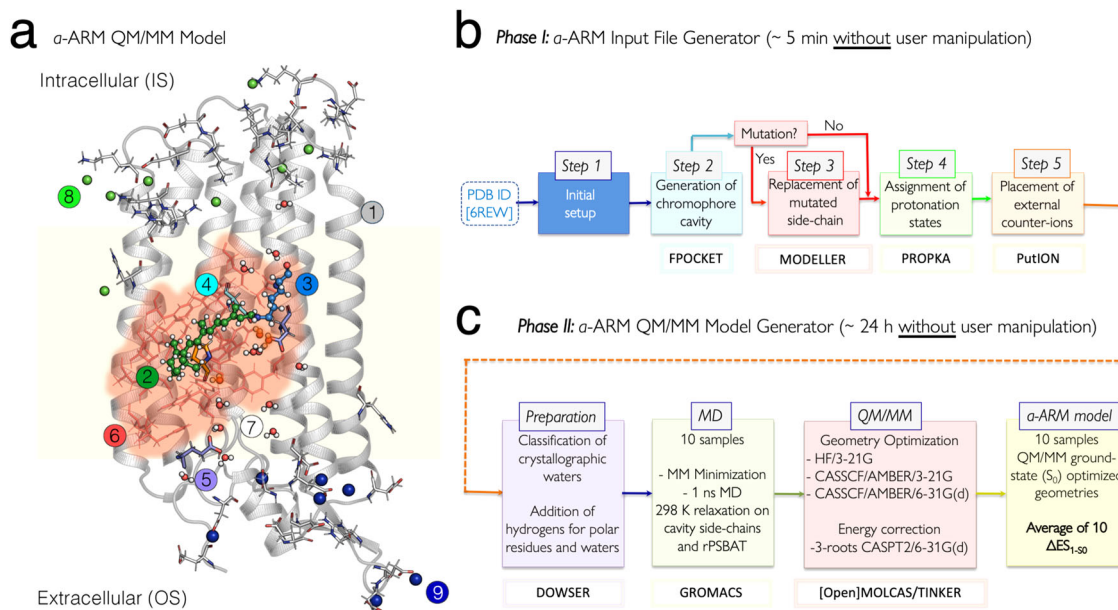
$\lambda_{\max}$  correlate with neither volume nor hydropathy, whereas the blue-shifted Arg is unique as it has both higher volume and lower hydropathy than Pro. The higher volume of Arg would suggest that the observed blue-shift is related to a steric effect. Below we show that, according to our computational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

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

**Construction of the WT model.** The QM/MM models (see Fig. 4a) were generated by using the *a*-ARM protocol, whose

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





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

replicas, see the “Materials and Methods” section) of about 127 nm (0.75 eV in terms of  $\Delta E_{S_1-S_0}$ ) blue-shifted with respect to the experimental data. In fact, as documented elsewhere<sup>9,30</sup>, the default models do not always replicate the correct electrostatics of the chromophore counterion complex in ion-pumping rhodopsins<sup>9,30</sup>. For this reason, a customized model was produced (see Supplementary Fig. 3 and Supplementary Table 1). The model customization employs the  $\alpha$ -ARM<sub>customized</sub> approach of ref. 9. (see also Supplementary Note 2) that guarantees reproducibility. While the default model predicts that both aspartic acid residues forming the counterion complex of the rPSBAT, namely D116 and D251<sup>31</sup>, are negatively charged, we have observed that the presence of two negative charges in the counterion complex would outbalance the positive charge of the rPSBAT, generating a large, blue-shifting effect<sup>9</sup> in the  $\lambda_{\text{max}}$  value. The  $\alpha$ -ARM<sub>customized</sub> approach allows to reassign the protonation states (step 4 of the input file generator in Fig. 4b) of D116 and D251. More specifically, as illustrated in Supplementary Fig. 3 and Supplementary Table 2, these protonation states are systematically scanned. It was found that a model featuring a protonated (i.e., neutral) secondary counterion (SC) D251 counterbalanced the charge in the vicinity of the rPSBAT by mitigating the overstabilization of the  $S_0$  positive charge of the chromophore Schiff base region yielding a smaller, blue-shifted error of about 14 nm (0.07 eV). Notice that in QM/MM modeling, it is a common practice to evaluate the protonation states of the rPSB counterion complex by looking, as a guidance, at the reproducibility of the experimental  $\lambda_{\text{max}}$  (see for instance refs. 13,32,33).

**Construction of the P219X mutant models.** The customized protonation states of the WT-KR2  $\alpha$ -ARM<sub>customized</sub> model were employed for the construction of the “default” P219X mutant

models. Notice that, as in the WT-KR2 case, the initial structure provided to the Input file generator phase for each mutant was the X-ray structure. As a result, 16 out of 19 models featured  $\Delta E_{S_1-S_0}$  values with an error, with respect to the observed value and therefore lower than 3.0 kcal mol<sup>-1</sup> (0.13 eV) threshold. Only 3 models, P219R, P219K, and P219H, were further customized. The default P219R model features a 67 nm (0.36 eV) blue-shifted  $\Delta E_{S_1-S_0}$  with respect to the observed value. This indicates that the positively charged arginine side-chain, located near the  $\beta$ -ionone ring of rPSBAT, leads to a too strong  $S_1$  destabilization (a positive charge  $\beta$ -ionone region is typical of the  $S_1$  state of the chromophore). In order to moderate such an effect, (i) the D116 was protonated (i.e., neutral) and the D251 was modeled as negatively charged, accordingly with the customization protocol reported in Supplementary Note 2, and (ii) the E160 residue in the R219 region, assigned to be neutral by step 4 of the default approach, was deprotonated assuming that the positive charge in position 219 increases the group pKa (see Supplementary Table 2). Such customization yields a model just 14 nm (0.07 eV) blue shifted. In the case of P219K, the lysine residue had to be assigned as neutral since the protonated lysine selected by the default protocol yielded a strongly blue-shifted  $\Delta E_{S_1-S_0}$  value with respect to the experiment. Finally, for P219H, the histidine residue had to be modeled with the neutral tautomer having the  $\epsilon$  nitrogen unprotonated (contrary to the default selection).

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

**Table 1**  $\alpha$ -ARM QM/MM models for the wild-type KR2 (WT-KR2) rhodopsin and 19 of its mutants (P219X, with X = A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y). First vertical excitation energy ( $\Delta E_{S_1-S_0}$ ), maximum absorption wavelength ( $\lambda_{\max}^a$ ), transition oscillator strength ( $f_{\text{Osc}}$ ), and difference between calculated and experimental data ( $\Delta\Delta E_{S_1-S_0}^{\text{Exp},\alpha\text{-ARM}}$ ), for the representative QM/MM model.

Variant	Experimental		$\alpha$ -ARM ( $N = 1$ ) <sup>a</sup>				Experimental vs $\alpha$ -ARM			
	$\Delta E_{S_1-S_0}^{\text{Exp}}$ (kcal mol <sup>-1</sup> )	(eV)	$\lambda_{\max}^{\text{Exp}}$ (nm)	$\Delta E_{S_1-S_0}^{\alpha\text{-ARM}}$ (kcal mol <sup>-1</sup> )	(eV)	$\lambda_{\max}^{\alpha\text{-ARM}}$ (nm)	$f_{\text{Osc}}$	$\Delta\Delta E_{S_1-S_0}^{\text{Exp},\alpha\text{-ARM}}$ (kcal mol <sup>-1</sup> )	(eV)	(nm)
WT	54.5	2.36	525	56.0	2.43	511	1.15	1.5	0.07	-14
P219A	53.4	2.31	536	55.8	2.42	512	1.16	2.4	0.11	-23
P219C	53.2	2.31	537	55.4	2.40	516	1.20	2.2	0.09	-21
P219D	53.1	2.30	539	55.0	2.39	520	1.22	2.0	0.09	-19
P219E	52.8	2.29	541	54.8	2.37	522	1.24	1.9	0.08	-19
P219F	53.1	2.30	538	54.9	2.38	521	1.22	1.8	0.08	-18
P219G	53.4	2.32	535	55.8	2.42	513	1.17	2.4	0.10	-23
P219H	52.5	2.28	545	54.5	2.36	525	1.22	2.0	0.09	-20
P219I	52.8	2.29	541	55.4	2.40	516	1.22	2.6	0.11	-25
P219K	53.3	2.31	536	54.8	2.38	521	1.22	1.5	0.07	-15
P219L	52.8	2.29	541	55.1	2.39	519	1.20	2.3	0.10	-22
P219M	53.1	2.30	538	54.8	2.37	522	1.21	1.7	0.07	-16
P219N	52.8	2.29	541	54.5	2.36	524	1.21	1.7	0.07	-17
P219Q	53.4	2.32	535	54.9	2.38	521	1.23	1.5	0.06	-14
P219R	55.5	2.41	515	57.0	2.47	501	1.12	1.5	0.06	-13
P219S	53.2	2.30	538	54.7	2.37	523	1.24	1.5	0.07	-15
P219T	52.8	2.29	541	55.5	2.40	516	1.19	2.6	0.11	-26
P219V	53.0	2.30	540	54.6	2.37	523	1.31	1.6	0.07	-16
P219W	53.0	2.30	539	55.1	2.39	519	1.22	2.1	0.09	-20
P219Y	53.3	2.31	537	55.2	2.39	518	1.19	1.9	0.08	-18
							MAE	1.9		
							AD <sub>max</sub>	2.6		
							MAD	0.3		

<sup>a</sup>Replica with  $\Delta E_{S_1-S_0}^{\alpha\text{-ARM}}$  closest to the average.

MAE: Mean absolute error, AD<sub>max</sub>: Maximum absolute error, MAD: Mean absolute deviation. See definitions in Supplementary Note 6.

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

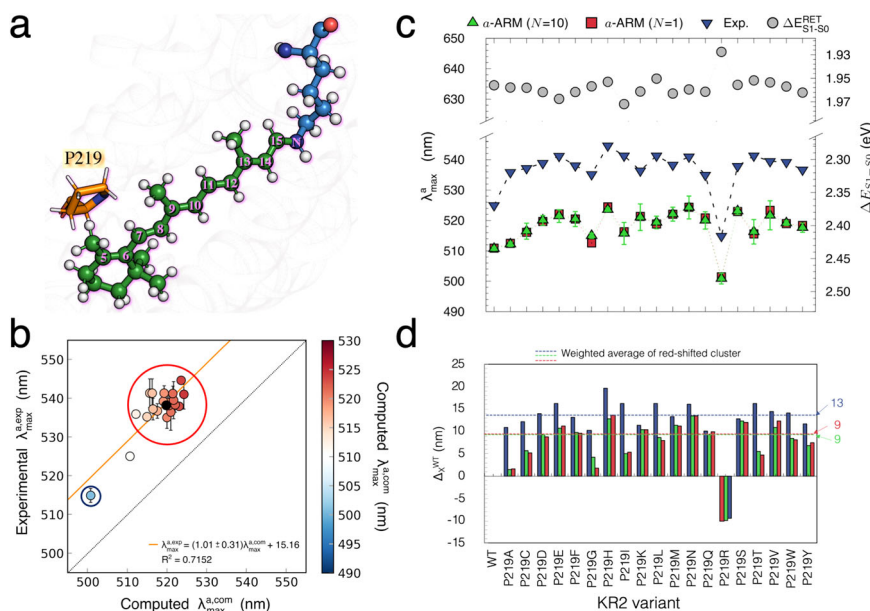
**Simulation of the  $\lambda_{\max}$  variation.** As mentioned above, the  $\alpha$ -ARM<sub>customized</sub> model of WT-KR2 has been selected as a suitable template for mutant modeling as it reproduces the observed  $\lambda_{\max}$  (525 nm) equivalent to a vertical excitation energy  $\Delta E_{S_1-S_0}^{\text{Exp}} = 54.5 \text{ kcal mol}^{-1}$  (2.36 eV). The discrepancy falls within the  $3.0 \text{ kcal mol}^{-1}$  (0.13 eV) error bar established for the  $\alpha$ -ARM protocol<sup>9,34,35</sup>. This result is reported in Table 1, together with both the computed ( $\Delta E_{S_1-S_0}^{\alpha\text{-ARM}}$ ) and observed ( $\Delta E_{S_1-S_0}^{\text{Exp}}$ ) vertical excitation energies, the corresponding oscillator strengths ( $f_{\text{Osc}}$ ), and their difference  $\Delta\Delta E_{S_1-S_0}^{\text{Exp},\alpha\text{-ARM}}$  for the representative model (i.e., replica with  $\Delta E_{S_1-S_0}^{\alpha\text{-ARM}}$  closest to the average). This is calculated as  $\Delta\Delta E_{S_1-S_0}^{\text{Exp},\alpha\text{-ARM}} = \Delta E_{S_1-S_0}^{\alpha\text{-ARM}} - \Delta E_{S_1-S_0}^{\text{Exp}}$ . The WT-KR2  $\Delta E_{S_1-S_0}^{\alpha\text{-ARM}}$  value is  $56.0 \pm 0.1 \text{ kcal mol}^{-1}$  (2.42 eV, 511 nm) and differs from the experimental data of just  $1.5 \text{ kcal mol}^{-1}$  (0.07 eV, -14 nm). However, as discussed below, such accuracy limit does not necessarily impact  $\lambda_{\max}$  trends that are not affected by systematic errors. Detailed information of the row data is provided in Supplementary Note 4 (see Supplementary Table 3).

The achieved WT-KR2 model shows that P219 is a suitable position for color tuning. Indeed, as displayed in Fig. 5a, it is close to the  $\beta$ -ionone ring of rPSBAT, so it is expected that the replacement of Pro by residues with different steric hindrance and/or polarity can affect the vertical excitation energy by stabilizing or destabilizing  $S_1$  with respect to  $S_0$ , as has been previously observed by some of the authors when modeling the

mutants P219G and P219T using the *original* version of the ARM protocol<sup>22</sup>. Although the conclusions derived from the study of Inoue et. al. can be qualitatively compared with the results presented in this work (see Supplementary Fig. 4 and Supplementary Table 4), notice that they cannot be quantitatively compared since (i) the QM/MM models were constructed from a different X-ray structure (i.e., 3X3C), using the *original* ARM<sup>9</sup> that featured (ii) manual input file generation (i.e., handmade and not reproducible counterion placement, different chromophore cavity, etc.) and (iii) a different methodological approach for the generation of the mutant side-chain, as described in Supplementary Note 5.

We begin the analysis of the experimental data by comparing the  $\Delta E_{S_1-S_0}^{\text{Exp}}$  trend with the corresponding computed values (see Supplementary Table 4), in the hope to learn how the P219X mutants shift their colors with respect to WT-KR2 (see blue down-triangles in Fig. 5c). As shown in Fig. 5b and reported in Table 1, the observed values vary from 545 nm to 515 nm. Remarkably, as reported in Supplementary Table 4, the P219R is the only variant that exhibit a blue-shifted effect with respect to the WT-KR2 of ca.  $1.1 \text{ kcal mol}^{-1}$ , whereas the other 18 mutants exhibit a red-shifting effect ranging from  $-1.0$  to  $-2.0 \text{ kcal mol}^{-1}$  suggesting that an electrostatic destabilization of the chromophore positive charge near its  $\beta$ -ionone is a key factor in the observed color tuning.

It is apparent that the WT-KR2 and P219R define a correlation line which is parallel to the perfect correlation line between observed and computed values, thus indicating a general systematic blue shifted error of ca. 15 nm (see Fig. 5d). On the other hand, the  $<1.0 \text{ kcal mol}^{-1}$  observed  $\Delta E_{S_1-S_0}^{\text{Exp}}$  variations among the 18 red-shifted mutants are, in most cases, far too small



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

for their trend to be reproduced by a  $a$ -ARM QM/MM model. These mutants form a cluster whose weighted average (the black circle in Fig. 5b) which represents the average of the measured and computed values weighted according to the more frequent deviation from the wild-type value, see Supplementary Note 6 and Supplementary Note 7) aligns with WT-KR2 and P219R supporting the general validity of the constructed models and confirming the presence of a systematic error (the row data reported in Table 1 shows a systematic blue-shifted error of ca.  $2.0 \text{ kcal mol}^{-1}$ ). Notice that for these models the mean absolute error (MAE) and the mean absolute deviation (MAD), calculated as indicated in Supplementary Note 6, are 1.9 and  $0.3 \text{ kcal mol}^{-1}$  respectively, consistently with those established for the  $a$ -ARM protocol<sup>9,30,35–37</sup>. For this reason, in the present work we avoid a detailed analysis of the members of the red-shifted cluster and focus on their weighted average features. In other words, we primarily focus on the effects that make: i) the P219R and ii) the center of the red-shifted cluster different from WT-KR2 in terms of  $\Delta E_{S_1-S_0}^{a-ARM}$ .

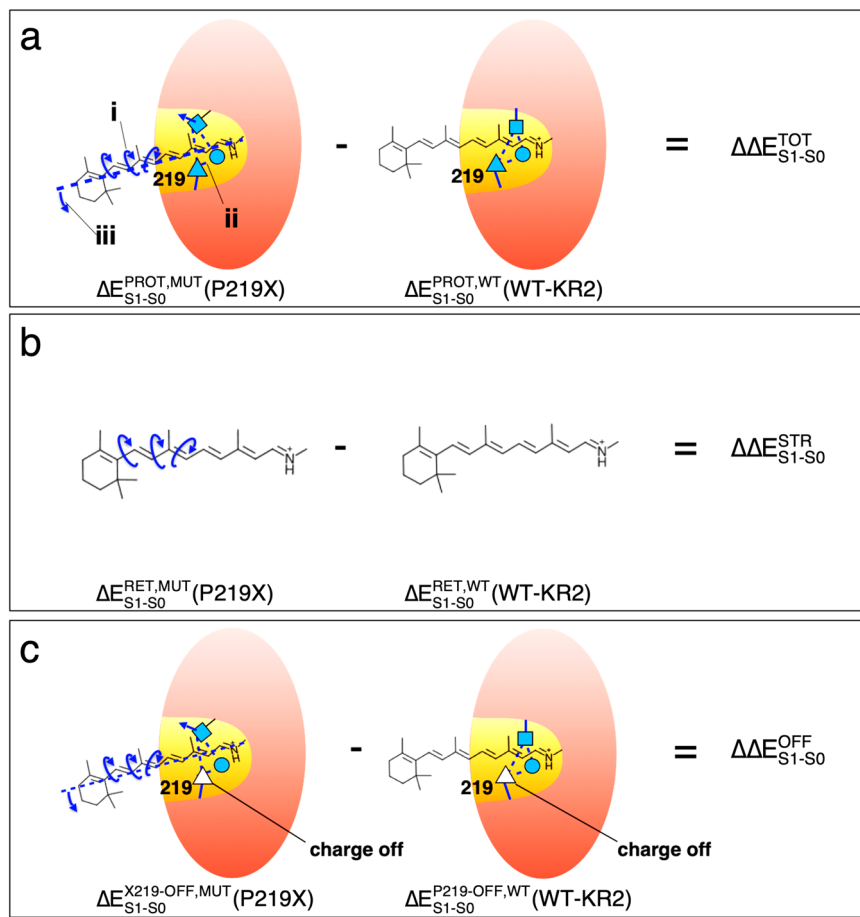
The observed  $\Delta E_{S_1-S_0}^{Exp}$  and calculated  $\Delta E_{S_1-S_0}^{a-ARM}$  values are reported in Fig. 5c, while their difference relative to WT-KR2 are given in Fig. 5d (see also Supplementary Table 5). These quantities were computed as the difference between observed  $\Delta E_{S_1-S_0}^{Exp}$  for each P219X mutant with respect to the observed value of the WT-KR2 ( $\Delta_{max,X}^{WT,Exp}$ ), as well as the difference between computed  $\Delta E_{S_1-S_0}^{a-ARM}$  of each of the P219X mutants with respect to the corresponding WT-KR2 ( $\Delta_{max,X}^{WT,a-ARM}$ ) values. Notice that in this figure the average values (see dashed horizontal lines) for the red-shifted cluster are also provided. In all the cases, blue or red direction of the shift is reproduced. More interestingly, in Fig. 5d it is evident that, in line with the observations (blue bars), the computed data (green and red bars) show that P219R is the

only mutant presenting a blue-shifted effect and P219H is the most red-shifted mutant.

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

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

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



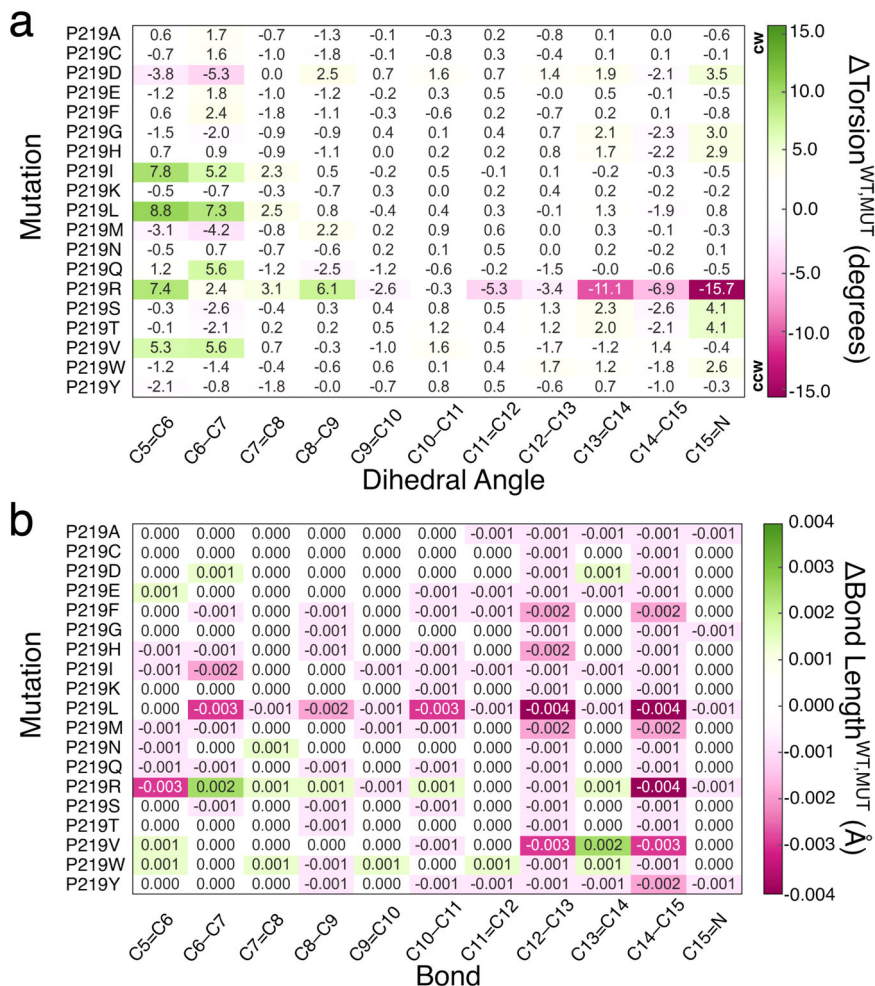
**Fig. 6 Pictorial illustration of the corresponding excitation energy analysis terms. (a)** Difference between the *a*-ARM models excitation energies (in protein environment,  $\Delta E_{S1-S0}^{PROT} = \Delta E_{S1-S0}^{a-ARM}$ ) of a P219X mutant with respect to the wild-type due to (i) structural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

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

“Direct electrostatic component” that is indirectly computed as  $\Delta \Delta E_{S1-S0}^{ELE(d)} = \Delta \Delta E_{S1-S0}^{ELE(t)} - \Delta \Delta E_{S1-S0}^{ELE(i)}$ . Further details are provided in Supplementary Note 8 (see Supplementary Table 6 and Supplementary Table 7).

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



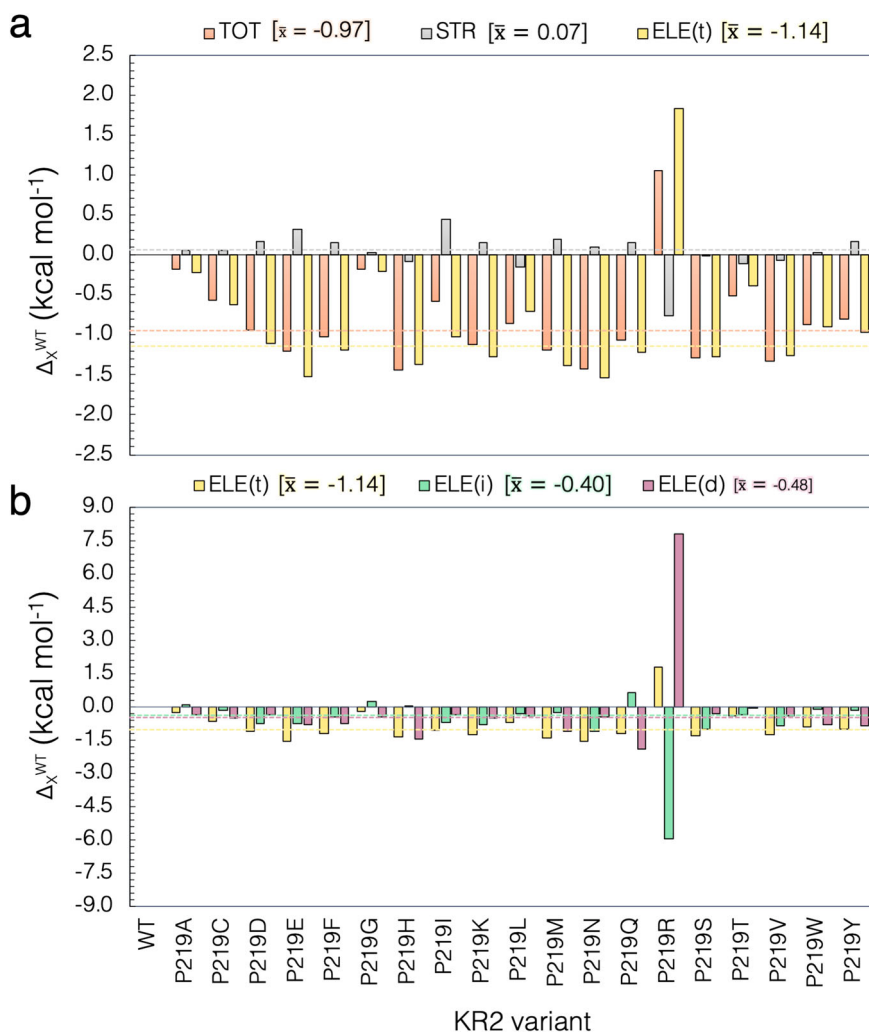


**Fig. 7** Heatmap representation of the variation of structural parameters of P219X (X= A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y) mutants with respect to the wild-type KR2 (WT-KR2). Difference between WT-KR2 and mutant (a) dihedral angles,  $\Delta\text{Torsion}^{\text{WT,MUT}}$  and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

above. To do so, we select a representative QM/MM structure for each mutant model (i.e., the replica with  $\Delta E_{\text{SI-SO}}^{\text{a-ARM}}$  closest to the average, see red squares in Fig. 5c) and compute the difference between the magnitude of its most relevant structural parameters (e.g., torsional dihedral angles and bond lengths) with respect to those corresponding to the representative structure of the WT-KR2 model. A visual structural comparison between each mutant and WT-KR2 is provided in Supplementary Figure 5. In Fig. 7 we report a heatmap visualization of such differences in terms of skeletal dihedral angles ( $\Delta\text{Torsion}^{\text{WT,MUT}}$ , Fig. 7a) and carbon-carbon bond lengths ( $\Delta\text{BondLength}^{\text{WT,MUT}}$ , Fig. 7b). Based on the range in which the latest quantities change, we establish an arbitrary threshold of 2.0 degrees and 0.01 Å, respectively, as meaningful variations. The dihedrals with the largest variation are C5 = C6 and C6-C7 that, as shown in Fig. 5a, belong to the rPSBAT framework geometrically closer to the 219 residue. In addition, for specific cases the “reactive” C13 = C14 dihedral as well as the C14-C15 and C15 = N dihedrals show variability. We start the analysis discussing the variants belonging to the red-shifted cluster. These are almost constantly accompanied by structural effect near the  $\beta$ -ionone ring (among them, P219I, P219L, P219M, P219V, P219F, P219Y featuring hydrophobic side-chain, and P219Q featuring large polar uncharged side-chain present the largest structural effects) making such a structural deformation a characteristic of the cluster center. On

the other hand, P219G, P219H and P219W show a limited change in the C15 = N region and P219D (in its neutral form), P219S and P219T (i.e., polar uncharged side-chains) present changes in both regions. As expected, the blue-shifted P219R mutant featuring a positively charged side-chain, induces a significant, more than 2.0 degrees, change of its dihedral angles near the  $\beta$ -ionone ring. More interestingly, this mutant exhibits a particular large variation of the “reactive” C13 = C14 dihedral as well as the C14-C15 and C15 = N dihedrals (see Fig. 7a).

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



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

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

We now look at the impact of the described geometrical changes on the  $\lambda_{\max}^a$  value. To this aim, we compare the computed  $\Delta E_{S1-S0}^{\text{RET,MUT}}$  values with the corresponding wild-type  $\Delta E_{S1-S0}^{\text{RET,WT}}$  for the isolated chromophore (in kcal mol<sup>-1</sup>), as reported in Supplementary Table 7. This analysis consists in extracting, for WT-KR2 and for each variant, the rPSBAT structure from the protein and computing its vertical excitation energy without relaxing the chromophore structure (see Fig. 6b). A small  $\Delta\Delta E_{S1-S0}^{\text{STR}} = \Delta E_{S1-S0}^{\text{RET,MUT}} - \Delta E_{S1-S0}^{\text{RET,WT}}$  value indicates that the geometrical distortion of the retinal due to the point mutation has only a limited effect on the excitation energy change. As mentioned above, we analyze the  $\Delta\Delta E_{S1-S0}^{\text{STR}}$  contribution for the red-shifted cluster center and P219R. To this aim, we computed the weighted average of the  $\Delta\Delta E_{S1-S0}^{\text{STR}}$  contribution for the members of the red-shifted cluster (see gray dashed line in Fig. 8a) and compare this value with the  $\Delta\Delta E_{S1-S0}^{\text{STR}}$  contribution of the blue-shifted one. Consistently with the small variations in

$\Delta E_{S1-S0}^{\text{RET,WT}}$  reported in Fig. 5c (see gray circles) for the red-shifted cluster, the weighted  $\Delta\Delta E_{S1-S0}^{\text{STR}}$  average of about 0.07 kcal mol<sup>-1</sup> reported in Fig. 8a and Supplementary Table 6, show that the impact of the rPSBAT geometrical deformation on the change in their excitation energy is very limited. On the other hand, as also confirmed by the geometrical data discussed above, the blue-shifted P219R variant features a considerable  $\Delta\Delta E_{S1-S0}^{\text{STR}}$  contribution of  $-0.8$  kcal mol<sup>-1</sup>. These results, also reported in Supplementary Table 7, give a first indication about the different behavior on the color tuning mechanism exhibited for the red-shifted clusters and P219R mutant.

**Electrostatic effects.** The QM/MM models also allow to investigate electrostatic effects. As anticipated above, the total electrostatic effect ( $\Delta\Delta E_{S1-S0}^{\text{ELE(t)}}$ ) can be decomposed in two parts: (i) the first can be considered as a *direct* component ( $\Delta\Delta E_{S1-S0}^{\text{ELE(d)}}$ ) due to the variation in number, magnitude, and position of the point charges of residue 219 caused by the P to X replacement and (ii) a more *indirect* component ( $\Delta\Delta E_{S1-S0}^{\text{ELE(i)}}$ ) produced from the reorganization of the local environment and hydrogen

bond network induced by the same replacement and due to the fact that conserved residues and water molecules change in position or orientation. Moreover, as discussed below, possible changes in protonation states of conserved residues, induced by P to X replacement, have a major contribution to the *indirect* component.

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

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

general arrangement of the charges in the cavity is not altered with respect to WT-KR2 (see Fig. 9b) or P219G (see Fig. 9c).

## Conclusions

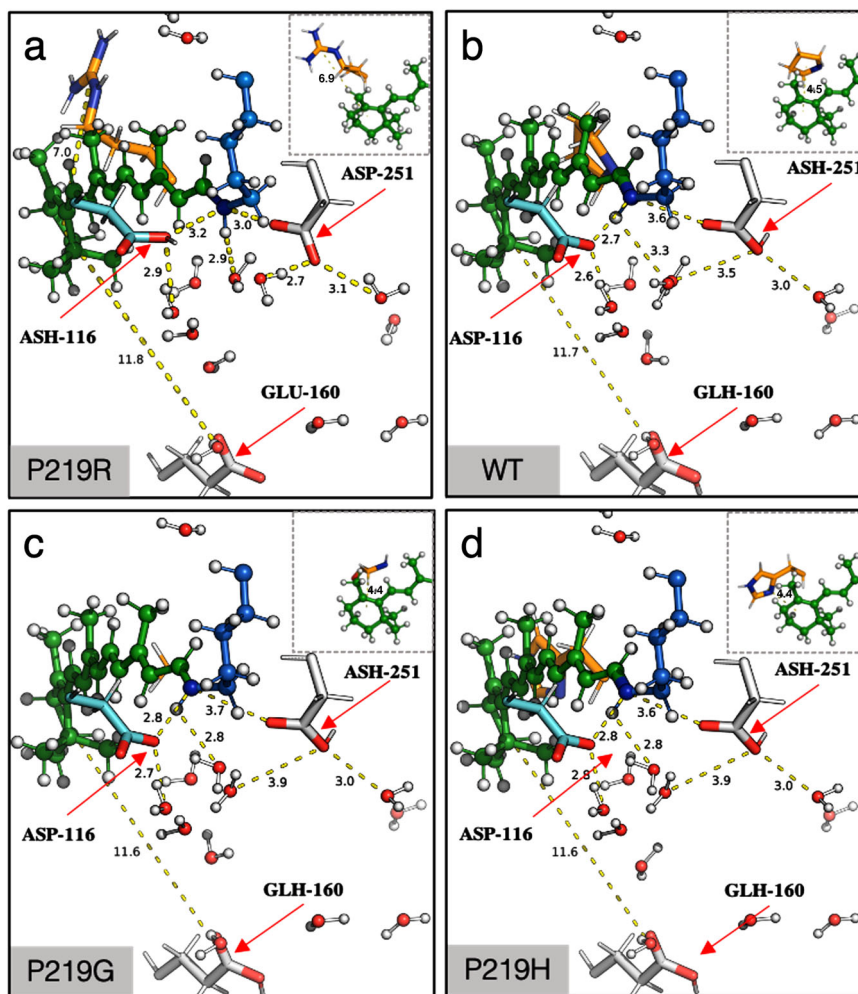
Above we have presented a combined experimental and computational investigation of the P219X mutants of KR2 supporting the existence of a G/P switch in microbial rhodopsins. We have also reported on a computational analysis of the mutant vertical excitation energy indicating that G/P is, essentially,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

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

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

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





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

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

## Materials and methods

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

**Measurement of  $\lambda_{\max}$  by hydroxylamine bleach.** The  $\lambda_{\max}$  of WT-KR2 and mutants was determined by bleaching the protein with hydroxylamine according to the previously reported method<sup>22</sup>. The *E. coli* cells expressing the rhodopsins were washed with a solution containing 100 mM NaCl, 50 mM  $\text{Na}_2\text{HPO}_4$  (pH 7) for three times. The washed cells were treated with 1 mM lysozyme for 1 h at room temperature, and then disrupted by sonication. To solubilize the rhodopsins, 3% DDM was added, and the samples were stirred for overnight at 4 °C. The rhodopsins were bleached with 500 mM hydroxylamine and subjected to illumination of yellow light ( $\lambda > 500$  nm) from the output of 1 kW tungsten–halogen projector lamp (Master HILUX-HR, Rikagaku) through a glass filter (Y-52, AGC Techno Glass, Japan). The absorption change upon the bleaching was measured by UV-visible spectrometer (V-730, JASCO, Japan).

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

**QM/MM modeling.** Congruous and reproducible computational models for WT-KR2 and its P219X (X = A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y) mutants were generated using the *a*-ARM version<sup>9</sup> of the Automatic Rhodopsin Modeling protocol<sup>34</sup>, a specialized user-friendly command-line oriented computational tool (i.e., Python3-based software package) for either the fully-automated (default) or semi-automated (i.e., customized) construction of basic hybrid QM/MM models of rhodopsins<sup>30,35–37,42</sup>, called *a*-ARM models. The produced *a*-ARM models illustrated in Fig. 4a are monomeric, “gas-phase”, and globally uncharged, based on electrostatic embedding and the hydrogen link-atom frontier between the QM and MM subsystems. The term “gas-phase” refers to the fact that the protein membrane and solvation effects are not explicitly modeled during the calculations. Nevertheless, the electrostatic effect of the protein external environment is effectively accounted for by an asymmetric distribution of external ions (i.e.,  $\text{Cl}^-$  or  $\text{Na}^+$ ) whose charge and position are determined, automatically, by a specific algorithm, near the most positively and/or negatively charged surface amino acids in both the intracellular (IS) and extracellular (OS) protein surfaces. Moreover, to



account the water-mediated hydrogen-bond network (HBN) in the protein cavity, the internal crystallographic waters are retained in the model, while the waters that are not experimentally detected are assumed to be extremely mobile or just absent in the chromophore hydrophobic protein cavity. Considering these approximations, during the construction of the *a*-ARM model the rhodopsin structure is subdivided into three sub-systems requiring layers of increasing level of accuracy for their description. The first MM-based sub-system, called protein environment, features residues (backbone and side-chain atoms) fixed at the crystallographic or comparative (homology) structure and incorporates external Cl<sup>-</sup> and/or Na<sup>+</sup> counterions fixed at preoptimized positions as well as some crystallographic/comparative waters (see below). The second sub-system, namely the chromophore cavity, constitutes the second MM layer of the model and contains amino acid residues with fixed backbone and relaxed side-chains, including the covalently linked lysine residue atoms (excluding the terminal NH<sub>2</sub>-C<sub>ε</sub> moiety), as well as flexible crystallographic/comparative waters in the vicinity (<4 Å) of the rPSB. The third sub-system, called Lys-QM, contains the MM atoms of the covalently linked lysine side-chain in contact (through C<sub>ε</sub>) with the QM/MM frontier (i.e., 9 atoms) and the entire QM sub-system which corresponds to a N-methylated retinal chromophore (i.e., 53 atoms). As illustrated in Fig. 4b, c, to obtain the QM/MM model automatically the only required input is either the PDB ID or a template PDB file (X-ray structure or comparative model) of the rhodopsin, that is processed by two subsequent phases called input file generator (Phase I, Fig. 4b) and QM/MM model generator (Phase II, Fig. 4c) correspondingly, previously reported in refs. <sup>9,30,34,35</sup> and summarized in Supplementary Note 2 and Supplementary Note 3, respectively. Briefly, Phase II produces 10 model replicas of the *a*-ARM model via S<sub>0</sub> geometry optimization of the input generated through Phase I starting from distinct molecular dynamics (MD) runs (see Supplementary Note 3). Such optimization relies on the use of the multi-configurational complete active space self-consistent field (CASSCF) at the 2-roots single-state, for modeling the QM sub-system while the MM subsystem is treated using the AMBER force field (CASSCF(12,12)/6-31 G(d)/AMBER level). The AMBER force field is not polarizable and only includes polarizability effect in a mean-field fashion possibly contributing the method systematic error.

The CASSCF(12,12) level of theory is followed by an energy correction at the multi-configurational second-order perturbation level (CASPT2) to recover the missing dynamical electron correlation. Thus, a 3-roots state-average CASPT2 that uses the three-root stage-average CASSCF(12,12)/6-31 G(d)/AMBER as the zero-order reference wavefunction, is computed (CASPT2(12,12)/6-31 G(d)/AMBER). Ultimately, each of the 10 model replicas correspond to an equilibrated gas-phase and globally uncharged monomer QM/MM model and it is associated with a ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> calculated between S<sub>0</sub> → S<sub>1</sub>. Finally, the average ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> is reported along with the corresponding standard deviation (see Section Supplementary Note 3).

**Validation, capabilities, and potential applications of *a*-ARM.** The *a*-ARM protocol, in its current version, does not represent a predictive tool but rather is designed to produce models useful for investigating the origin of trends in spectroscopic/photochemical properties (e.g., between sequence variability and function) emerging from different sets of experimental data. Furthermore, the models are used assuming that single amino acid replacements will change the protein conformation at the level of the chromophore cavity. Accordingly, after demonstrating that the models reproduce, within the protocol error bar (see below), the trend in observed vertical excitation energy ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> (equivalent to the λ<sup>a</sup><sub>max</sub> value), the QM/MM models are employed to study the factors (i.e., steric and electrostatic contributions to ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub>) determining the ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> change of each mutant with respect to the WT-KR2. This is only possible because the level of protocol automation guarantees, regardless of the user ability or computational facility, reproducible *a*-ARM QM/MM models when the same parameters are employed in Phases I and II of Fig. 4b, c. Indeed, the production of the default *a*-ARM model (see *a*-ARM<sub>default</sub> approach in Supplementary Note 2) is fully automated and does not require human intervention thus preventing the generation of biased models. Furthermore, the absence of parameters required at the input stage, prevents overfitting. On the other hand, in specific cases, mostly related to the difficulties in predicting the correct ionization state of ionizable groups surrounding the chromophore it is necessary to customize the default model by following a well-defined protocol (see *a*-ARM<sub>customized</sub> approach in Supplementary Note 2). Such a systematic (non-arbitrary) protocol, documented in refs. <sup>9,30</sup>, and summarized in Supplementary Note 2, involves three steps which only concern the ionization states of the ionizable residues that, as reported in refs. <sup>9,30</sup> and anticipated above, is the most frequent cause of inconsistencies between the computed and observed excitation energies. The customization procedure does not represent an overfitting of the model but a systematic ionization state scan also used by other groups working on rhodopsin simulations (see for instance refs. <sup>13,32,33</sup>).

The *a*-ARM models described above (see Supplementary Note 10 and Supplementary Fig. 6) have been successfully benchmarked by reproducing the trends in ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> of a set of 44 rhodopsins (25 of wild type and 19 mutants)<sup>9,30,35-37,42</sup>. More specifically, the *a*-ARM<sub>default</sub> approach proved to be capable of reproducing the observed ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> values for 79% (35/44) of the models, whereas the other 21% were successfully obtained with the *a*-ARM<sub>customized</sub> approach (i.e., customizing the protonation states pattern). The estimated trend deviation, relative to the probed set using as a reference bovine rhodopsin (Rh), has

been reported as 0.7 ± 0.5 kcal mol<sup>-1</sup> (0.03 ± 0.02 eV) and the mean absolute error (MAE) (see definition in Supplementary Note 6) as 1.0 kcal mol<sup>-1</sup> (0.04 eV)<sup>9</sup>. Both computed and observed trends in ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> for the benchmark set, that comprises microbial and animal rhodopsins whose structures were obtained from either X-ray crystallography or comparative modeling, is presented in Supplementary Figure 7.

Considering that in this study we are interested in trends in vertical excitation energies, and more specifically red- or blue-shifts of these values, the results are presented and discussed in terms of mutation-induced shifts with respect to WT-KR2 rather than differences between observed and calculated values. On the other hand, the observed ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> displays a remarkable experimental “asymmetry” with 18 over 19 single site variants featuring a red-shifted ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> values compressed in a 1.0 kcal mol<sup>-1</sup> (0.04 eV or ca. 10 nm) wide range. Such relatively small difference has prevented a robust individual quantitative analysis of the color variation of these mutants. For these reasons the red-shifted mutants have been collected in a single cluster and the analysis is mainly performed by focusing on the origin of the ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> change when going from WT-KR2 to the center of the cluster. Still, the less and most red-shifted mutants falling parallel to the correlation line (P219G and P219H respectively) are individually discussed. The limitations and pitfalls of the protocol are summarized in Supplementary Note 12.

**Automation of the mutant rotamer selection.** The QM/MM modeling of the mutants required a second step of experimentally-driven customization dealing with the residue side-chain conformation (i.e., rotamer selection). That procedure consisted in evaluating the model performance for a set of automatically chosen rotamers and select the one that better reproduces the observed ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> value (we assume ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp</sup> = hc/λ<sup>a,Exp</sup><sub>max</sub> and use ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp</sup> and λ<sup>a,Exp</sup><sub>max</sub> interchangeably in the following). To this aim, an update version of the mutation routine of the input generator (see step 3 in Fig. 4b), that uses Modeller<sup>43</sup> instead of the default rotamer library SCWRL4<sup>44</sup>, is proposed as part of the present work. A detailed description of the new approach is provided in Supplementary Note 13. Briefly, in order to explore the performance of different rotamers of the mutated side-chain, three *a*-ARM QM/MM models featuring the three highest scored mutated side-chain rotamers selected by Modeller<sup>43</sup>, are produced and their ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> evaluated (see Supplementary Fig. 7). Then, the model that better reproduces the observed ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> is selected (see Supplementary Fig. 8). To perform such selection, we use, as a baseline, the difference between the computed and observed ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> of the WT-KR2, hereafter referred to as ΔΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp,WT</sup>. The equivalent quantity calculated for each rotamer (ΔΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp,rotX</sup>, with X=1, 2, 3) is then contrasted with the ΔΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp,WT</sup> via the equation rotX = (ΔΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp,rotX</sup> - ΔΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub><sup>Exp,WT</sup>) (see Supplementary Fig. 9). The rotamer that features the lower rotX value (preferring blue-shifted values) is chosen as the representative *a*-ARM model (see Supplementary Fig. 10). Although this approach relies on experimental information and does not represent a predictive tool, it automates the side-chain conformation selection during the construction of mutant models aimed at the reproduction of experimental trends in properties. Some limitations and pitfalls of the proposed approach are provided in Supplementary Note 13.

**Statistics and reproducibility.** As further described in “QM/MM Modeling” section and Supplementary Note 3, each *a*-ARM QM/MM model is composed of n = 10 independent replicas that correspond to an equilibrated gas-phase and globally uncharged monomer, and it is associated with a ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> calculated between S<sub>0</sub> → S<sub>1</sub>. As shown in Fig. 5c, for each model the average ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> is reported (see green triangles) along with the corresponding standard deviation (see green error bars). Finally, the replica with ΔE<sub>S<sub>1</sub>-S<sub>0</sub></sub> closest to the average (see red squares) is selected as a representative model for the color tuning analyses. The production of the models, as well as the computation of the average and standard deviation, and the posterior selection of the representative replica is performed automatically by the *a*-ARM protocol illustrated in Fig. 4. Therefore, the reproducibility of the data is guaranteed when selecting the same initial seeds for the n = 10 MD runs. On the other hand, the statistical parameters employed for the different computational analyses, i.e., mean absolute error (MAE), mean absolute deviation (MAD), and weighted average ( $\bar{x}$ ) are described in Supplementary Note 6 and the equations are provided. Furthermore, the methodology employed for the calculation of the weighted average is provided in Supplementary Note 6, along with an illustrative example.

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

## Data availability

All experimental data shown in main figures were deposited in Supplementary Data 1. XYZ Coordinates for the optimized S<sub>0</sub> *a*-ARM QM/MM models are provided as Supplementary Data 2-21 as follows: P219A (Supplementary Data 2.xyz), P219C (Supplementary Data 3.xyz), P219D (Supplementary Data 4.xyz), P219E (Supplementary Data 5.xyz), P219F (Supplementary Data 6.xyz), P219G (Supplementary Data 7.xyz), P219H (Supplementary Data 8.xyz), P219I (Supplementary Data 9.xyz), P219K (Supplementary Data 10.xyz), P219L (Supplementary Data 11.xyz), P219M (Supplementary Data 12.xyz), P219N (Supplementary Data 13.xyz), P219Q

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

### Code availability

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

Received: 3 January 2021; Accepted: 19 September 2021;

Published online: 13 October 2021

### References

- Ernst, O. P. et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2013).
- Devine, E. L., Oprian, D. D. & Theobald, D. L. Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proc. Natl. Acad. Sci. USA.* **110**, 13351–13355 (2013).
- Shimono, K., Ikeura, Y., Sudo, Y., Iwamoto, M. & Kamo, N. Environment around the chromophore in pharaonis phorbodopsin: Mutation analysis of the retinal binding site. *Biochim. Biophys. Acta* **1515**, 92–100 (2001).
- Yokoyama, S. Evolution of dim-light and color vision pigments. *Annu. Rev. Genomics Hum. Genet.* **9**, 259–282 (2008).
- Sudo, Y. E. A. A blue-shifted light-driven proton pump for neural silencing. *J. Biol. Chem.* **288**, 20624–20632 (2013).
- Engqvist, M. K. E. A. Directed evolution of Gloeobacter violaceus rhodopsin spectral properties. *J. Mol. Biol.* **427**, 205–220 (2015).
- Hayashi, S. E. A. Structural determinants of spectral tuning in retinal proteins-bacteriorhodopsin vs sensory rhodopsin II. *J. Phys. Chem. B* **105**, 10124–10131 (2001).
- Fujimoto, K., Hayashi, S., Hasegawa, J. Y. & Nakatsuji, H. Theoretical studies on the color-tuning mechanism in retinal proteins. *J. Chem. Theory Comput.* **3**, 605–618 (2007).
- Pedraza-González, L., De Vico, L., Marín, M. D. C., Fanelli, F. & Olivucci, M. *a*-ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *J. Chem. Theory Comput.* **15**, 3134–3152 (2019).
- Béja, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
- Katayama, K. & Sekharan, S. S. Y. *Optogenetics* (eds. Yawo, H., Kandori, H. & Koizumi, A.) vol. 7, 89–107 (Springer, 2015).
- Inoue, K., Tsukamoto, T. & Sudo, Y. Molecular and evolutionary aspects of microbial sensory rhodopsins. *Biochim. Biophys. Acta* **1837**, 562–577 (2013).
- Broser, M. et al. Neor, a near-infrared absorbing rhodopsin. *Nat. Commun.* **11**, 5682 (2020).
- Gozem, S., Luk, H. L., Schapiro, I. & Olivucci, M. Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chem. Rev.* **117**, 13502–13565 (2017).
- Kato, H. E. et al. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nat. Commun.* **6**, 1–10 (2015).
- Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. & Tsien, R. Y. Reachr: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nat. Neurosci.* **16**, 1499–1508 (2013).
- Oda, K. E. A. Crystal structure of the red light-activated channelrhodopsin Chrimson. *Nat. Commun.* **9**, 3949 (2018).
- Man, D. E. A. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
- Bielawski, J. P., Dunn, K. A., Sabehi, G. & Beja, O. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl. Acad. Sci. USA.* **101**, 14824–14829 (2004).
- Ozaki, Y., Kawashima, T., Abe-Yoshizumi, R. & Kandori, H. A color-determining amino acid residue of proteorhodopsin. *Biochemistry* **53**, 6032–6040 (2014).
- Sudo, Y. E. A. A microbial rhodopsin with a unique retinal composition shows both sensory rhodopsin II and bacteriorhodopsin-like properties. *J. Biol. Chem.* **286**, 5967–5976 (2011).
- Inoue, K. et al. Red-shifting mutation of light-driven sodium-pump rhodopsin. *Nat. Commun.* **10**, 1993 (2019).
- Melaccio, F., Ferré, N. & Olivucci, M. Quantum chemical modeling of rhodopsin mutants displaying switchable colors. *Phys. Chem. Chem. Phys.* **14**, 12485–12495 (2012).
- Huntress, M. E. A. Towards an understanding of the retinal chromophore in rhodopsin mimics. *J. Phys. Chem. B* **117**, 10053–10070 (2013).
- Orozco-Gonzalez, Y., Kabir, M. P. & Gozem, S. Electrostatic spectral tuning maps for biological chromophores. *J. Phys. Chem. B* **129**, 4813–4824 (2019).
- Tang, K., Lin, D. L. K. & Jiao, N. Draft genome sequence of *Parvularcula oceani* JLT2013T, a rhodopsin-containing bacterium isolated from deep-sea water of the southeastern Pacific. *Mar. Genomics* **24**, 211–213 (2015).
- Choi, D. H. Y. H., Chun, J. & Cho, B. C. *Jannaschia seosinensis* sp. Nov., isolated from hypersaline water of a solar saltern in Korea. *Int. J. Syst. Evol. Microbiol.* **56**, 45–49 (2006).
- Inoue, K. E. A. A natural light-driven inward proton pump. *Nat. Commun.* **7**, 13415 (2016).
- Kovalev, K. E. A. Structure and mechanisms of sodium-pumping KR2 rhodopsin. *Sci. Adv.* **5**, eaav2671 (2019).
- Pedraza-González, L. et al. On the automatic construction of QM/MM models for biological photoreceptors: rhodopsins as model systems, In *QM/MM Studies of Light-responsive Biological Systems*. vol 31 (Springer, Cham, 2021).
- Skopintsev, P. E. A. et al. Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature* **583**, 314–318 (2020).
- Adam, S., Wiebeler, C. & Schapiro, I. Structural factors determining the absorption spectrum of channelrhodopsins: A Case study of the chimera C1C2. *J. Chem. Theory Comput.* (2021).
- Kaufmann, J. C. et al. Modulation of light energy transfer from chromophore to protein in the channelrhodopsin ReaChR. *Biophys. J.* **119**, 705–716 (2020).
- Melaccio, F. et al. Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *J. Chem. Theory Comput.* **12**, 6020–6034 (2016).
- Aquilante, F. et al. Modern quantum chemistry with [Open]Molcas. *J. Chem. Phys.* **152**, 214117 (2020).
- Pedraza-González, L. et al. Web-ARM: a web-based interface for the automatic construction of QM/MM models of rhodopsins. *J. Chem. Inf. Model.* **60**, 1481–1493 (2020).
- Mroginski, M. A. et al. Frontiers in multiscale modeling of photoreceptor proteins. *Photochem. Photobiol.* **97**, 243–269 (2021).
- Mao, J. et al. Structural basis of the green-blue color switching in proteorhodopsin as determined by NMR spectroscopy. *J. Am. Chem. Soc.* **136**, 17578–17590 (2014).
- Gieseking, R. L., Risko, C. & Bredas, J. L. Distinguishing the effects of bond-length alternation vs bond-order alternation on the nonlinear optical properties of  $\pi$ -conjugated chromophores. *J. Phys. Chem. Lett.* **6**, 2158–2162 (2015).
- Ho Choi, C., Kertesz, M. & Karpfen, A. The effects of electron correlation on the degree of bond alternation and electronic structure of oligomers of polyacetylene. *J. Chem. Phys.* **107**, 6712–6721 (1997).
- Inoue, K. E. A. A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **4**, 1678 (2013).
- Gholami, S. et al. Multistate multi-configuration quantum chemical computation of the two-photon absorption spectra of bovine rhodopsin. *J. Phys. Chem. Lett.* **11**, 6293–6300 (2019).
- Web, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5–6 (2016).
- Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. J. Improved prediction of protein side-chain conformations with SCWRLA. *Proteins Struct. Funct. Bioinformatics* **77**, 778–795 (2009).

### Acknowledgements

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

### Author contributions

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

### Competing interests

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02684-z>.

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

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

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021