PNAS
www.pnas.org

.

Supplementary Information for

High-Resolution Structural Insights into the Heliorhodopsin Family

K. Kovalev[1,2,3,4,5†], D. Volkov[2,3†], R. Astashkin[1,4†], A. Alekseev[2,3,4,5†], I. Gushchin[4], J. M. Haro-Moreno[10], I. Chizhov[11], S. Siletsky[12], M. Mamedov[12], A. Rogachev[4,8], T. Balandin[2,3], V. Borshchevskiy[4], A. Popov[7], G. Bourenkov[9], E. Bamberg[4,6], F. Rodriguez-Valera[10], G. Büldt[4] and V. Gordeliy[1,2,3,4*]

Corresponding author: Valentin Gordeliy
Email: valentin.gordeliy@ibs.fr

This PDF file includes:

Figures S1 to S22
Table S1

www.pnas.org/cgi/doi/10.1073/pnas.1915888117

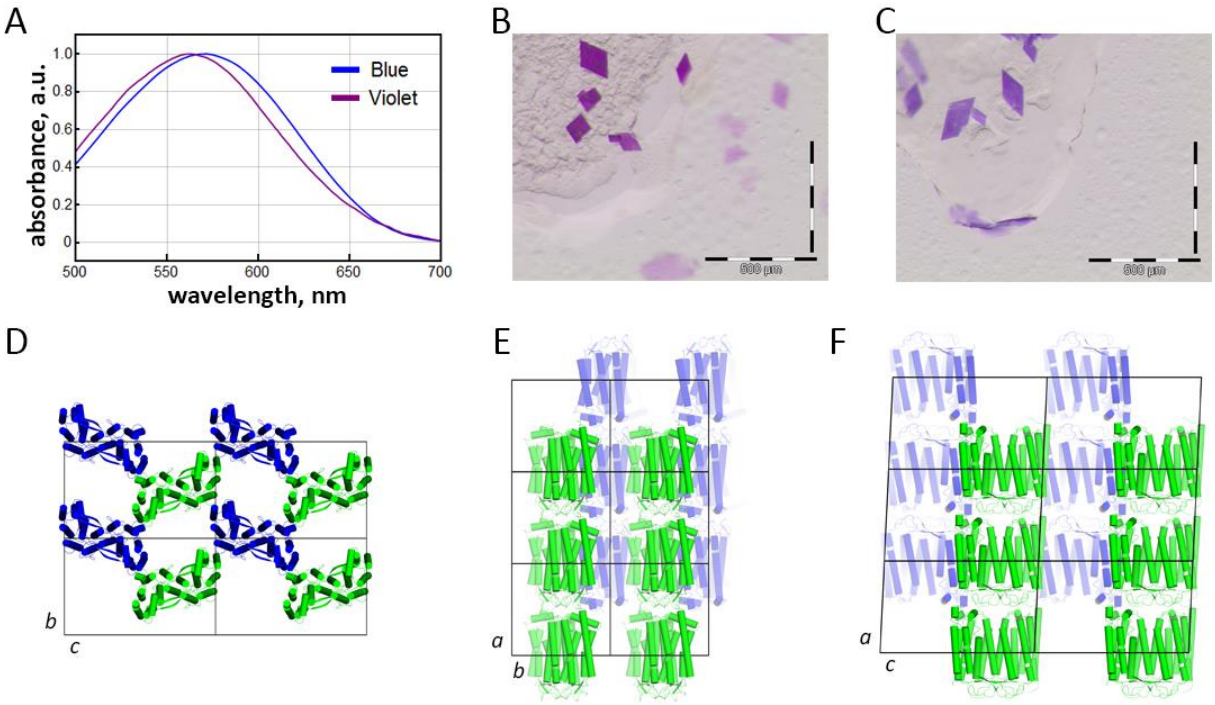

Fig. S1. Crystals and crystal packing of 48C12. A. Absorption spectra from 48C12 crystals of violet and blue forms. **B.** Example of violet crystals of 48C12. **C.** Example of violet crystals of 48C12. **D-F.** Crystal packing of both violet and blue forms of 48C12.

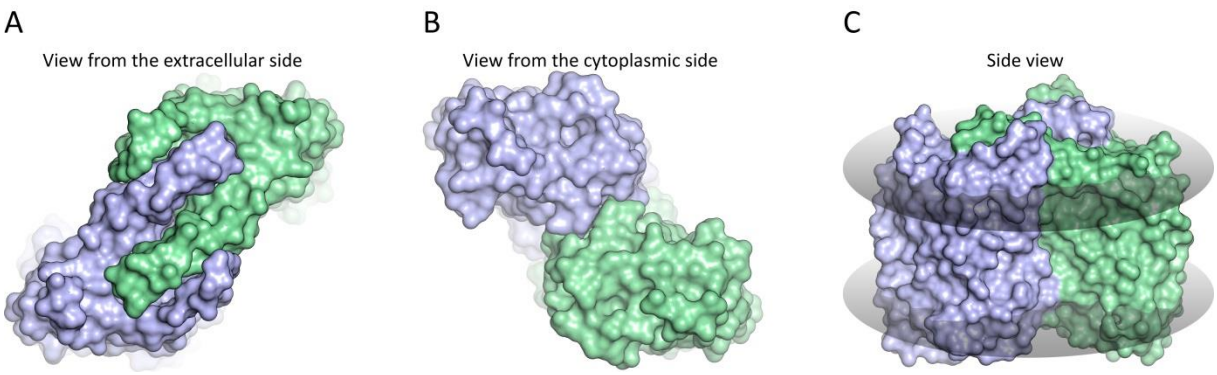

Fig. S2. Surface representation of the 48C12 dimer. A. View from the extracellular side. **B.** View from the cytoplasmic side. **C.** Side view in the membrane.

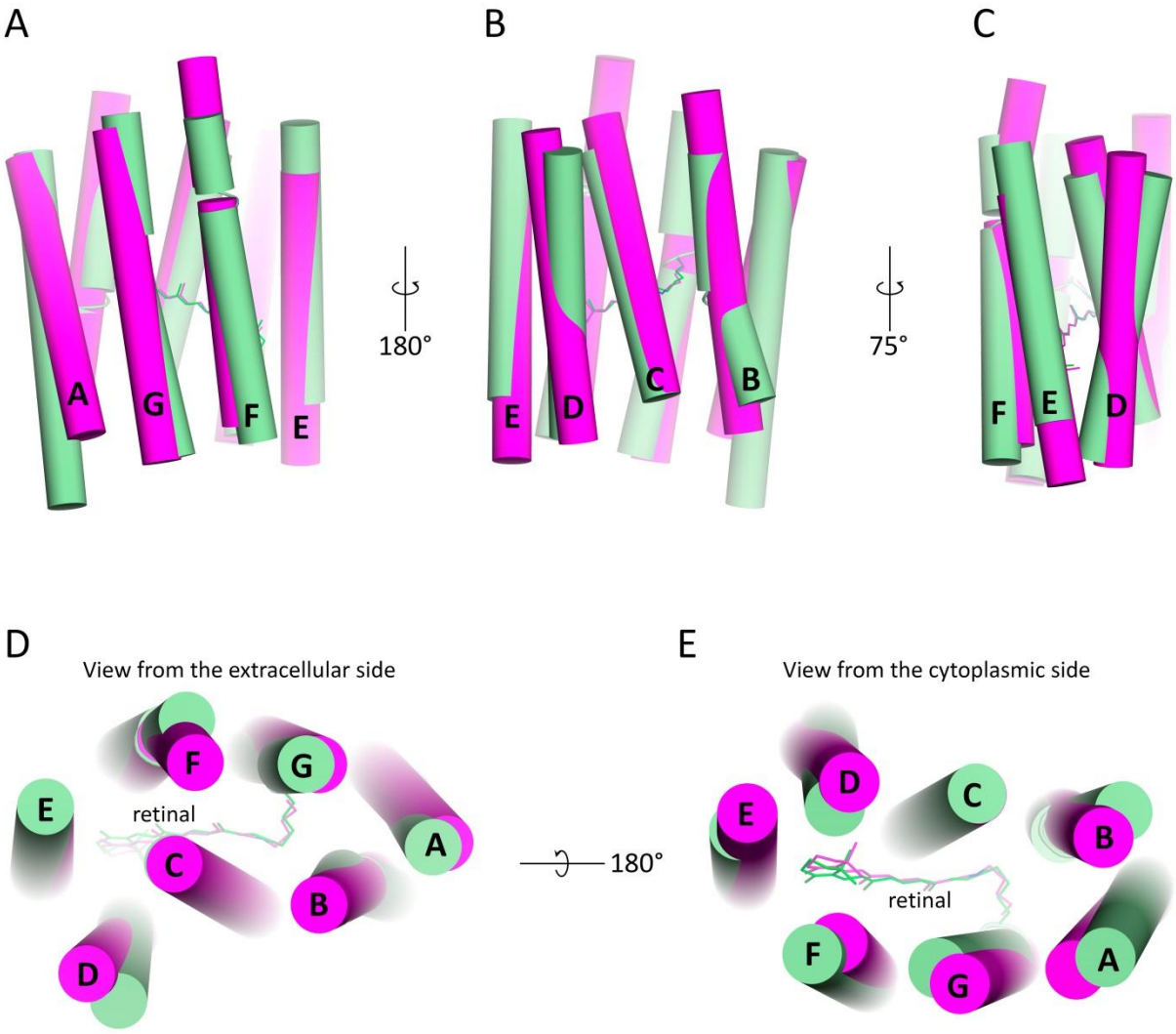

Fig. S3. Structure alignment of 48C12 (green) and bR (purple, PDB ID 1C3W). A. Side view from the side of helices F and G. **B.** Side view from the side of helices B and C. **C.** Side view from the side of helices D and E. **D.** View from the extracellular side. **E.** View from the cytoplasmic side.

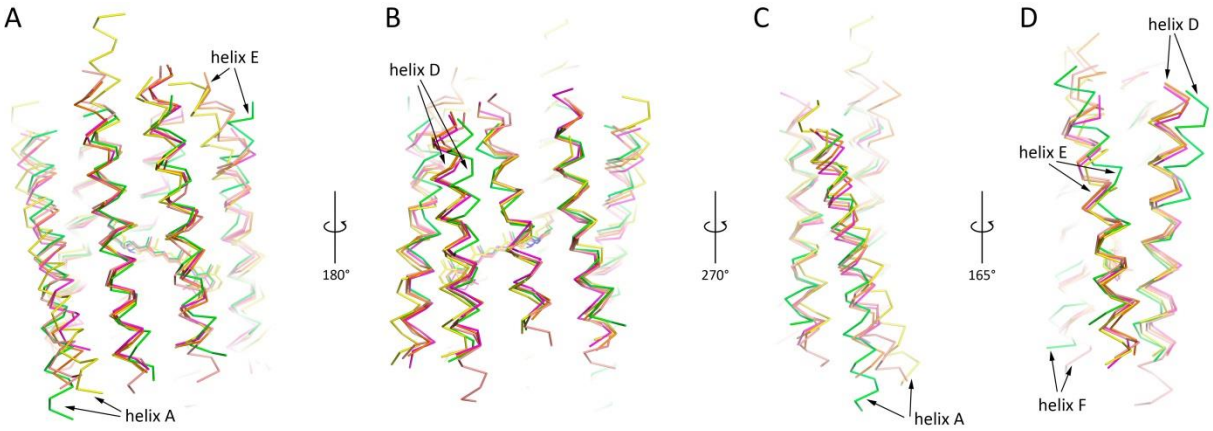

Fig. S4. Structure alignment of 48C12 (green), bR (purple, PDB ID 1C3W), NpSRII (orange, PDB ID 3QAP), ChR2 (yellow, PDB ID 6EID) and ESR (salmon, PDB ID 4HYJ). A. Side view from the side of helices F and G. **B.** Side view from the side of helices B and C. **C.** Side view from the side of helix A. **D.** Side view from the side of helices D and F. Black arrows indicate the most significant differences in helices location between 48C12 heliorhodopsin and representatives of type 1 rhodopsins. Extracellular side for 48C12 is at the top, cytoplasmic side is at the bottom of each figure section.

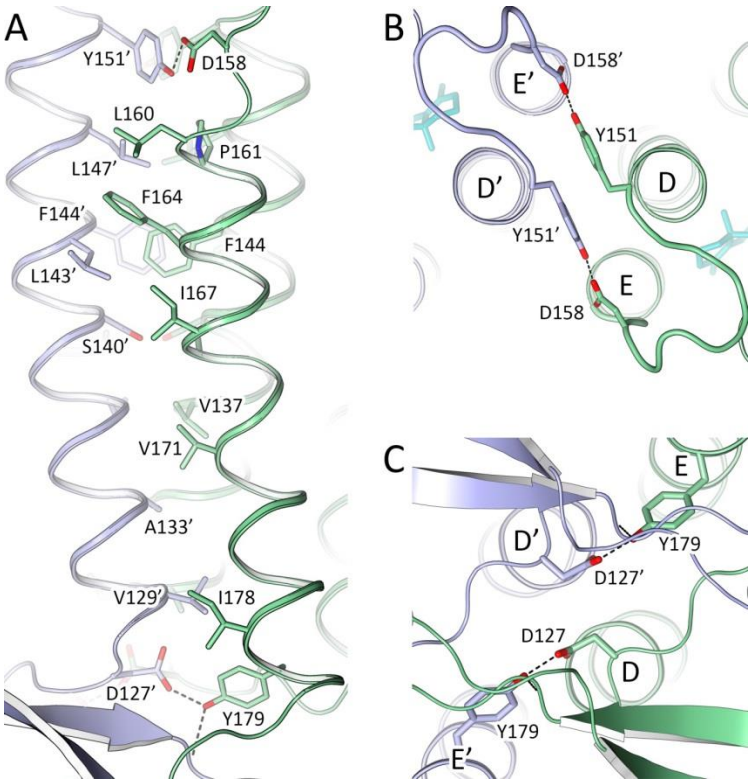

Fig. S5. Dimerization interface of 48C12. A. Side view of the oligomerization interface. Extracellular side is on the bottom. **B.** View from the cytoplasmic side. **C.** View from the extracellular side. The strongest contacts by hydrogen bonding are shown with dashed lines.

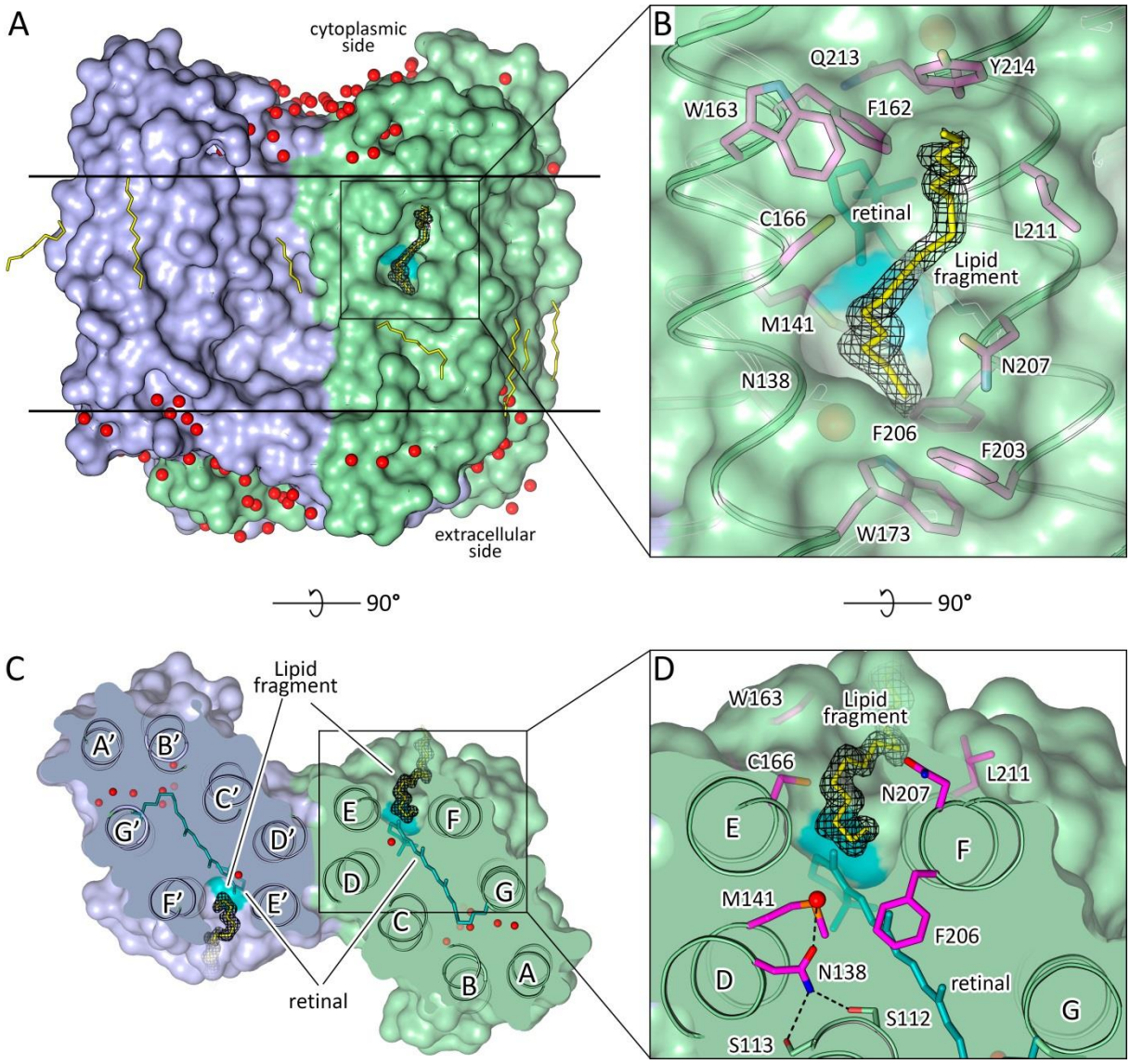

Fig. S6. Lipid molecule permeating into 48C12 protomer. A. Surface view of the 48C12 dimer in the membrane with lipid molecules. Cofactor retinal is colored teal. Membrane core boundaries are shown with black lines. **B.** Detailed side view of the lipid molecule inside the protomer. Residues, comprising the pocket for the molecule are colored violet. **C.** Section view from the extracellular side. Lipid molecule is deepened into the protomer between helices E and F. **D.** Detailed section view from the extracellular side. Hydrogen bonds are shown with black dashed lines. 2FoFc electron density maps around lipid molecule are contoured at the level of 1.5σ

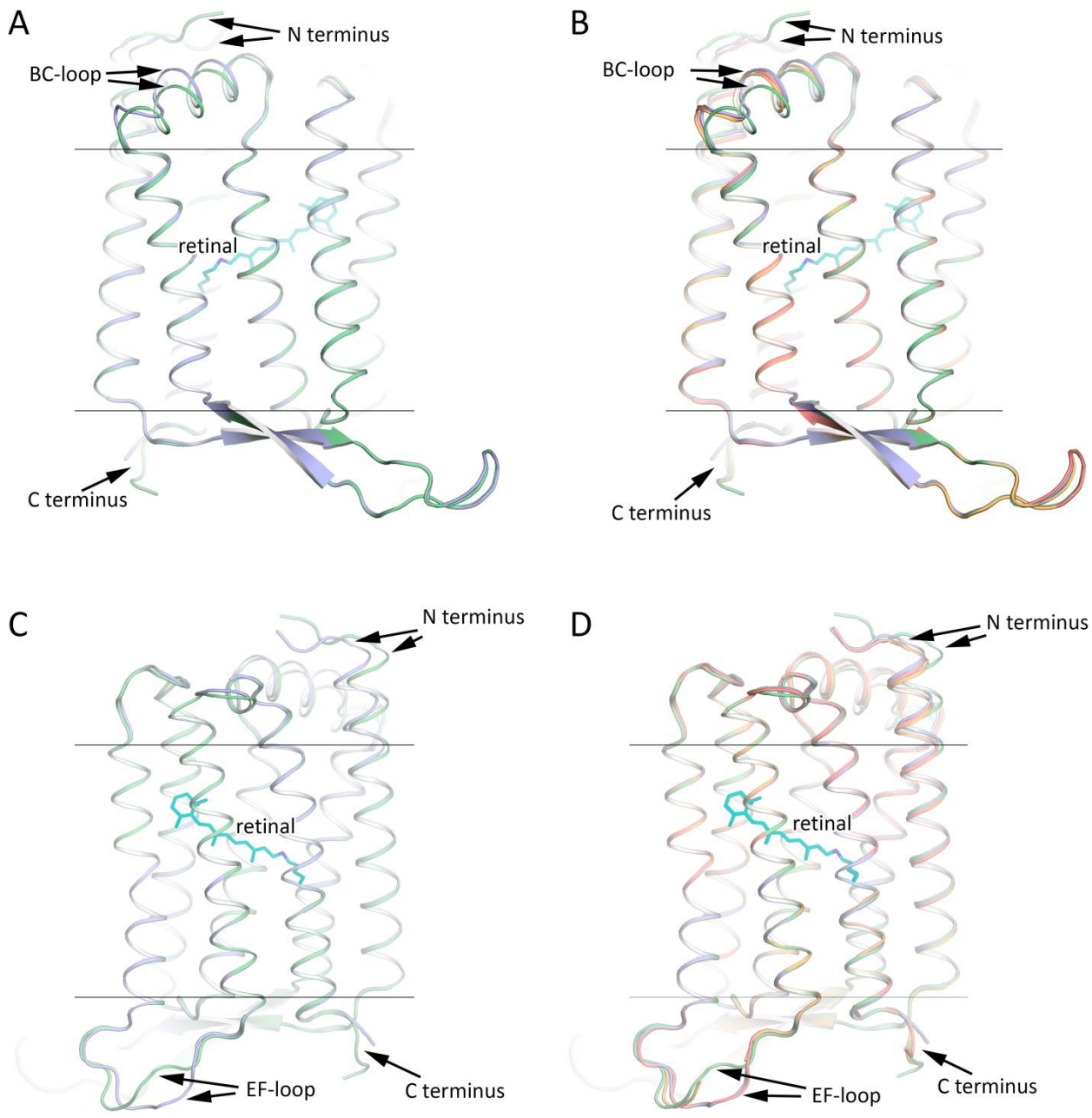

Fig. S7. Structure alignment of the backbones of 48C12 protomers. Alignment of protomer A (green) and protomer B (blue) at neutral pH. Side view from the side of helices B and C (**A**) and from the side of helices F and G (**C**). Alignment of protomer A (green) and protomer B (blue) at neutral pH and protomer A (orange) and protomer B (red) at acidic pH. Side view from the side of helices B and C (**B**) and from the side of helices F and G (**D**). Extracellular side for 48C12 is at the top, cytoplasmic side is at the bottom of each figure section.

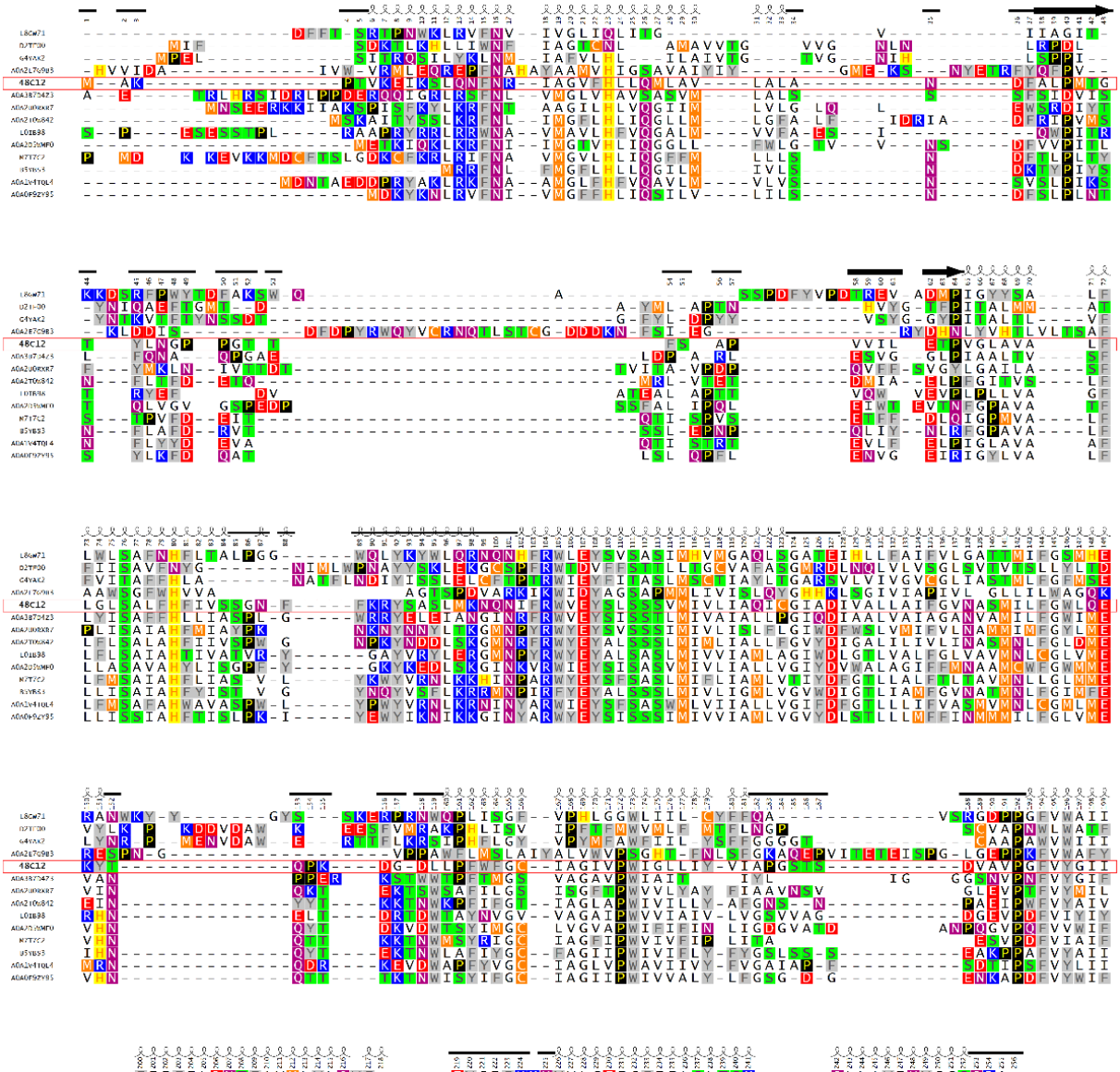

Fig. S8. Sequence alignment of heliorhodopsins of different classes (Heliorhodopsin 48C12 numbering). Proteins represent a different part of a phylogenetic tree and different domains of life. Heliorhodopsin 48C12 (A0A2R4S913 Uniprot ID), A0A3B7D4Z3 (Uniprot ID) and A0A2D5WMF0 - Bacteria>Terrabacteria group>Actinobacteria both, L8GW71 –Eukaryota>Amoebozoa, A0A2E7G9B3 - Bacteria>Proteobacteria, D2TF00 and G4YAK2 – Viruses both, L0IB98 - Archaea>Euryarchaeota, A0A2T0W842 - Bacteria>Terrabacteria group>Firmicutes, A0A2U0RXR7 - Archaea>TACK group, A0A0F9ZY95 - Bacteria>unclassified, M7T7C2 and A0A1V4TQL4 - Archaea>Euryarchaeota both, B5YBS3 - Bacteria>Dictyoglomi. The

shown region corresponds to the alignment part where heliorhodopsin 48C12 is fully represented, C and N termini of other heliorhodopsins are truncated. Helices for heliorhodopsin 48C12 are shown as coils, β-sheets as bold arrows, loops as plain lines.

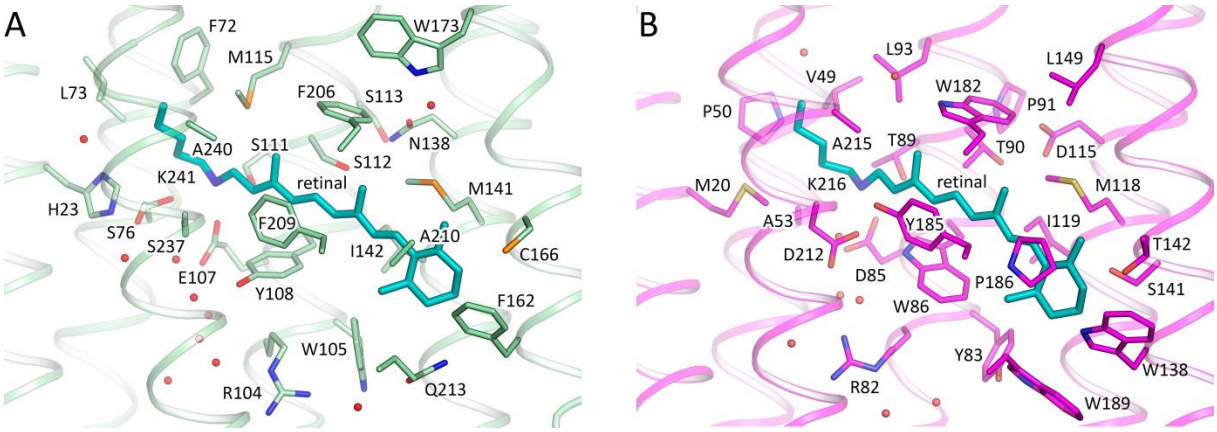

Fig. S9. Retinal binding pocket of 48C12 (**A)** and bR (**B)**. Cofactor retinal is colored teal.

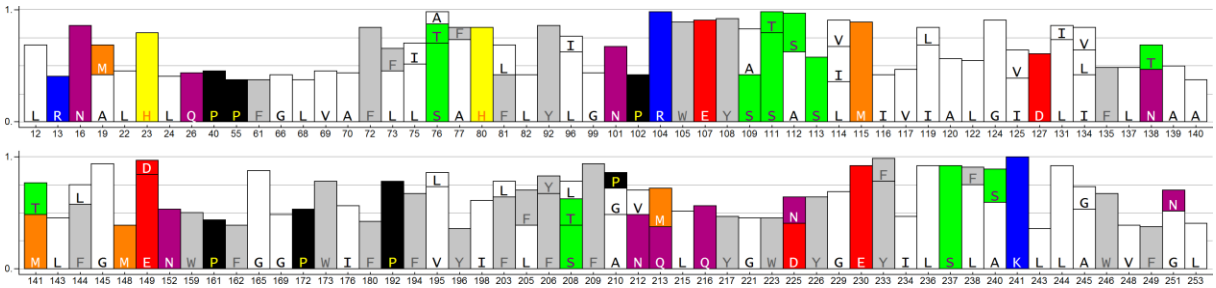

Fig. S10. Most conservative residues in whole heliorhodopsins family. Whole base of heliorhodopsins (n = 479) was first processed to remove all sequences with more that 50% identity (n = 64) not to exaggerate conservativity of residues that are conservative among big subclasses of heliorhodopsins.

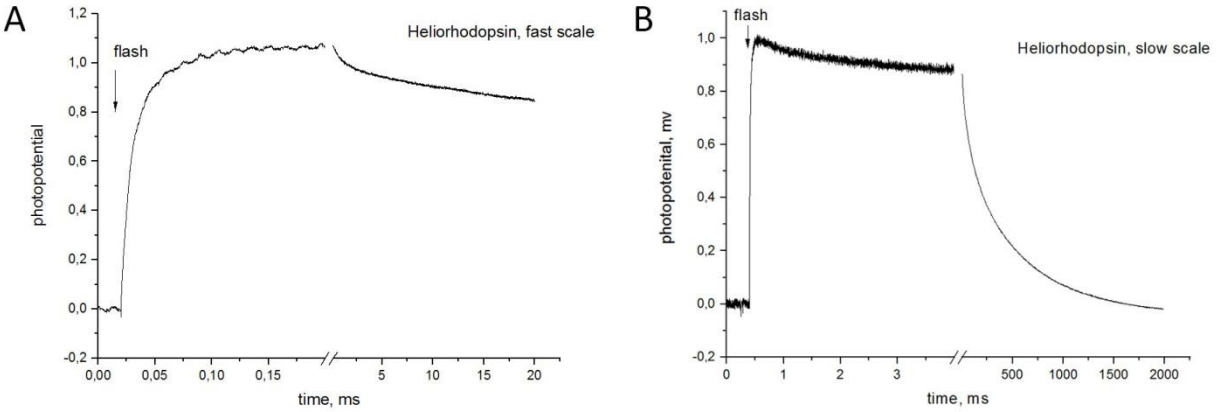

Fig. S11. Kinetics of light-induced changes in membrane potential difference, ΔΨ, in proteoliposomes containing 48C12 heliorhodopsin. **A:** Fast time scale. **B:** Slow time scale. Measurements with colloidin film-adhered liposomes with heliorhodopsin were conducted in the buffer containing 20 mM Hepes (pH 7.5, NaOH).

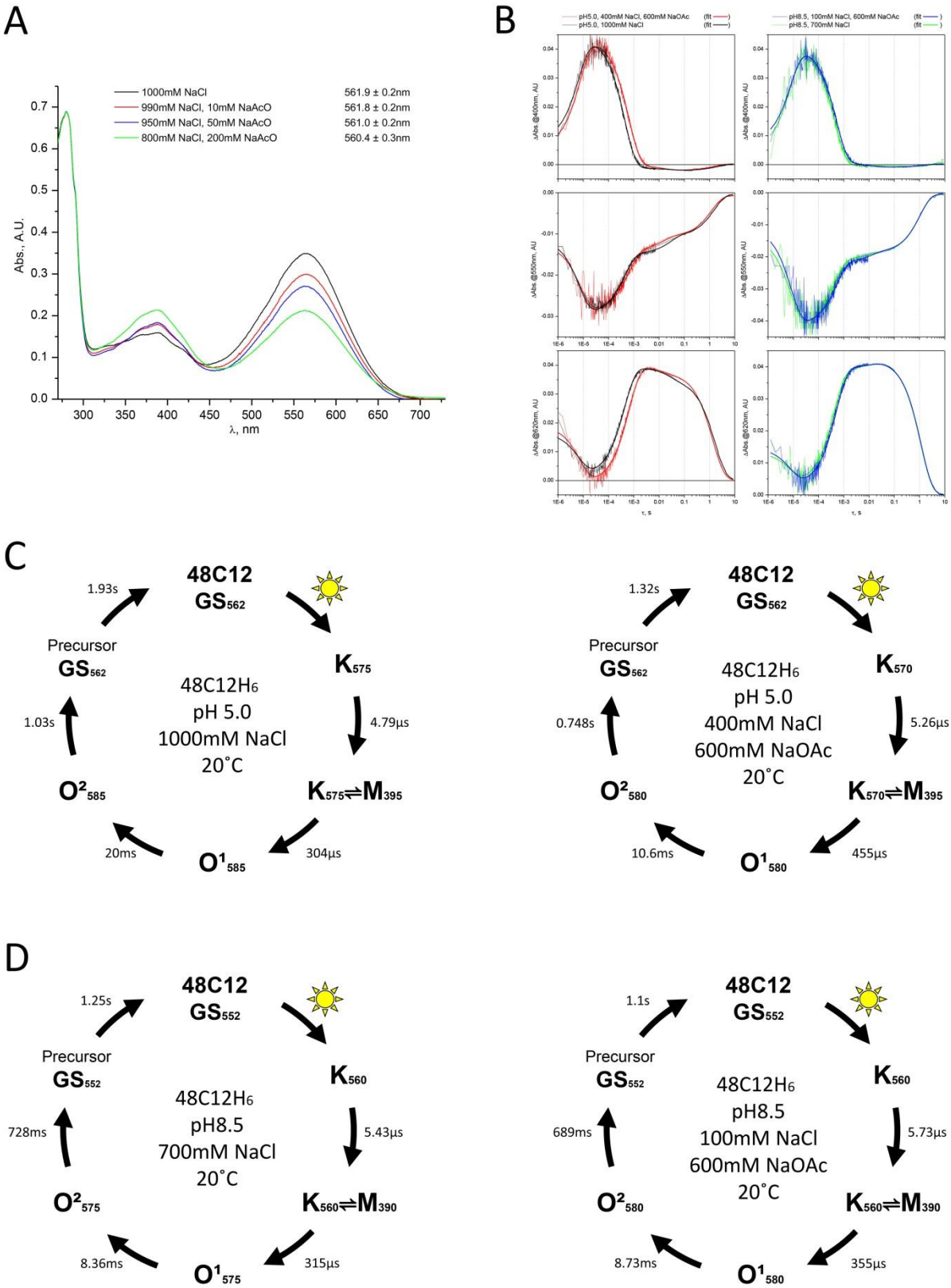

Fig. S12. Spectroscopy of 48C12. A. UV-Vis absorption spectra of the 48C12 at pH 5.0 in the presence of acetate. Retinal dissociation (left peak) is also observed at acidic pH

during dialysis against buffers containing acetate. **B.** Time evolution of the transient absorption changes of photo-excited 48C12 at pH 5.0 (black, red) and 8.5 (green, blue) with (red, blue) and without (black, green) acetate. **C, D.** Proposed photocycles of the 48C12 at pH 5.0 and 8.5 with and without acetate.

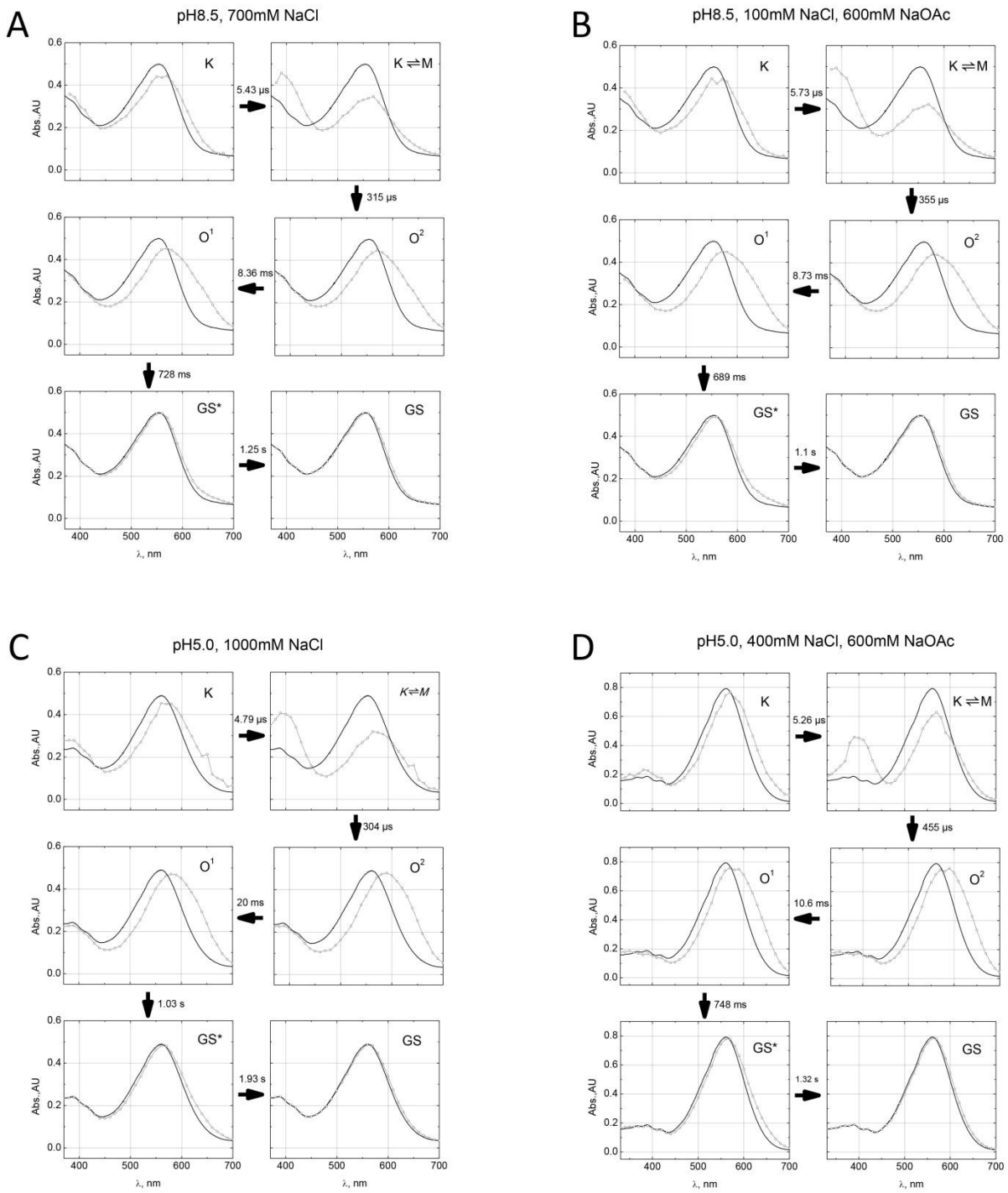

Fig. S13. UV-Vis spectra of intermediate states of 48C12 photocycle at pH 8.5 without acetate (**A**), with acetate (**B**) and at pH 5.0 without acetate (**C**) and with acetate (**D**). Five kinetically distinct protein states (gray lines) are obtained via global multi exponential analysis of the flash photolysis data exemplified in the Fig.S12. An each panel contains for the reference the correspondent spectrum on unexcited protein (black lines). The spectra of intermediate states were calculated from correspondent spectra of exponents, which were further converted to the differential spectra of the states

assuming the sequential irreversible model of the photocycle. The half-times of reactions are depicted between the panels and are shown with black arrows.

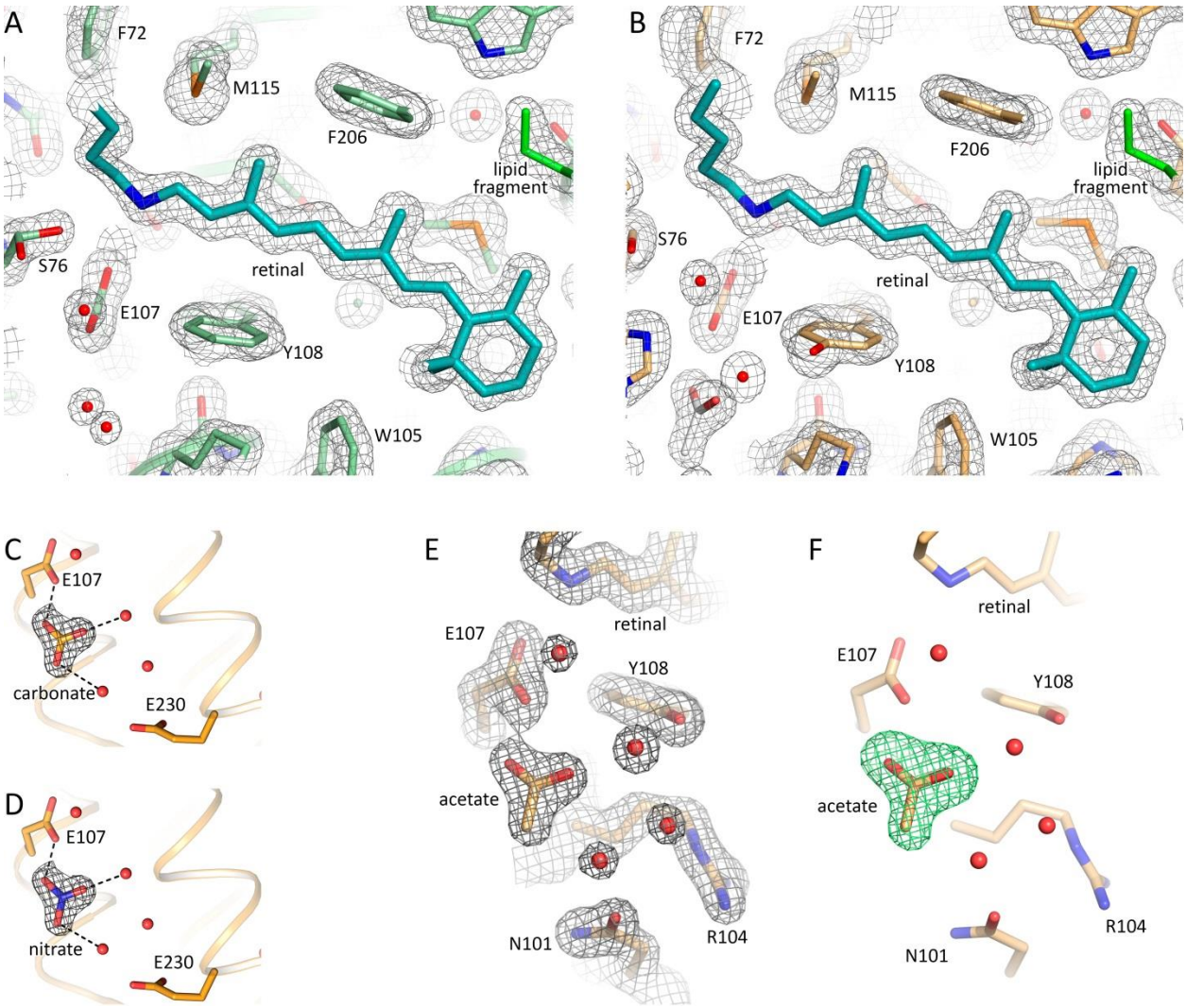

Fig. S14. Examples of electron density maps of 48C12 model. A, B. 2Fo-Fc electron density maps near the retinal of the chain A of violet and blue forms, respectively. The maps are contoured at the level of 1.5σ. **C, D.** Examples of 2Fo-Fc electron density maps of the active site of the protein after hypothetical fitting of the triangle difference electron densities with carbonate and nitrate molecules, respectively. Carbonate and nitrate molecules were placed for modelling at the position of the acetate in the blue form of 48C12. Putative hydrogen bonds are shown with black dashed lines. The maps are contoured at the level of 1.5σ. **E.** 2Fo-Fc electron density maps of the active site and acetate molecule in the blue form of the 48C12 heliorhodopsin. The maps are contoured at the level of 1.5σ. **F.** Simulated annealing omit Fo-Fc difference electron density maps calculated using *phenix.polder* omitting the acetate molecule. The maps are contoured at the level of 3σ.

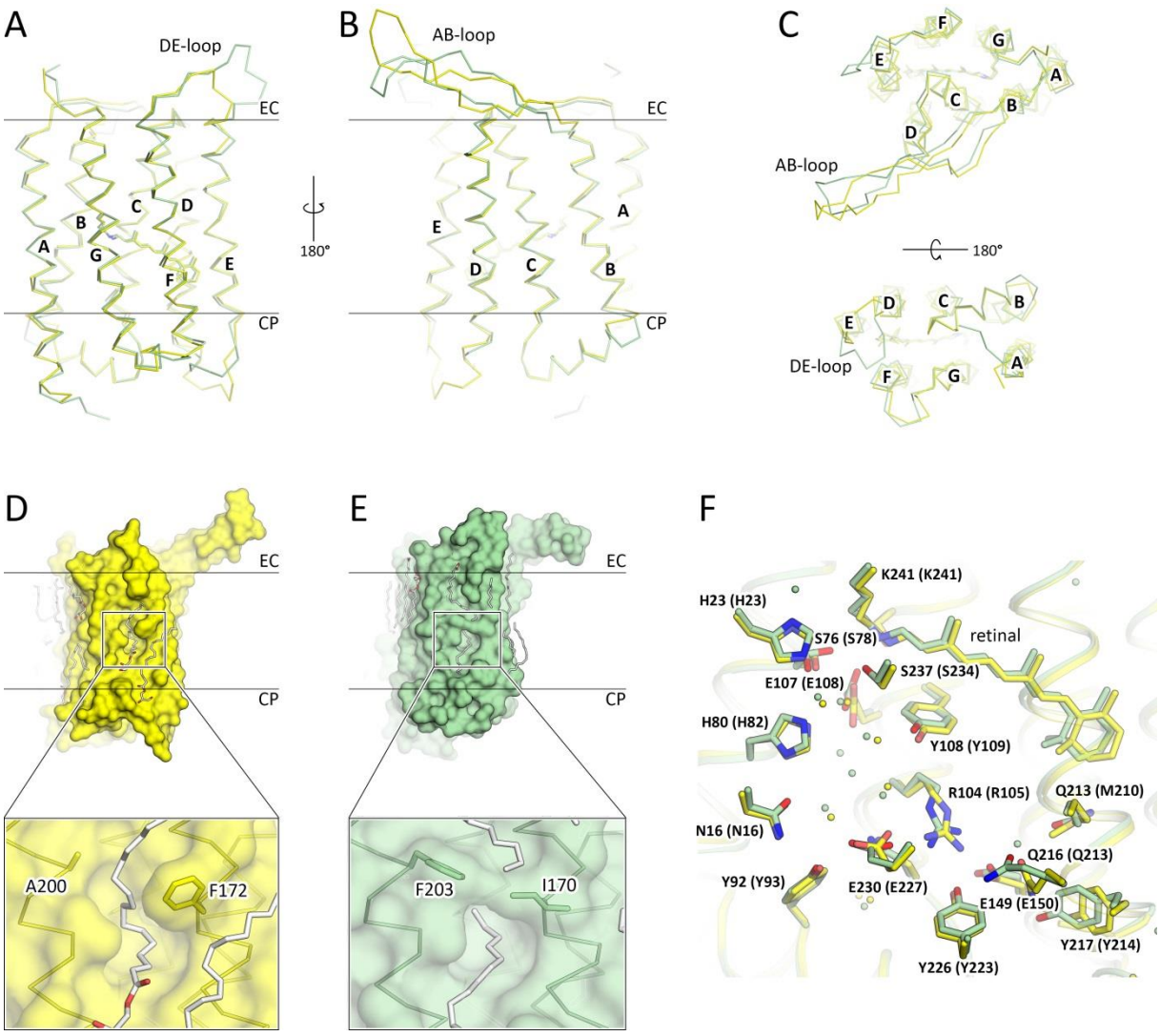

Fig. S15. Comparison of 48C12 and *Ta*HeR structures. A-C. Structural alignment of 48C12 (green, present work) and *Ta*HeR (yellow, PDB ID: 6IS6). Side views (**A, B**) and view from the extracellular (**C**, top) and cytoplasmic (**C**, bottom) sides. **D-E.** Comparison of the fenestrations of the 48C12 and *Ta*HeR surfaces near retinal cofactor. Surface of only one protomer of protein dimer is shown. **F.** Detailed view of the structural alignment of the 48C12 and *Ta*HeR cytoplasmic inner region. Residues are labeled according to the 48C12 (*Ta*HeR) numbering. Membrane core boundaries are shown with gray lines. Waters are shown with green and yellow spheres for 48C12 and *Ta*HeR, respectively.

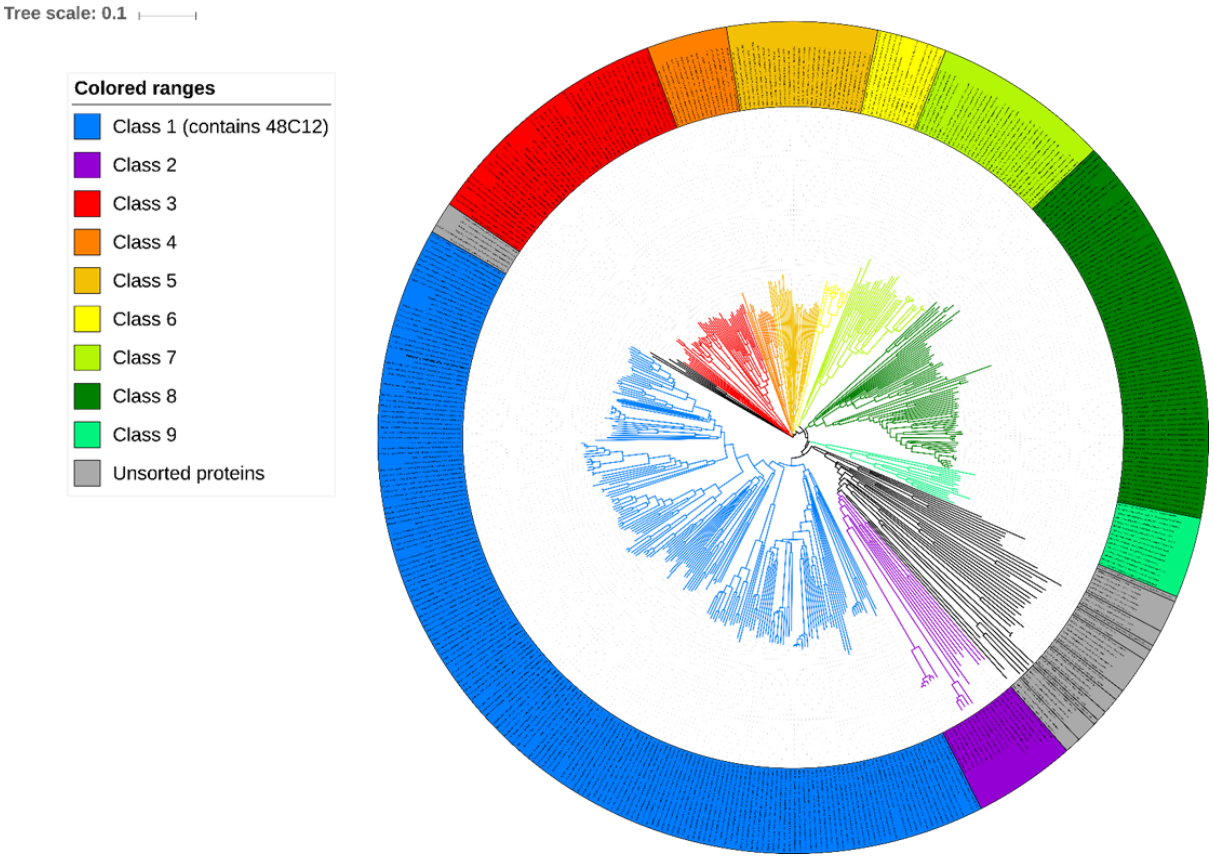

Fig. S16. Phylogenetic tree for heliorhodopsins. Clades whose average branch length distance to leaves if below 0.397 are collapsed into classes. Classes containing more than 10 proteins assigned with index number and color (see legend). Classes containing less than 10 proteins assigned with gray color and “Unsorted proteins” label. Analyzed base of heliorhodopsins consists of proteins from original article[25] and proteins predicted with InterPro.

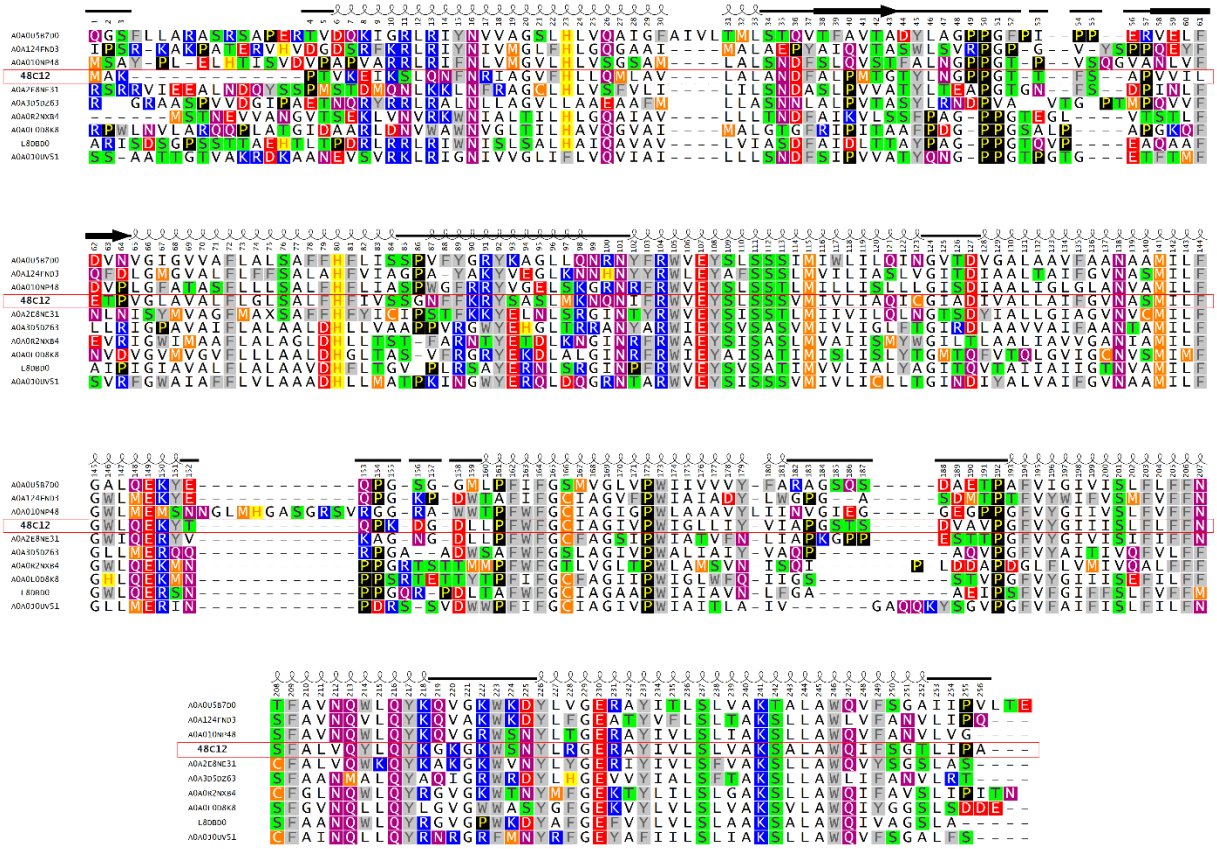

Fig. S17. **Sequence alignment of heliorhodopsin subfamily containing 48C12 heliorhodopsin (Heliorhodopsin 48C12 numbering).** Proteins presented on the alignment are the most distant members of the subfamily 1. A0A0L0D8K8 – eukaryota, 48C12, A0A010NP48, A0A0J0UV51, A0A0R2NXB4, A0A0U5B7D0, A0A124FND3, A0A2E8NE31, A0A3D5DZ63 and L8DBD0 - Bacteria>Terrabacteria group>Actinobacteria all. The shown region corresponds to the alignment part where heliorhodopsin 48C12 is fully represented, C and N termini of other heliorhodopsins are truncated. Helices for heliorhodopsin 48C12 are shown as coils, β-sheets as bold arrows, loops as plain lines.

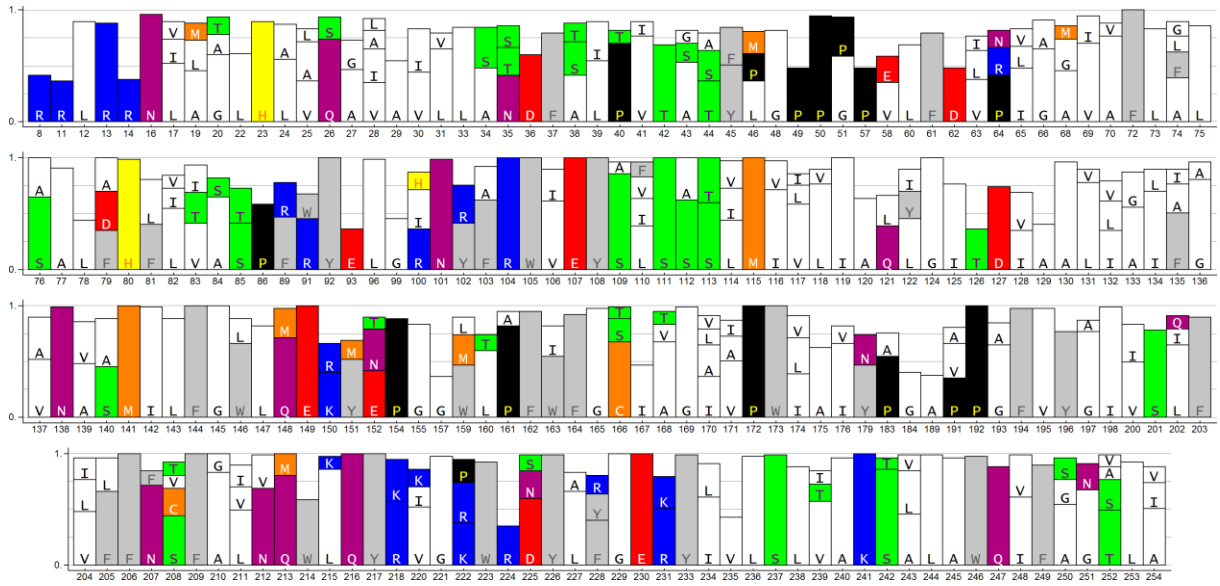

Fig. S18. Most conservative residues in heliorhodopsins subfamily containing 48C12. Subfamily base (n = 195) was firstly processed to remove all sequences with more that 75% identity (n = 77) not to exaggerate conservativity of residues conservative in big subclasses of this subclass.

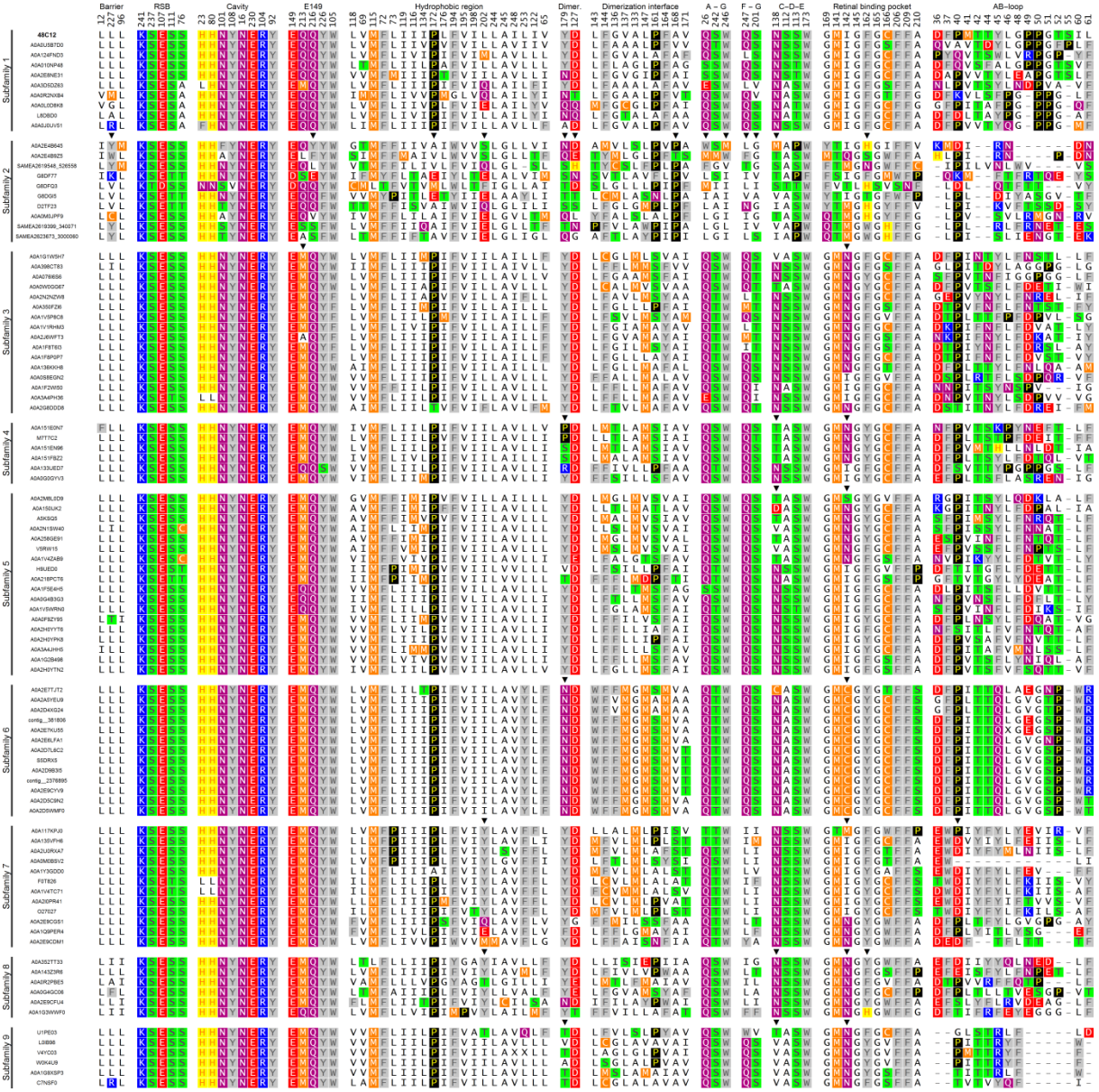

Fig. S19. Alignment of HeRs from different classes. Shown positions chosen for structurally important regions: “Hydrophobic barrier between inner polar cavity and cytoplasm”, “Retinal and RSB-interacting residues”, “Residues, comprising inner polar cavity at the cytoplasmic part”, “Polar region at the cytoplasmic side near E149”, “Polar residues at the dimerization interface, responsible for contacts”, “Hydrophobic residues at the dimerization interface inside the membrane”, ”Cluster between A and G helices at the extracellular side of the protein”, “Cluster between F and G helices at the extracellular side of the protein”, “Cluster near β-ionone ring of retinal between helices C, D, and E with 1 water molecule”, “Retinal binding pocket”, “AB-loop”

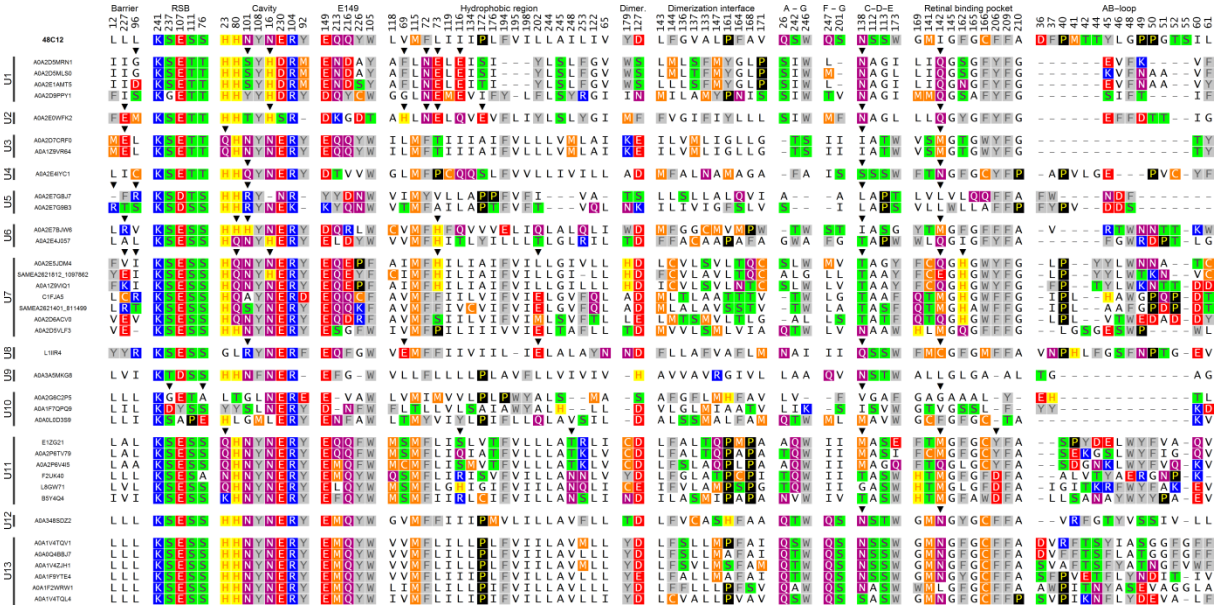

Fig. S20. Alignment of HeRs from different classes from “Unsorted proteins”. Shown positions chosen for structurally important regions: “Hydrophobic barrier between inner polar cavity and cytoplasm”, “Retinal and RSB-interacting residues”, “Residues, comprising inner polar cavity at the cytoplasmic part”, “Polar region at the cytoplasmic side near E149”, “Polar residues at the dimerization interface, responsible for contacts”, “Hydrophobic residues at the dimerization interface inside the membrane”, ”Cluster between A and G helices at the extracellular side of the protein”, “Cluster between F and G helices at the extracellular side of the protein”, “Cluster near β-ionone ring of retinal between helices C, D, and E with 1 water molecule”, “Retinal binding pocket”, “AB-loop”.

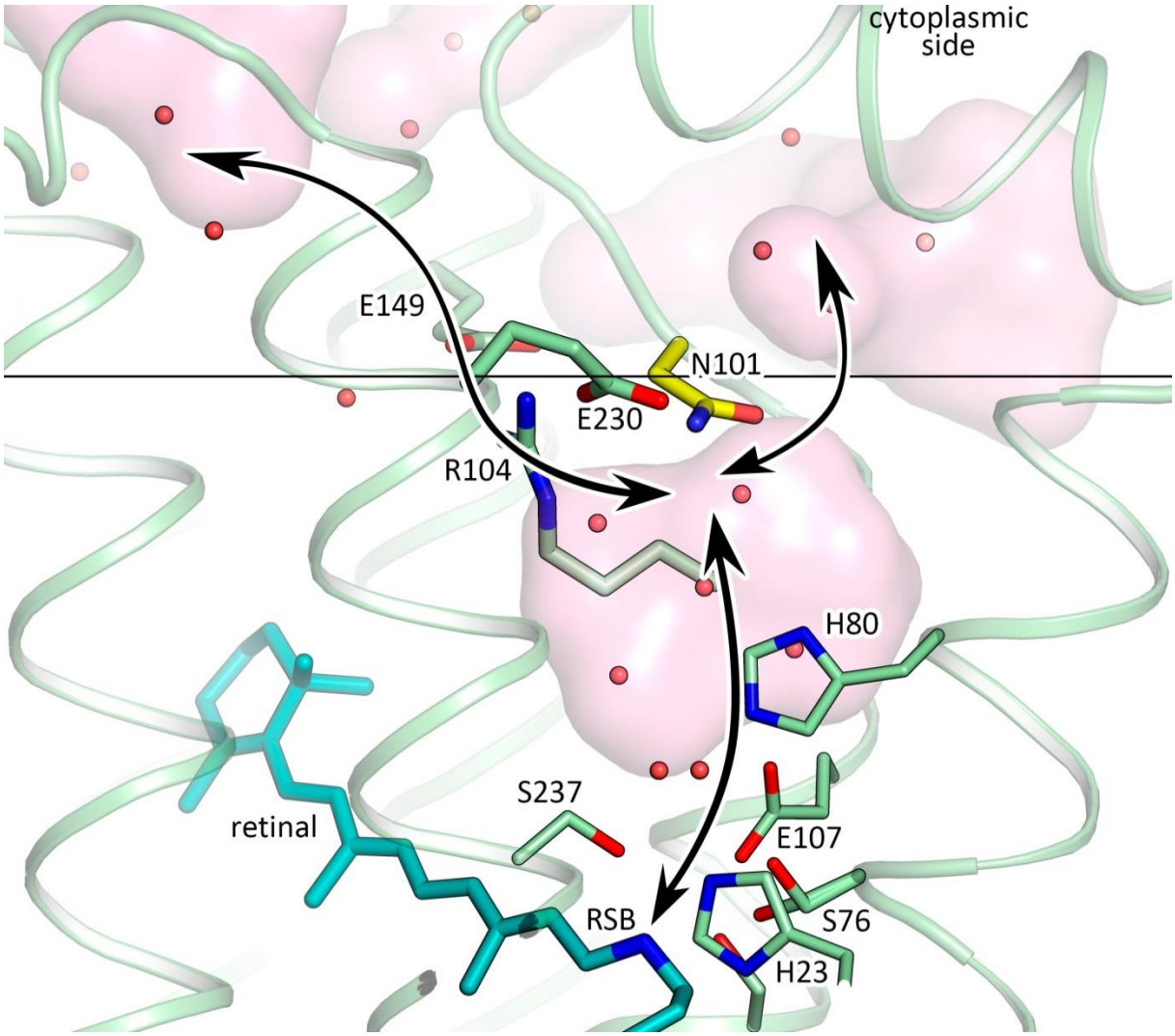

Fig. S21. Putative pathway of the proton in 48C12 heliorhodopsin. Black arrows indicate the putative way of the protons. Membrane core boundary at the cytoplasmic side is shown with black line. Asn101 residue is colored yellow.

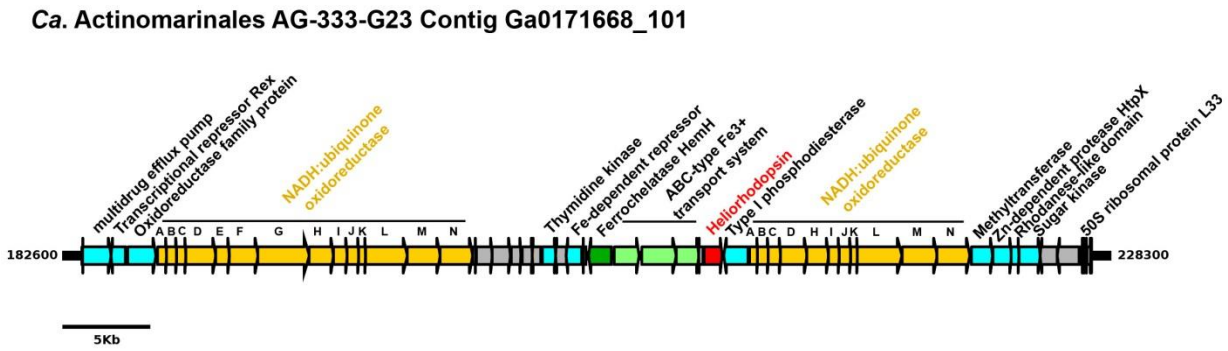

Fig. S22. Linear representation of a genome fragment of 45.7 Kb retrieved from the marine *Ca.* Actinomarinales AG-333-G23 contig Ga0171668_101. Proteins were annotated against the NCBI-nr and InterPro databases. Genes coloured in grey represent hypothetical proteins. In red is coloured and highlighted the heliorhodopsin protein. A scale of 5 Kb long is represented in a black line below.

Table S1. Data collection and refinement statistics of the 48C12.

	Violet form	*Blue form*
Data collection		
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	56.12, 59.75, 94.32	56.18, 60.74, 92.91
α,β,γ (°)	90, 92.30, 90	90, 92.02, 90
Wavelength (Å)	0.978	0.978
Resolution (Å)	49.06-1.50 (1.53-1.50)	48.81-1.50 (1.53-1.50)
R_{merge} (%)	5.3 (113.8)	7.0 (158.6)
R_{pim} (%)	3.4 (71.1)	4.1 (93.3)
$I/\sigma I$	8.3 (0.8)	8.9 (0.7)
$CC_{1/2}$ *(%)*	99.9 (58.3)	99.9 (33.3)
Completeness (%)	98.6 (97.4)	99.6 (95.0)
Multiplicity	3.4 (3.5)	3.8 (3.8)
Unique reflections	98,436 (4795)	99,767 (4713)
Refinement		
Resolution (Å)	20-1.50	20-1.50
No. reflections	93,455	94,710
R_{work}/ R_{free} (%)	15.3/19.9	17.3/20.1
No. atoms		
Protein	4066	4009
Water	233	162
Lipid fragments	409	420
Retinal	40	40
Sulfate	5	10
Acetate	-	8
B factors (Å^2)		
Protein	25	27
Water	36	35
Lipid fragments	36	53
Retinal	18	17
Sulfate	61	90
Acetate	-	29
R.m.s deviations		
Protein bond lengths (Å)	0.005	0.004
Protein bond angles (°)	1.19	1.16
Ramachandran analysis		
Favored (%)	98.0	97.5
Outliers (%)	0.4	0.4