nature portfolio

https://doi.org/10.1038/s41594-022-00783-x

## Supplementary information

# Rhodopsin-bestrophin fusion proteins from unicellular algae form gigantic pentameric ion channels

In the format provided by the authors and unedited

# Supplementary Information

# Rhodopsin-bestrophin fusion proteins from unicellular algae form gigantic pentameric ion channels

Andrey Rozenberg[1†], Igor Kaczmarczyk[2†], Donna Matzov[2†], Johannes Vierock[3†], Takashi Nagata[4,5], Masahiro Sugiura[6], Kota Katayama[5,6,7], Yuma Kawasaki[4], Masae Konno[4,5], Yujiro Nagasaka[4], Mako Aoyama[6], Ishita Das[8], Efrat Pahima[9], Jonathan Church[9], Suliman Adam[9], Veniamin A. Borin[9], Ariel Chazan[1], Sandra Augustin[3], Jonas Wietek[10], Julien Dine[10], Yoav Peleg[11], Akira Kawanabe[12], Yuichiro Fujiwara[12], Ofer Yizhar[10], Mordechai Sheves[8], Igor Schapiro[9], Yuji Furutani[6,7], Hideki Kandori[6,7], Keiichi Inoue[4], Peter Hegemann[3], Oded Béjà[1]*, Moran Shalev-Benami[2]*

[†]These authors contributed equally to this work: Andrey Rozenberg, Igor Kaczmarczyk, Donna Matzov, Johannes Vierock
*Corresponding authors. beja@technion.ac.il (O.B.); moransb@weizmann.ac.il (M.S.-B.)

[1]Faculty of Biology, Technion — Israel Institute of Technology, Haifa 3200003, Israel.
[2]Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 761001, Israel.
[3]Institute for Biology, Experimental Biophysics, Humboldt-Universität zu Berlin, Berlin 10115, Germany.
[4]The Institute for Solid State Physics, The University of Tokyo, Kashiwa, Chiba 277-8581, Japan.
[5]PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan.
[6]Department of Life Science and Applied Chemistry, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan.
[7]OptoBioTechnology Research Center, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan.
[8]Department of Molecular Chemistry and Materials Science, Weizmann Institute of Science, Rehovot 761001, Israel.
[9]Fritz Haber Center for Molecular Dynamics Research Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel.
[10]Department of Neurobiology, Weizmann Institute of Science, Rehovot 761001, Israel.
[11]Structural Proteomics Unit (SPU), Life Sciences Core Facilities (LSCF), Weizmann Institute of Science, Rehovot 7610001, Israel.
[12]Laboratory of Molecular Physiology & Biophysics, Faculty of Medicine, Kagawa University, Miki-cho, Kagawa 761-0793, Japan.

## This document contains:

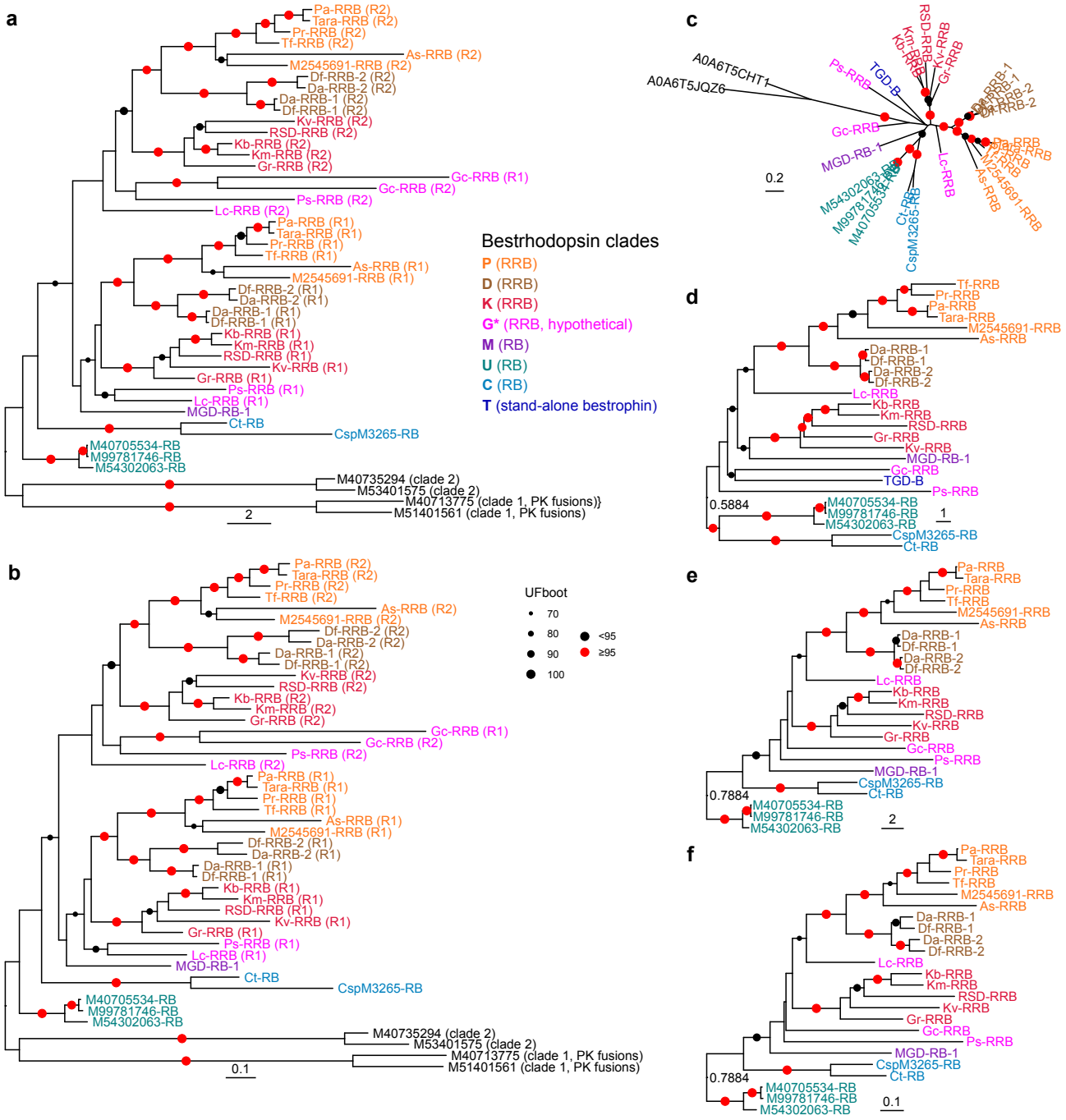
Supplementary Figures 1-13

## Other supplementary material includes:

Supplementary Data File 1 (separate spreadsheet file). ***Sheet "Genes"***: Bestrhodopsin and bestrhodosin-related genes collected. ***Sheet "Searched data"***. Dinoflagellates and other algae searched for the presence of bestrhodopsin genes.

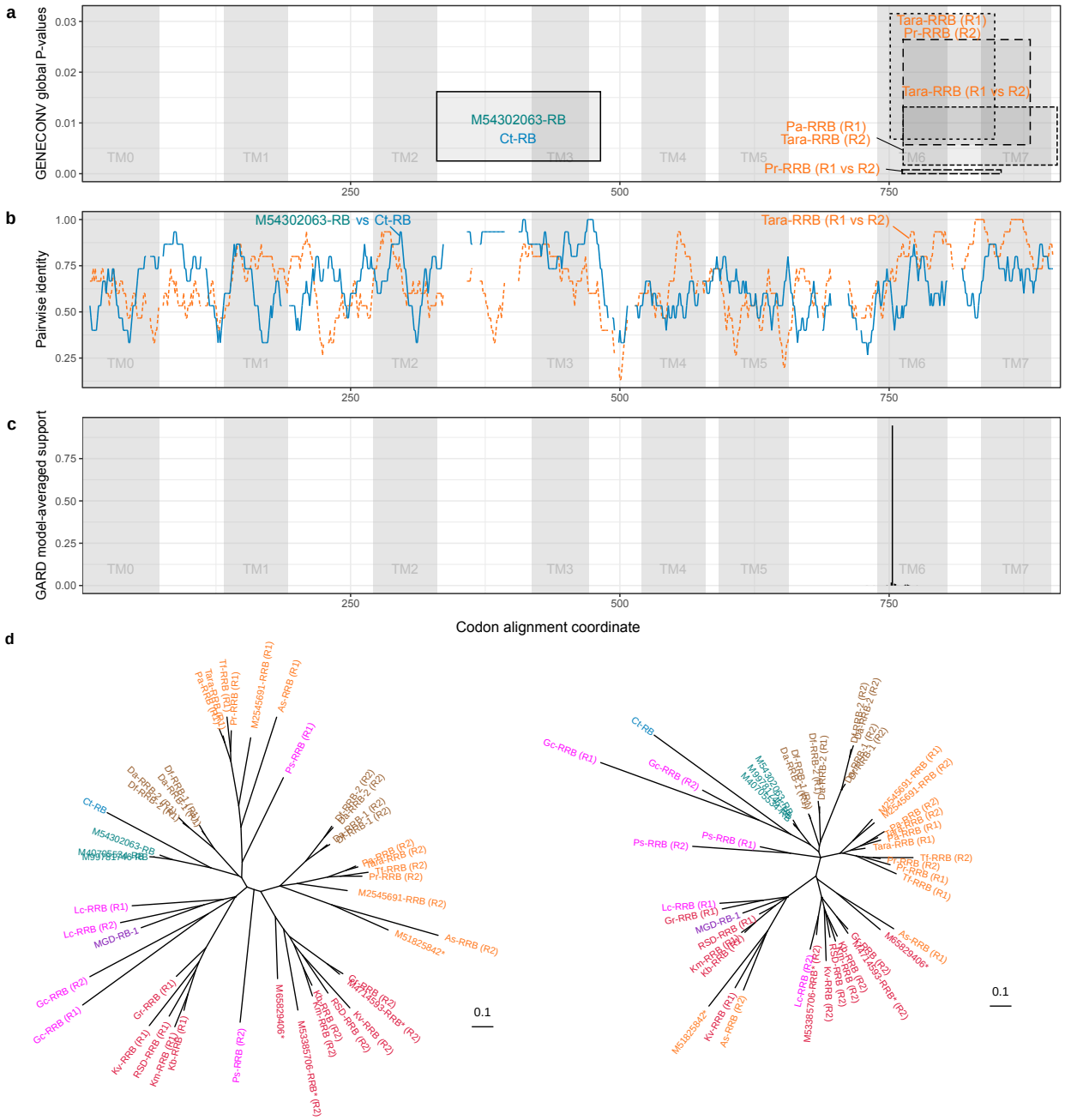
Supplementary Data File 2 (separate file in Genbank format). Collected sequences of the bestrhodopsin genes with descriptions of data sources and annotated CDS regions.

Supplementary Data File 3 (separate file in Genbank format). Collected sequences of bestrhopdopsin-related genes with annotated CDS regions.



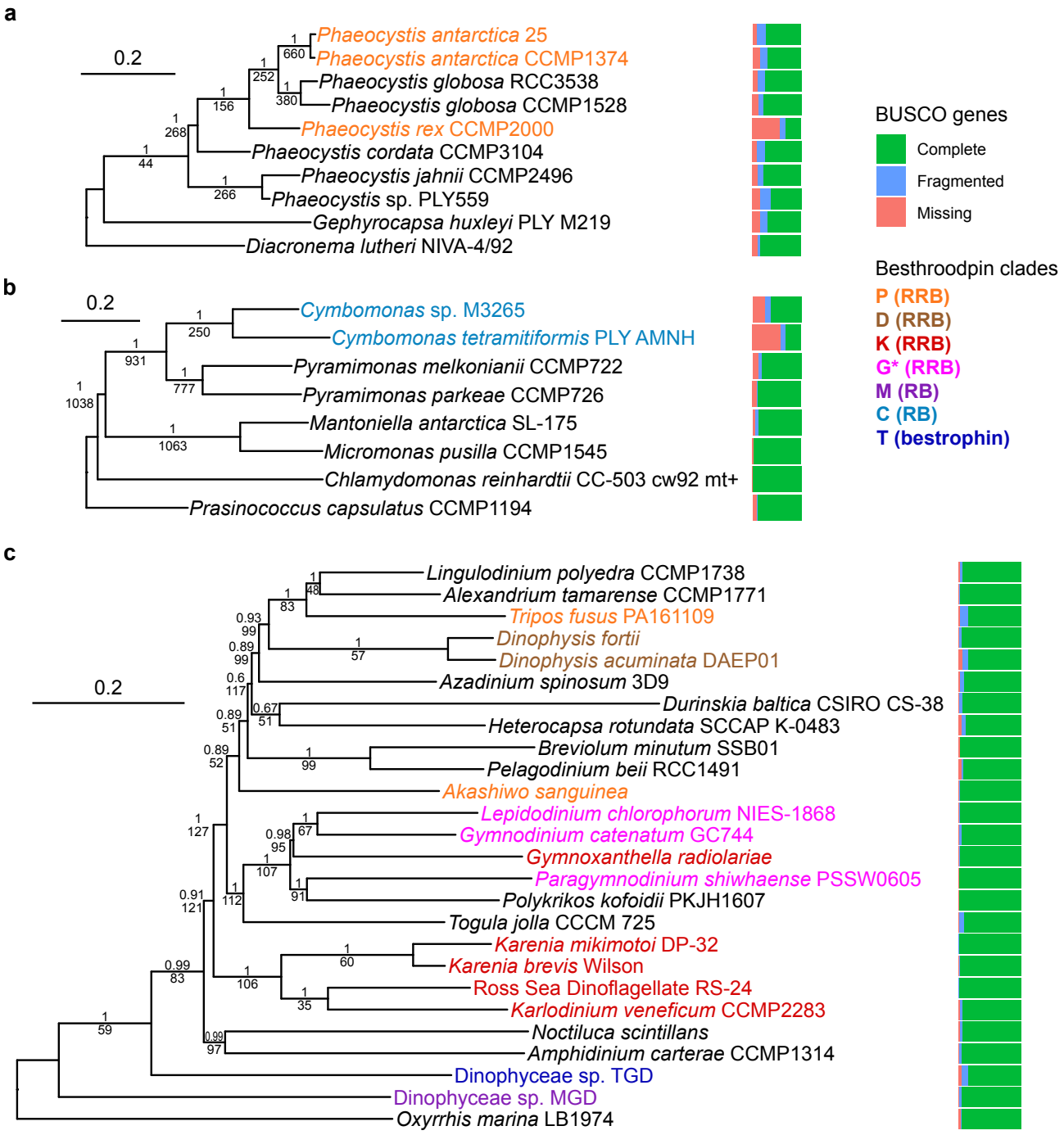

**Supplementary Fig. 1**. **Phylogenetic relationships between rhodopsin and bestrophin domains in bestrhodopsins. a, b**, Phylogeny of the rhodopsin domains based on the codon alignment (KOSI07+F+R5 as the best-fit substitution model) (**a**) and the same topology with branch lengths optimized based on the protein alignment (LG+G4 model) (**b**). The trees are outgroup-rooted using two of the three metatranscriptomic rhodopsin clades of dinoflagellate origin (judging by the presence of dinoflagellate spliced leader RNA in the transcripts), including a clade of 8TM rhodopsins fused with a protein kinase domain (PK). The outgroup status of the three clades is based on the global phylogeny of 8TM rhodopsins (see Extended Data Fig. 1). **c-f**, Phylogeny of the bestrophin domains. **c,** Attempt at outgroup-rooting using the two bestrophin proteins from *Azadinium spinosum* (A0A6T5JQZ6 and A0A6T5CHT1) that are most similar to

bestrhodopsin bestrophin domains. Protein alignment, LG+G4 model. **d**, Phylogeny based on the codon alignment including the stand-alone bestrophin from Dinophyceae sp. TGD (TGD-B) (KOSI07+F3X4+R4 model). **e, f**, Phylogeny of the bestrophin domains based on the codon alignment excluding TGD-B (KOSI07+F3X4+R3 model) (**e**) and the same topology with branch lengths optimized based on the protein alignment (LG+G4 model) (**f**). The trees in (**d-f**) were rooted using Root Digger with the numbers next to the root indicating Likelihood Weight Ratio of the root placement.



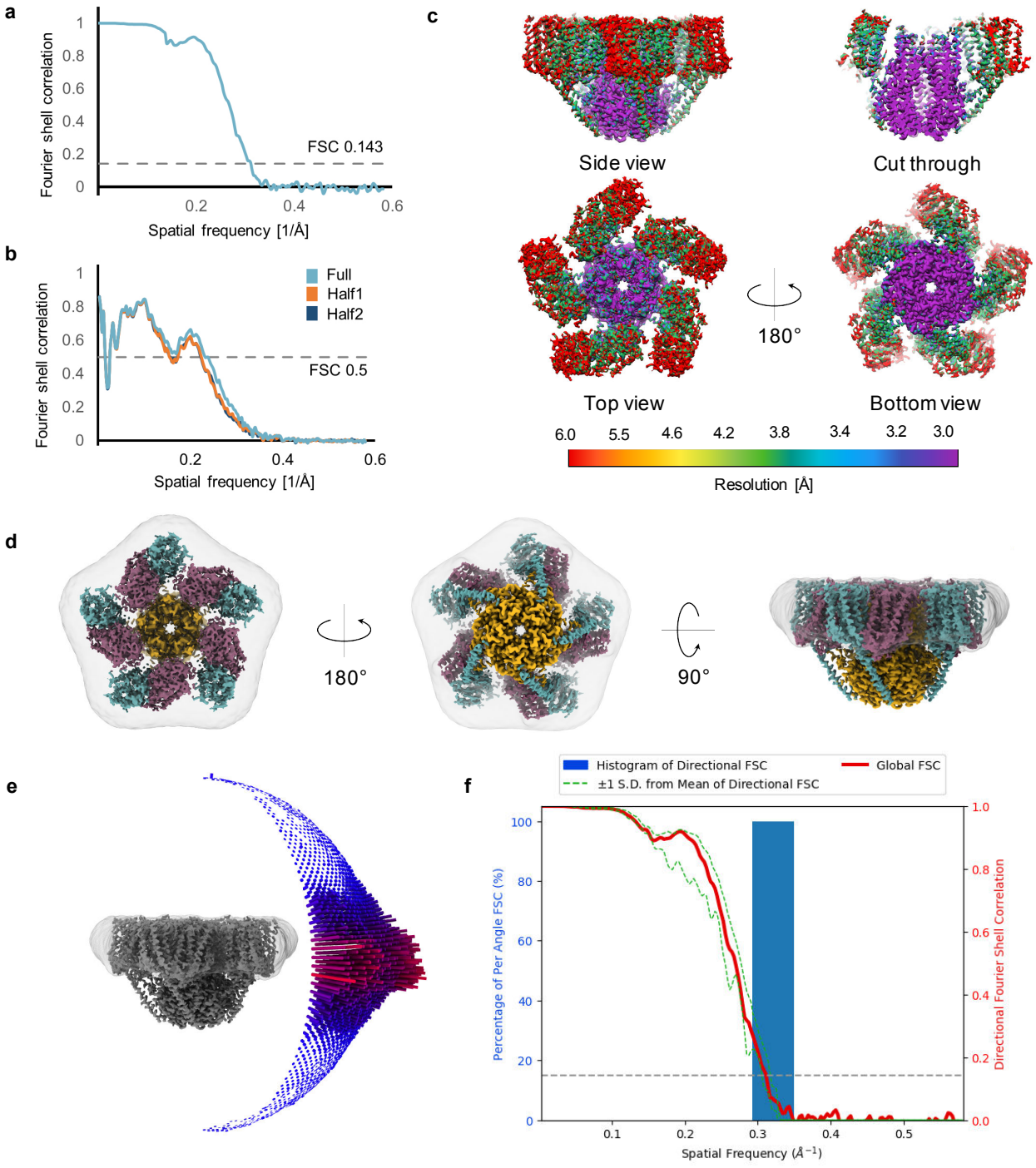

**Supplementary Fig. 2. Signatures of recombination between rhodopsin domains in bestrhodopsins based on codon alignment of complete domain sequences. a**, Recent recombination as revealed by GENECONV: significant global inner fragments (for each pair, upper bounds correspond to Bonferroni-corrected Karlin-Altschul P-values, the lower bounds correspond to simulated P-values) and **b**, examples of the corresponding pairwise nucleotide identity profiles (averaged over sliding windows of 15 bp). **c, d**, Phylogeny-based analysis of breakpoints by GARD that identifies one breakpoint at the alignment position 729 (falling in amino acid positions L292 (R1) and L695 (R2) of Tara-RRB) with Δc-AIC of 63.65 when contrasted against the null model and Δc-AIC 169.95 when contrasted against the single tree multiple partition model. Trees in (**d**) correspond to the two parts of the alignment, before and after the breakpoint,

respectively. The alignment coordinates corresponding to the transmembrane helices are highlighted in (**a-c**). Sequence labels are colored by their group. Asterisks indicate rhodopsin domains from fragmented sequences.



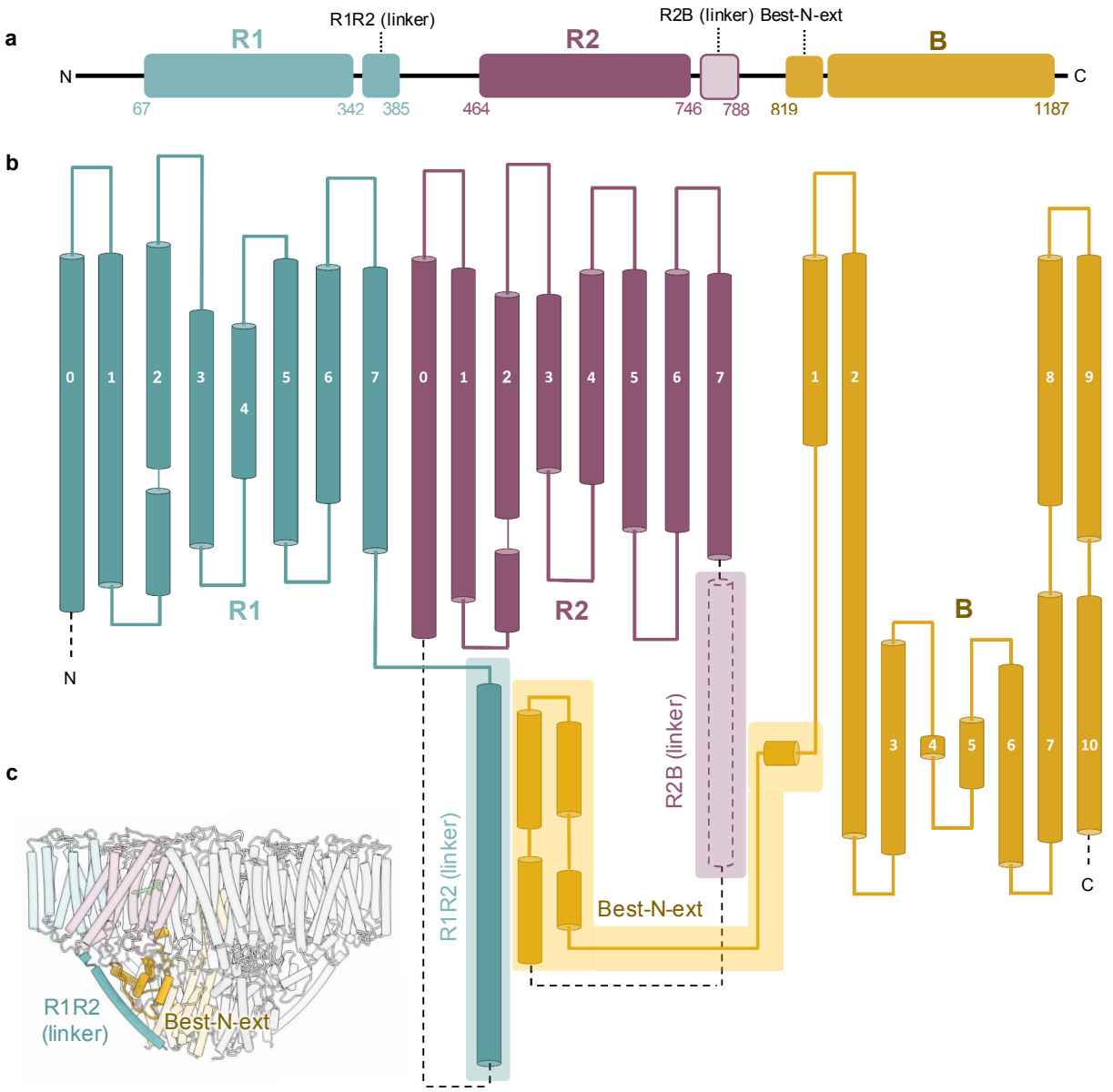

**Supplementary Fig. 3. Reference algal species tree and gene set completeness.** Multigene species trees reconstructed for charting the distribution of bestrhodopsin genes: haptophytes Phaeocystales and selected outgroups (**a**), chlorophytes Pyramimonadales and selected outgroups (**b**) and dinoflagellates (rooted by *Oxyrris marina*) (**c**). Numbers above branches indicate ASTRAL posterior probabilities, numbers below indicate effective numbers of genes. The branch lengths are in units of substitutions per site as calculated with ERaBLE. Completeness of the gene sets according to BUSCO is depicted to the right with

eukaryota_odb10, chlorophyta_odb10 and alveolata_odb10 databases used, respectively. Detailed descriptions of the assemblies are provided in Supplementary Data File 1.




**Supplementary Fig. 4. Map and Model Quality. a,** Gold standard Fourier shell correlation (FSC) curves of half maps that were individually calculated indicating an averaged resolution of 3.21 Å at 0.143 FSC. **b,** FSC curves of the model *vs.* the final map (cyan) and two half maps (orange and dark blue). Plots were calculated with EMAN2. **c,** Final density map colored by local resolution. Color scheme for resolution values is presented in Å. Model orientation is indicated below each panel. **d,** EM map colored according to domains, with R1 in teal, R2 in purple and the channel in yellow. Detergent micelle is presented in light gray. Top and bottom views are in the left and middle panels. Side view is presented to the right. **e,** Euler angle distribution

of the cryo-EM reconstruction as observed from the side view representation [85]. **f,** FSC curve of RRB half maps (red) with the spread of directional resolution values at +/− 1σ from the mean (green dotted lines), and a histogram of one hundred values sampled from the 3D FSC (blue bars, left axis). Black dotted line indicates a 0.143 threshold [112,113].

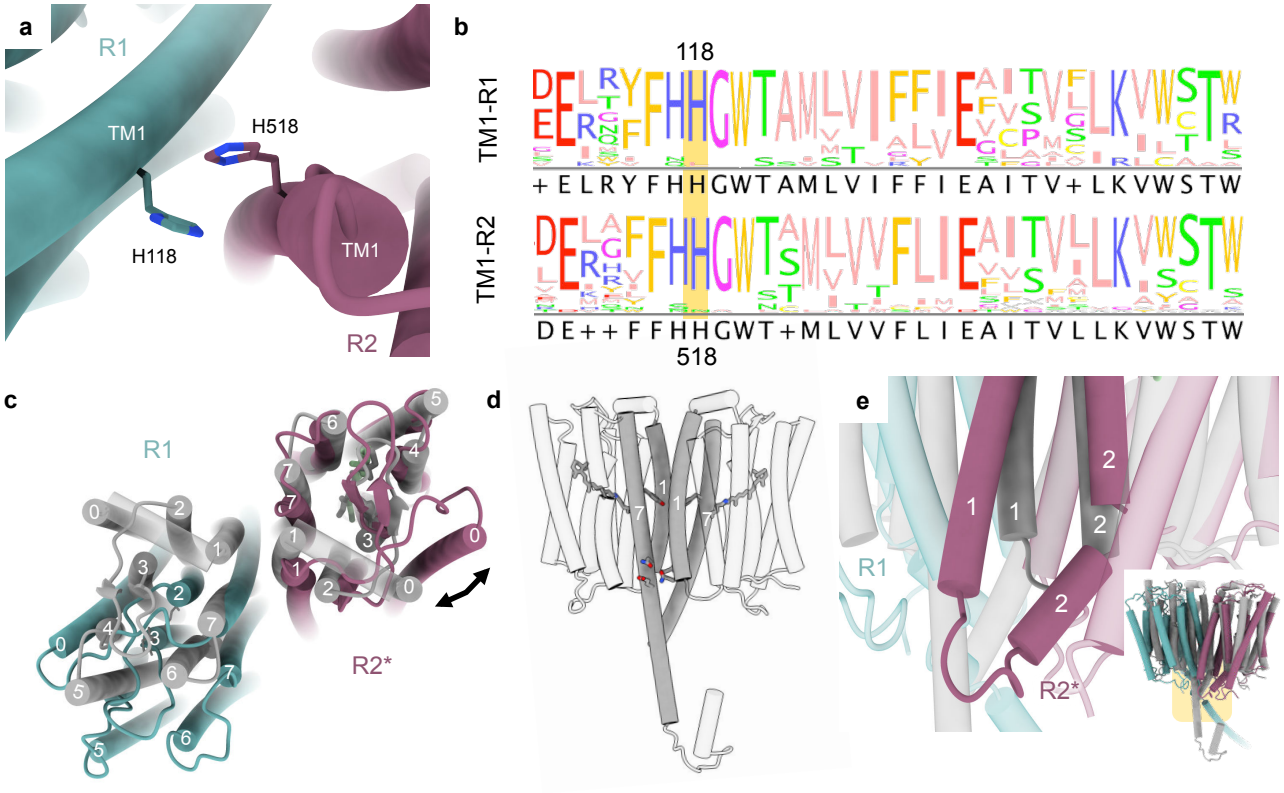



d

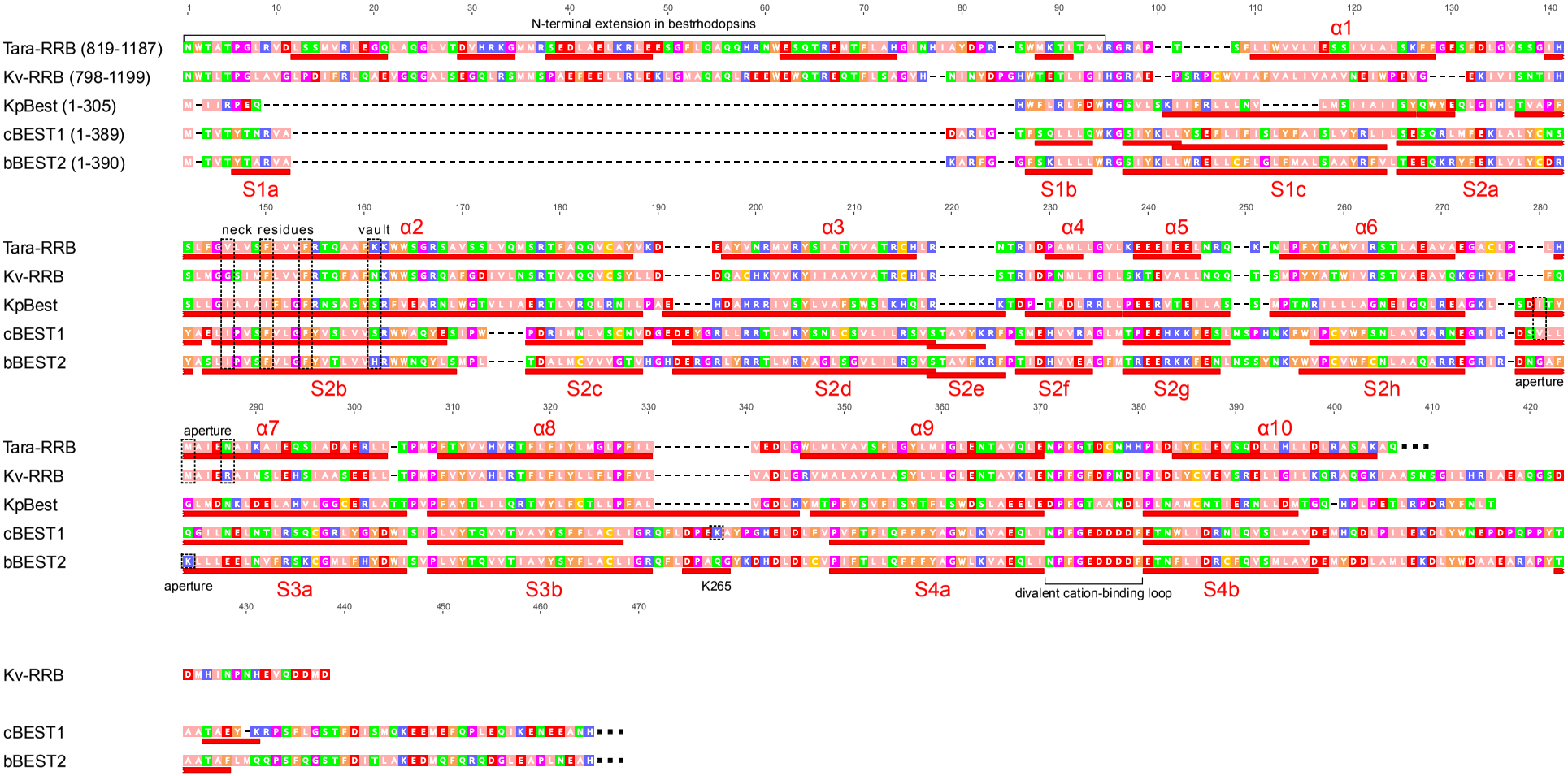
```
  1 MASSTAAPPA AVPAIETAAP PSPEPRDSVS VNEGGDASSA QAGAAVGSTI IVGAPELEKE YSDLNFAQVA
 71 RDEGRRCLLM CVAFAIAIAH LYIYPALFGV RIVDQAEVPA EERTYFHHGW TAMLVIFFIE GVTVFLKVCS
141 TRKTRWLEKA VLQKLDGNIG VLIGEYIVVA ATYIIMGANL IPVFEEPRSG RRVYAVRYME WTIDACGLVY
211 LDCRILFGMP FSKFRMLLVY SVLYMLFGLW AALASTWMWY AIFLSASWFF FGLVCYYYWT FHRQNPSPLQ
281 QFGRAPIKQA ILVFVIVWWV LYGVLFMLCF QAPDVVPQWL EQLLWTGMDV VMKLSHTVVL MAWRETQWEI
351 DAVVDRQKVE AGRAIAQLDH QRAIHERDLV RLRSRVYYGE HIKSEEEIKS EEVISRSLRA RKSRQGQDGT
421 EPSSALSGAS STKEPPALEE NSGSAAPSWT AVLAKGKATS SPFARVNKIF MREAGLCLVL CLAFVVALLH
491 LPVYSEWFGV EVLDAEAVPH DELGFFHHGW TTMLVVFLIE SITVLLKVWS TWHDPRLAEN VAQQLSGNLG
561 VLIAEYLVVG ATYVILGYNL MPVFVVHRPG VASRRVYAVR YMEWAVDATG LIWLDCHCLF SRNFNEFRMA
631 IVWTVAYMLF GLWSALASTW AWYWAFLLAS WAAFLIVCLI LVRFLRQDPY PHQPFGKTSV KPCILAFIIG
701 WWVLYGILFM VCFQAPDAVP QWLEQFLWTG MDVVMKLSHT VVLMAWRTTE WNVCELHGRN QAEKSKKSLL
771 NEGSHARQEK QQLEAIMLEG GGVEAAGRVG SVSFAPSSSR RRKESDSTNW TATPGLRVDL SSMVRLEGQL
841 AQGLVTDVHR KGMMRSEDLA ELKRLEESGF LQAQQHRNWE SQTREMTFLA HGINHIAYDP RSWMKTLTAV
911 RGRAPTSFLL WVVLIESSIV LALSKFFGES FDLGVSSGIH SLFGVLVSFL VVFRTQAAFK KWWSGRSAVS
981 SLVQMSRTFA QQVCAYVKDE AYVNRMVRYS IATVVATRCH LRNTRIDPAM LLGVLKEEEI EELNRQKNLP
1051 FYTAWVIRST LAEAVAEGAC LPLHMAIENA IKAIEQSIAD AERLLTPMPF TYVVHVRTFL FIYLMGLPFI
1121 LVEDLGWLML VAVSFLGYLM IGLENTAVQL ENPFGTDCNH HPLDLYCLEV SQDLLHLLDL RASAKAQ
```

**Supplementary Fig. 5. Domain organization in Tara-RRB. a,** Tara-RRB domain organization with residue numbers. Domains are colored according to their respective colors in main Fig. 2. The helical regions R1R2

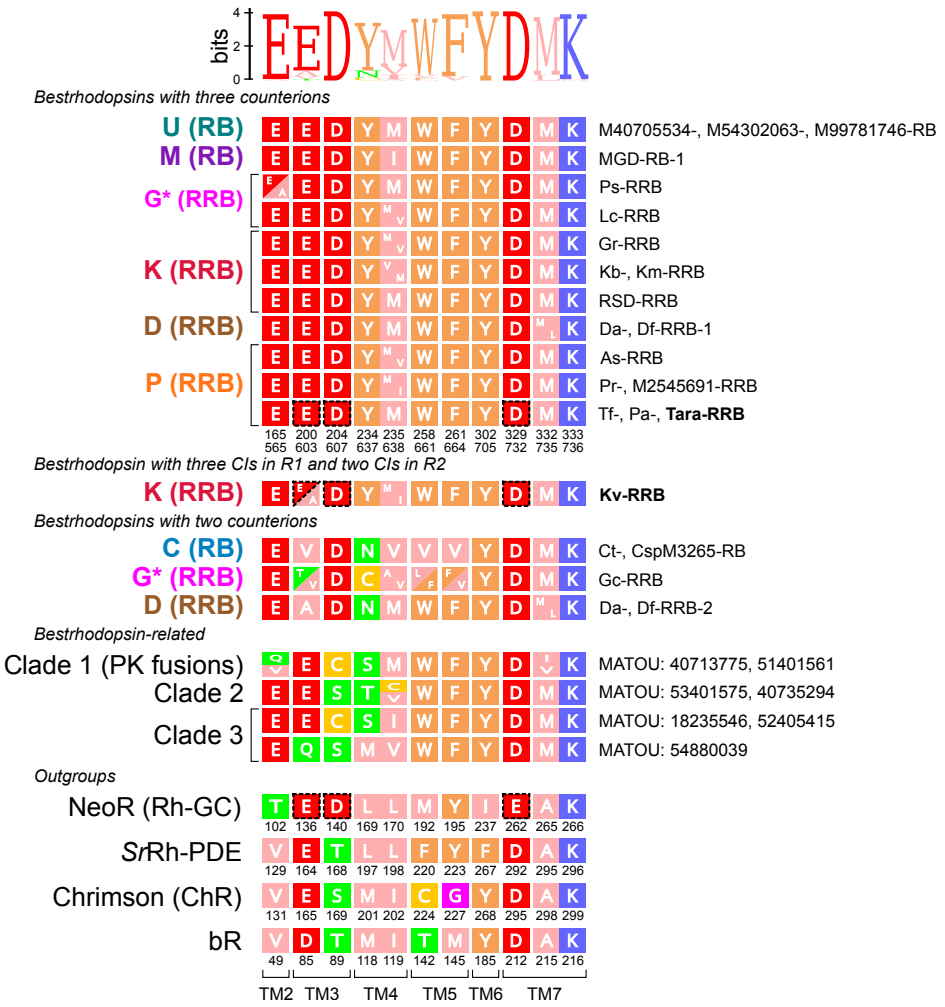
and R2B, as well as bestrophin's N-terminal extensions (Best-N-ext) are highlighted. **b,** Schematic 2D domain organization according to the 3D structure. Dashed lines indicate areas that were not modeled in the 3D structure due to local flexibility. **c,** 3D structures with highlighted linker domains. R1R2 linker is in teal, and the N-terminus of the bestrophin channel is in yellow. **d,** Protein sequence of Tara-RRB with residues colored by domains. Linkers are underlined. Regions that were not modeled in the deposited structure due to segmental flexibility are highlighted in gray.



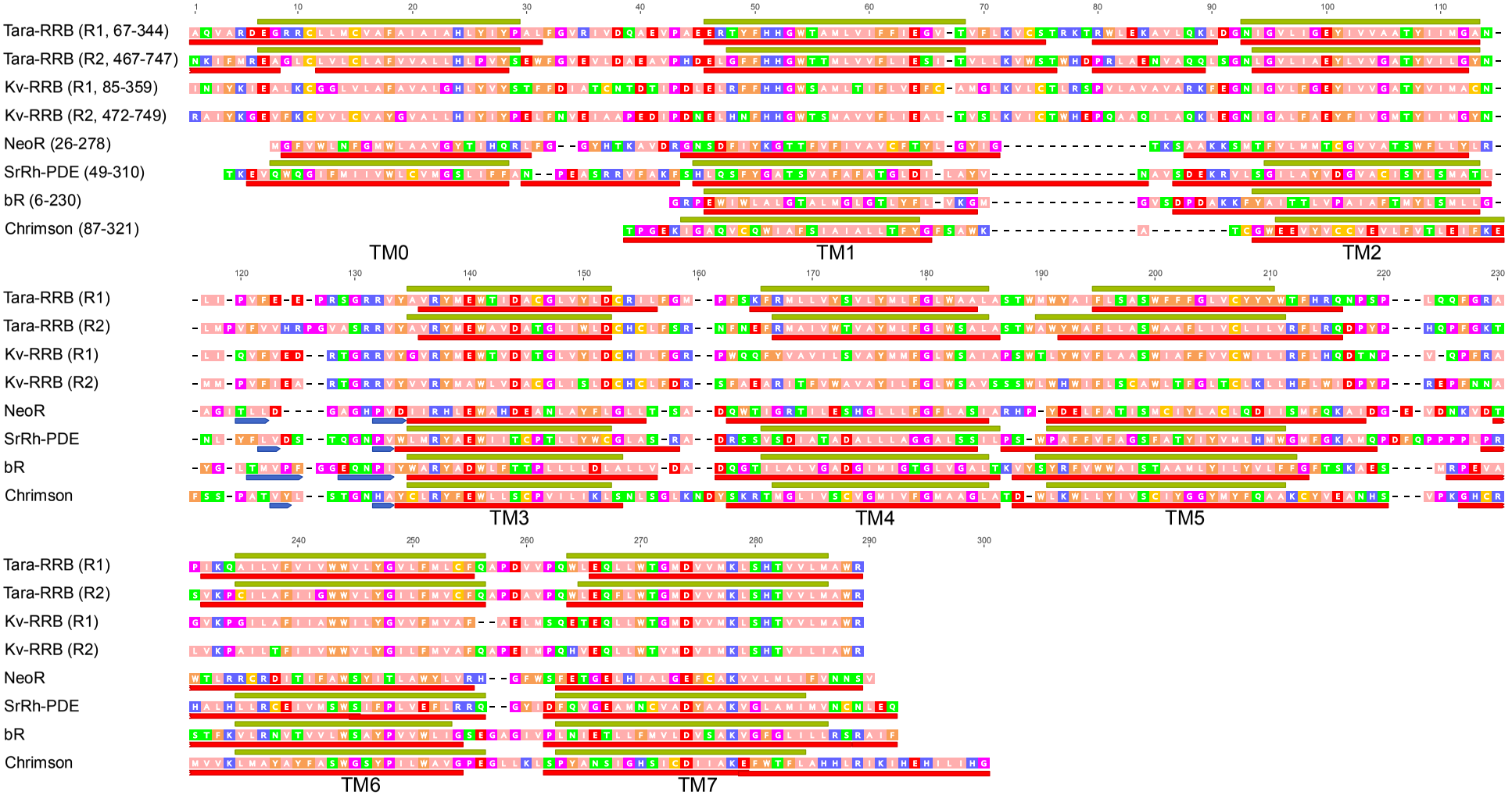

**Supplementary Fig. 6. Rhodopsins in RRB contain eight TMs and highly resemble enzymerhodopsins. a,** The interface between R1 and R2 in RRBs involves interactions between two conserved histidies in TM1 (H118 and H518 in R1 and R2, respectively). A close up view of the interface is with R1 and R2 in teal and purple, respectively. **b,** Sequence conservation in TM1 of RBs and RRBs with conserved histidines highlighted in yellow. **c,** Superposition of RRB with *Sr*Rh-PDE (PDB ID: 7D7Q). R1 and R2 are in teal and purple, respectively, rhodopsin phosphodiesterase in gray. Arrow indicates TM0 localization which has a 20 deg tilt compared with the enzymerhodopsin structure. As such, in Rh-PDE the interaction of TM0 is limited to helices 2 and 3 only, and no contact is observed with TM4 as indicated in the RRB structure (main Fig. 3a). Asterix indicates that the alignment was performed on R2. **d,** Enzymerhodopsins operate as obligate dimers, and in the *Sr*Rh-PDE structure the interface between the homo-monomers involves residues from TM1 and TM7. This is in contrast to RRBs where TM1-TM1 interactions only are participating in the dimerization interface between the two rhodopsin units. Coordinated correspond to the *Sr*Rh-PDE structure (PDB ID: 7D7Q). Helix numbers participating in the interface are indicated and residues involved are presented as sticks. **e,** TM1 and TM2 in RRB are significantly longer compared with the *Sr*Rh-PDE structure (PDB ID: 7D7Q), as a result the helical domain extension in RRB is directed opposite to that observed in enzymerhodopsins.


N-terminal extension in bestrhodopsins
α1
Tara-RRB (819-1187)
Kv-RRB (798-1199)
KpBest (1-305)
cBEST1 (1-389)
bBEST2 (1-390)
S1a
S1b
S1c
S2a
neck residues
vault
α2
α3
α4
α5
α6
Tara-RRB
Kv-RRB
KpBest
cBEST1
bBEST2
S2b
S2c
S2d
S2e
S2f
S2g
S2h
aperture
α7
α8
α9
α10
S3a
S3b
K265
S4a
divalent cation-binding loop
S4b
Kv-RRB
cBEST1
bBEST2

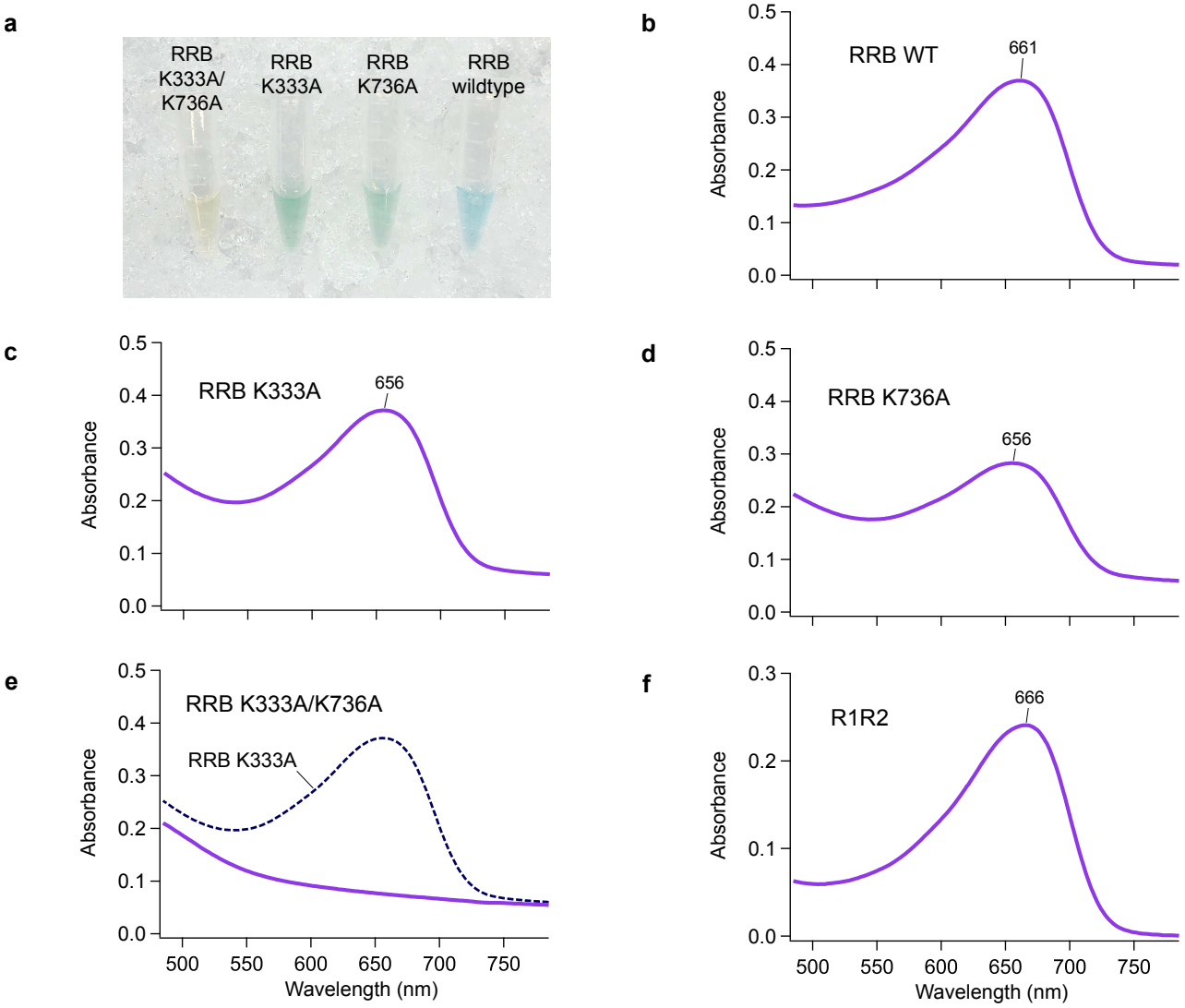
**Supplementary Fig. 7. Alignment of bestrhodopsin bestrophin domains and reference bestrophins.** Curated 3D-Coffee alignment of bestrophins. The non-structured C-termini are generally not aligned and are trimmed as indicated. Helical regions are indicated in red for proteins with structures available and labeled (for KpBest above, for animal bestrophins below). The reference bestrophins are as follows: KpBest (PDB 4WD8), cBEST1 (PDB 6N23) and bBEST2 (PDB 6VX9). Functionally important positions are highlighted with frames.



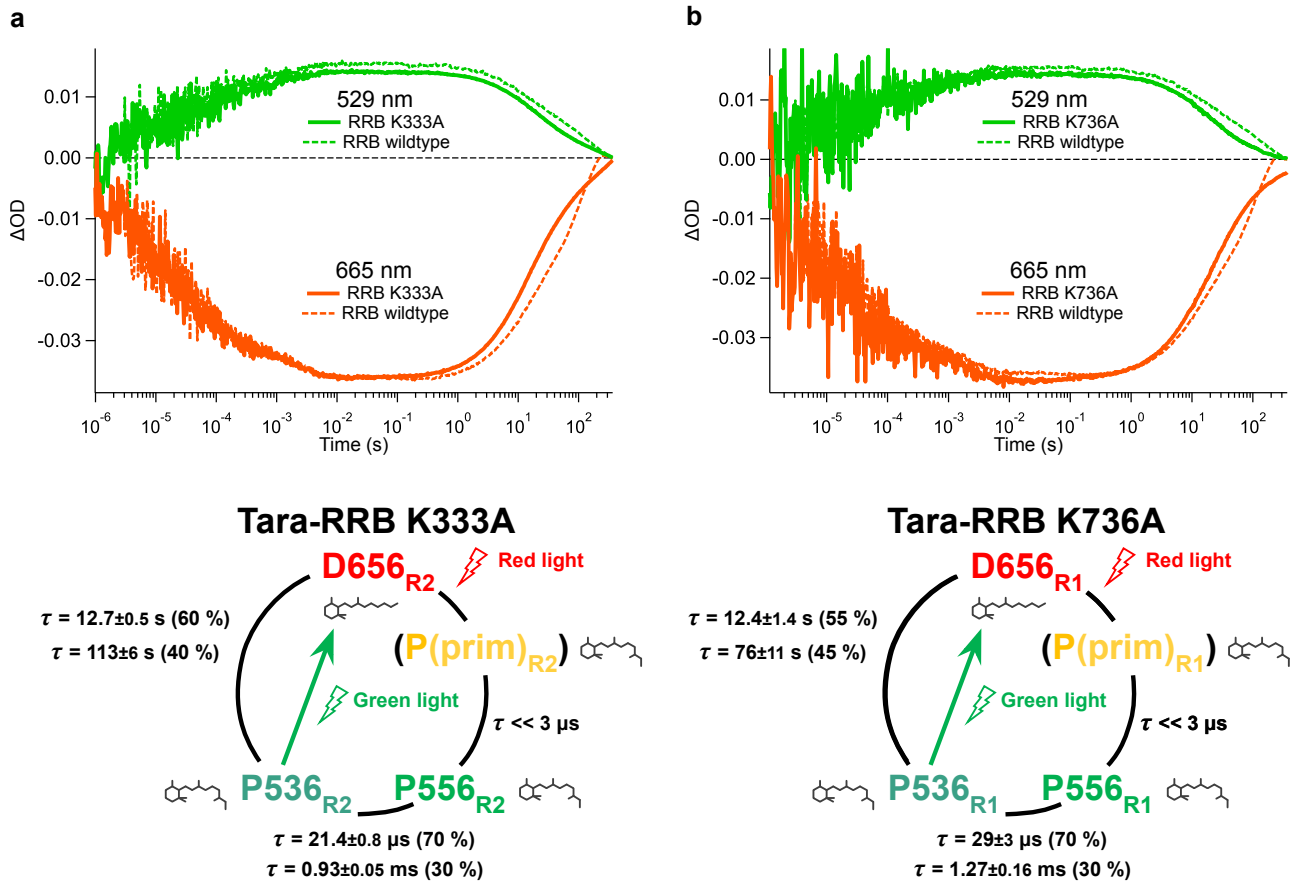

**Supplementary Fig. 8. Key residue in the retinal binding pocket of bestrhodopsins.** Bestrhodopsins are categorized according to the number of the counterions and further grouped by phylogenetic group (see Fig. 1a-c). Genes representing each residue composition are listed to the right. For RRBs the residues in the R1 and R2 domains, when different, are provided across the diagonal. Three clades of bestrhodopsin-related proteins, representatives of enzymerhodopsins (NeoR and SrRh-PDE), as well as more distantly related 7TM rhodopsins are provided for cross-comparison. Residue numbers corresponding to homologous positions are indicated.


Tara-RRB (R1, 67-344)
Tara-RRB (R2, 467-747)
Kv-RRB (R1, 85-359)
Kv-RRB (R2, 472-749)
NeoR (26-278)
SrRh-PDE (49-310)
bR (6-230)
Chrimson (87-321)
TM0
TM1
TM2
Tara-RRB (R1)
Tara-RRB (R2)
Kv-RRB (R1)
Kv-RRB (R2)
NeoR
SrRh-PDE
bR
Chrimson
TM3
TM4
TM5
Tara-RRB (R1)
Tara-RRB (R2)
Kv-RRB (R1)
Kv-RRB (R2)
NeoR
SrRh-PDE
bR
Chrimson
TM6
TM7

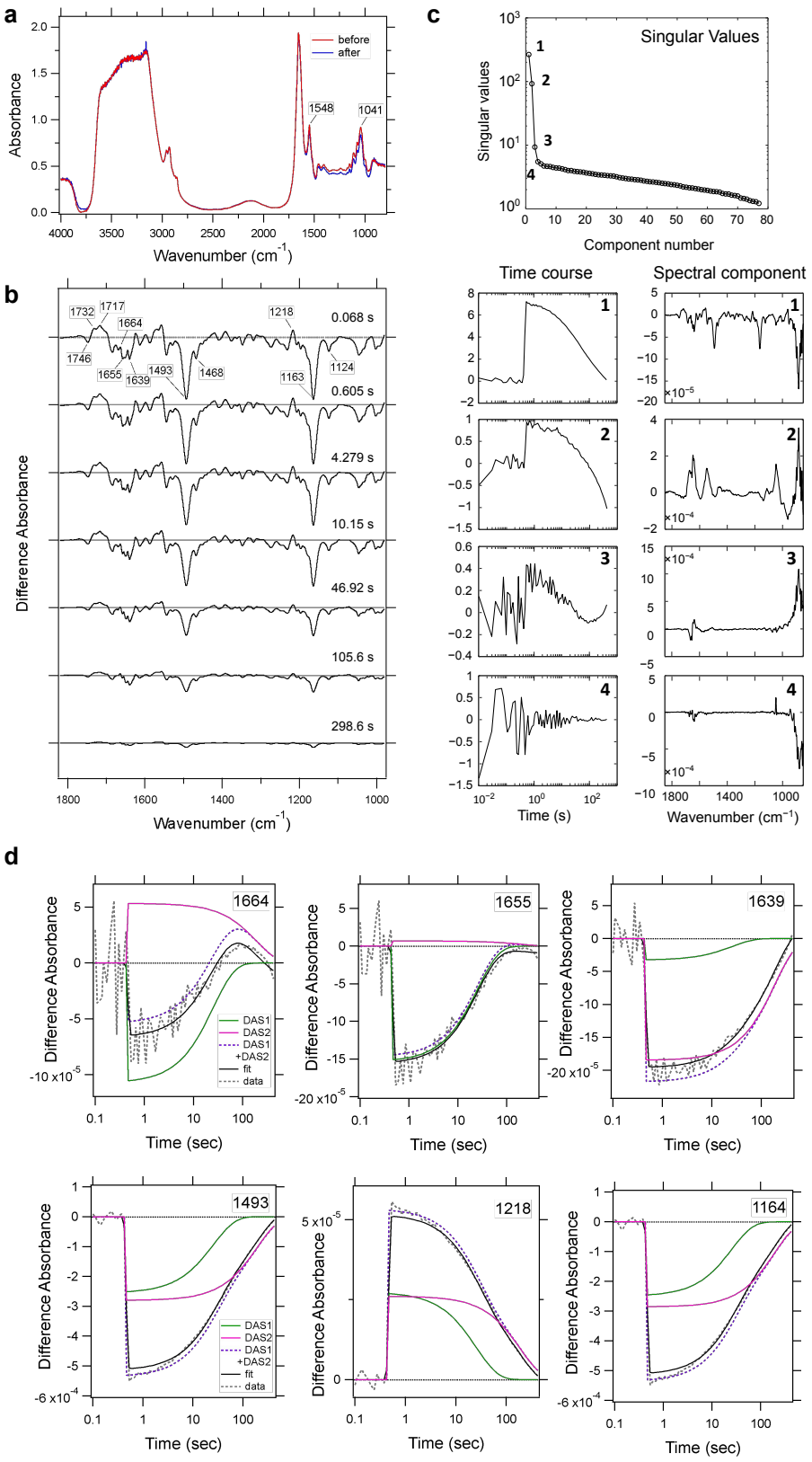
**Supplementary Fig. 9. Alignment of bestrhodopsin rhodopsin domains and reference rhodopsins.** Curated 3D-Coffee alignment of stand-alone rhodopsins and rhodopsin domains. Secondary structure features are indicated for rhodopsins with structures available: α-helices (red lines below sequences) and β-sheets (blue arrows). Transmembrane regions (according to PPM/OPM) are numbered TM0-TM7 and are shown as green lines above sequences. The reference rhodopsins are as follows: NeoR (Uniprot A0A1Y2CSJ0), SrRh-PDE (PDB 7CJ3), bR (bacteriorhodopsin, PDB 1M0L), GPR (green-absorbing proteorhodopsin, PDB 7B03), KR2 sodium pump (PDB 4XTN).




**Supplementary Fig. 10. Absorption spectra of Tara-RRB lysine mutants and isolated domains.** A picture of Tara-RRB mutants in which retinal-binding lysines were substituted with alanine, and wildtype RRB **(a)**. Absorption spectra (purple lines) of Tara-RRB WT **(b)**, K333A **(c)**, K736A **(d)**, K333A/K736A **(e)** mutants, and isolated Tara-R1R2 **(f)** domain. For comparison, the spectrum of Tara-RRB K333A was overlaid on that of Tara-RRB K333A/K736A in d (dashed line).

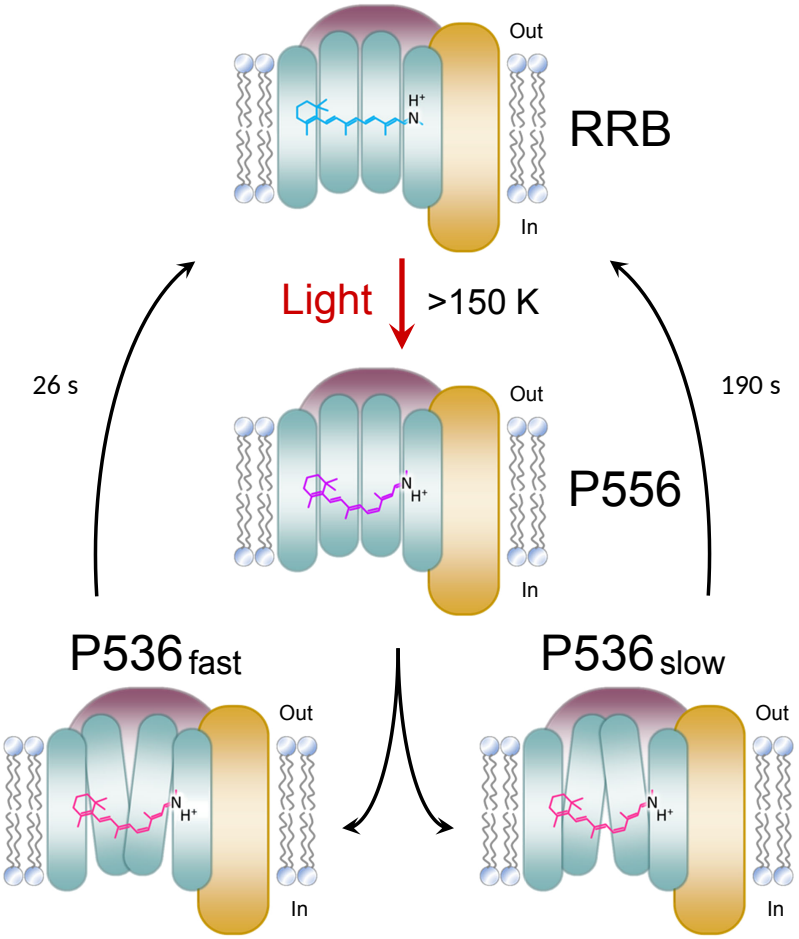



**Supplementary Fig. 11. Photocycles of Tara-RRB K333A and Tara-RRB K736A mutants.** Time evolutions transient absorption change (top) and the photocycle models (bottom) of Tara-RRB K333A **(a)**, and Tara-RRB K736A **(b)** mutants (solid lines) at specific wavelengths of Tara-RRB excited at $\lambda$ = 650 nm in 0.1% (w/v) DDM, 0.02% (w/v) CHS, 20 mM HEPES pH = 7.5, 200 mM NaCl. The time evolution of Tara-RRB wildtype was overlaid (dashed lines). While the formation of the P556 and P536 intermediates of Tara-RRB K333A and K736A mutants was similar to that of Tara-RRB wildtype, significant difference was found in the recovery process of the initial state. The recovery process was reproduced by two exponential functions. For Tara-RRB WT, the slower component was dominant (Fig. 5b), whereas in both lysine mutants, the contribution of the faster component was larger, suggesting that the simultaneous excitation of both domains may slow down the photocycle of each rhodopsin domain.




**Supplementary Fig. 12. Time-resolved FTIR spectroscopy of Tara-RRB. a,** Infrared absorption spectra of Tara-RRB solubilized by LMNG/GDN detergents which are the same detergents used for cryo-EM experiments. The red and blue spectra were recorded before and after time-resolved FTIR measurement, respectively. The 1548-cm$^{-1}$ band is amide II mode of protein and the 1041-cm$^{-1}$ band is C-O stretching mode

of sugar head groups of LMNG. **b,** Light-induced difference infrared spectra of Tara-RRB after singular value decomposition analysis. The spectra recorded just after the laser flash (0.068 s) and at six selected time points are shown, in which the original dark state recovers within ~5 min (298.6 s). **c,** Singular value decomposition analysis on time-resolved FTIR spectral data. Three components (#1, # 2, and #3) have significantly large singular values compared to those of component numbers larger than 4. Among them, the spectral shape of the component #2 would have originated from baseline distortion. Thus, we omitted this component for global exponential fitting analysis in the next step. **d,** Infrared absorbance changes at six specific wavenumbers and fitting curves obtained by global exponential fitting analysis. Time traces of infrared absorption changes at six specific wavenumbers (dotted gray lines) are shown with exponential curves of DAS1 (green), DAS2 (magenta), and DAS1+DAS2 (purple), and the total fitting curves obtained by global exponential fitting analysis (black). The experimental data (dotted gray lines) were treated by singular value decomposition analysis for removing the spectral distortion.




**Supplementary Fig. 13. Proposed structural changes during the photocycle of a rhodopsin domain in Tara-RRB.** Photoactivation is highly temperature-dependent in Tara-RRB, where no isomerization takes place at ≤150 K. The P556 intermediate is stabilized at 170-230 K, which is converted to the P536 intermediate in sub-ms at room temperature. The P536 intermediate is stabilized at >230 K, whose lifetimes are 10-200 s at room temperature. Both P556 and P536 contain 11-*cis* retinal with different colors. Time-resolved FTIR study showed different helical structural changes between fast and slow decaying P536 intermediates ($P536_{fast}$ and $P536_{slow}$, respectively), whose spectral changes in amide-I are canceled with each other (Fig. 5d). Consequently, spectral changes in amide-I region (1700-1600 $cm^{-1}$; 250 K and 298 K in Fig. 5f) look much smaller in Tara-RRB than in bacteriorhodopsin and retinochrome (Supplementary Fig. 13c). In this model, the reaction is branched from the P556 intermediate, whereas strong temperature dependence may suggest structural heterogeneity in the unphotolyzed state of RRB. Yellow cylinder represents the bestrophin domain.