1

Figure 1

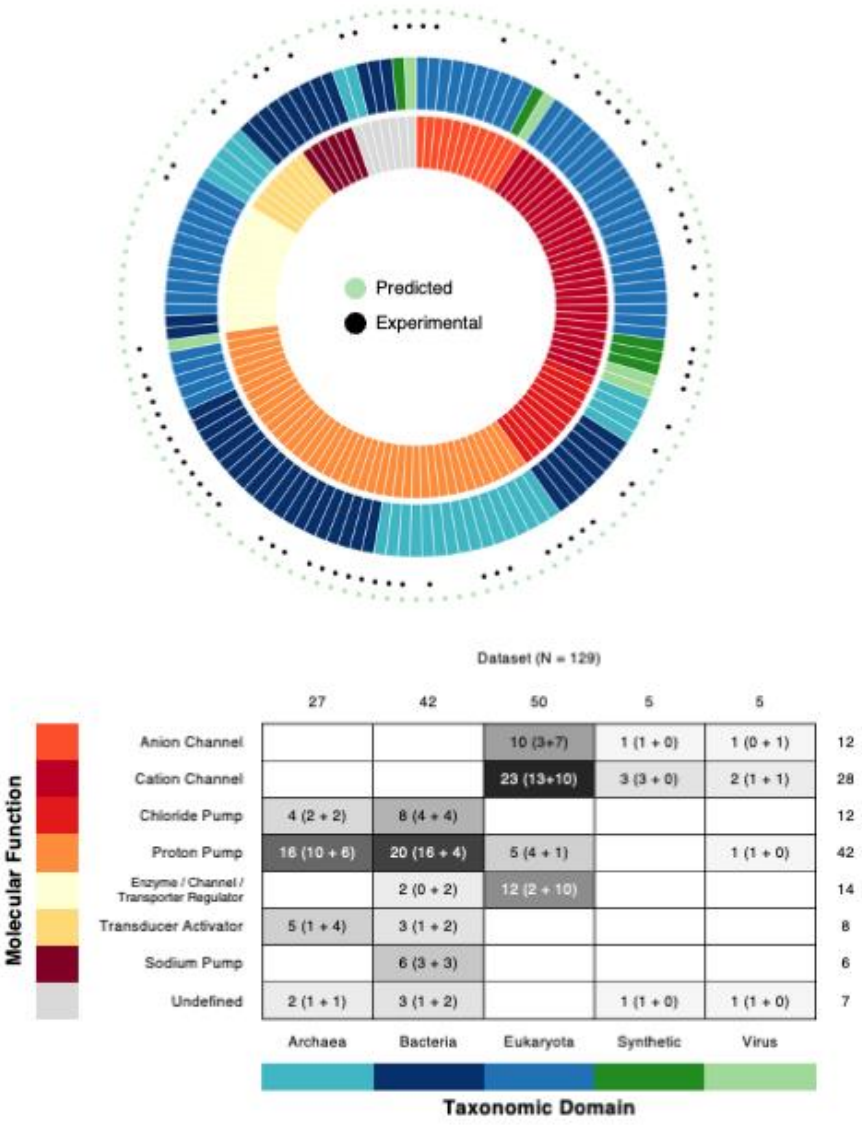


Fig. 1 | **Dataset of microbial rhodopsin structures.** The colors in the inner and outer rings indicate molecular function and taxonomic domain respectively (as shown in the table). Experimental structures and computational models are marked with black and cyan circles respectively. The table shows the composition of the dataset; each cell shows "total number of structures (experimental + predictions)" in that category.

2

Figure 2

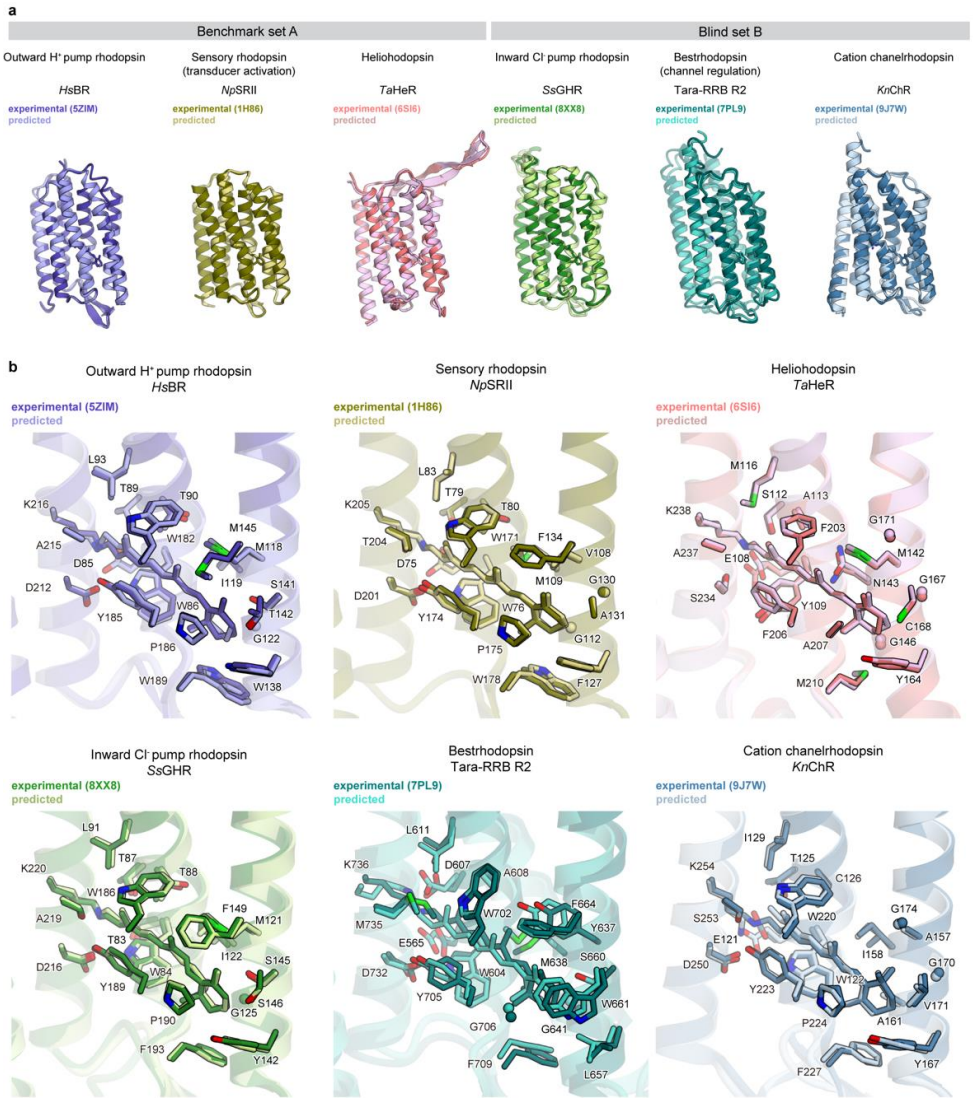


Fig. 2 | Accuracy of the predicted models. a Structural superposition between experimental (solid colors) and predicted (light colors) structures of representative rhodopsins: from benchmark set A, the outward proton pump rhodopsin *Hs*BR (PDB ID: 5ZIM), the transducer-activating rhodopsin *Np*SRII (PDB ID: 1H86), and the functionally uncharacterized heliorhodopsin *Ta*HeR (PDB ID: 6SI6); and from blind set B, the inward chloride pump rhodopsin *Ss*GHR (PDB ID: 8XX8), the channel-regulating rhodopsin Tara-RRB (PDB ID: 7PL9), and the cation-conducting channelrhodopsin *Kn*ChR (PDB ID: 9J7W). **b** Detailed view of the retinal-binding pockets. Transmembrane segments are depicted as translucent ribbons and retinal and protein side chains as solid sticks. TM 5, 6 and 7 are omitted for clarity.

Figure 3

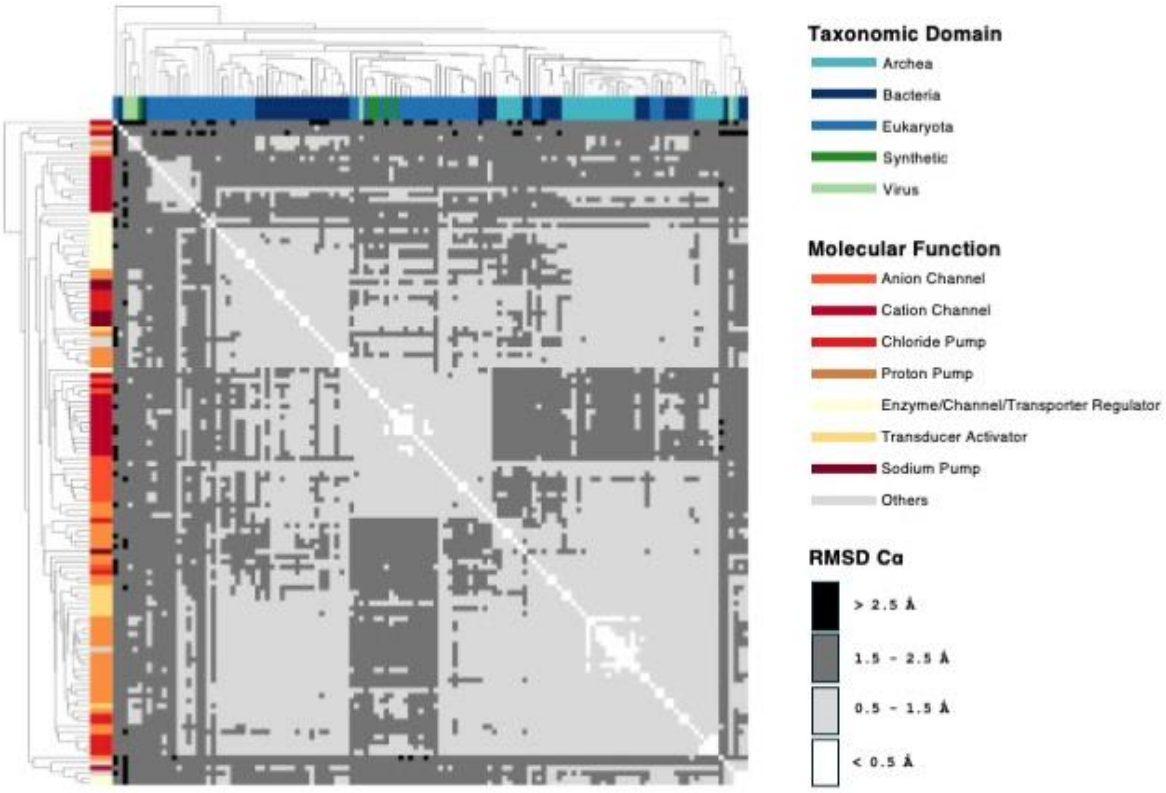


Fig. 3 | Structural landscape of microbial rhodopsins. The heatmap displays pairwise RMSD Cα values for all 129 dataset members. Rhodopsins are sorted according to hierarchical clustering based on structural similarity. Annotation bars display taxonomic domain (x-axis; green/blue) and molecular function (in the y-axis; orange/red). The greyscale bar indicates the RMSD values (light grey = 0.5-1.5 Å, dark grey = 1.5–2.5 Å, black ≥2.5 Å).

4

5

Figure 4

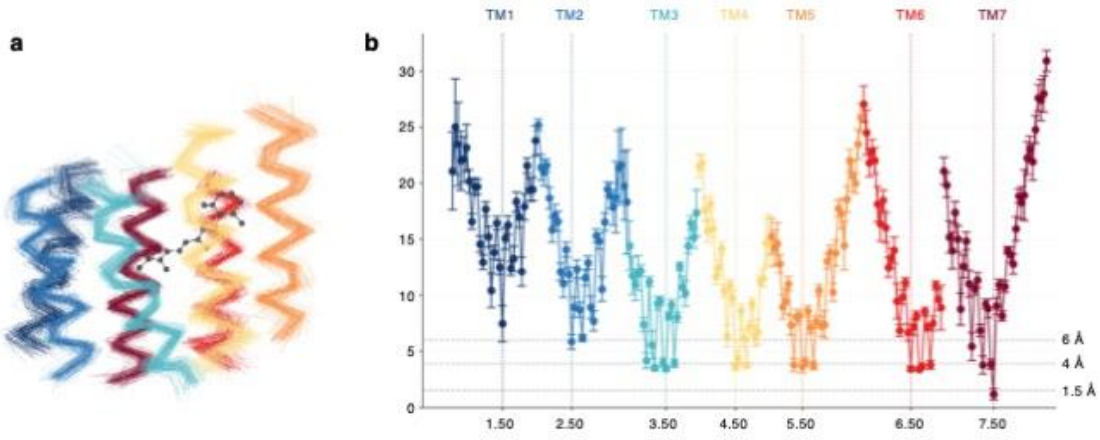


Fig. 4 | Location of the MO numbering anchor positions. a Structural alignment of the 129 dataset members allows selecting the consensus residues closest to retinal in each helix. **b** Average minimum distance from each structurally equivalent residue position to the retinal chromophore calculated across the dataset (closest side-chain atom, excluding hydrogen; an analogous graph for Cα distances is shown in **Supplementary Fig. 2a**). Vertical lines indicate the defined X.50 anchor position for each helix (1.50–7.50). Horizontal lines indicate key interaction distances: 1.5 Å approximates a covalent bond length and is relevant for the Schiff base linkage; 4 Å represents a typical cutoff for non-covalent interactions, delineating the direct binding pocket; and 6 Å encompasses the broader binding-site environment and is the radius used for iRMSD calculations.

Figure 5

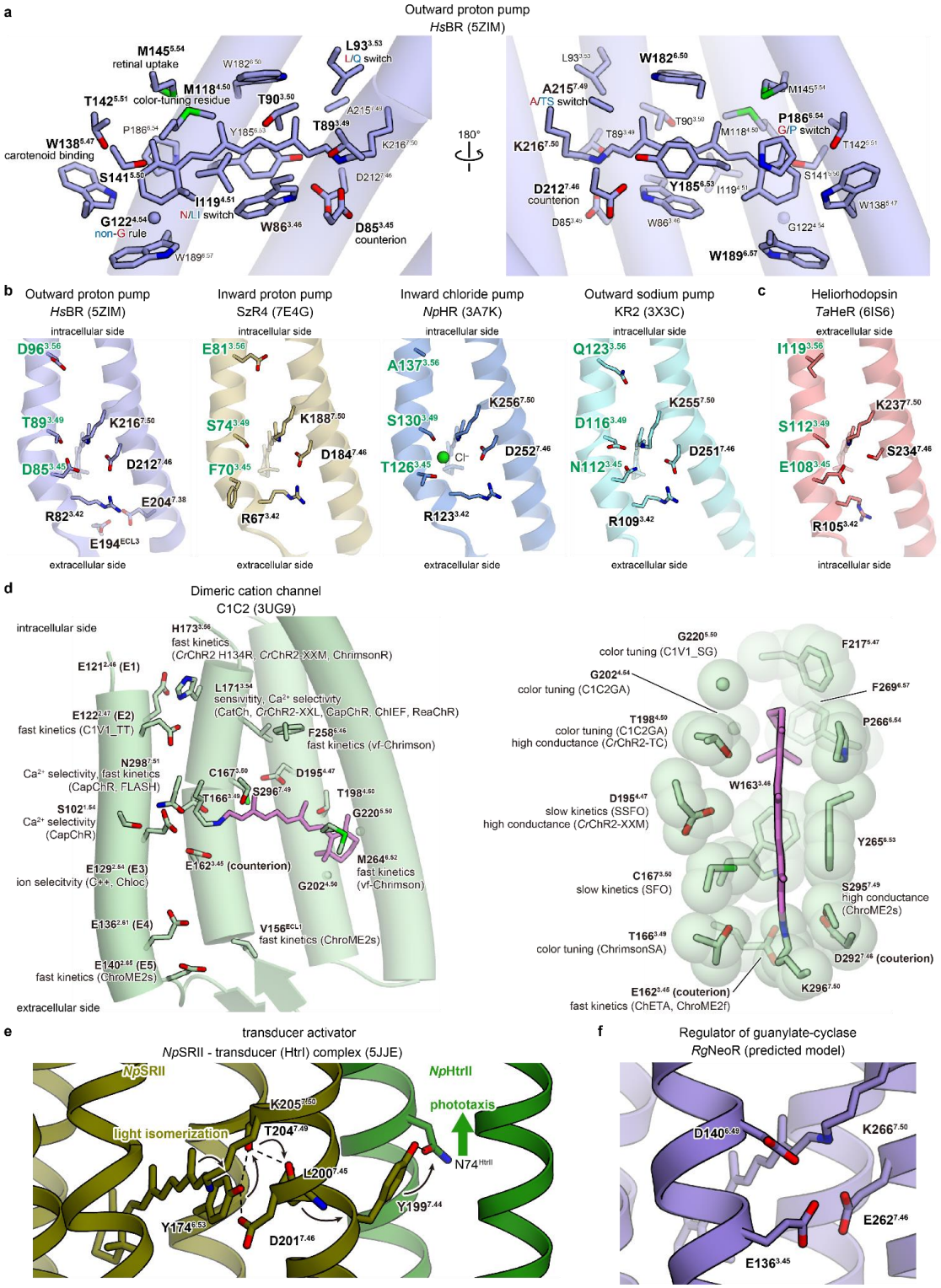

a
Outward proton pump
HsBR (5ZIM)
M145 5.54 retinal uptake
W182 6.50
L93 3.53 L/Q switch
M118 4.50 color-tuning residue
T142 5.51
T90 3.50
A215 7.49
T89 3.49
P186 6.54
Y185 6.53
K216 7.50
W138 5.47 carotenoid binding
S141 5.50
D212 7.46
I119 4.51 N/LI switch
G122 4.54 non-G rule
W86 3.46
D85 3.45 counterion
W189 6.57
180°
A215 7.49 A/TS switch
P186 6.54 G/P switch
K216 7.50
D212 7.46 counterion
b
Outward proton pump HsBR (5ZIM)
Inward proton pump SzR4 (7E4G)
Inward chloride pump NpHR (3A7K)
Outward sodium pump KR2 (3X3C)
intracellular side
extracellular side
D96 3.56
T89 3.49
K216 7.50
D85 3.45
D212 7.46
E204 7.38
R82 3.42
E194 ECL3
E81 3.56
S74 3.49
K188 7.50
F70 3.45
D184 7.46
R67 3.42
A137 3.56
S130 3.49
K256 7.50
Cl⁻
T126 3.45
D252 7.46
R123 3.42
Q123 3.56
D116 3.49
K255 7.50
N112 3.45
D251 7.46
R109 3.42
c
Heliorhodopsin TaHeR (6IS6)
extracellular side
intracellular side
I119 3.56
S112 3.49
K237 7.50
E108 3.45
S234 7.46
R105 3.42
d
Dimeric cation channel
C1C2 (3UG9)
intracellular side
extracellular side
H173 3.56 fast kinetics (CrChR2 H134R, CrChR2-XXM, ChrimsonR)
E121 2.46 (E1)
L171 3.54 sensivitity, Ca2+ selectivity (CatCh, CrChR2-XXL, CapChR, ChIEF, ReaChR)
E122 2.47 (E2) fast kinetics (C1V1_TT)
F258 5.46 fast kinetics (vf-Chrimson)
N298 7.51 Ca2+ selectivity, fast kinetics (CapChR, FLASH)
C167 3.50
D195 4.47
T166 3.49
S296 7.49
T198 4.50
S102 1.54 Ca2+ selectivity (CapChR)
G220 5.50
E129 2.54 (E3) ion selecivity (C++, Chloc)
E162 3.45 (counterion)
M264 6.52 fast kinetics (vf-Chrimson)
G202 4.50
E136 2.61 (E4)
V156 ECL1 fast kinetics (ChroME2s)
E140 2.65 (E5) fast kinetics (ChroME2s)
G220 5.50 color tuning (C1V1_SG)
F217 5.47
G202 4.54 color tuning (C1C2GA)
F269 6.57
T198 4.50 color tuning (C1C2GA) high conductance (CrChR2-TC)
P266 6.54
W163 3.46
D195 4.47 slow kinetics (SSFO) high conductance (CrChR2-XXM)
Y265 6.53
C167 3.50 slow kinetics (SFO)
S295 7.49 high conductance (ChroME2s)
T166 3.49 color tuning (ChrimsonSA)
D292 7.46 (couterion)
E162 3.45 (couterion) fast kinetics (ChETA, ChroME2f)
K296 7.50
e
transducer activator
NpSRII - transducer (HtrI) complex (5JJE)
NpSRII
NpHtrII
K205 7.50
light isomerization
T204 7.49
phototaxis
N74 HtrII
L200 7.45
Y199 7.44
Y174 5.63
D201 7.46
f
Regulator of guanylate-cyclase
RgNeoR (predicted model)
D140 6.49
K266 7.50
E262 7.46
E136 3.45

Fig. 5 | Demonstration of the MO numbering system. a Retinal-binding pocket of the outward proton pump *Hs*BR (PDB ID: 5ZIM), shown from two different orientations. Residues constituting the retinal-binding pocket are depicted as sticks. Bold labels indicate residues positioned on the front side of the retinal chromophore. Annotations below each label indicate the functional role of each residue. Blue and red annotations denote shifts toward longer and shorter wavelengths, respectively. **b** Schiff-base vicinity of representative ion-pumping rhodopsins: the outward proton pump *Hs*BR (PDB ID: 5ZIM), the inward proton pump SzR4 (PDB ID: 7E4G), the inward chloride pump *Np*HR (PDB ID: 3A7K), and the outward sodium pump KR2 (PDB ID: 3X3C) (left to right). Only TM3 and TM7 are shown as ribbons for clarity. The residues displayed correspond to functionally important positions. **c** Schiff-base vicinity of the heliorhodopsin *Ta*HeR (PDB ID: 6IS6). In panels b and c, the residues of the DTD motif are highlighted in green. **d** Distribution of functionally important residues identified across various dimeric channelrhodopsin variants, mapped onto the overall structure (left) and the retinal-binding pocket (right) of C1C2 (PDB ID: 3UG9). Variant names associated with each mapped residue are shown adjacent to the corresponding positions. Annotations below each label indicate the functional properties or engineered mutations reported for that variant. **e** Transducer-activating rhodopsin *Np*SRII (khaki) and its cognate transducer *Np*HtrII (green) from the archaeon *Natronomonas pharaonis* (PDB ID: 5JJE). Residues critical for signal activation are shown in stick representation. TM1 is omitted for clarity. **f** Three key residues implicated in long-wavelength absorption in the guanylate cyclase–fused rhodopsin *Rg*NeoR. The structure shown is a model predicted using Boltz1.

8

9

Figure 6

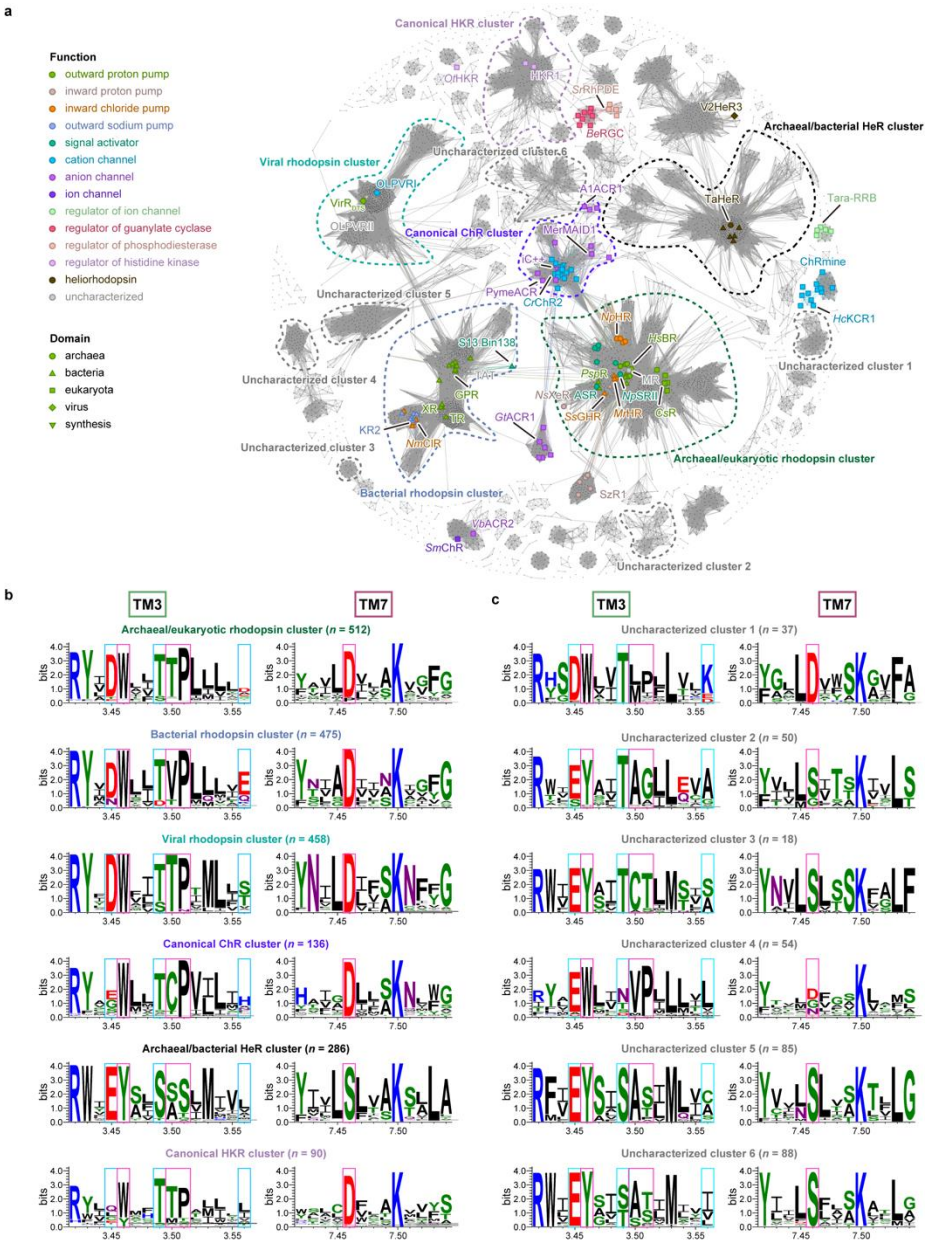


Fig. 6 | Application of the MO numbering system to a large dataset of microbial rhodopsins. a Sequence similarity network composed of approximately 40,000 non-redundant microbial rhodopsin sequences, including a large collection of uncharacterized rhodopsin-like sequences. Six major clusters containing functionally characterized rhodopsins and six of the 17 clusters containing only uncharacterized rhodopsins are highlighted with dashed circles. Representative rhodopsins are mapped onto the clusters, and the colors and shapes of the labels indicate function

and domain, respectively. **b** Sequence logos of TM3 and TM7 for six major clusters containing functionally characterized rhodopsins and six of the 17 clusters containing only uncharacterized rhodopsins. Cyan rectangles indicate residues of the DTD motif and magenta rectangles indicate highly conserved residues within each cluster.

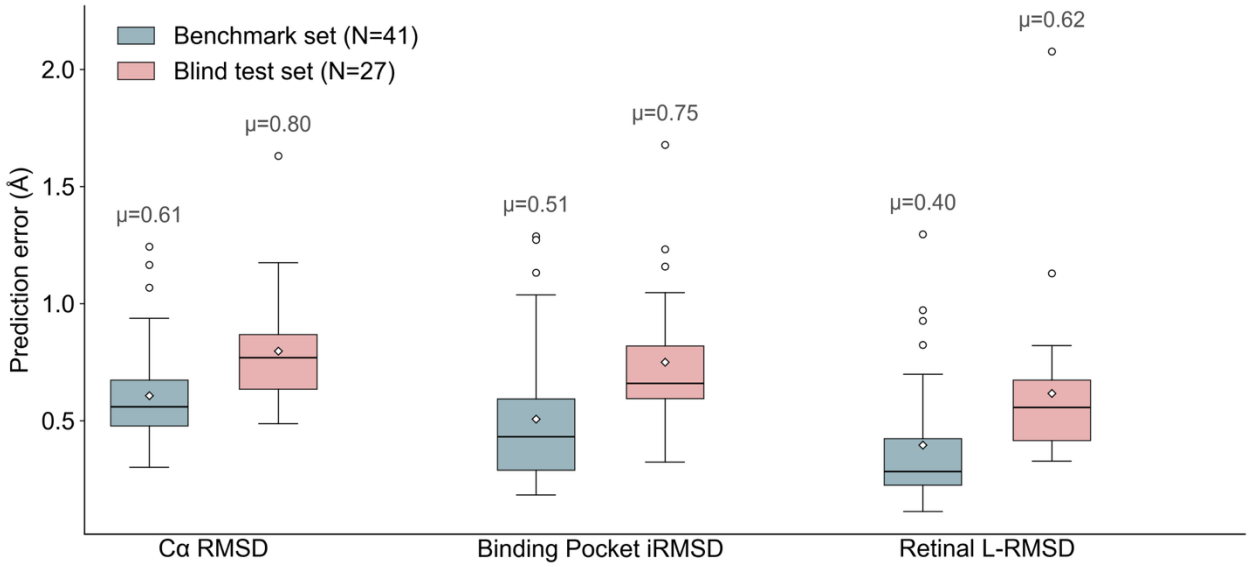
11

12

13

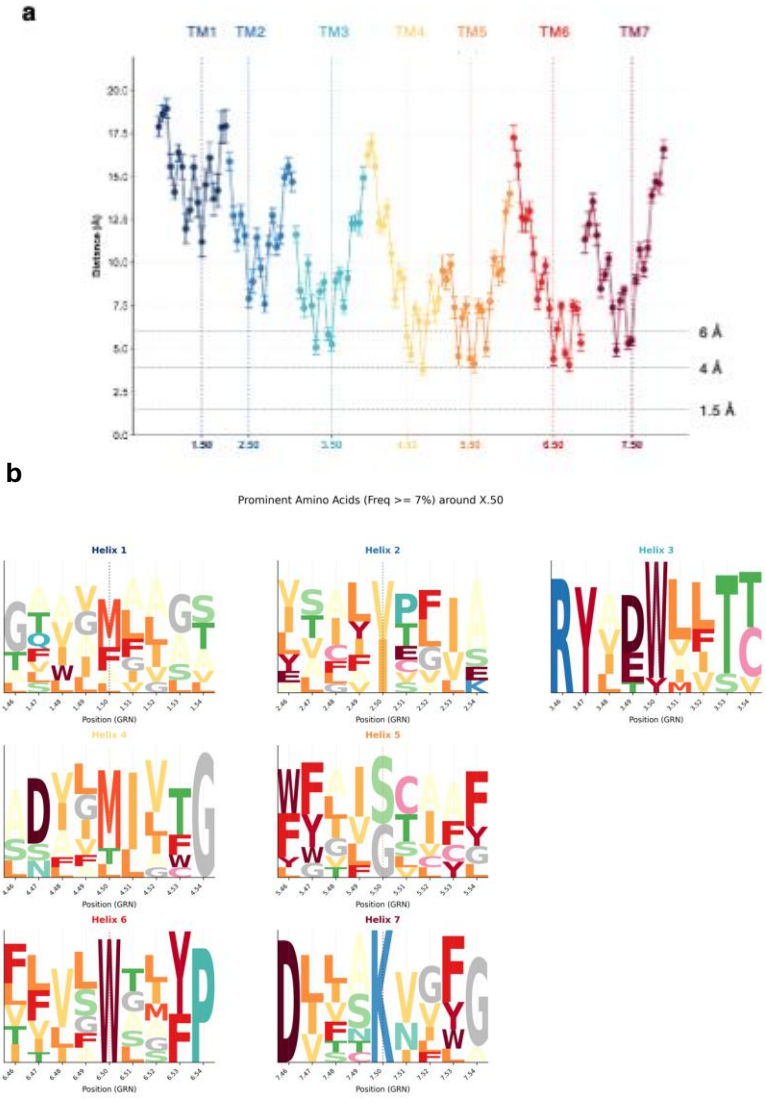
14

Supplementary Figure 1



Supplementary Fig. 1 | **Accuracy of the predicted models.** Prediction errors for the two validation sets: benchmark set A (blue; n=42) and blind set B (red; n=27 for backbone, n=26 for the binding pocket). The prediction errors are presented separately for the entire structure (global Cα RMSD), for the binding pocket (iRMSD), and for retinal (L-RMSD). Means are shown over each box. Only four predicted models in benchmark set A showed an error larger than 1 Å in the binding-pocket residues (iRMSD): TR (PDB ID: 5AZD); *Ns*XeR (PDB ID: 6EYU); Chrimson (PDB ID: 5ZIH), and XR (PDB ID: 3DDL). Four models in blind set B also exceeded this threshold: *Er*NaR (PDB ID: 8QLF) and three potassium channelrhodopsins (8H86, 8H87, and 8JXO). The most challenging cases were in blind set B: B1ChR2 (PDB ID: 8JXO, pocket iRMSD 1.68 Å, L-RMSD 2.08 Å) and Tara-RRB (PDB ID: 7PL9), a rhodopsin functionally coupled to the bestrophin chloride channel (excluded from pocket analysis).

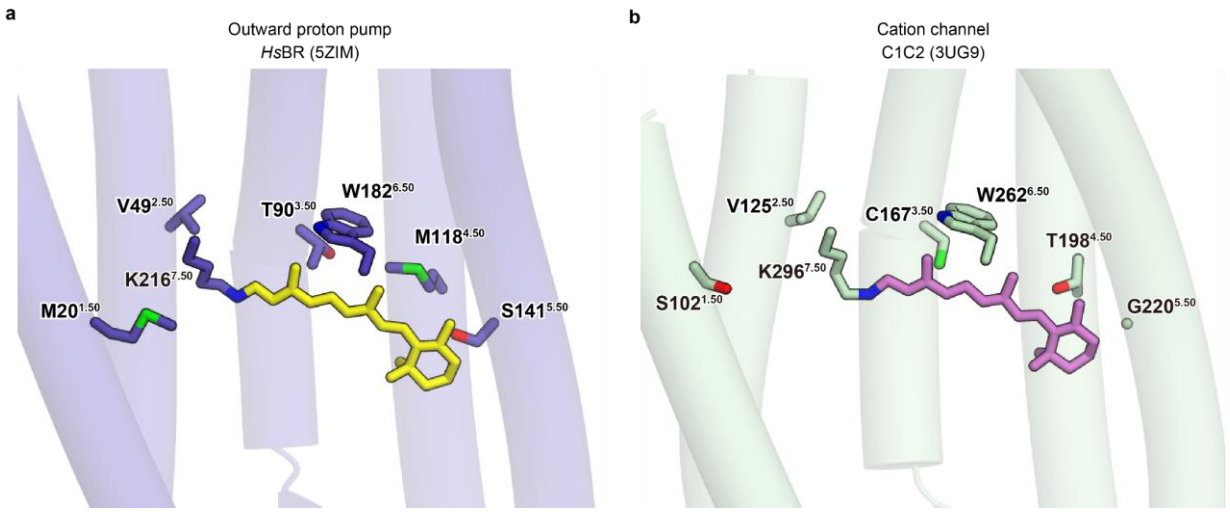
Supplementary Figure 2



Supplementary Fig. 2 | a Average minimum distance from each structurally equivalent residue position to the retinal chromophore (closest Cα atom), calculated across the dataset. Vertical lines indicate the defined X.50 anchor position for each helix (1.50–7.50). Horizontal lines indicate key interaction distances for side chains (shown for comparison with Figure 4b): 1.5 Å approximates a covalent bond length and is relevant for the Schiff base linkage; 4 Å represents a typical cutoff for non-covalent interactions, delineating the direct binding pocket; and 6 Å encompasses the broader binding-site environment and is the radius used for iRMSD calculations. **b** Sequence logos depicting the variability around the anchor positions in TM1–TM7 for the rhodopsins in the structural dataset (n = 129).

19

Supplementary Figure 3



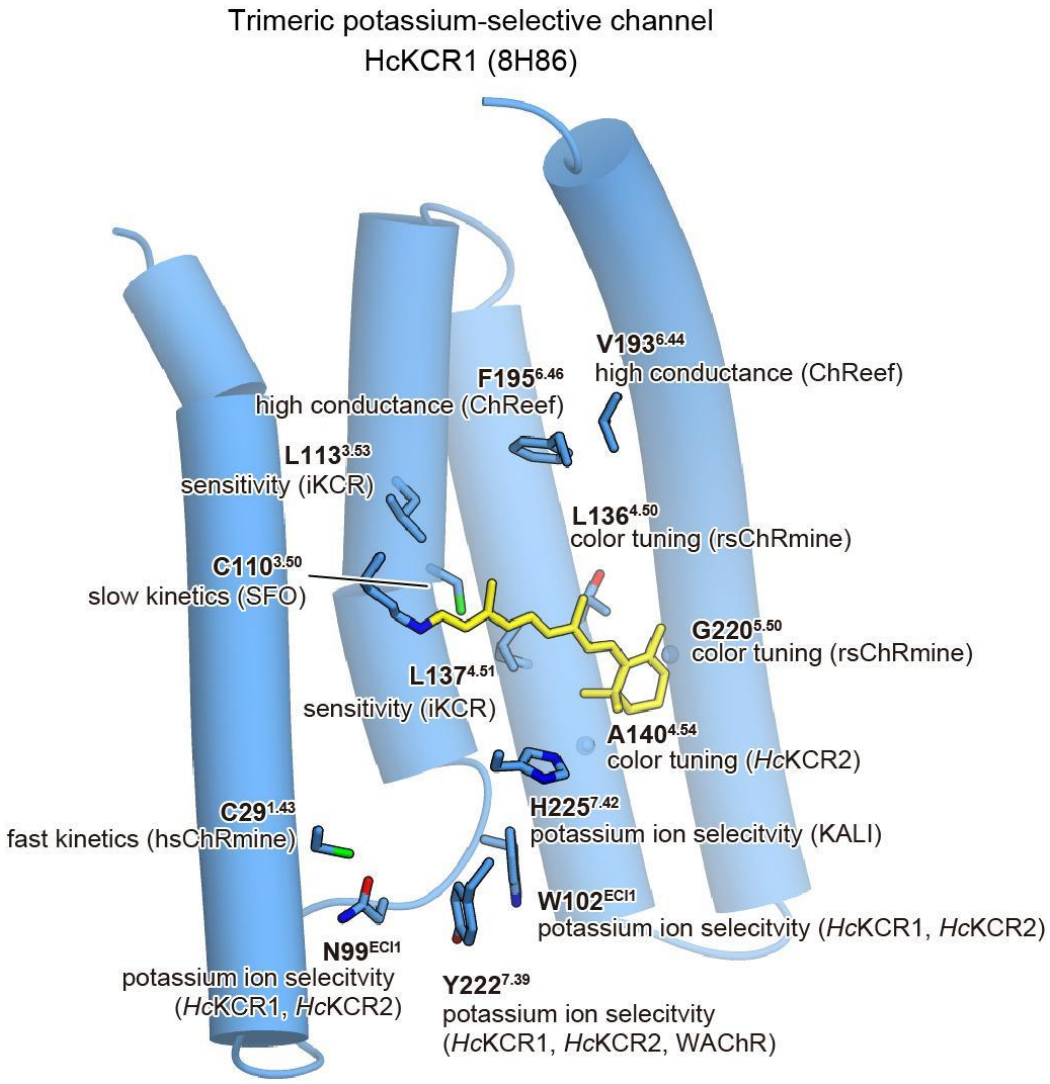
Supplementary Fig. 3 | Examples of x.50 residue positions. The x.50 residues are mapped onto the structures of **a** the representative outward proton-pumping rhodopsin *Hs*BR and **b** the cation-conducting channelrhodopsin C1C2.

20

21

22

Supplementary Figure 4



Supplementary Fig. 4 | Distribution of functionally important residues identified across various trimeric channelrhodopsin variants. Residues are mapped onto the overall structure of the potassium-selective channelrhodopsin *Hc*KCR1 (PDB ID: 8H86).

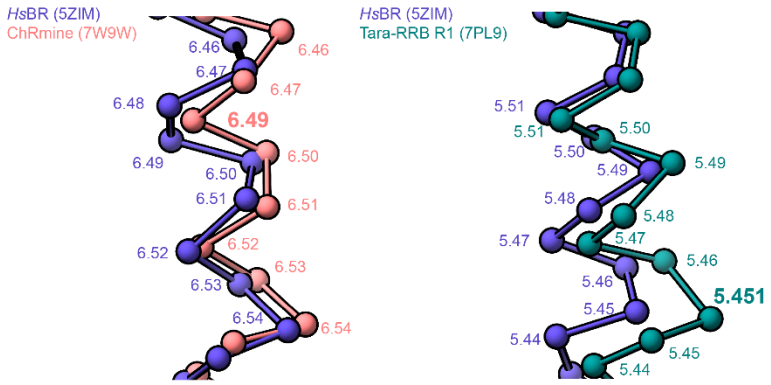
Supplementary Figure 5

a

*Hs*BR_no	179	180	181	182	183	184	185	186	187
ChRmine_no	221	-	222	223	224	225	226	227	228
MO_no	6.47	6.48	6.49	6.50	6.51	6.52	6.53	6.54	6.55
*Hs*BR	V	V	L	W	S	A	Y	P	V
ChRmine	M	-	L	W	I	L	F	P	L

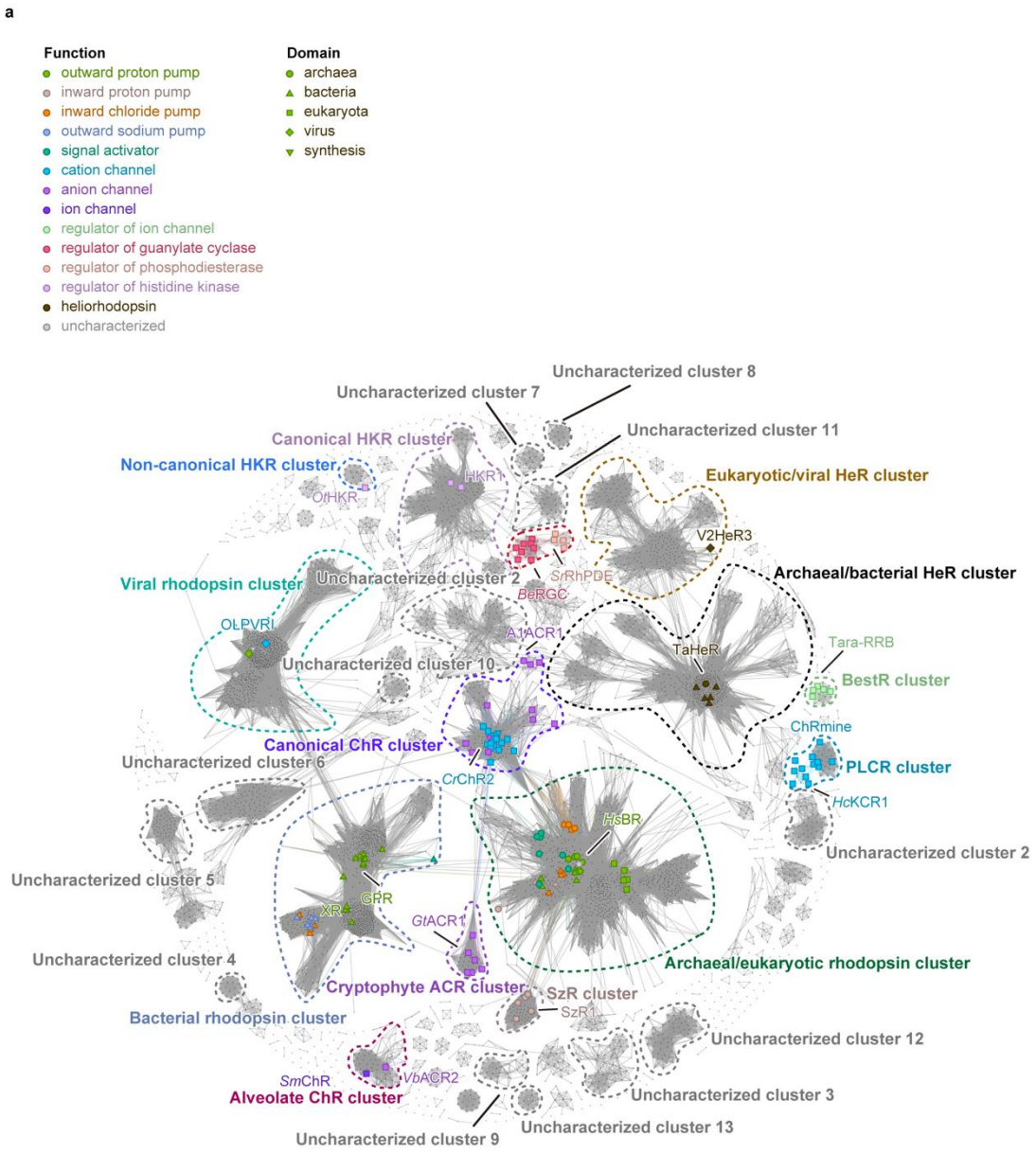
*Hs*BR_no	135	136	-	137	138	139	140	141	142
TaraRRB_no	250	251	252	253	254	255	256	257	258
MO_no	5.44	5.45	5.451	5.46	5.47	5.48	5.49	5.50	5.51
*Hs*BR	F	V	-	W	W	A	I	S	T
TaraRRB R1	Y	A	I	F	L	S	A	S	W

b



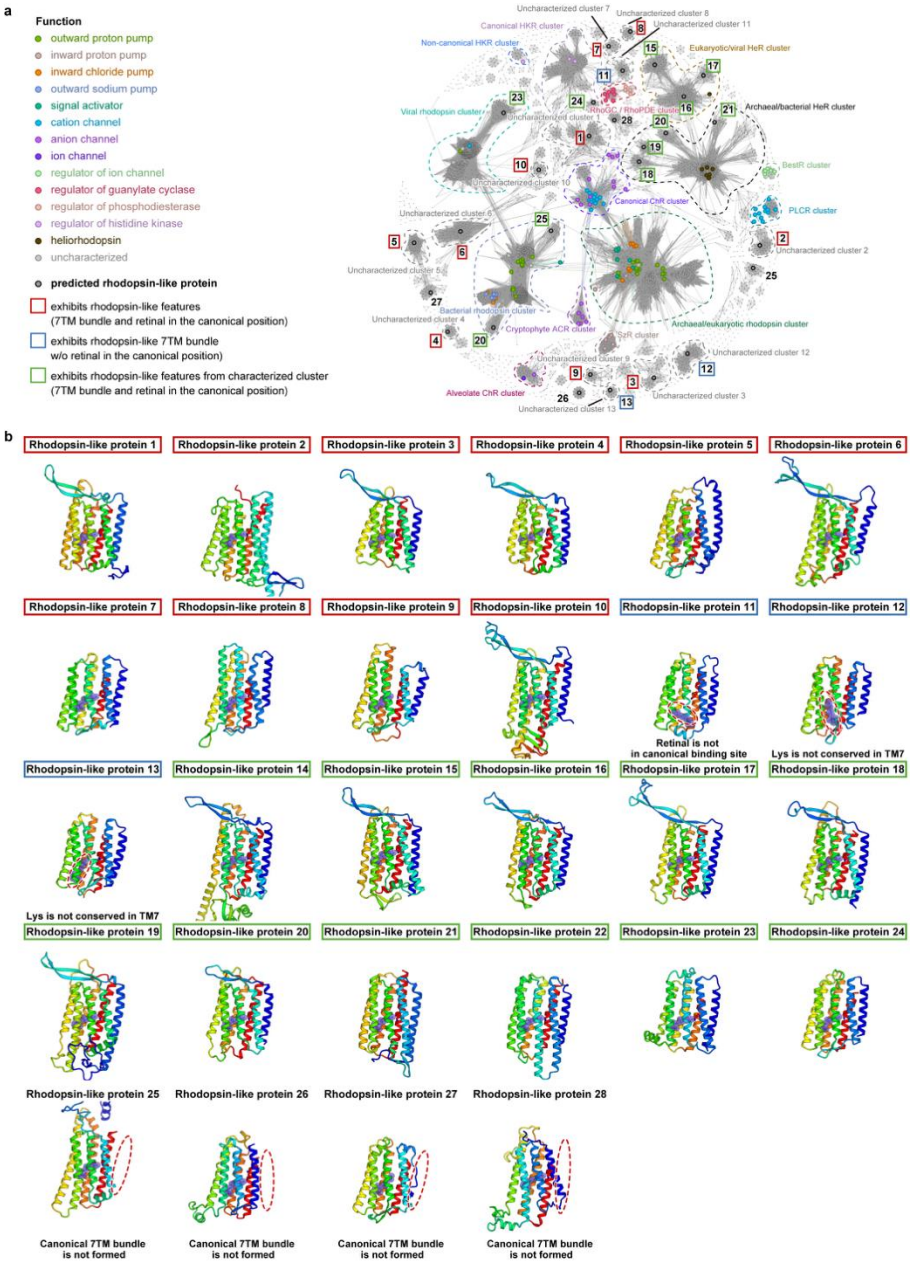
Supplementary Fig. 5 | Rules for insertions and gaps in the MO numbering system. a Structure-based alignment of TM6 between the reference *HsBR* and ChRmine (top) and of TM5 between *Hs*BR and TaraRRB-R1 (bottom). *Hs*BR_no and ChRmine_no/TaraRRB_no indicate the residue numbers in each protein, and the MO numbering row shows the corresponding numbers in the MO numbering system. In the gap in the ChRmine sequence, the GRN is skipped. The insertion in TaraRRB-R1 is assigned a fractional GRN (5.451). **b** Superposition of Cα backbone traces of the reference *Hs*BR (PDB ID: 5ZIM; blue) with ChRmine (PDB ID: 7W9W; pink, left) and with TaraRRB-R1 (PDB ID: 7PL9; green, right). Numbers along the helices indicate the GRNs, highlighting the positions of the gap (6.49, left) and the fractional GRN (5.451, right) corresponding to the alignment shown in panel a.

Supplementary Figure 6



Supplementary Fig. 6 | Sequence similarity network of microbial rhodopsins. Sequence similarity network composed of approximately 40,000 non-redundant microbial rhodopsin sequences, including a large collection of uncharacterized rhodopsin-like sequences. Six major and eight minor clusters containing functionally characterized rhodopsins, as well as 13 clusters containing only uncharacterized rhodopsins, are highlighted with dashed circles. Representative rhodopsins are mapped onto the clusters, and the colors and shapes of the labels indicate function and domain (biological taxonomy), respectively.

Supplementary Figure 7



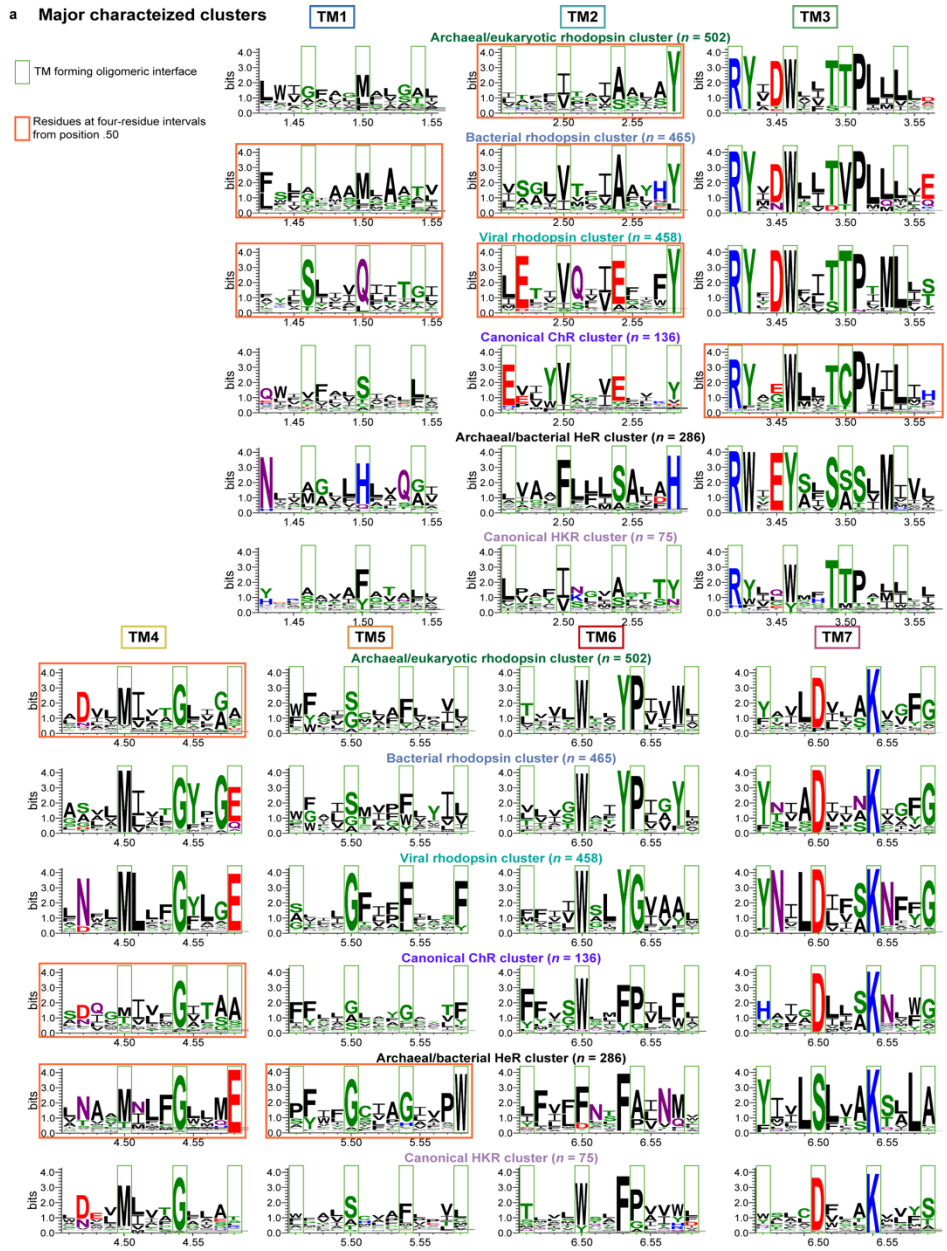
Supplementary Fig. 7 | a Sequence similarity network composed of approximately 40,000 non-redundant microbial rhodopsin sequences, including a large collection of uncharacterized rhodopsin-like sequences. Six major and eight minor clusters containing functionally characterized rhodopsins, as well as 13 clusters containing only uncharacterized rhodopsins, are highlighted with dashed circles. Representative rhodopsins are mapped onto the clusters, and the colors and shapes of the labels indicate function and domain (biological taxonomy),

respectively. In addition, 28 rhodopsin-like sequences selected from the peripheral regions of characterized and uncharacterized clusters are highlighted. **b** Predicted structures of the 28 rhodopsin-like sequences selected from the peripheral regions of characterized and uncharacterized clusters. Four of the 28 sequences (25–28) exhibit a non-rhodopsin fold, one (11) shows non-canonical binding of the retinal chromophore, and two (12 and 13) lack the conserved lysine on TM7 that covalently binds the retinal chromophore.

29

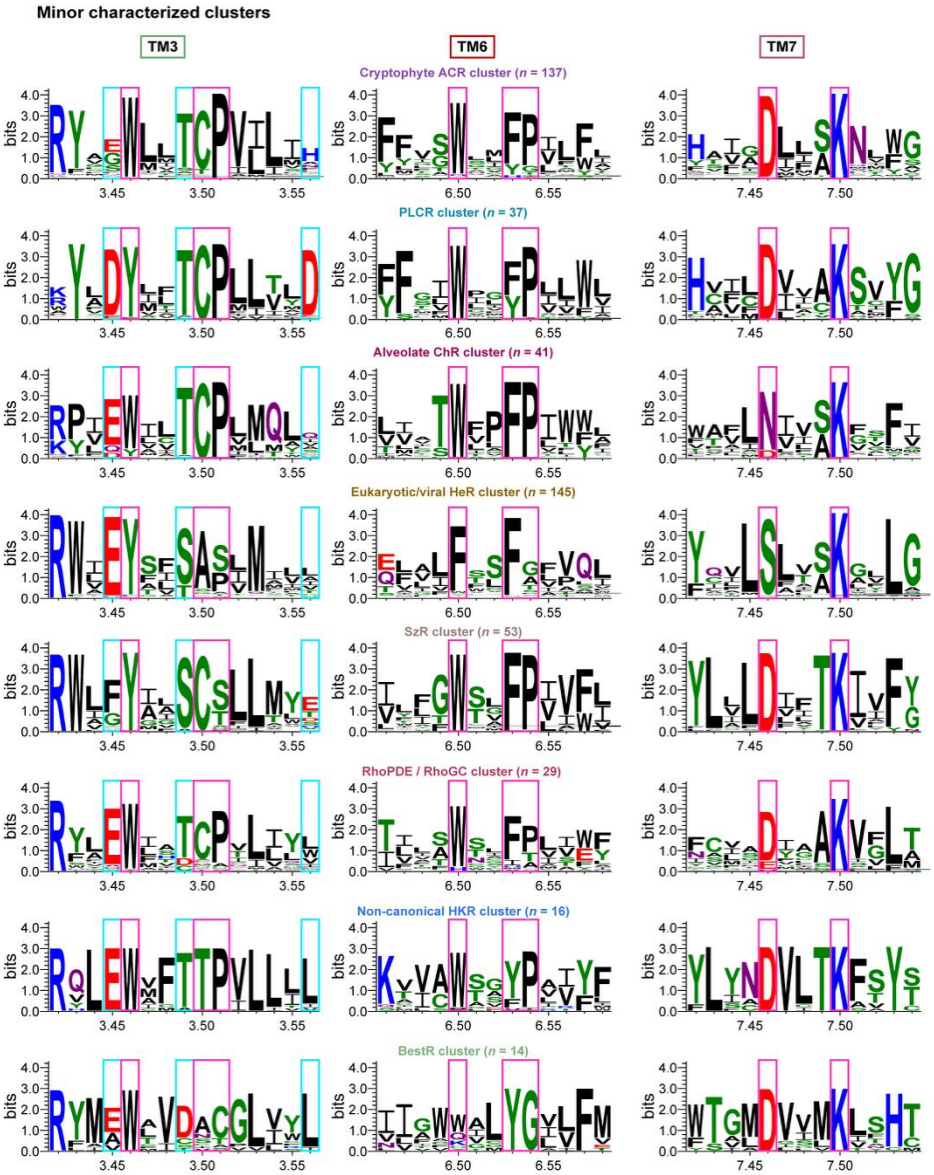
30

Supplementary Figure 8



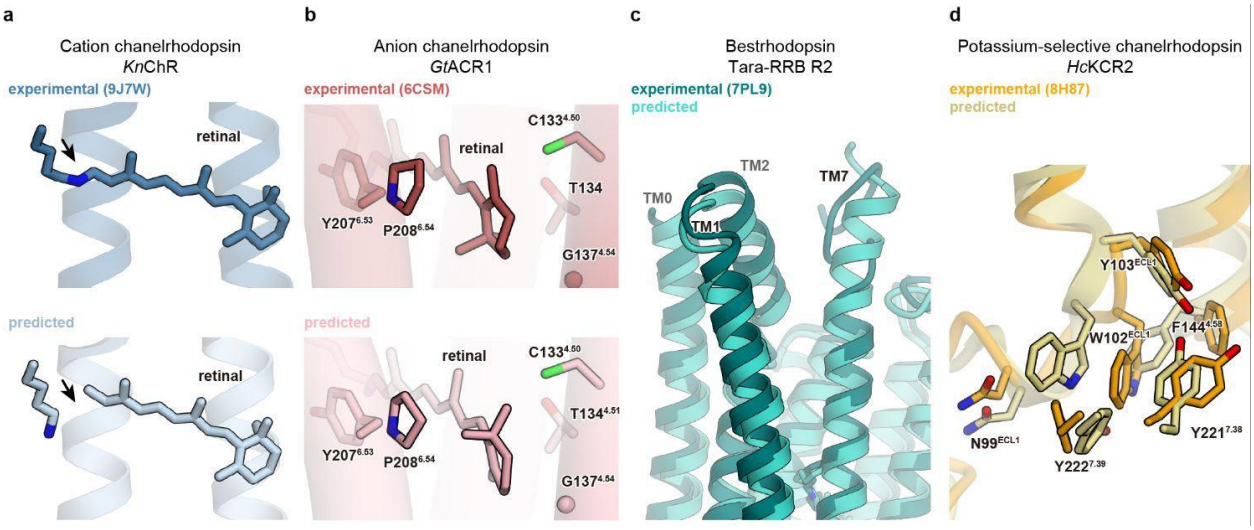
Supplementary Fig. 8 | Sequence logos of TM1–TM7 for the six major clusters containing functionally characterized rhodopsins. Transmembrane helices forming oligomeric interfaces and residues at four-residue intervals from position x.50 are highlighted with orange and green rectangles, respectively.

Supplementary Figure 9



Supplementary Fig. 9 | Sequence logos of TM3, TM6, and TM7 for eight minor clusters containing functionally characterized rhodopsins. Cyan rectangles indicate residues of the DTD motif, and magenta rectangles indicate highly conserved residues within each cluster.

Supplementary Figure 10



Supplementary Fig. 10 | Current limitations of structural predictions. Representative examples highlighting structural discrepancies between experimentally determined and computationally predicted structures. **a** In the predicted structure of *Kn*ChR, the retinal is not covalently linked to the Schiff base lysine. **b** In *Gt*ACR1, the β-ionone ring of the retinal adopts an incorrect twisted conformation. **c** In Tara-RRB R2, the cytoplasmic ends of TM0, 1, 2, and TM7 are poorly aligned. **d** In *Hc*KCR2, the architecture of the potassium selectivity filter is incorrectly predicted.