

A Generic Residue-Numbering system for Microbial Rhodopsins

Flurin S. Hidber^{1,2,10#} Seiya Tajima^{3,4#}, Koichiro E. Kishi^{3,4,#}, Yosuke Nishimura^{5,6}, Susumu Yoshizawa^{5,7,8}, Xavier Deupi^{1,9,10,*}, Hideaki E. Kato^{3,4,11,12,13,*}

¹ Condensed Matter Theory Group, PSI Center for Scientific Computing, Theory, and Data, 5232 Villigen-PSI, Switzerland

² Department of Biology, ETH Zürich, Zürich, Switzerland

³ Research Center for Advanced Science and Technology, The University of Tokyo, Meguro, Tokyo, Japan.

⁴ Department of Life Sciences, School of Arts and Sciences, The University of Tokyo, Meguro, Tokyo, Japan.

⁵ Atmosphere and Ocean Research Institute, The University of Tokyo, Chiba, Japan

⁶ Research Center for Bioscience and Nanoscience (CeBN), Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa, Japan

⁷ Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

⁸ Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo, Japan

⁹ Laboratory of Biomolecular Research, PSI Center for Life Sciences, 5232 Villigen-PSI, Switzerland

¹⁰ Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

¹¹ Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo, Japan.

¹² FOREST, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan.

¹³ CREST, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan.

Flurin S. Hidber, Seiya Tajima, and Koichiro E. Kishi contributed equally to this work.

*Correspondence: Xavier Deupi (xavier.deupi@psi.ch) and Hideaki E. Kato (c-hekato@g.ecc.u-tokyo.ac.jp).

Abstract

Microbial rhodopsins are photoreceptors with a seven-transmembrane architecture that use the chromophore retinal—a vitamin A derivative—to convert light absorption into ion transport, signaling, and enzymatic activity. Since their discovery half a century ago, genome sequencing has uncovered thousands of these proteins across archaea, bacteria, eukaryotes, and viruses, revealing a diverse functional repertoire that includes ion pumps and channels, transducer activators, and enzyme/channel regulators. Although the structure of the seven-transmembrane bundle and the retinal-binding pocket is highly conserved, the sequence identity across major classes is low. This low sequence identity, combined with the rapidly expanding number of microbial rhodopsins identified by metagenomics, makes direct comparisons across the superfamily increasingly difficult. Here, we introduce a generic residue-numbering system for microbial rhodopsins (MO numbering system), a structure-based framework anchored to the retinal-binding pocket that enables consistent residue comparison across even the most divergent family members. Analyzing 129 structures and approximately 40,000 sequences, we demonstrate its utility by analyzing sequence–structure–function relationships across the superfamily, including previously uncharacterized members. The MO numbering system provides a common language for organizing the growing repertoire of microbial rhodopsins, integrating structural, spectroscopic, and electrophysiological data, and engineering next-generation optogenetic tools.

Introduction

The microbial rhodopsin superfamily comprises light-sensitive proteins that use the chromophore retinal—a vitamin A derivative—to convert photon absorption into concerted cellular events. Since the discovery of the proton pump bacteriorhodopsin in *Halobacterium salinarum* (*HsBR*) over fifty years ago¹, microbial rhodopsins have revealed a striking diversity of functions. These proteins are found across all domains of life—Archaea, Bacteria, and Eukaryota—and viruses^{2, 3, 4} functioning as proton, chloride, or sodium pumps, cation and anion channels, activators of downstream signal transducers, and even as enzyme/channel/transporter regulators. Their photochemical properties are equally varied: while in most microbial rhodopsins light isomerizes all-*trans* retinal to the 13-*cis* form, some produce 9-*cis* or 11-*cis* isomers as photoproducts^{2, 5, 6, 7}.

Despite this functional and photochemical diversity, all microbial rhodopsins share a core architecture: a tight bundle of seven transmembrane helices (TMs) surrounding the retinal-binding pocket, with retinal bound via a Schiff base to a conserved lysine on TM7. Some families form oligomers ranging from dimers to hexamers^{6, 8, 9, 10, 11, 12, 13, 14}, while others have an additional TM or an inverted membrane topology^{6, 10, 15, 16}. Remarkably, this structural conservation persists even though sequence identity between major families can be as low as 10–12%¹⁷, with a nearly identical orientation of retinal in the binding pocket. Such conservation despite extreme sequence divergence¹⁸ suggests powerful evolutionary constraints on the architecture required for retinal binding and photoisomerization.

Microbial rhodopsins are often classified, annotated, and compared by tracking characteristic sequence motifs that are coupled to their function. However, the high sequence variability within the family complicates the analysis of large datasets obtained in metagenomics studies, particularly the comparison of distant subfamilies. Researchers in the field often resort to developing *ad hoc* sequence numbering systems using a reference rhodopsin, such as *HsBR*, channelrhodopsin-2 from *Chlamydomonas reinhardtii* (*CrChR2*, the most common optogenetics tool)¹⁹, or C1C2 also from *Chlamydomonas reinhardtii* (the first channelrhodopsin structure, a chimera between channelrhodopsin-1 and 2)¹¹. This makes it challenging to recognize equivalent positions and conserved motifs in different rhodopsins throughout the scientific literature. Other fields have overcome this hurdle by adopting generic residue-numbering (GRN) systems. A prime example is the field of G protein-coupled receptors (GPCRs), in which the Ballesteros–Weinstein GRN system²⁰—or its descendant GPCRdb system²¹—assign a unique index to equivalent positions in different proteins. Similar systems developed for microbial rhodopsins are limited to a few specific families²² and cannot be applied across the entire superfamily.

In this work, we develop a GRN system for microbial opsins (MO numbering system) based on the structure of the retinal-binding site, rather than on sequence conservation. This numbering system provides a consistent framework for comparing residues across even the most divergent rhodopsins. Using a dataset of 129 structures and approximately 40,000 sequences, we demonstrate its utility by mapping functionally important residues across ion pumps, channels, and sensory rhodopsins, and by identifying conserved sequence signatures in both characterized and previously uncharacterized rhodopsin families.

Results

Assembly of a dataset of microbial rhodopsin structures

Comparative structural analysis of microbial rhodopsins has been somewhat hindered by limited structural coverage of certain functional classes and taxonomic groups. To address this limitation, we assembled a comprehensive dataset of 129 structures, combining 69 experimentally determined structures and 60 computationally predicted models (**Fig. 1**). The dataset includes rhodopsins from all domains of life—archaea (27, 21%), bacteria (42, 33%), and eukaryota (50, 39%)—and also includes synthetic variants (5, 4%) and viral rhodopsins (5, 4%). Functionally, the dataset spans proton pumps (42, 33%), cation channels (28, 22%), regulatory rhodopsins with fused domains (14, 11%), anion channels (12, 9%), chloride pumps (12, 9%), transducer activators (8, 6%), sodium pumps (6, 5%), and rhodopsins of undefined function (7, 5%). Due to biases in available structural databases, proton pumps are overrepresented in the experimentally determined structures (31/69, 45%). To balance functional coverage, we enriched underrepresented families through predictive structural modeling: enzyme/channel/transporter regulators (12 predictions), cation channels (11), proton pumps (11), anion channels (8), chloride pumps (6), transducer activators (6), sodium pumps (3), and rhodopsins of undefined function (3).

The dataset reflects known taxonomic distributions of rhodopsin functions. Channelrhodopsins—both cation-conducting (23/28) and anion-conducting (10/12)—are concentrated in eukaryotes, consistent with their restriction to algae and related protists²³. Similarly, enzyme/channel/transporter regulatory rhodopsins are predominantly eukaryotic (12/14), consistent with enzyme rhodopsins having been identified mostly in fungi, green algae, and choanoflagellates¹⁰. Sodium pumps appear exclusively in bacteria (6/6), consistent with their known distribution in marine flavobacteria and related lineages²⁴. Chloride pumps are found in both archaea and bacteria, reflecting independent evolutionary origins in these lineages²⁵. Proton pumps show the broadest distribution across Archaea, Bacteria, and Eukaryota, consistent with ancestral sequence reconstruction studies suggesting that the last common ancestor of microbial rhodopsins functioned as a light-driven proton pump²⁶.

Validation of the structural predictions

As nearly half of our dataset consists of predicted models, we first assessed their accuracy using two complementary validation sets of experimental structures: benchmark set A ($n = 42$), composed of structures deposited before September 2021 (which may have been used in the training of the structure prediction methods), and blind set B ($n = 27$), composed exclusively of structures deposited after the model training cutoff. We assessed the accuracy of the models by comparing the protein backbone, the retinal-binding pocket, and retinal (**Supplementary Fig. 1**). The predicted models have high accuracy. Models in benchmark set A had a mean backbone C α RMSD of 0.61 ± 0.23 Å, binding-pocket iRMSD of 0.51 ± 0.28 Å, and retinal L-RMSD of 0.40 ± 0.24 Å. Models in blind set B had a mean backbone C α RMSD of 0.80 ± 0.24 Å, binding-pocket iRMSD of 0.75 ± 0.28 Å, and retinal L-RMSD of 0.62 ± 0.27 Å. The errors in blind set B are consistently higher, but the mean differences between the predicted models and the reference structures remain below 1 Å.

To illustrate the accuracy of the predicted models, we show experimental and predicted structures for six representative rhodopsins spanning different functional classes (**Fig. 2**). In all cases, the overall fold was accurately reproduced, with loops largely correctly positioned (**Fig. 2a**). The retinal-binding pocket was also consistently modeled at sub-angstrom precision (**Fig. 2b**), including accurate handling of unusual 6-s-cis β-ionone ring conformations (as observed in the cation channelrhodopsin *KnChR*), consistent with the non-G rule^{27, 28, 29}. These results show that current structure prediction methods reliably capture both the global microbial rhodopsin fold and the details of the retinal-binding pocket, supporting the inclusion of predicted structures to increase the diversity of our dataset.

Overall structural conservation of the microbial rhodopsin fold

To analyze structural conservation across the dataset, we calculated pairwise RMSDs between C α atoms of the transmembrane helices for all structure pairs. The resulting RMSD matrix (**Fig. 3**) reveals substantial structural conservation across the superfamily. However, sequence identity is low, averaging $21 \pm 9\%$ in the TM bundle (min = 7%) and $19 \pm 9\%$ in the full sequence (min = 6%). Hierarchical clustering of the RMSD matrix identified a single main structural cluster, confirming that microbial rhodopsin fold is highly conserved across the superfamily. A generic residue-numbering system for microbial rhodopsins should therefore be based on structural features rather than sequence.

Development of the generic residue-numbering system

The MO numbering system is inspired by the Ballesteros-Weinstein system for G protein-coupled receptors but adapted to the unique architecture of microbial rhodopsins. MO numbering codes consist of two numbers: the first denotes the helix (1-7) and the second the position within the helix relative to the residue closest to retinal, defined as position 50. For example, position 7.46 denotes a residue in TM7, four residues before the Schiff base-forming lysine (defined as 7.50).

The first task in developing the MO numbering system was identifying the closest residues to retinal in each helix (**Fig. 4a**), which serve as the ‘anchors’ of the numbering system. The distance profiles along each helix (**Fig. 4b**) exhibit the expected systematic variations, reaching their minima at central positions and increasing toward the N- and C-termini. The binding pocket (<4 Å around retinal) is formed by residues from TM3-7. Only two residues in TM2 are part of the extended pocket (<6 Å), while TM1 is at a peripheral location within the 7-TM bundle and further from retinal. The closest residue to retinal in each helix is designated as an anchor (**Fig. 4b, vertical lines**). The sequence variability around the anchor positions (**Supplementary Fig. 2b**) shows that they lie within or adjacent to regions of elevated sequence conservation despite the overall sequence divergence of the superfamily. These data support the choice of a ligand-centric approach to capture structurally and functionally critical positions across diverse microbial rhodopsins (**Supplementary Fig. 3**).

Application of the generic residue-numbering system

To demonstrate the application of the MO numbering system, we first focus on residues that form the retinal-binding pocket, using *HsBR* as a reference (**Fig. 5a**). Retinal is covalently bound to K216^{7.50} (GRNs are shown as superscripts) via a Schiff base linkage and is surrounded by 18 additional residues from TM3-7 (5 from TM3, 3 from TM4, 4 from TM5, 4 from TM6, and 2 from TM7). Two highly conserved carboxylates, D^{3.45} and D^{7.46}, located near K^{7.50} (D85^{3.45} and D212^{7.46} in *HsBR*), are historically referred to as the “Schiff base counterions” and stabilize the positive charge of the protonated Schiff base. Residues at positions 3.53, 4.51, 6.54, and 7.49—known as the L/Q, N/LI, G/P, and A/TS switches, respectively—contribute to spectral tuning in microbial rhodopsins^{30, 31, 32, 33}. In addition, position 4.54 (G122^{4.54} in *HsBR*) plays an important role in determining retinal planarity: when a non-glycine residue occupies this position, steric clash with the retinal β-ionone ring can induce rotation and favor a 6-s-cis retinal conformation (the “non-G rule”)^{27, 28, 29}. Two other notable positions near the β-ionone ring are 5.44 and 5.47. Although retinal is enclosed by TM3-7, many microbial rhodopsins have a small lateral opening between TM5 and TM6 that connects the retinal pocket to the lipid bilayer^{11, 13, 34, 35, 36}. The shape of this opening varies among microbial rhodopsins: in some bacterial ion-pumping rhodopsins, residue 5.47 contributes to its formation, whereas in heliorhodopsins (HeR) both 5.44 and 5.47 contribute. Functionally, 5.44 has been implicated in retinal uptake from the membrane—the G171^{5.44}W mutation in *TsHeR* inhibits this process¹³—while 5.47 is involved in binding the carotenoid antenna pigment, as evidenced by the G178^{5.47}W mutation in *Gloeoebacter* rhodopsin, which inhibits carotenoid binding^{37, 38}.

The MO numbering system also facilitates comparative analyses across subfamilies. As an example, in ion-pumping rhodopsins, the identity of the transported substrate and the direction of transport are primarily

influenced by three residues at positions 3.45, 3.49, and 3.56, which are located on one face of TM3 (**Fig. 5b**)^{2, 4, 25}. Ion-pumping rhodopsins containing D^{3.45}, T^{3.49}, and D^{3.56}—known as the DTD motif—predominantly function as outward proton pumps. Upon photoactivation, the proton is released from the Schiff base to the extracellular bulk solvent via T^{3.49}, D^{3.45}, and the proton release group composed of E^{ECL3} and E^{7.38}, accompanied by a conformational change of the key arginine residue R^{3.42}. A new proton is subsequently supplied from the intracellular bulk solvent to the Schiff base via D^{3.56}, thereby accomplishing net transport of a single proton from the intracellular to the extracellular side. In contrast, inward proton-pumping rhodopsins, which comprise schizorhodopsins (SzRs) and xenorhodopsins (XeRs), lack one of the two counterions near the Schiff base. Specifically, D85^{3.45} in HsBR is replaced by F^{3.45} in SzRs, whereas D212^{7.46} in HsBR is replaced by P^{7.46} in XeRs. The D^{3.45}–T^{3.49}–D^{3.56} motif is replaced by FSE in SzRs and by DSA, DTL, or DTS in XeRs, respectively, and the proton release group conserved in outward proton pumps is absent in both SzRs and XeRs^{39, 40}. Inward chloride pumps have T^{3.45}–S^{3.49}–A/D^{3.56} or N^{3.45}–T^{3.49}–Q^{3.56} motifs, and the loss of negative charge caused by replacement of D^{3.45} with T^{3.45} or N^{3.45} is compensated by binding of the chloride substrate^{41, 42, 43}. Outward sodium pumps share the N^{3.45}–D^{3.49}–Q^{3.56} motif, in which transient proton transfer from the Schiff base to D^{3.49} enables the sodium ion to pass through the Schiff base region^{24, 35, 44, 45}.

The MO numbering system is equally applicable to rhodopsins with non-canonical architectures. Heliorhodopsins (HeRs), for instance, have an inverted topology, with an intracellular N-terminus and an extracellular C-terminus (**Fig. 5c**). Despite this inversion, our structural alignment and annotation approach remains effective. At the sequence level, HeRs retain a TM3 counterion at position 3.45, whereas the TM7 counterion at 7.46 is replaced by serine. While the molecular functions of most HeRs remain unclear, a recent study suggested that V2HeR3 functions as a proton transporter⁴⁶, and the MO numbering system may facilitate future comparative analyses as more functional data become available.

Channelrhodopsins, widely used in optogenetics, present a challenge for systematic comparison: mutagenesis and protein engineering have generated numerous functional variants, while metagenomics continues to uncover natural diversity. The MO numbering system enables mapping of this diversity onto a common structural framework, as illustrated for C1C2 structure (**Fig. 5d**). Unlike ion-pumping rhodopsins, the TM3 residues at positions 3.45, 3.49, and 3.56 are not essential for channel function in canonical dimeric channelrhodopsins; substitutions at these and nearby positions, such as E^{3.45}T or H^{3.56}R, primarily affect channel kinetics^{47, 48}. Instead, ion selectivity and kinetics are largely determined by five glutamates (E1–E5, corresponding to E^{2.46}, E^{2.47}, E^{2.54}, E^{2.61}, and E^{2.65}) aligned along TM2^{49, 50, 51, 52}. The DC gate, formed by D^{4.47} and C^{3.50}, controls channel closure kinetics and light sensitivity, enabling bistable optical control of neuronal activity^{53, 54, 55}. **Fig. 5d** summarizes additional positions implicated in spectral tuning, conductance, and kinetics^{27, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68}, and **Supplementary Fig. 4** summarizes positions that affect properties in trimeric channelrhodopsins^{14, 28, 69, 70, 71, 72, 73, 74, 75}.

Transducer-activating rhodopsins (sensory rhodopsins) relay light signals to downstream effectors. Structural, spectroscopic, and mutational studies have clarified the activation mechanism of *NpSRII*^{76, 77, 78} (**Fig. 5e**). *NpSRII* forms a 2:2 complex with its transducer HtrII through an interface formed by TM6 and TM7 of *NpSRII* and TM1 and TM2 of HtrII. In *NpSRII*, retinal photoisomerization induces structural changes in residues Y174^{6.53}, L200^{7.45}, D201^{7.46}, and T204^{7.49}, which are transmitted to HtrII via the interaction between Y199^{7.44} and N74 of HtrII^{78, 79}. Notably, the key residue T204^{7.49} in *NpSRII* is not conserved in *H. salinarum* SRI (*HsSRI*), suggesting that this signaling pathway may be specific to *NpSRII* rather than a general feature of transducer-activating rhodopsins⁸⁰.

Enzyme-fused rhodopsins, such as rhodopsin-guanylyl cyclases and rhodopsin-phosphodiesterases, couple light detection to enzymatic activity. A characteristic feature of these proteins—shared with some channel-fused rhodopsins—is an additional TM0 helix at the N-terminus of the seven-transmembrane core^{6, 16}. The MO numbering system annotates the seven helices forming the retinal-binding domain, excluding TM0. For NeoR, the longest-wavelength rhodopsin reported to date, functional studies have identified three acidic residues

essential for activity: E136^{3,45}, D140^{3,49}, and E262^{7,46 81} (**Fig. 5f**). As structural and spectroscopic data on enzyme-fused rhodopsins accumulate, the MO numbering system will facilitate comparative analysis of their activation mechanisms.

The MO numbering system also accommodates microbial rhodopsins that exhibit gaps or insertions arising from local structural distortions (**Supplementary Fig. 5**). For example, ChRmine, a trimeric cation-conducting channelrhodopsin⁸², contains a constricted 3₁₀ helical turn within TM6 that results in a gap in the sequence alignment¹⁴ (**Supplementary Fig. 5, salmon**). In this case, the ‘missing’ position (6.48) is simply skipped. Conversely, TaraRRB-R1, a channel-regulating rhodopsin⁶, contains a bulged π-helical turn in TM5 that introduces an additional residue between 5.45 and 5.46 (**Supplementary Fig. 5, green**), which is designated 5.451.

Together, these applications demonstrate that the MO numbering system provides a unified framework for comparing rhodopsins across functional classes, taxonomic origins, and structural variations.

Application of the MO numbering system to a large dataset of microbial rhodopsins

To test the MO numbering system at scale, we applied it to a large sequence dataset that includes both characterized and uncharacterized rhodopsins. We combined the 129 functionally characterized rhodopsins in our curated dataset with a large collection of uncharacterized rhodopsin-like sequences, yielding an initial set of over 300,000 sequences from archaea, bacteria, eukaryotes, and viruses (Nishimura et al., in preparation). After curation, we obtained approximately 40,000 non-redundant sequences. From pairwise sequence similarities, we constructed a protein similarity network in which nodes represent individual sequences and edges connect pairs exceeding a defined similarity threshold (**Fig. 6a**). The network resolved 14 clusters containing functionally characterized rhodopsins and 17 clusters containing only uncharacterized rhodopsins (**Supplementary Fig. 6**). The 14 characterized clusters comprise six major groups—archaeal/eukaryotic, bacterial, viral (VirR)^{83, 84}, canonical channelrhodopsins (ChRs), canonical histidine kinase-fused rhodopsins (HKRs)^{85, 86}, and canonical heliorrhodopsins (HeRs)⁸⁷—and eight minor groups: schizorhodopsins (SzRs)⁸⁸, cryptophyte anion-conducting channelrhodopsins (ACRs)⁸⁹, alveolate ChRs²³, pump-like channelrhodopsins (PLCRs)⁹⁰, guanylate cyclase-/phosphodiesterase-fused rhodopsins (RhoGC/RhoPDEs)¹⁶, bestrophin-fused rhodopsins (BestRs)⁶, noncanonical HKRs⁹¹, and eukaryotic/viral HeRs⁴⁶ (**Supplementary Fig. 6**).

To validate the clustering and explore uncharacterized rhodopsins, we predicted structures for 11 sequences from the periphery of 5 of the 14 functionally characterized clusters, along with one representative member from each of the 17 uncharacterized clusters (**Supplementary Fig. 7a**). In total, we generated structural models for 28 rhodopsin-like proteins (**Supplementary Fig. 7b**). All 11 characterized and 13 of the 17 uncharacterized sequences (uncharacterized clusters 1–13) adopted a rhodopsin fold and were retained for further analysis. Of the 13 uncharacterized rhodopsins, two lacked the conserved Lys^{7,50} and one exhibited an incorrectly predicted retinal-binding pose. Several uncharacterized clusters exhibited distinctive structural features: for example, the rhodopsin in cluster 2 has an unusually long TM1, whereas the rhodopsin in cluster 10 contains β-strands in ECL2 and ECL3 that form an extracellular β-sheet (**Supplementary Fig. 7b**).

To examine sequence conservation, we generated sequence logos in selected clusters (**Fig. 6b,c; Supplementary Fig. 8 and 9**). Conserved residues grouped near the X.50 anchor positions at approximately four-residue intervals, consistent with enrichment on the retinal-facing side of each helix; lipid-exposed positions showed lower conservation. Across all clusters, residues in TM3, TM6, and TM7 were more conserved than those in other helices. The arginine at position 3.42 is highly conserved across all functionally characterized clusters—with the exception of the PLCR cluster—consistent with its known key functional roles^{92, 93, 94, 95, 96}. Residues forming the retinal-binding pocket (positions 3.46, 6.50, 6.53, and 6.54) and the Schiff base counterions (positions 3.45 and 7.46) are well conserved within characterized clusters; however, position 7.46 is more conserved than 3.45. Most characterized clusters have a highly conserved Asp at 7.46, but the alveolata ChR, bacterial/archaeal HeR, and eukaryotic/viral HeR clusters show highly conserved Asn, Ser, and Ser, respectively. Positions 3.50 and 3.51

vary markedly between characterized clusters—while remaining conserved within each—suggesting that these positions are linked to cluster-specific functions and serve as characteristic sequence signatures for these rhodopsin families. Similar conservation patterns were observed in uncharacterized clusters, although position 3.50 in cluster 1 and position 7.46 in cluster 4 showed reduced conservation.

Discussion

The repertoire of microbial rhodopsins continues to expand. With estimates of global microbial diversity ranging from 10^6 ⁹⁷ to 10^{12} species⁹⁸, and over two million bacterial and archaeal genomes now sequenced⁹⁹, the number of identified rhodopsin genes—already exceeding 10,000¹⁰⁰—will undoubtedly grow further. This expanding diversity, combined with ongoing optogenetics engineering efforts, has made a unified residue-numbering system increasingly necessary.

The limitations of existing numbering approaches become apparent when comparing across functional classes. Ion-pumping rhodopsins have historically been described using BR-based residue numbering, whereas channelrhodopsins are typically numbered according to ChR2- or C1C2-based systems. The three-residue TM3 motif introduced in 2013–2014 provides a useful signature for classifying ion-pump functions^{24, 25}; however, in channelrhodopsins, functional properties are determined primarily by other residues, including five glutamates in TM2, rather than by the TM3 motif. These differences underscore the need for a broadly applicable numbering system, which motivated the development of the MO numbering system.

The sequence similarity network provides insight into the evolutionary relationships among microbial rhodopsins (**Fig. 6a**, **Supplementary Fig. 6**). Rhodopsins from archaea and eukaryotes form a single major cluster (the archaeal/eukaryotic rhodopsin cluster), whereas bacterial rhodopsins occupy a distinct cluster (the bacterial rhodopsin cluster). This pattern indicates that eukaryotic microbial rhodopsins are more closely related to archaeal than to bacterial rhodopsins, consistent with previous phylogenetic and structural analyses^{101, 102, 103, 104}.

Rhodopsins within the same cluster but with different functions likely diverged from a common ancestor, whereas functionally similar rhodopsins in different clusters likely arose through convergent evolution. For example, both the archaeal/eukaryotic and bacterial clusters include chloride-pumping rhodopsins, yet their TM3 motifs differ ($T^{3.45}-S^{3.49}-A/D^{3.56}$ versus $N^{3.45}-T^{3.49}-Q^{3.56}$, respectively), suggesting that chloride-transport function evolved independently in these lineages. Within clusters, oligomeric assembly and sequence motifs are conserved despite functional diversification. This conservation likely explains why researchers have been able to interconvert functions through relatively few mutations, even among rhodopsins sharing only ~30% sequence identity^{51, 52, 79, 105, 106, 107, 108}.

The conservation of oligomeric assembly within clusters raises a question: what determines oligomeric state? All characterized microbial rhodopsins form specific oligomers, ranging from dimers to hexamers, and rhodopsins within the same clade consistently adopt the same oligomeric state^{8, 9, 13, 14, 88, 109, 110}. Yet residues on the outer, lipid-exposed surface are generally poorly conserved, even in helices that participate in oligomeric interfaces. This suggests that oligomeric state is not dictated by a fixed set of conserved interface residues. In some rhodopsins, such as *NpHR*, outer-facing residues do contribute to interprotomer contacts and stabilize oligomerization¹¹¹, but these residues are not conserved across clusters, implying that they fine-tune oligomer stability within a family rather than specifying oligomeric state. Instead, we propose that the type of oligomer formed is largely determined by helix-packing geometry, specifically how the transmembrane helices are tilted and arranged relative to one another. Helix tilt can be influenced by glycine and proline residues within helices, as well as loop architecture. In particular, the length and orientation of the extracellular loop 1 (ECL1) have been proposed to distinguish trimeric from pentameric or hexameric assemblies¹⁰¹. However, pentameric rhodopsins such as GR and TR form trimers in DDM micelles^{101, 109}, suggesting that the ECL1 may bias helix tilt toward particular arrangements without rigidly fixing protomer number. This observation implies that the membrane or detergent environment also contributes to stabilizing a given assembly.

Beyond evolutionary and oligomeric insights, our study also assessed the utility of structure prediction for rhodopsin analysis. Structure prediction methods performed well for rhodopsins. Predictions captured the

overall fold and the detailed architecture of the retinal-binding pocket, often correctly reproducing the retinal configuration and orientation, including whether the chromophore adopts an all-*trans* geometry or a 6-*s-cis* conformation consistent with the non-G rule^{28, 29, 35}. Our backbone RMSDs (0.61 Å for Set A and 0.80 Å for Set B) are comparable to those reported for Boltz-1 in the Runs N' Poses benchmark¹¹² (0.56 Å) and lower than AlphaFold2 predictions of GPCRs¹¹³ (1.12–1.55 Å). The accuracy of retinal placement is particularly notable: our mean retinal RMSDs of 0.41 Å (Set A) and 0.62 Å (Set B) are substantially lower than the 2.30 Å mean ligand RMSD that Boltz-1 achieved on diverse protein-ligand complexes in the Runs N' Poses benchmark. This likely reflects the strong structural constraints imposed by the retinal-binding pocket, including the covalent Schiff base linkage to the conserved lysine.

However, inaccuracies were frequently observed in fine structural details, including the placement of loop regions and side-chain rotamers. For example, in the predicted structure of *KnChR*, the retinal chromophore is not covalently linked to the Schiff base-forming lysine (K^{7,50}) (**Fig. 2b** and **Supplementary Fig. 10a**). In *GtACR1*, the β-ionone ring of the retinal adopts an incorrectly twisted conformation¹² (**Supplementary Fig. 10b**). In *Tara-RRB R2*, the cytoplasmic ends of TM0, TM1, TM2, and TM7 show marked deviations from the experimentally determined structure (**Fig. 2a** and **Supplementary Fig. 10c**). In *HcKCR2*, the architecture of the potassium selectivity filter is incorrectly predicted²⁸ (**Supplementary Fig. 10d**). These issues may become problematic when predicted structures are used e.g. as inputs for molecular dynamics simulations or for interpreting spectroscopic measurements, where accurate loop conformations and side-chain rotamers are critical. Thus, while current structure prediction methods are sufficiently accurate for comparative analyses and for capturing the overall architecture and retinal environment, they do not replace experimental structure determination when precise atomic detail or characterization of conformational dynamics is required.

The MO numbering system offers the microbial rhodopsin field a common vocabulary. As the repertoire of rhodopsins continues to expand through metagenomics and structural prediction, we anticipate that it will become a standard tool for comparing residues, interpreting functional data, and guiding optogenetic engineering.

Methods

Structural Dataset

We assembled a comprehensive structural dataset of 129 microbial rhodopsins spanning all major functional classes—proton pumps, chloride pumps, sodium pumps, cation channels, anion channels, sensory rhodopsins, and enzyme rhodopsins—with representatives from archaea, bacteria, and eukaryotes. Experimentally determined structures were obtained from the Protein Data Bank, supplemented by eight recently solved channelrhodopsin structures (Kishi et al., *in preparation*) ($n = 69$). The experimental structures were partitioned into two validation cohorts based on PDB deposition date relative to the Boltz-1¹¹⁴ training data cutoff (September 2021): a benchmark set (Set A, $n = 42$) potentially overlapping with training data, and a blind test set (Set B, $n = 27$) comprising structures deposited after the cutoff. For each experimentally characterized protein, we generated Boltz-1 predictions to enable systematic assessment of prediction accuracy. We additionally predicted structures for 60 functionally characterized rhodopsins lacking experimental structures. All analyses were performed on chain A, with retinal chromophores identified by residue name and validated by proximity ($\leq 6.0 \text{ \AA}$ to protein atoms).

Structure Prediction and Validation

For Boltz-1 predictions, input sequences were trimmed to transmembrane domain boundaries when exceeding 400 residues. Retinal was specified as all-trans-retinal without the aldehyde oxygen (SMILES code: CC=C(C)C=CC=C(C)C=CC1=C(CCCC1(C)C)C), allowing Boltz-1 to predict the covalent Schiff base linkage to the conserved lysine. Prediction accuracy was assessed by comparing each model to its experimental counterpart using the combinatorial extension alignment algorithm (CEalign)¹¹⁵ (window size = 8, maximum gap = 30) with a two-pass refinement procedure: an initial global alignment using all C α atoms, followed by refinement restricted to residue pairs with inter-C α distance $\leq 3 \text{ \AA}$ after the initial superposition to exclude flexible regions with large structural divergence. Accuracy was quantified using three metrics: backbone C α RMSD (overall fold accuracy), binding pocket RMSD (C α atoms within 6.0 \AA of retinal; local accuracy in the chromophore environment), and ligand RMSD (mean closest-atom distance between experimental and predicted retinal coordinates).

Structure Conservation Analysis

We first aligned the shared TM bundle of all members of our dataset. Pairwise structural similarities (calculated using CEalign) were computed as C α RMSD across TM 1–7 to generate a symmetric distance matrix. Our structure analysis leverages the conserved microbial rhodopsin fold by using the closest residue in each helix (TM1-TM6) and the Schiff-base lysine in TM7 as anchors. We then performed hierarchical clustering. For each GRN position, two distance profiles were calculated: (i) side-chain distances, from heavy atoms to the nearest retinal atom, and (ii) C α distances, from backbone C α to the nearest retinal atom. Equivalent residue pairs for all pairwise alignments were stored to create the blueprint of the MO numbering system.

Generic Residue-Numbering System

We developed the MO numbering system by assigning GRNs based on structural equivalence relative to a reference structure, 7BMH, selected based on its low mean RMSD to all other structures. In the MO numbering system, residues are numbered relative to the helix-specific anchors using the format <Helix>.<Position> (e.g., 3.49, 3.50, 3.51). After annotating the reference structure, we used the stored residue pair equivalences to assign GRNs to all other structures. The resulting preliminary MO numbering system was curated to correct alignment artifacts, including helix truncations and register shifts arising from the repetitive nature of α -helical structures. Then, the less conserved outer parts of the helices were annotated up until the end of the helices by

determining helical residues based on phi-psi angles of the backbone. The non-helical loop regions and tails received systematic identifiers.

Microbial Rhodopsin Sequence Search

To construct Hidden Markov Models (HMMs) for rhodopsin sequence exploration, known rhodopsins and their homologs were collected from NCBI, the UniProt Archive¹¹⁶, and the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP)¹¹⁷. One collection (subAEV: 3,173 sequences) was constructed mainly from archaeal-, eukaryotic-, and viral-derived rhodopsins¹¹⁸. Another collection (subB: 2,275 sequences) was mainly composed of bacterial-derived rhodopsins. A third collection (subH: 720 sequences) was composed of heliorrhodopsins and their homologs. Profile HMMs for the subAEV, subB, and subH collections were constructed from sequence alignments computed using MAFFT (v7.453)¹¹⁹ with the "--genafpair" and "--maxiterate 1000" options. Using these HMMs, we searched the UniParc database (as of November 2020), the marine eukaryotic reference catalog¹²⁰, the 1000 Plant Transcriptomes Initiative dataset¹²¹, MMETSP, and 10,032 metagenomic assemblies derived from various environments (Nishimura et al., in preparation). Rhodopsin homologs were identified using hmmsearch (HMMER v3.3)¹²² with an e-value threshold of $<1e^{-5}$ and a length threshold of ≥ 174 amino acids, corresponding to the HMM length. We clustered 309,570 protein sequences using CD-HIT¹²³ (v.4.8.1) at 90% global identity (-c 0.90, -G 1) without additional coverage constraints (no -aS/-aL), resulting in 41,136 non-redundant sequences.

Network Analysis Visualization

We collected reference rhodopsins ($n = 147$) that were functionally and/or structurally characterized, along with the 41,136 non-redundant, uncharacterized rhodopsin sequences. Because enzymatic rhodopsins contain additional domains, the rhodopsin regions of both reference and uncharacterized sequences were extracted using aligned regions based on hmmsearch (HMMER v3.3)¹²² with the three HMMs (subAEV, subB, and subH). To further remove redundancy within the uncharacterized dataset, protein clustering was performed using MMseqs2 (ver. 15-6f452)¹²⁴ at 60% identity and 50% coverage, with options "--min-seq-id 0.6 -c 0.5 --cluster-mode 2 --cov-mode 0". Cluster representatives ($n = 5,168$) were combined with the reference rhodopsins (total $n = 5,315$).

Normalized pairwise similarity within the resultant sequences was computed using CPAP (<https://github.com/yosuken/cpap>), which performed all-against-all BLASTp with options "-evalue 1e-2 -dbsize 100000000" and quantified normalized similarity scores between all possible protein pairs based on high-scoring segment pairs (HSPs) that have $\geq 20\%$ identity and are ≥ 20 amino acids in length, as follows:

Let $n_{AB,i}$ be the length-normalized bit score (i.e., bit score per position) of the HSP covering position i of protein A , where BLASTp is performed using A as the query and B as the subject. If multiple HSPs cover position i , the maximum value is used.

$$M_{AB,i} = \max(n_{AB,i})$$

Let L_A be the length of protein A . Let N_{AB} ($0 \leq N_{AB} \leq 1$) be the normalized similarity score between proteins A and B .

$$U_{AB} = \frac{\sum_i M_{AB,i}}{\sum_i M_{AA,i}}, 0 \leq i \leq L_A$$

$$N_{AB} = \frac{1}{2}(U_{AB} + U_{BA})$$

The resulting network was visualized using Gephi (ver. 0.10.1). Edges were drawn between protein pairs with normalized similarity scores ≥ 0.2 .

Sequence Logo Generation

Sequence logos were generated using WebLogo3¹²⁵. Sequences were aligned according to their MO numbering positions, and logos were computed for each transmembrane helix. For cluster-specific analyses, sequences were grouped based on their assignment in the sequence similarity network, and separate logos were generated for each TM regions (TM1-7). Amino acid frequencies are displayed as information content in bits.

Statistical Analysis and Data Availability

Correlations between structural and sequence metrics were computed using Pearson's correlation with two-tailed significance testing. All analyses (structural alignment, superposition) were performed in Python 3.10 using Biopython. Dataset management and CIF parsing were performed using ProtOS (Hidber et al., manuscript in preparation, <https://github.com/flurinh/protos>). Structure coordinates, alignments, RMSD matrices, GRN tables and all analysis code are provided at <https://github.com/flurinh/mogrn>. The structure analysis data is available at <https://doi.org/10.5281/zenodo.18147121>.

Acknowledgments

We would like to thank Y. Oheda and A. Kojima (UTokyo) for technical and K. Hasegawa and A. Ohira (UTokyo) for administrative support. We acknowledge ChatGPT, a multimodal large language model created by OpenAI, for providing guidance to improve the readability of this manuscript. After using this tool, we reviewed and edited the content as needed and took full responsibility for the content of the publication. This work was supported by the Nakajima Foundation to H.E.K., Takeda Science Foundation to H.E.K., Japan Agency for Medical Research and Development (AMED) (grant no. 24bm1123057h0001 to H.E.K.), Japan Society for the Promotion of Science (JSPS) KAKENHI (grant no. 24KJ0788 to S.T., 22K15089 to Y.N., 22H00557 to S.Y. and 22K19265, 22H00400, and 25H01338 to H.E.K.), JST FOREST (grant no. JPMJFR204S to H.E.K.), JST CREST (grant no. JPMJCR21P3 and JPMJCR23B1 to H.E.K.), European Union's Horizon 2020 research and innovation programme (grant agreement no. 951644, supporting F.S.H.), and Swiss National Science Foundation (grant no. 192780 to X.D.). The authors dedicate this paper to G.F.X. Schertler, H. Kandori, and P. Hegemann: many years of insightful and exciting scientific discussions partly inspired this work.

Author Contributions

F.S.H., S.T., K.E.K., X.D., and H.E.K. discussed and defined the MO numbering system. Under the supervision of X.D., F.S.H. performed the structure comparison analysis and built the MO numbering system framework. Under the supervision of H.E.K., S.T. analyzed sequence alignments and structures. Y.N. and S.Y. mined in-house metagenome and transcriptome databases to identify rhodopsin-like genes. Under the supervision of S.Y. and H.E.K., S.T. and Y.N. constructed and analyzed the sequence similarity network. F.S.H., S.T., X.D., and H.E.K. wrote the manuscript and prepared the figures with input from all authors. X.D. and H.E.K. supervised the project.

Competing Interests

The authors declare no competing interests.

References

1. Oesterhelt D, Stoeckenius W. Rhodopsin-like Protein from the Purple Membrane of *Halobacterium halobium*. *Nat Biol* **233**, 149-152 (1971).
2. Ernst OP, Lodowski DT, Elstner M, Hegemann P, Brown LS, Kandori H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem Rev* **114**, 126-163 (2014).
3. Kato Y, Hasunuma T. Carotenoids: Biosynthetic and Biofunctional Approaches. *Adv Exp Med Biol* **1261**, 121-135 (2021).
4. Rozenberg A, Inoue K, Kandori H, Béjà O. Microbial Rhodopsins: The Last Two Decades. *Annu Rev Microbiol* **75**, 1-21 (2021).
5. Sudo Y, et al. A Microbial Rhodopsin with a Unique Retinal Composition Shows Both Sensory Rhodopsin II and Bacteriorhodopsin-like Properties*. *J Biol Chem* **286**, 5967-5976 (2011).
6. Rozenberg A, et al. Rhodopsin-bestrophin fusion proteins from unicellular algae form gigantic pentameric ion channels. *Nat Struct Mol Biol* **29**, 592-603 (2022).
7. Sugiura M, et al. Unusual Photoisomerization Pathway in a Near-Infrared Light Absorbing Enzymerhodopsin. *J Phys Chem Lett* **13**, 9539-9543 (2022).
8. Shibata M, et al. Oligomeric states of microbial rhodopsins determined by high-speed atomic force microscopy and circular dichroic spectroscopy. *Sci Rep* **8**, 8262 (2018).
9. Mannen K, et al. Multiple Roles of a Conserved Glutamate Residue for Unique Biophysical Properties in a New Group of Microbial Rhodopsins Homologous to TAT Rhodopsin. *J Mol Biol* **436**, 168331 (2024).
10. Ikuta T, et al. Structural insights into the mechanism of rhodopsin phosphodiesterase. *Nat Commun* **11**, 5605 (2020).
11. Kato HE, et al. Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **482**, 369-374 (2012).
12. Kim YS, et al. Crystal structure of the natural anion-conducting channelrhodopsin GtACR1. *Nature* **561**, 343-348 (2018).
13. Shihoya W, et al. Crystal structure of heliorhodopsin. *Nature* **574**, 132-136 (2019).
14. Kishi KE, et al. Structural basis for channel conduction in the pump-like channelrhodopsin ChRmine. *Cell* **185**, 672-689.e623 (2022).
15. Pushkarev A, et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**, 595-599 (2018).
16. Mukherjee S, Hegemann P, Broser M. Enzymerhodopsins: novel photoregulated catalysts for optogenetics. *Curr Opin Struct Biol* **57**, 118-126 (2019).
17. Grip WJD, Ganapathy S. Rhodopsins: An Excitingly Versatile Protein Species for Research, Development and Creative Engineering. *Front Chem* **10**, 879609 (2022).
18. Olivella M, Gonzalez A, Pardo L, Deupi X. Relation between sequence and structure in membrane proteins. *Bioinformatics* **29**, 1589-1592 (2013).
19. Nagel G, et al. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc Natl Acad Sci* **100**, 13940-13945 (2003).
20. Ballesteros JA, Weinstein H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci* **25**, 366-428 (1995).

21. Isberg V, et al. Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharmacol Sci* **36**, 22-31 (2015).
22. Bertalan Ev, et al. Hydrogen-Bonding and Hydrophobic Interaction Networks as Structural Determinants of Microbial Rhodopsin Function. *J Phys Chem B* **128**, 7407-7426 (2024).
23. Govorunova EG, et al. Cation and Anion Channelrhodopsins: Sequence Motifs and Taxonomic Distribution. *mBio* **12**, 10.1128/mbio.01656-01621 (2021).
24. Inoue K, et al. A light-driven sodium ion pump in marine bacteria. *Nat Commun* **4**, 1678 (2013).
25. Yoshizawa S, et al. Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc Natl Acad Sci* **111**, 6732-6737 (2014).
26. Sephus CD, Fer E, Garcia AK, Adam ZR, Schwieterman EW, Kacar B. Earliest Photic Zone Niches Probed by Ancestral Microbial Rhodopsins. *Mol Biol Evol* **39**, msac100 (2022).
27. Kato HE, et al. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nat Commun* **6**, 7177 (2015).
28. Tajima S, et al. Structural basis for ion selectivity in potassium-selective channelrhodopsins. *Cell* **186**, 4325-4344.e4326 (2023).
29. Wang YZ, et al. Cryo-EM structure of a blue-shifted channelrhodopsin from Klebsormidium nitens. *Nat Commun* **16**, 5297 (2025).
30. Man D, et al. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J* **22**, 1725-1731 (2003).
31. Inoue K, et al. Red-shifting mutation of light-driven sodium-pump rhodopsin. *Nat Commun* **10**, 1993 (2019).
32. Sugiura M, Tsunoda SP, Hibi M, Kandori H. Molecular Properties of New Enzyme Rhodopsins with Phosphodiesterase Activity. *ACS Omega* **5**, 10602-10609 (2020).
33. Wang WW, Sineshchekov OA, Spudich EN, Spudich JL. Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J Biol Chem* **278**, 33985-33991 (2003).
34. Luecke H, et al. Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc Natl Acad Sci* **105**, 16561-16565 (2008).
35. Kato HE, et al. Structural basis for Na⁺ transport mechanism by a light-driven Na⁺ pump. *Nature* **521**, 48-53 (2015).
36. Tzilil G, et al. Structural insights into light harvesting by antenna-containing rhodopsins in marine Asgard archaea. *Nat Microbiol* **10**, 1484-1500 (2025).
37. Imasheva ES, Balashov SP, Choi AR, Jung K-H, Lanyi JK. Reconstitution of *Gloeobacter violaceus* Rhodopsin with a Light-Harvesting Carotenoid Antenna. *Biochemistry* **48**, 10948-10955 (2009).
38. Chuon K, et al. Carotenoid binding in *Gloeobacteria* rhodopsin provides insights into divergent evolution of xanthorhodopsin types. *Commun Biol* **5**, 512 (2022).
39. Shevchenko V, et al. Inward H⁺ pump xenorhodopsin: Mechanism and alternative optogenetic approach. *Sci Adv* **3**, e1603187 (2017).
40. Higuchi A, et al. Crystal structure of schizorhodopsin reveals mechanism of inward proton pumping. *Proc Natl Acad Sci* **118**, e2016328118 (2021).
41. Kolbe M, Besir Hs, Essen L-O, Oesterhelt D. Structure of the Light-Driven Chloride Pump Halorhodopsin at 1.8 Å Resolution. *Science* **288**, 1390-1396 (2000).
42. Kim K, et al. Crystal structure and functional characterization of a light-driven chloride pump having an NTQ motif. *Nat Commun* **7**, 12677 (2016).

43. Yun J-H, et al. Structure-Based Functional Modification Study of a Cyanobacterial Chloride Pump for Transporting Multiple Anions. *J Mol Biol* **432**, 5273-5286 (2020).
44. Kato HE, Inoue K, Kandori H, Nureki O. The light-driven sodium ion pump: A new player in rhodopsin research. *Bioessays* **38**, 1274-1282 (2016).
45. Skopintsev P, et al. Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature* **583**, 314-318 (2020).
46. Hososhima S, et al. Proton-transporting heliorhodopsins from marine giant viruses. *eLife* **11**, e78416 (2022).
47. Nagel G, Brauner M, Liewald JF, Adeishvili N, Bamberg E, Gottschalk A. Light Activation of Channelrhodopsin-2 in Excitable Cells of *Caenorhabditis elegans* Triggers Rapid Behavioral Responses. *Curr Biol* **15**, 2279-2284 (2005).
48. Gunaydin LA, Yizhar O, Berndt A, Sohal VS, Deisseroth K, Hegemann P. Ultrafast optogenetic control. *Nat Neurosci* **13**, 387-392 (2010).
49. Sugiyama Y, et al. Photocurrent attenuation by a single polar-to-nonpolar point mutation of channelrhodopsin-2. *Photochem Photobiol Sci* **8**, 328-336 (2009).
50. Eisenhauer K, et al. In Channelrhodopsin-2 Glu-90 Is Crucial for Ion Selectivity and Is Deprotonated during the Photocycle*. *J Biol Chem* **287**, 6904-6911 (2012).
51. Berndt A, Lee SY, Ramakrishnan C, Deisseroth K. Structure-Guided Transformation of Channelrhodopsin into a Light-Activated Chloride Channel. *Science* **344**, 420-424 (2014).
52. Wietek J, et al. Conversion of Channelrhodopsin into a Light-Gated Chloride Channel. *Science* **344**, 409-412 (2014).
53. Berndt A, Yizhar O, Gunaydin LA, Hegemann P, Deisseroth K. Bi-stable neural state switches. *Nat Neurosci* **12**, 229-234 (2009).
54. Nack M, et al. The DC gate in Channelrhodopsin-2: crucial hydrogen bonding interaction between C128 and D156. *Photochem Photobiol Sci* **9**, 194-198 (2010).
55. Yizhar O, Fieno Lief E, Davidson Thomas J, Mogri M, Deisseroth K. Optogenetics in Neural Systems. *Neuron* **71**, 9-34 (2011).
56. Berndt A, et al. High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels. *Proc Natl Acad Sci* **108**, 7595-7600 (2011).
57. Kleinlogel S, et al. Ultra light-sensitive and fast neuronal activation with the Ca²⁺-permeable channelrhodopsin CatCh. *Nat Neurosci* **14**, 513-518 (2011).
58. Lin JY, Knutsen PM, Muller A, Kleinfeld D, Tsien RY. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nat Neurosci* **16**, 1499-1508 (2013).
59. Klapoetke NC, et al. Independent optical excitation of distinct neural populations. *Nat Methods* **11**, 338-346 (2014).
60. Dawydow A, et al. Channelrhodopsin-2-XXL, a powerful optogenetic tool for low-light applications. *Proc Natl Acad Sci U S A* **111**, 13972-13977 (2014).
61. Rajasethupathy P, et al. Projections from neocortex mediate top-down control of memory retrieval. *Nature* **526**, 653-659 (2015).
62. Mager T, et al. High frequency neural spiking and auditory signaling by ultrafast red-shifted optogenetics. *Nat Commun* **9**, 1750 (2018).
63. Oda K, et al. Crystal structure of the red light-activated channelrhodopsin Chrimson. *Nat Commun* **9**, 3949 (2018).
64. Sridharan S, et al. High-performance microbial opsins for spatially and temporally precise perturbations of large neuronal networks. *Neuron* **110**, 1139-1155.e1136 (2022).

65. Scholz N, et al. Mechano-dependent signaling by Latrophilin/CIRL quenches cAMP in proprioceptive neurons. *eLife* **6**, (2017).
66. Kato HE, et al. Structural mechanisms of selectivity and gating in anion channelrhodopsins. *Nature* **561**, 349-354 (2018).
67. Fernandez Lahore RG, et al. Calcium-permeable channelrhodopsins for the photocontrol of calcium signalling. *Nat Commun* **13**, 7844 (2022).
68. Mardinly AR, et al. Precise multimodal optical control of neural ensemble activity. *Nat Neurosci* **21**, 881-893 (2018).
69. Vierock J, et al. WiChR, a highly potassium-selective channelrhodopsin for low-light one- and two-photon inhibition of excitable cells. *Sci Adv* **8**, eadd7729 (2022).
70. Morizumi T, et al. Structures of channelrhodopsin paralogs in peptidiscs explain their contrasting K(+) and Na(+) selectivities. *Nat Commun* **14**, 4365 (2023).
71. Tucker K, Sridharan S, Adesnik H, Brohawn SG. Cryo-EM structures of the channelrhodopsin ChRmine in lipid nanodiscs. *Nat Commun* **13**, 4842 (2022).
72. Duan X, et al. Stabilized Ion Selectivity Corrects Activation Drift in Kalium Channelrhodopsins. *Adv Sci (Weinh)*, e09180 (2025).
73. Duan X, et al. Suppression of epileptic seizures by transcranial activation of K(+) -selective channelrhodopsin. *Nat Commun* **16**, 559 (2025).
74. Alekseev A, et al. Efficient and sustained optogenetic control of sensory and cardiac systems. *Nat Biomed Eng*, (2025).
75. Tose AJ, Nava AA, McGrath SN, Mardinly AR, Naka A. WACRs are excitatory opsins sensitive to indoor lighting. *bioRxiv*, 2025.2009.2012.675947 (2025).
76. Gordeliy VI, et al. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **419**, 484-487 (2002).
77. Moukhambetianov R, et al. Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. *Nature* **440**, 115-119 (2006).
78. Ishchenko A, et al. New Insights on Signal Propagation by Sensory Rhodopsin II/Transducer Complex. *Sci Rep* **7**, 41811 (2017).
79. Sudo Y, Furutani Y, Kandori H, Spudich JL. Functional Importance of the Interhelical Hydrogen Bond between Thr204 and Tyr174 of Sensory Rhodopsin II and Its Alteration during the Signaling Process*. *J Biol Chem* **281**, 34239-34245 (2006).
80. Matsunami-Nakamura R, et al. Key determinants for signaling in the sensory rhodopsin II/transducer complex are different between *Halobacterium salinarum* and *Natronomonas pharaonis*. *FEBS Lett* **597**, 2334-2344 (2023).
81. Broser M, et al. NeoR, a near-infrared absorbing rhodopsin. *Nat Commun* **11**, 5682 (2020).
82. Marshel JH, et al. Cortical layer-specific critical dynamics triggering perception. *Science* **365**, (2019).
83. Philosof A, Beja O. Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environ Microbiol Rep* **5**, 475-482 (2013).
84. Needham DM, et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc Natl Acad Sci U S A* **116**, 20574-20583 (2019).
85. Kateriya S, Nagel G, Bamberg E, Hegemann P. "Vision" in single-celled algae. *News Physiol Sci* **19**, 133-137 (2004).

86. Luck M, et al. A photochromic histidine kinase rhodopsin (HKR1) that is bimodally switched by ultraviolet and blue light. *J Biol Chem* **287**, 40083-40090 (2012).
87. Pushkarev A, et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**, 595-599 (2018).
88. Inoue K, et al. Schizorhodopsins: A family of rhodopsins from Asgard archaea that function as light-driven inward H(+) pumps. *Sci Adv* **6**, eaaz2441 (2020).
89. Govorunova EG, Sineshchekov OA, Janz R, Liu X, Spudich JL. NEUROSCIENCE. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **349**, 647-650 (2015).
90. Kishi KE, Kato HE. Pump-like channelrhodopsins: Not just bridging the gap between ion pumps and ion channels. *Curr Opin Struct Biol* **79**, 102562 (2023).
91. Luck M, et al. Photoreactions of the Histidine Kinase Rhodopsin Ot-HKR from the Marine Picoalga Ostreococcus tauri. *Biochemistry* **58**, 1878-1891 (2019).
92. Balashov SP, et al. Effect of the arginine-82 to alanine mutation in bacteriorhodopsin on dark adaptation, proton release, and the photochemical cycle. *Biochemistry* **32**, 10331-10343 (1993).
93. Ikeura Y, Shimono K, Iwamoto M, Sudo Y, Kamo N. Arg-72 of pharaonis Phoborhodopsin (Sensory Rhodopsin II) is Important for the Maintenance of the Protein Structure in the Solubilized State. *Photochem Photobiol* **77**, 96-100 (2003).
94. Vogt A, Guo Y, Tsunoda SP, Kateriya S, Elstner M, Hegemann P. Conversion of a light-driven proton pump into a light-gated ion channel. *Sci Rep* **5**, 16450 (2015).
95. Buhl E, et al. Assessing the Role of R120 in the Gating of CrChR2 by Time-Resolved Spectroscopy from Femtoseconds to Seconds. *J Am Chem Soc* **145**, 21832-21840 (2023).
96. Kubo M, et al. Role of Arg123 in light-driven anion pump mechanisms of pharaonis halorhodopsin. *Photochem Photobiol* **85**, 547-555 (2009).
97. Louca S, Mazel F, Doebeli M, Parfrey LW. A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS Biol* **17**, e3000106 (2019).
98. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci* **113**, 5970-5975 (2016).
99. Shen W, Lees JA, Iqbal Z. Efficient sequence alignment against millions of prokaryotic genomes with LexicMap. *Nat Biotechnol*, 1-8 (2025).
100. Nagata T, Inoue K. Rhodopsins at a glance. *J Cell Sci* **134**, (2021).
101. Morizumi T, et al. X-ray Crystallographic Structure and Oligomerization of Gloeobacter Rhodopsin. *Sci Rep* **9**, 11283 (2019).
102. Brown LS. Fungal rhodopsins and opsin-related proteins: eukaryotic homologues of bacteriorhodopsin with unknown functions. *Photochem Photobiol Sci* **3**, 555-565 (2004).
103. Sharma AK, Spudich JL, Doolittle WF. Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* **14**, 463-469 (2006).
104. Zabelskii D, et al. Structure-based insights into evolution of rhodopsins. *Commun Biol* **4**, 821 (2021).
105. Sasaki J, et al. Conversion of bacteriorhodopsin into a chloride ion pump. *Science* **269**, 73-75 (1995).
106. Hasemi T, Kikukawa T, Kamo N, Demura M. Characterization of a Cyanobacterial Chloride-pumping Rhodopsin and Its Conversion into a Proton Pump. *J Biol Chem* **291**, 355-362 (2016).

107. Inoue K, Nomura Y, Kandori H. Asymmetric Functional Conversion of Eubacterial Light-driven Ion Pumps. *J Biol Chem* **291**, 9883-9893 (2016).
108. Takeno Y, et al. Structural basis for color tuning and passive ion conductance in red-shifted pump-fold channelrhodopsin ChR024. *bioRxiv*, 2025.2009.2016.675330 (2025).
109. Shionoya T, et al. High Thermal Stability of Oligomeric Assemblies of Thermophilic Rhodopsin in a Lipid Environment. *J Phys Chem B* **122**, 6945-6953 (2018).
110. Ikuta T, et al. Structural insights into the mechanism of rhodopsin phosphodiesterase. *Nat Commun* **11**, 5605 (2020).
111. Sasaki T, et al. Halorhodopsin from Natronomonas pharaonis Forms a Trimer Even in the Presence of a Detergent, Dodecyl- β -d-maltoside. *Photochem Photobiol* **85**, 130-136 (2009).
112. Škrinjar P, Eberhardt J, Durairaj J, Schwede T. Have protein-ligand co-folding methods moved beyond memorisation? *bioRxiv*, 2025.2002.2003.636309 (2025).
113. Heo L, Feig M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins* **90**, 1873-1885 (2022).
114. Wohlwend J, et al. Boltz-1 Democratizing Biomolecular Interaction Modeling. *bioRxiv*, (2025).
115. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* **11**, 739-747 (1998).
116. UniProt C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531 (2023).
117. Keeling PJ, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol* **12**, e1001889 (2014).
118. Hasegawa-Takano M, et al. Cyanorhodopsin-II represents a yellow-absorbing proton-pumping rhodopsin clade within cyanobacteria. *ISME J* **18**, (2024).
119. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772-780 (2013).
120. Steinegger M, Mirdita M, Soding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nat Methods* **16**, 603-606 (2019).
121. Carpenter EJ, et al. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience* **8**, (2019).
122. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res* **46**, W200-W204 (2018).
123. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).
124. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).
125. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: A Sequence Logo Generator. *Genome Res* **14**, 1188-1190 (2004).

Figure 1

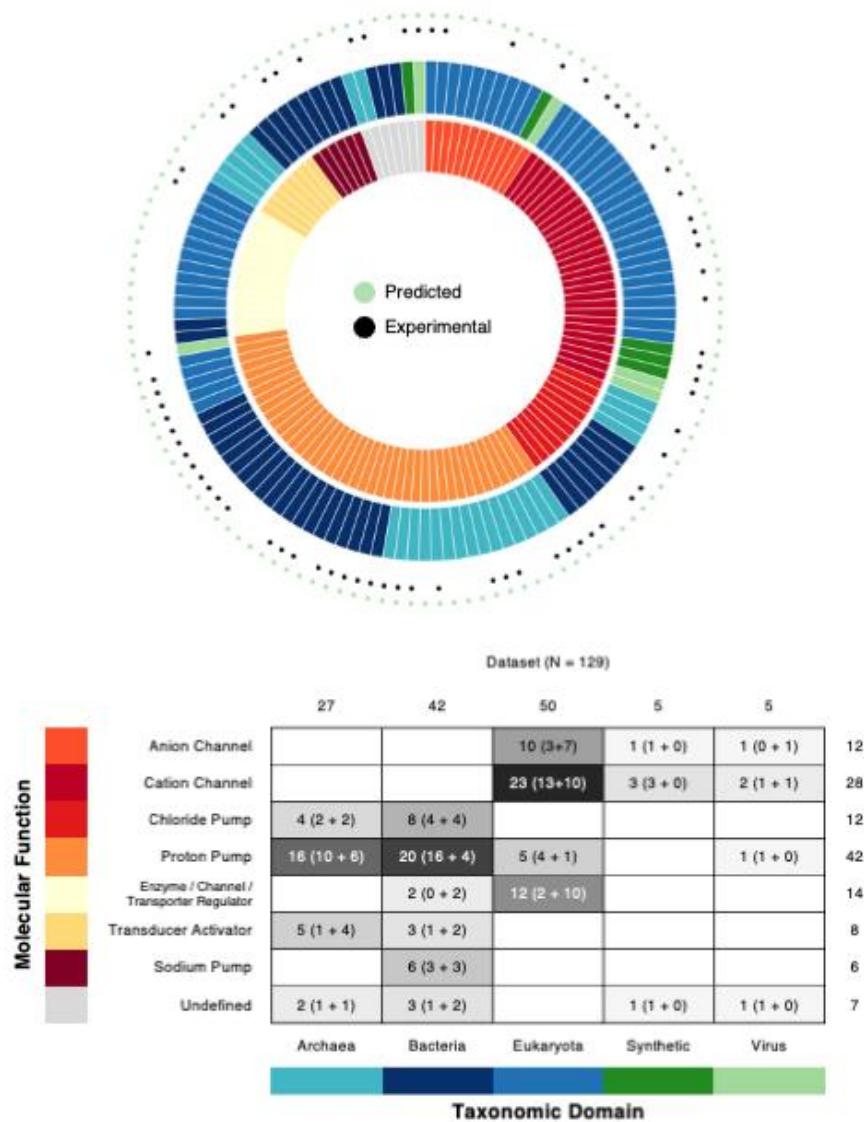


Fig. 1 | Dataset of microbial rhodopsin structures. The colors in the inner and outer rings indicate molecular function and taxonomic domain respectively (as shown in the table). Experimental structures and computational models are marked with black and cyan circles respectively. The table shows the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

Figure 2

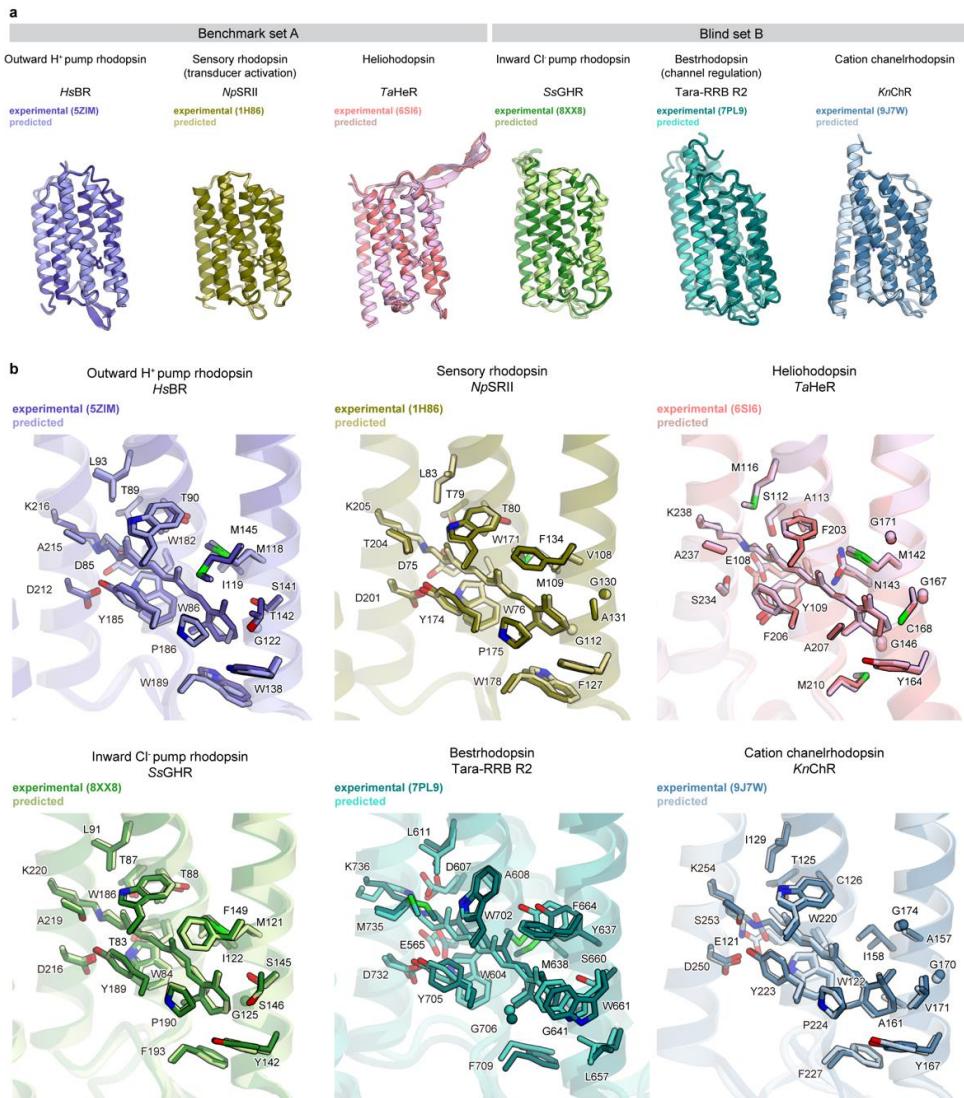


Fig. 2 | Accuracy of the predicted models. **a** Structural superposition between experimental (solid colors) and predicted (light colors) structures of representative rhodopsins: from benchmark set A, the outward proton pump rhodopsin *HsBR* (PDB ID: 5ZIM), the transducer-activating rhodopsin *NpSRII* (PDB ID: 1H86), and the functionally uncharacterized heliorhodopsin *TaHeR* (PDB ID: 6SI6); and from blind set B, the inward chloride pump rhodopsin *SsGHR* (PDB ID: 8XX8), the channel-regulating rhodopsin *Tara-RRB* (PDB ID: 7PL9), and the cation-conducting channelrhodopsin *KnChR* (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

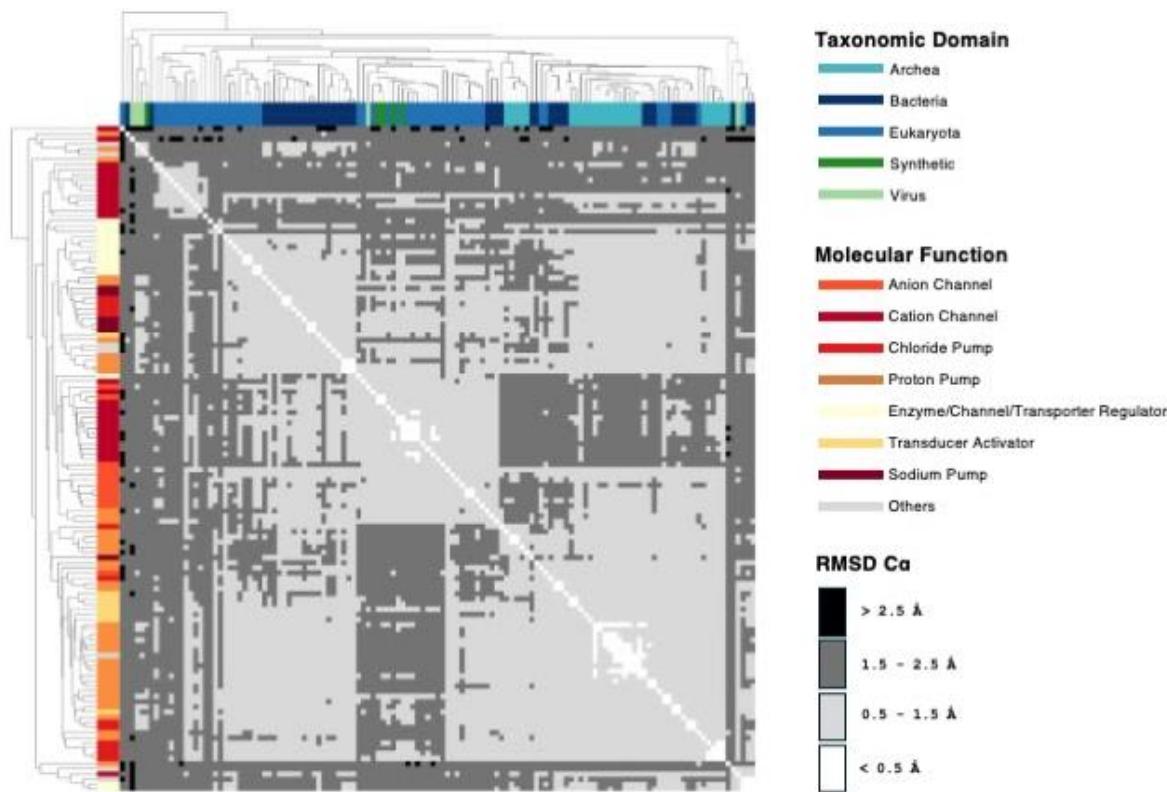


Fig. 3 | Structural landscape of microbial rhodopsins. The heatmap displays pairwise RMSD C_α values for all 129 dataset members. Rhodopsins are sorted according to hierarchical clustering based on structural similarity. Annotation bars display taxonomic domain (x-axis; green/blue) and molecular function (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

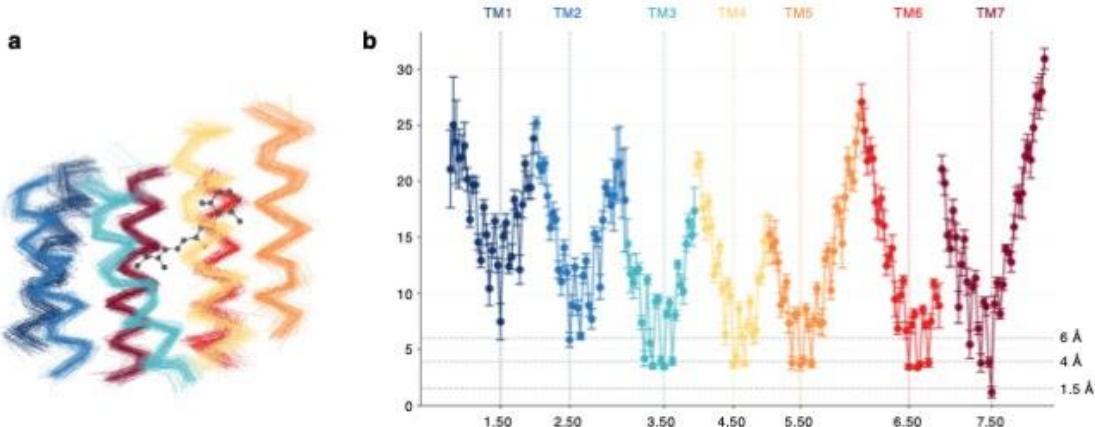


Fig. 4 | Location of the MO numbering anchor positions. **a** Structural alignment of the 129 dataset members allows selecting the consensus residues closest to retinal in each helix. **b** Average minimum distance from each structurally equivalent residue position to the retinal chromophore calculated across the dataset (closest side-chain atom, excluding hydrogen; an analogous graph for $\text{C}\alpha$ 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.

6

7

Figure 5

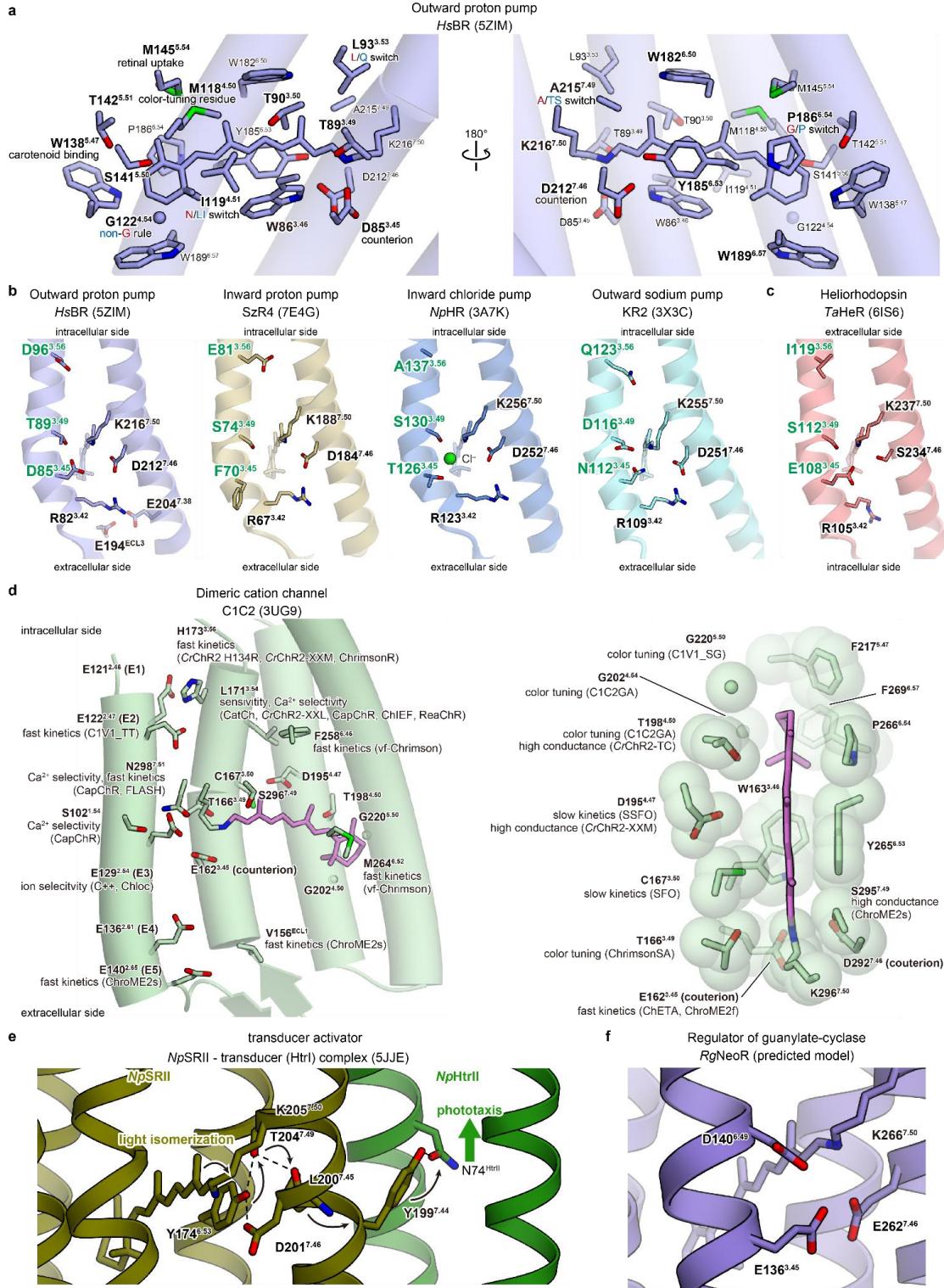


Fig. 5 | Demonstration of the MO numbering system. **a** Retinal-binding pocket of the outward proton pump *HsBR* (PDB ID: 5ZIM), shown from two different orientations. Residues constituting the retinal-binding pocket are depicted as sticks. Bold labels indicate residues positioned on the front side 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 *HsBR* (PDB ID: 5ZIM), the inward proton pump *SzR4* (PDB ID: 7E4G), the inward chloride pump *NpHR* (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 *TaHeR* (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 *NpSRII* (khaki) and its cognate transducer *NpHtrII* (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 *RgNeoR*. The structure shown is a model predicted using Boltz1.

8

9

Figure 6

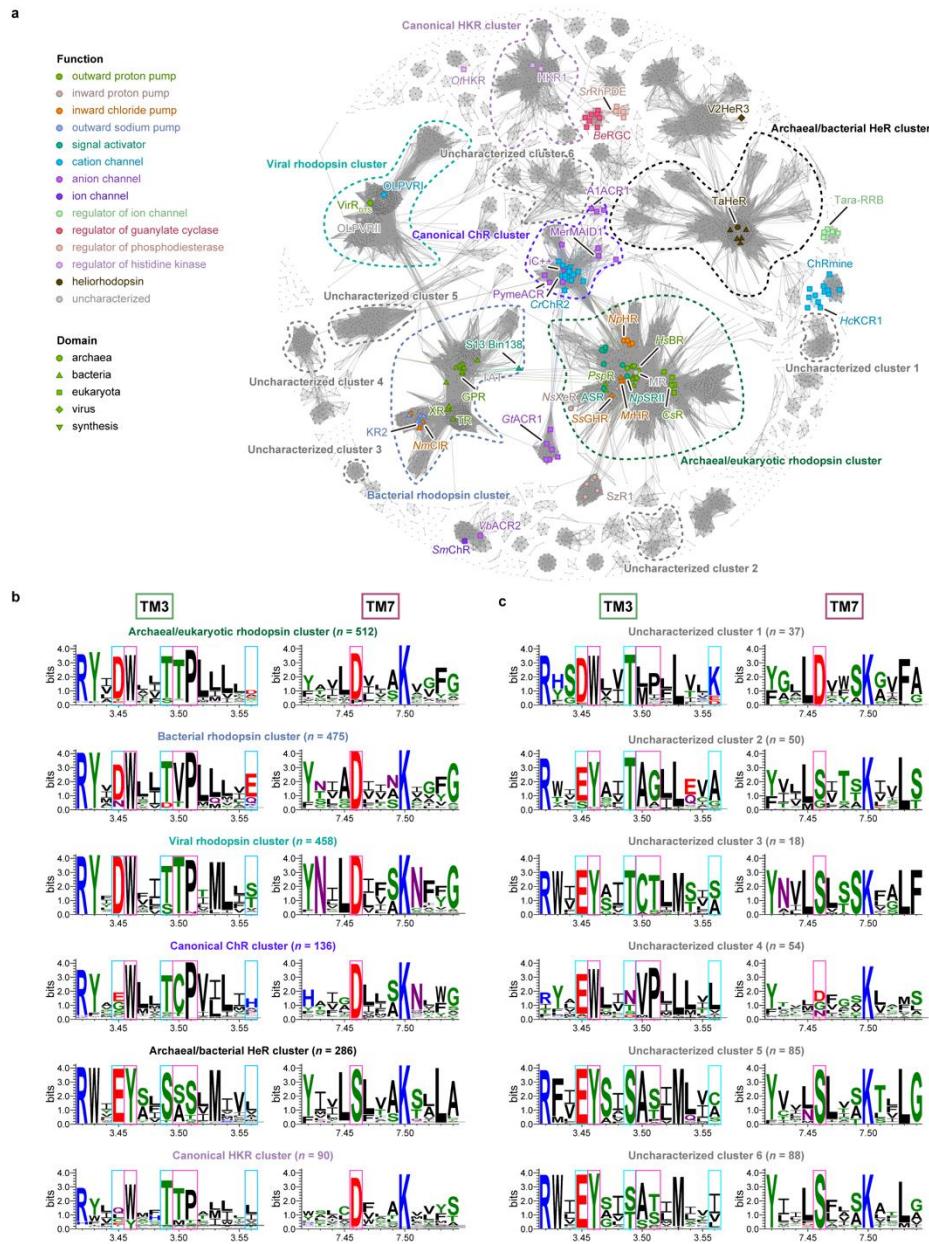


Fig. 6 | Application of the MO numbering system to a large dataset of microbial rhodopsins. **a** Sequence similarity network composed of approximately 40,000 non-redundant microbial rhodopsin sequences, including a large collection of uncharacterized rhodopsin-like sequences. Six major clusters containing 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.

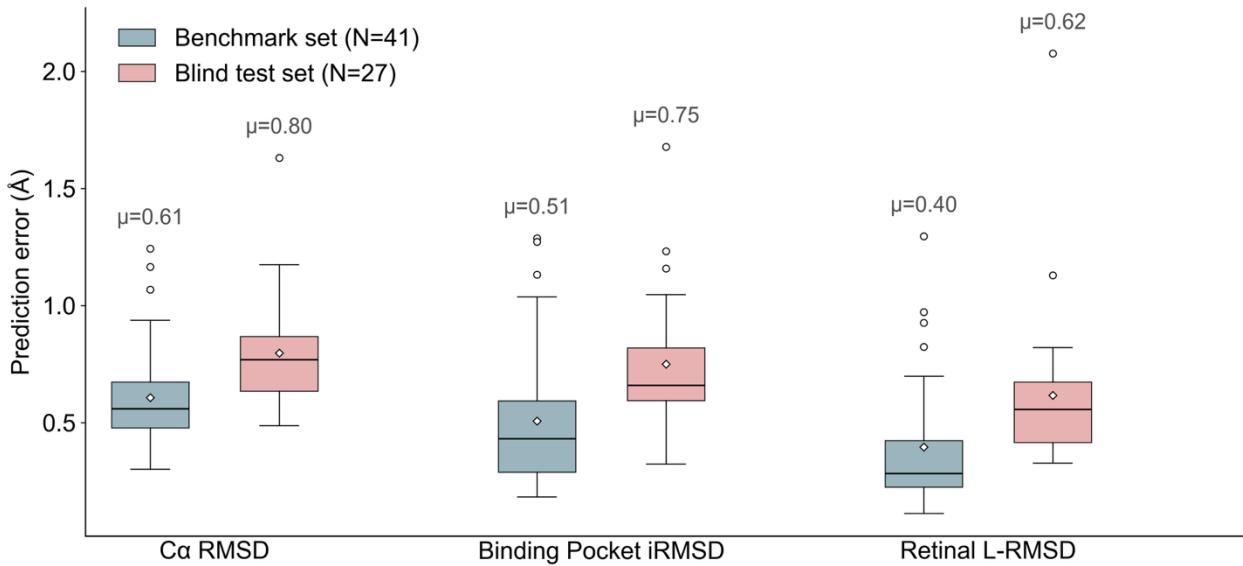
11

12

13

14

Supplementary Figure 1

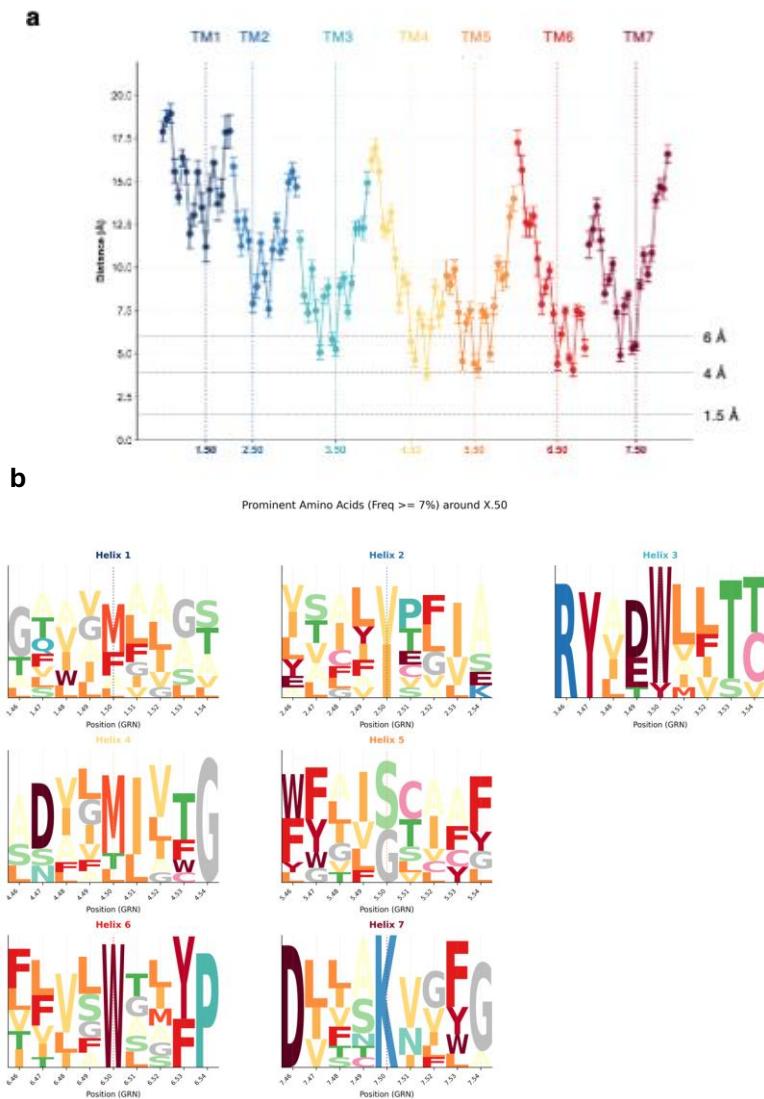


Supplementary Fig. 1 | Accuracy of the predicted models. Prediction errors for the two validation sets: benchmark set A (blue; n=42) and blind set B (red; n=27 for backbone, n=26 for the binding pocket). The prediction errors are presented separately for the entire structure (global C α RMSD), for the 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); NsXeR (PDB ID: 6EYU); Chrimson (PDB ID: 5ZIH), and XR (PDB ID: 3DDL). Four models in blind set B also exceeded this threshold: ErNaR (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).

15

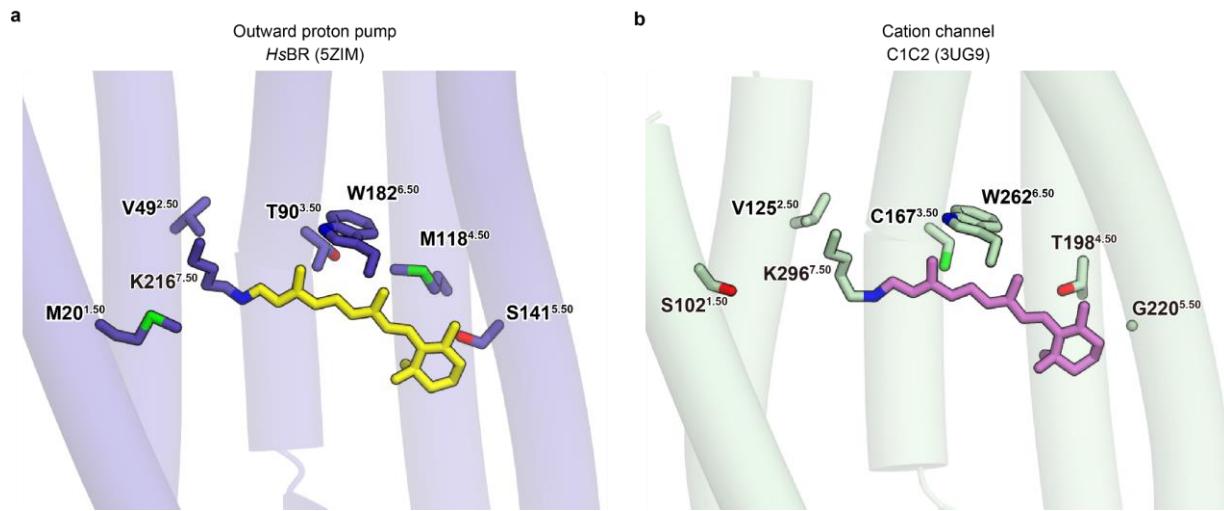
16

Supplementary Figure 2



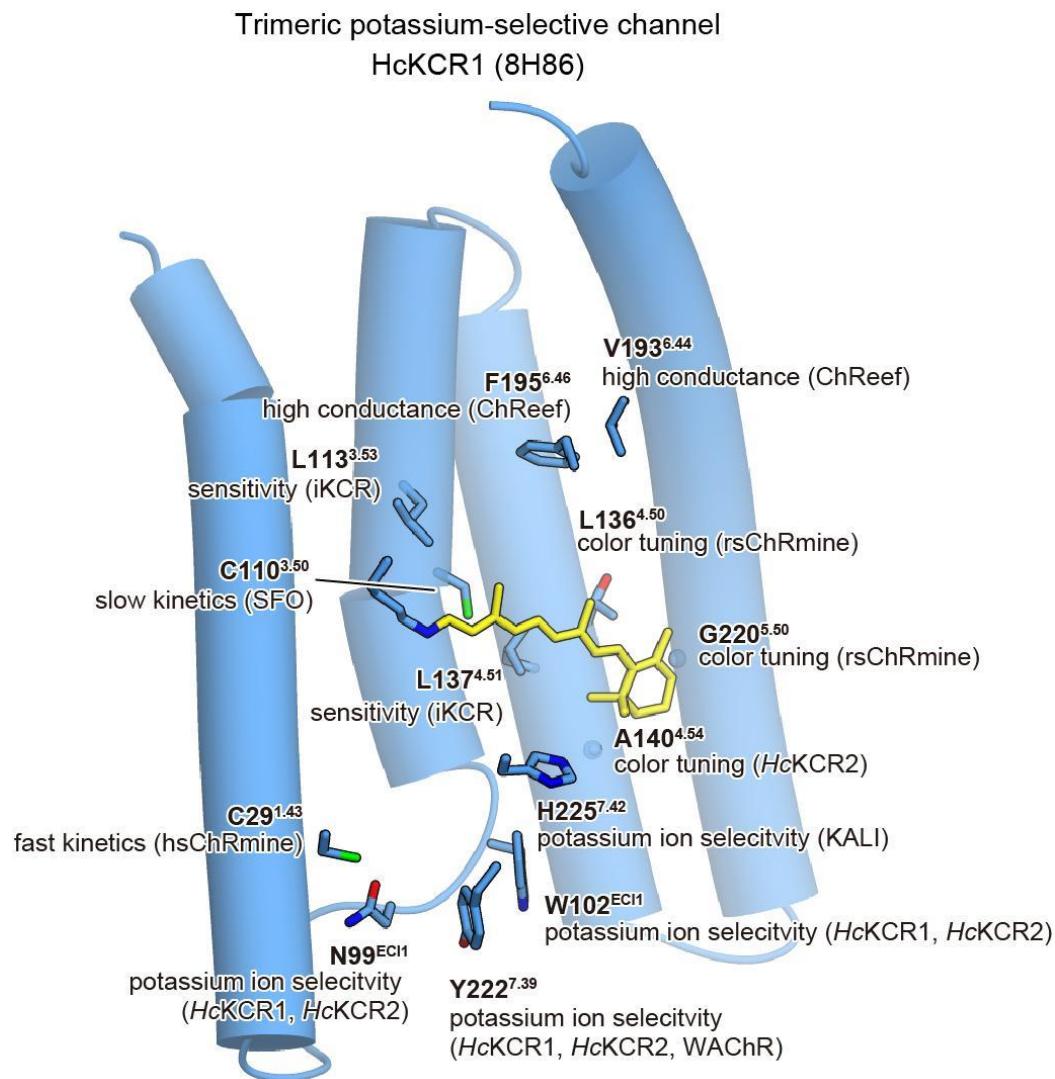
Supplementary Fig. 2 | a Average minimum distance from each structurally equivalent residue position to the retinal chromophore (closest C_α atom), calculated across the dataset. Vertical lines indicate the defined X.50 anchor position for each helix (1.50–7.50). Horizontal lines indicate key interaction 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).

Supplementary Figure 3



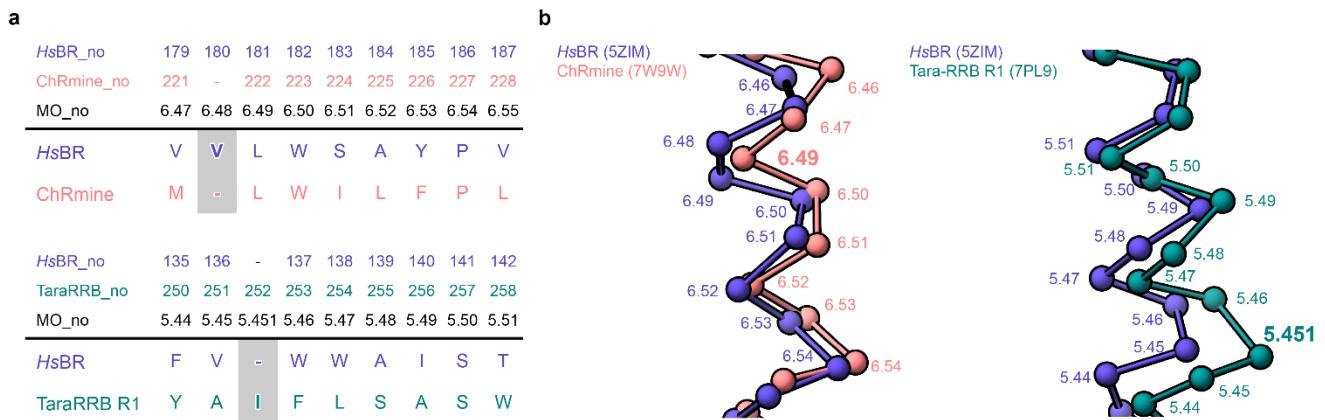
Supplementary Fig. 3 | Examples of x.50 residue positions. The x.50 residues are mapped onto the structures of **a** the representative outward proton-pumping rhodopsin *HsBR* and **b** the cation-conducting channelrhodopsin C1C2.

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 HcKCR1 (PDB ID: 8H86).

Supplementary Figure 5



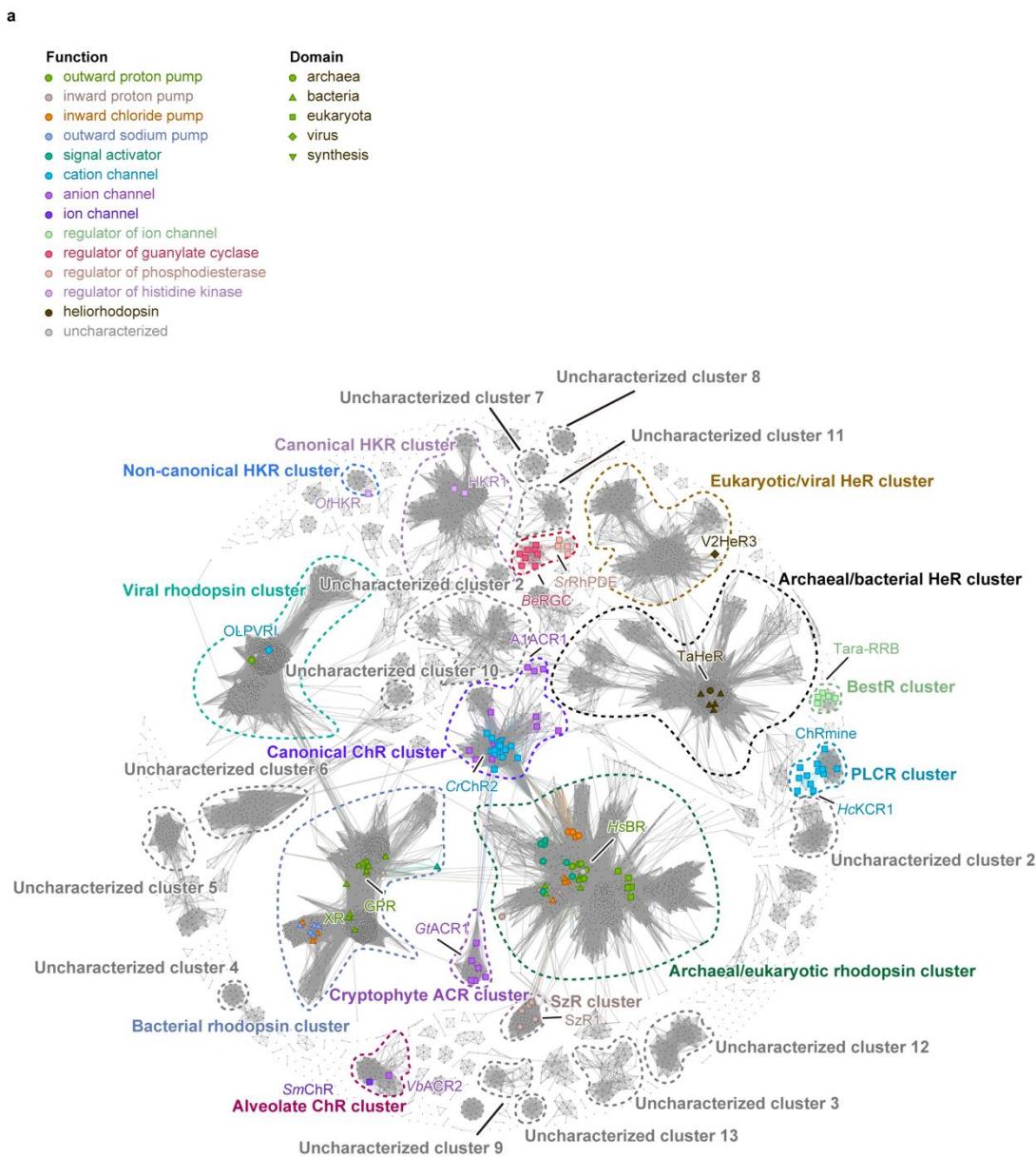
Supplementary Fig. 5 | Rules for insertions and gaps in the MO numbering system. **a** Structure-based alignment of TM6 between the reference *HsBR* and *ChRmine* (top) and of TM5 between *HsBR* and *TaraRRB-R1* (bottom). *HsBR_no* and *ChRmine_no/TaraRRB_no* indicate the residue numbers in each protein, 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 $\text{C}\alpha$ backbone traces of the reference *HsBR* (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.

24

25

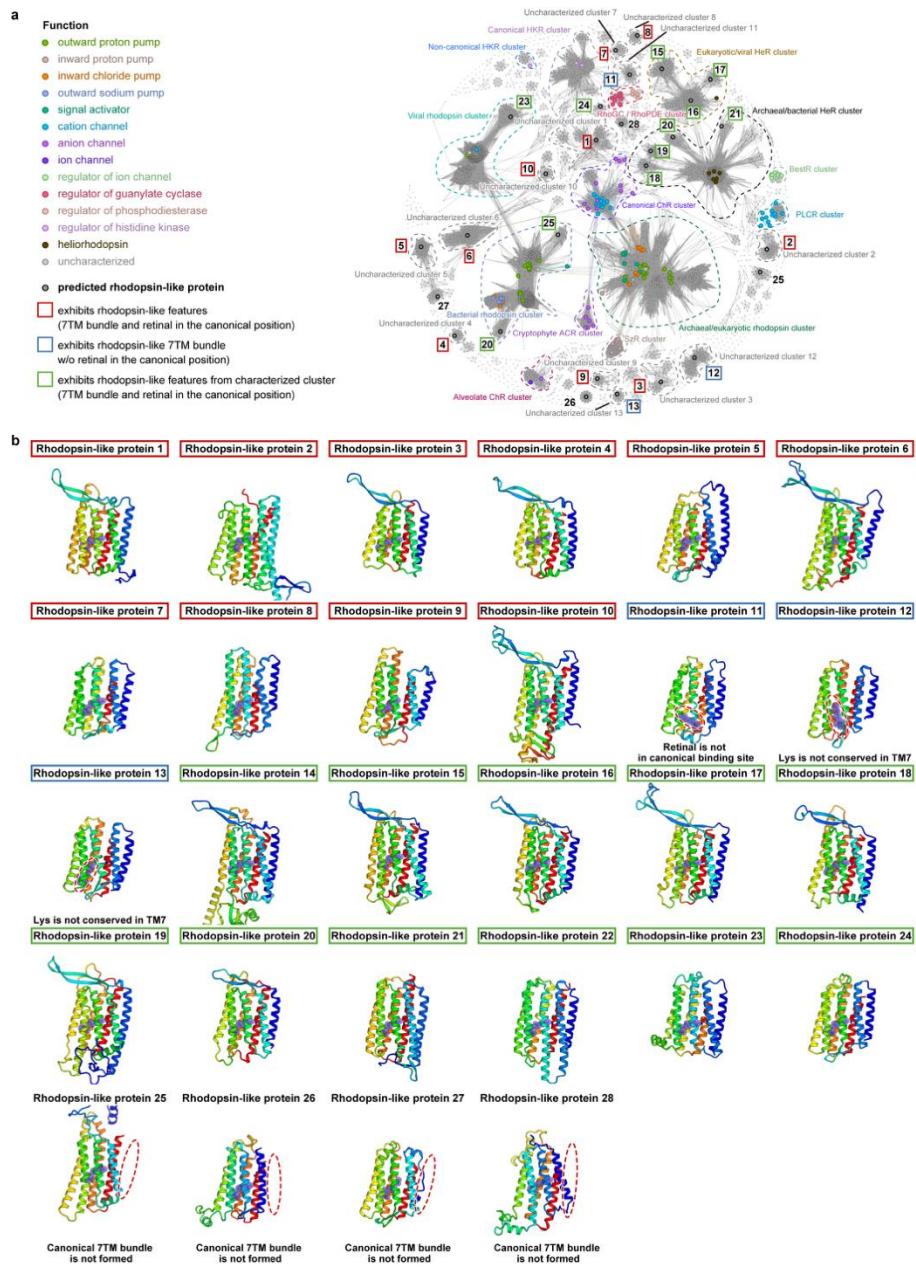
26

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



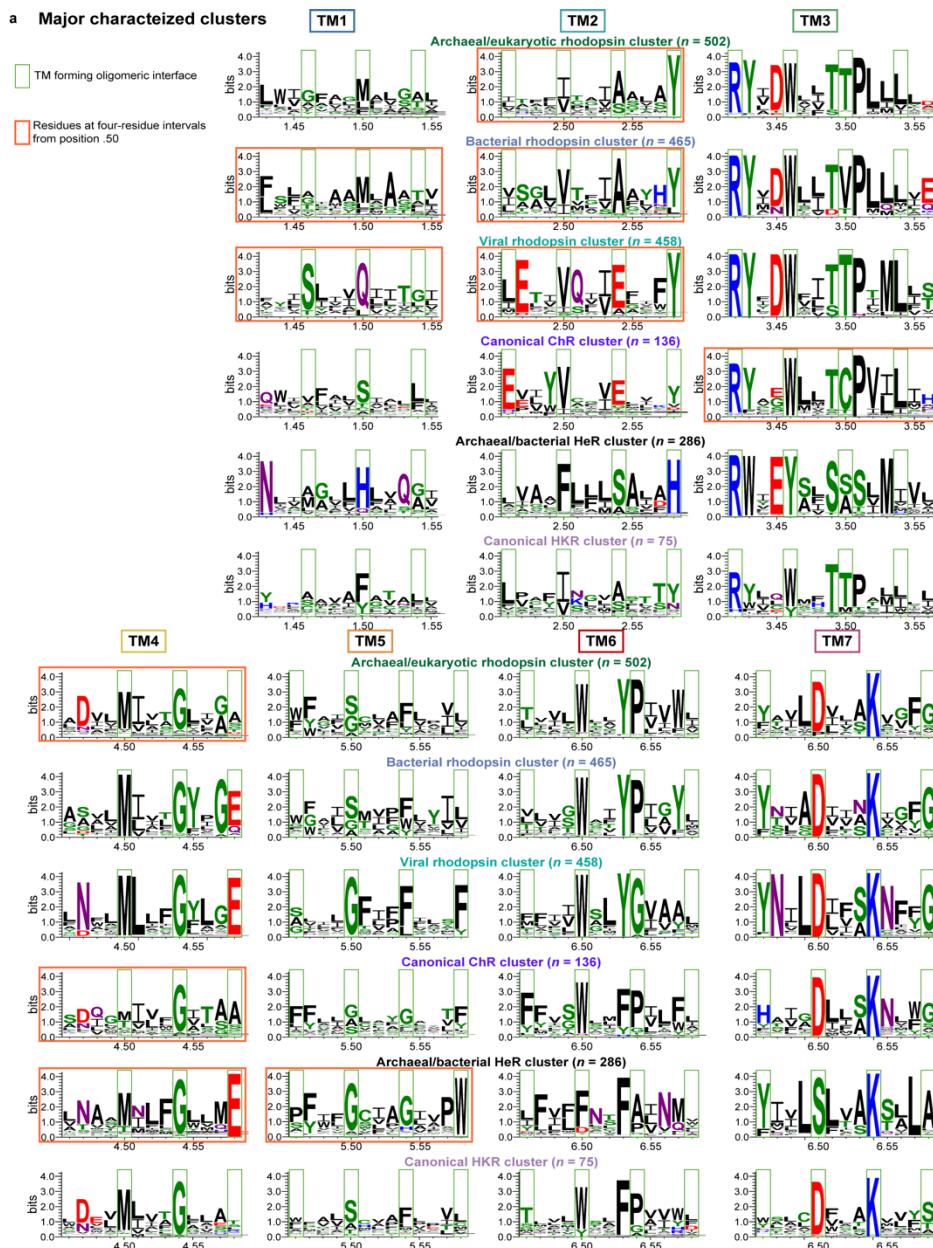
Supplementary Fig. 7 | a Sequence similarity network composed of approximately 40,000 non-redundant microbial rhodopsin sequences, including a large collection of uncharacterized rhodopsin-like sequences. Six major and eight minor clusters containing functionally characterized rhodopsins, as well as 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

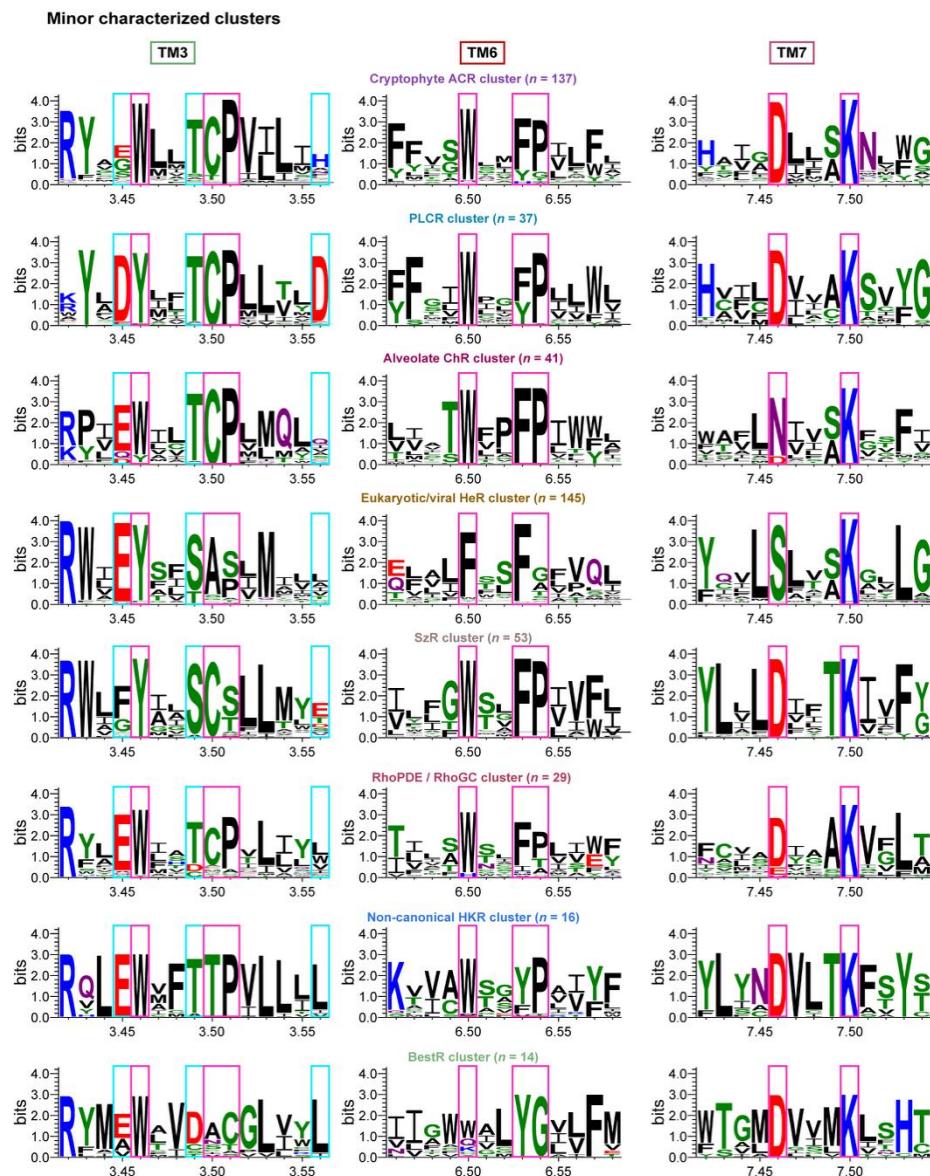
30

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



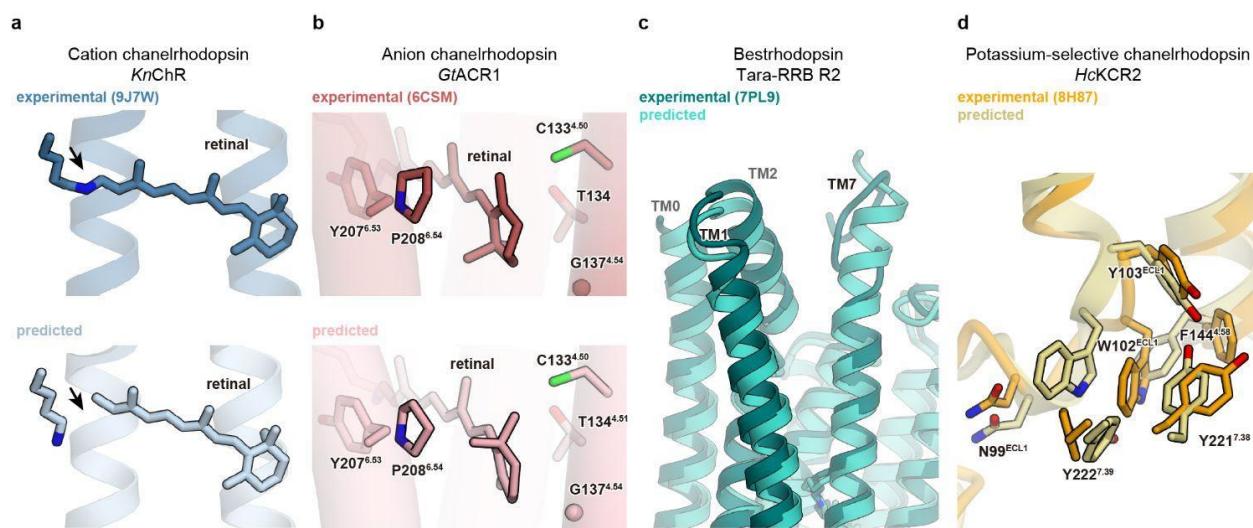
Supplementary Fig. 9 | Sequence logos of TM3, TM6, and TM7 for eight minor clusters containing functionally characterized rhodopsins. Cyan rectangles indicate residues of the DTD motif, and magenta rectangles indicate highly conserved residues within each cluster.

34

35

Supplementary Figure 10

36



37

38 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 *KnChR*, the retinal is not covalently linked to the Schiff base lysine. **b** In *GtACR1*, 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 *HcKCR2*, the architecture of the potassium selectivity filter is incorrectly predicted.

43

44