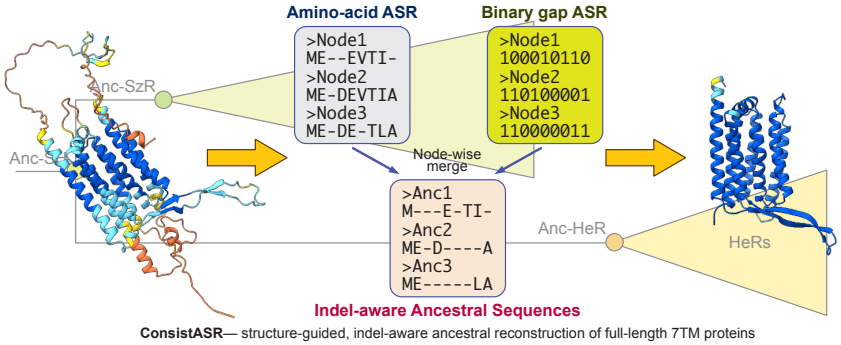

Graphical Abstract

A Structure-guided, Indel-aware Framework for Ancestral Reconstruction of Full-length Seven-transmembrane Proteins

Haruto Ishikawa, Yasuhisa Mizutani

Highlights

A Structure-guided, Indel-aware Framework for Ancestral Reconstruction of Full-length Seven-transmembrane Proteins

Haruto Ishikawa, Yasuhisa Mizutani

- Indel-aware ASR reconstructs full-length seven-transmembrane proteins.
- Structure-guided alignments and mixture models reduce ASR alignment bias.
- Indel-based masking restores realistic loop and terminal architectures.
- Ancestral schizorhodopsin and heliorhodopsin are compact, high-confidence 7TM folds.
- Anc-SzR and Anc-HeR express as stable, colored holoproteins in *E. coli*.

A Structure-guided, Indel-aware Framework for Ancestral Reconstruction of Full-length Seven-transmembrane Proteins

Haruto Ishikawa, Yasuhisa Mizutani

[a] *Department of Chemistry, Graduate School of Science, The University of Osaka, 1-1 Machikaneyama, Toyonaka, 560-0043, Osaka, Japan*

Abstract

Ancestral sequence reconstruction (ASR) of seven-transmembrane (7TM) proteins is hindered by low sequence identity, compositional bias, and heterogeneous extra-membrane (EM) architectures, which exacerbate alignment ambiguity and make indel treatment a critical source of error. Consequently, most membrane-protein ASR studies restrict analyses to trimmed transmembrane (TM) cores and handle EM segments through manual curation. Here, we propose a structure-guided, indel-aware ASR framework that facilitates reliable reconstruction of full-length 7TM proteins from untrimmed alignments. The approach integrates structure-consistent multiple sequence alignments, profile-mixture evolutionary models, reproducible indel refinement, and AlphaFold-based structural evaluation. We further define a node-level reliability index that summarizes posterior support and predicted structural confidence. Applied to microbial rhodopsins, this framework yields compact, full-length ancestors of schizorhodopsins (Anc-SzR) and heliorhodopsins (Anc-HeR) whose TM cores and EM segments display uniformly high confidence. Indel-aware refinement restores realistic loop and terminal architectures, including characteristic EM β-strands and short helices. Reliability patterns vary among nodes: Anc-SzR is consistently robust across alignments, Anc-HeR is structurally well resolved but topologically alignment-sensitive, and

⋆Correspondence
Email address: haruto@chem.sci.osaka-u.ac.jp (Haruto Ishikawa)

the deeper their common ancestor retains a confident global fold despite diffuse residue-wise support. Finally, both Anc-SzR and Anc-HeR express in *E. coli* as stable, colored holoproteins, demonstrating that full-length reconstructions are not only computationally coherent but also experimentally tractable. Together, these results provide a practical strategy for structure-informed, indel-aware ASR of full-length 7TM proteins and offer a general foundation for expanding such reconstructions to diverse 7TM families, encompassing G protein-coupled receptors.

Keywords: ancestral sequence reconstruction, seven-transmembrane proteins, microbial rhodopsins, indels, AlphaFold

---

Introduction

Ancestral sequence reconstruction (ASR) infers the sequences of extinct proteins at internal nodes of a phylogeny, enabling experimental tests of how structure, stability, and function have changed over evolutionary time [1, 2]. By combining phylogenetic models with multiple sequence alignments, ASR has been used to study the evolution of ligand specificity, catalytic mechanisms, and conformational dynamics, and to design historical variants for biochemical characterization. However, the reliability of reconstructed sequences is highly sensitive to technical choices, including multiple sequence alignment [3, 4], evolutionary model specification [5, 6], and the treatment of insertions and deletions (indels) [7, 8]. These factors can alter both inferred tree topologies and the posterior probabilities (PPs) of ancestral residues, often without obvious diagnostics in standard ASR workflows.

For soluble proteins, the impact of alignment uncertainty, model choice, and indel treatment on ASR has been examined in detail [2, 4]. Conversely, seven-transmembrane (7TM) membrane proteins present a distinct set of challenges that remain less well explored. 7TM proteins combine a conserved hydrophobic core of helices with extra-membrane (EM) segments of substantial diversity, including N- and C-terminal tails, peri-membrane helix termini, amphipathic

elements, β-strands, and flexible coils [9, 10]. To mitigate alignment difficulties, many studies trim away most EM regions and termini, effectively restricting ASR to the TM core and handling indels implicitly by manual curation. At the same time, low overall identity, compositional bias, and topological constraints complicate both alignment and model fitting for 7TM families [11, 12]. These issues have discouraged systematic evaluation of full-length 7TM ancestors and have limited the use of EM architecture as a source of evolutionary information.

Previous work has shown that alignment errors can directly propagate into ASR errors and even distort downstream structural and functional inferences, often with a stronger impact than phylogenetic uncertainty itself [13]. In particular, Aadland and Kolaczkowski demonstrated that insertion–deletion errors are a major contributor to ASR inaccuracy across a range of realistic simulation conditions. They further argued that most current ASR workflows still treat indels only implicitly via a single chosen alignment rather than modelling them explicitly. Consequently, 7TM proteins with long, gap-rich EM regions are regarded as particularly challenging test cases for alignment- and indel-sensitive ASR.

Microbial rhodopsins provide an attractive testbed for addressing these issues. Schizorhodopsins (SzRs) and heliorhodopsins (HeRs) have been shown to share a conserved 7TM architecture and a retinal-binding lysine in the seventh transmembrane helix (TM7), but differ in membrane topology and oligomeric state: SzRs adopt an N-out/C-in orientation and form trimers, whereas HeRs display an inverted N-in/C-out topology and form dimers [14, 15, 16, 17, 18, 19]. Here and below, TM1–TM7 denote the seven canonical transmembrane helices numbered from the N terminus. Both families also contain characteristic EM secondary-structure elements, including short and long β-strands and short helices that contribute to intersubunit packing and function [15, 17, 18, 19]. These contrasts in membrane topology, oligomeric state and EM secondary-structure elements provide a controlled setting in which to compare alignment and model choices, to evaluate the impact of indels on full-length ASR, and to test whether reconstructed ancestors can recover not only the 7TM core but also

EM architectural motifs with high confidence.

Recent advances now make it possible to revisit these questions in a more systematic way. Profile-based models such as Q.pfam+R can capture the compositional and site-specific constraints characteristic of membrane proteins more effectively than classical empirical matrices [11], and fast likelihood-based phylogenetic programs such as IQ-TREE support these models in routine analyses. In parallel, explicit binary gap models implemented in RAxML and related tools provide a principled way to reconstruct ancestral indel patterns on a fixed topology. Finally, deep-learning-based structure predictors such as AlphaFold deliver residue-wise confidence measures (pLDDT) that can be combined with ASR posterior probabilities to assess the consistency between sequence-level and structure-level evidence for a given ancestor.

In this study, we develop a structure-guided, indel-aware framework for ancestral reconstruction of full-length 7TM microbial rhodopsins and use it to evaluate the reliability of both TM and EM regions. Our approach integrates (i) structure-consistent multiple sequence alignments (PSI/TM-Coffee and MAFFT L-INS-i), (ii) profile-based evolutionary models (Q.pfam+R) for ASR in IQ-TREE, (iii) an indel-aware refinement procedure that maps binary gap reconstructions from RAxML onto IQ-TREE ancestral states to produce node-specific, indel-corrected sequences, and (iv) AlphaFold-based structural evaluation of monomeric and oligomeric models. We further define a composite node-level reliability metric that combines posterior probabilities, ultrafast bootstrap support, and pLDDT into a single index.

When applied to SzRs and HeRs, this framework enables the quantification of the impact of alignment and model choice on the stability of key ancestral nodes. It also facilitates the determination of the extent to which indel-aware refinement restores realistic sequence lengths and improves confidence in EM regions. Furthermore, it allows for the assessment of how evolutionary, statistical, and structural signals diverge or converge across different ancestral nodes. Finally, we heterologously express and purify the indel-corrected Anc-SzR and Anc-HeR in *Escherichia coli* (*E. coli*), showing that full-length ancestors produced by

this pipeline fold as retinal-binding 7TM proteins with characteristic visible absorption spectra. More broadly, our results provide a practical strategy for performing and interpreting full-length ASR in 7TM proteins, and suggest that EM architecture can be treated as an informative component of membrane-protein evolution rather than as material to be routinely trimmed away.

Results

Alignment strategy and model selection for HeR–SzR phylogenies

Multiple sequence alignment had a decisive impact on the stability of HeR–SzR phylogenies. We therefore compared two structure-aware or structure-compatible strategies: PSI/TM-Coffee [20], which incorporates transmembrane-specific constraints and homology information, and MAFFT L-INS-i [21], which is widely used for protein families where alignment is difficult. The present study demonstrated that PSI/TM-Coffee consistently preserved TM topology and EM registers, whereas L-INS-i produced alignments that were generally coherent but introduced localized uncertainty around the HeR TM1–TM2 region. These differences in alignment quality foreshadowed the topological and ASR stability patterns that are examined below.

For each alignment, the best-fit amino-acid substitution model was determined under the Bayesian information criterion (BIC) using ModelFinder in IQ-TREE [22, 23]. Across the full HeR+SzR+outgroup dataset and the two reduced subsets (HeR+outgroup and SzR+outgroup), profile models of the Q.pfam family with FreeRate heterogeneity (Q.pfam+R5–R7) were consistently preferred (Table 1). For the PSI/TM-Coffee alignment of the full dataset, Q.pfam+R7 had the lowest BIC (BIC $\approx 1.06 \times 10^5$), outperforming the best LG-based free-rate model (LG+F+R8) by ΔBIC ≈ 210. For the MAFFT L-INS-i alignment, Q.pfam+R7 was again preferred, with ΔBIC > 250 relative to the best non-profile empirical model (LG+F+R7). The substantial discrepancy in BIC values indicates a marked preference for Q.pfam+R in comparison to conventional LG/WAG/JTT-type matrices.

A single notable exception occurred for the MAFFT L-INS-i alignment of the HeR+outgroup dataset, where an LG+F+R model was favored and the HeR ancestral node was collapsed into a near-polytomy in the UFBoot2 consensus tree, indicating locally unstable ultrafast bootstrap support around this split [24]. This exception highlights that model choice is often secondary to alignment quality for 7TM proteins: when the alignment introduces local inconsistencies—particularly in the vicinity of the HeR TM1–TM2 registers—the preferred substitution model can shift and internal nodes become unstable. The preference for Q.pfam+R models remained stable when moderately gapped terminal segments were trimmed with trimAl `gappyout`, indicating that it reflects underlying signal present in both membrane-embedded and extra-membrane regions rather than an artifact of gap-rich columns (Table S1).

In contrast, SzR clades demonstrated robustness across all alignments and consistently supported Q.pfam+R models with high UFBoot2 support at the SzR ancestral node (typically 90–95%), underscoring their relative insensitivity to alignment variation (Table 2). Based on these observations, all downstream ASR analyses used PSI/TM-Coffee alignments coupled with Q.pfam+R models as the default and most reliable configuration. This combination consistently produced stable tree topologies and coherent ancestral posterior-probability profiles for both HeR and SzR lineages. It is important to note that the consistent preference for Q.pfam+R models also dictated the methodological choice of performing ASR in IQ-TREE, as profile models of this class are not available in classical pipelines based on RAxML or PAML, and a direct comparison with a conventional RAxML+PAML workflow under LG-based models is provided in Tables S2 and S3.

For completeness, we also evaluated more parameter-rich empirical profile-mixture models. In particular, LG+C60+F+R8 substantially improved the overall fit relative to Q.pfam+R7 (e.g. for the PSI/TM-Coffee alignment of the full dataset, LG+C60+F+R8 achieved BIC $\approx 1.03 \times 10^5$ and $\ln L \approx -5.02 \times 10^4$ compared with BIC $\approx 1.06 \times 10^5$ and $\ln L \approx -5.16 \times 10^4$ for Q.pfam+R7; Table S1). However, these improvements in BIC and log-likelihood did not

resolve the alignment-sensitive behaviour around the HeR ancestral node: under LG+C60+F+R8, the Anc-HeR branch still showed only moderate or unstable UFBoot2 support (no consensus branch for the PSI/TM-Coffee alignment and 47% for MAFFT L-INS-i), whereas Anc-SzR and Anc-SH remained strongly supported. It is evident that LG+C60+F+R8 is computationally more demanding and less accessible for routine ASR workflows. Furthermore, it did not qualitatively change the topology or node-level reliability patterns described below. Therefore, we adopted Q.pfam+R7 as the primary model for ancestral reconstructions and used LG+C60+F+R8 analyses as a robustness check (Tables S2 and S3).

These patterns are summarized in Table 1 and visualized in Figure. 1, highlighting the strong dependence of the HeR ancestral node on alignment quality and the marked robustness of the SzR clade across all conditions.

Indel-aware refinement prevents overextended, low-confidence ancestors

Indel treatment emerged as a major source of uncertainty in full-length ASR. When ASR was performed directly in IQ-TREE without explicit indel modeling, ancestral sequences for HeRs and SzRs became substantially longer than extant homologues, particularly in extra-membrane regions. For the PSI/TM-Coffee alignment, Anc-SzR, Anc-HeR and Anc-SH reached 401, 434 and 408 residues, respectively, compared with 202 residues for SzR4 and 256 residues for HeR 48C12. These over-extended segments lowered mean site-wise PP (to ∼38–66%) and, when evaluated by AlphaFold, produced models that were globally 7TM-like but contained elongated, poorly supported tails and EM segments with depressed mean pLDDT values (typically ∼60–70; computed from per-residue Cα pLDDT; see Materials and Methods). Similar behavior was observed for the MAFFT L-INS-i alignment, suggesting that inflated ancestral length and structurally unreliable termini are a generic consequence of gap-unaware ASR in these datasets (Figure 2A and Table 3).

To address this, we implemented a two-stage indel-aware refinement pipeline inspired by binary gap coding strategies (Figure 2B) [25]. In our implemen-

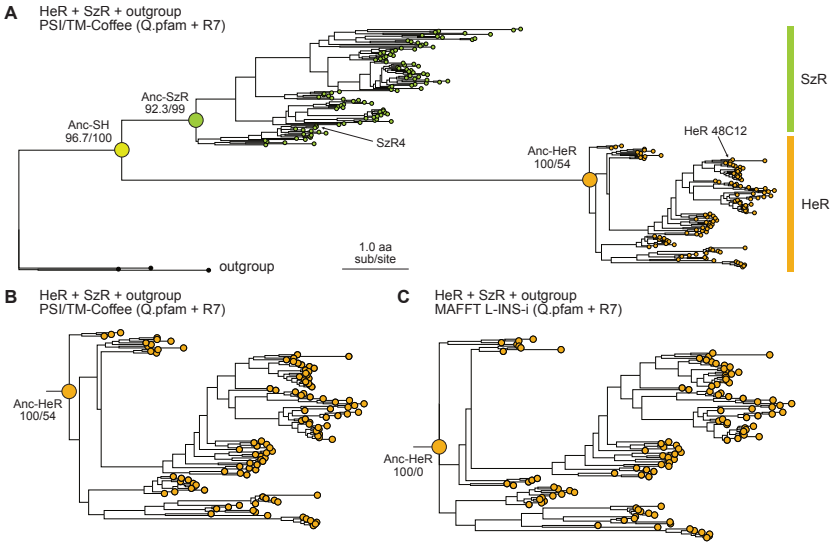

Figure 1: **Phylogenetic context of HeR and SzR ancestors and alignment sensitivity of the HeR clade.** (A) Maximum-likelihood tree for the HeR+SzR+outgroup dataset inferred from the PSI/TM-Coffee alignment under Q.pfam+R7. HeR (orange), SzR (green) and outgroup clades are indicated, and the three focal ancestral nodes (Anc-SH, Anc-SzR and Anc-HeR) are labelled with their SH-aLRT/UFBoot2 support. Extant structural representatives (SzR4 and HeR 48C12) are also marked. (B) Enlargement of the HeR clade from panel (A), showing internal branching patterns and SH-aLRT/UFBoot2 support around Anc-HeR under PSI/TM-Coffee. (C) Corresponding enlargement of the HeR clade inferred from the MAFFT L-INS-i alignment under the same model, illustrating the near-polytomous resolution and collapse of the Anc-HeR branch. In all panels, support values are shown as SH-aLRT/UFBoot2; when a split is absent from the UFBoot2 consensus (i.e. collapsed into a local polytomy), the UFBoot2 value is reported as 0 (e.g. Anc-HeR in panel (C)).

tation, amino-acid ASR is performed in IQ-TREE under profile-based models (Q.pfam+R), whereas RAxML is used exclusively for binary gap inference on the fixed IQ-TREE topology. First, the amino-acid alignment was converted to a binary gap/non-gap representation and analyzed with RAxML under a two-state model to infer ancestral gap patterns at each internal node. This step provides an explicit probabilistic model of indel presence that is not available in IQ-TREE's standard ASR framework. Second, inferred gap states were mapped back onto the IQ-TREE `.state` files using a node-mapping script that reconciles RAxML and IQ-TREE node labels. Columns designated as gaps for a given ancestor were then masked, and the remaining residues were concatenated to produce indel-corrected ancestral sequences. Rather than applying a global gap filter, we therefore remove sites in a node-specific manner according to the inferred ancestral indel state, allowing each ancestor to contract towards a length that is compatible with its own gap history (Figure 2C).

Indel-aware correction had a pronounced impact on both statistical and structural metrics (Figure 2D and Table 3). For the PSI/TM-Coffee alignment, mean PP increased from 51–66% to 82–91% for Anc-SzR and Anc-HeR, and mean pLDDT rose from ∼62–69 to >93. Concurrently, the sequence lengths converged to values that closely matched extant structures: 206 residues for Anc-SzR and 259 residues for Anc-HeR, compared with 202 residues for SzR4 and 256 residues for HeR 48C12. Comparable improvements were observed for the MAFFT L-INS-i alignment in an independent implementation of the same pipeline, with Anc-SzR and Anc-HeR contracting to 203 and 260 residues, respectively. Anc-SzR, Anc-HeR and Anc-SH all conserved the lysine residues capable of binding retinal. The FSE motif, which is present in TM3 of many SzRs, was conserved in Anc-SzR but was replaced by a DSE sequence in Anc-SH. AlphaFold models of the indel-corrected ancestors formed compact 7TM folds with uniformly high pLDDT and retained structured extra-membrane motifs, including the short β-strands between TM2 and TM3 in SzR4 and the combination of a short α-helix between TM2 and TM3 and long β-strands between TM1 and TM2 in HeR 48C12 [18, 17]. The visual impact of indel-aware

refinement is striking: models predicted from uncorrected ancestors exhibit elongated, low-confidence extra-membrane tails, whereas the indel-corrected ancestors collapse into compact 7TM folds whose lengths closely match extant rhodopsins and whose EM motifs are recovered with uniformly high pLDDT. Consequently, the node-specific removal of sites by the binary indel model is not merely a cosmetic truncation of termini but a critical step for obtaining full-length 7TM ancestors that are both statistically well supported and structurally coherent.

Beyond the global statistics, the patterns of EM secondary structure between TM1–TM2 and TM2–TM3 further illustrate how the indel-aware pipeline captures branch-specific gains and losses of architectural motifs (Figure 3). In extant structures, SzR4 carries short β-strands between TM2 and TM3 and no extended secondary structure between TM1 and TM2, whereas HeR 48C12 combines a short α-helix between TM2 and TM3 with long β-strands between TM1 and TM2 (Figure 3A and B). The indel-corrected ancestors mirror these patterns in a lineage-specific manner: Anc-SzR retains SzR4-like short β-strands between TM2 and TM3 and a loop-like TM1–TM2 region, whereas Anc-HeR retains the HeR 48C12-like combination of a short α-helix between TM2 and TM3 and long β-strands between TM1 and TM2. Anc-SH, by contrast, preserves only the SzR-like short β-strands between TM2 and TM3 and lacks the HeR-specific α-helix and long β-strands. Given that Anc-SH is phylogenetically closer to Anc-SzR than to Anc-HeR, this pattern is consistent with a gain of the HeR-specific EM helix/β-strand combination along the HeR branch after the divergence from Anc-SH (Figure 3C). It also shows that the binary indel reconstruction does not simply trim termini uniformly, but can represent node-specific absence of EM motifs that are present in more derived nodes and extant sequences.

We further examined how similar the indel-corrected ancestors are across alignment strategies and to extant structural templates. Pairwise amino-acid identities between PSI/TM-Coffee- and L-INS-i–based ancestors were 88.7%, 88.8% and 82.8% for Anc-SzR, Anc-HeR and Anc-SH, respectively, indicating that the indel-aware pipeline converges on highly similar solutions despite differences

in the underlying alignments. The ancestors also retained moderate similarity to extant rhodopsins: Anc-SzR (PSI/TM-Coffee) shared 54.0% identity with SzR4 (52.5% for the FFT-NS-1–based reconstruction), and Anc-HeR shared 51.0% and 47.1% identity with HeR 48C12 for PSI/TM-Coffee- and L-INS-i–based ancestors, respectively. These values are characteristic of ASR targets at this phylogenetic depth and establish a quantitative foundation for the structural comparisons that follow.

Region-level reliability: TM versus EM segments

To quantify the reliability of different structural compartments, we partitioned each ancestral model into transmembrane (TM) and extra-membrane (EM) regions using membrane boundaries derived from the Orientations of Proteins in Membranes (OPM) database [26]. For SzR, we used the OPM entry for SzR4; for HeR, we used the entry for HeR 48C12. The term "TM regions" is used to denote membrane-embedded portions of α-helices within the hydrophobic core of the bilayer. In contrast, "EM regions" refer to all residues located outside the bilayer, including helix termini, β-strands, and short coils. Thus, in our convention "TM" refers strictly to the membrane-embedded cores of helices, whereas helical residues outside the OPM membrane slab are classified as EM.

OPM-derived TM/EM labels were projected onto Anc-SzR and Anc-HeR via the structure-consistent alignments used in ASR, and region-level metrics were computed from indel-corrected sequences and AlphaFold models (Figure 4A, Table 4 and Tables S4–S7). For each ancestor and region, we summarized the mean PP, the mean pLDDT, their product (PP$\times$pLDDT) and their difference (PP$-$pLDDT, with both PP and pLDDT scaled to 0–1). Anc-SH was excluded from Table 4 due to the complexity of its hybrid HeR/SzR character, which hinders direct mapping to a single extant template. However, its global patterns are described qualitatively below.

Three consistent patterns emerged. First, TM regions are uniformly high-confidence. In Anc-SzR and Anc-HeR, TM regions displayed high mean PP and very high mean pLDDT (Figure 4B and Table 4). For example, TM segments in

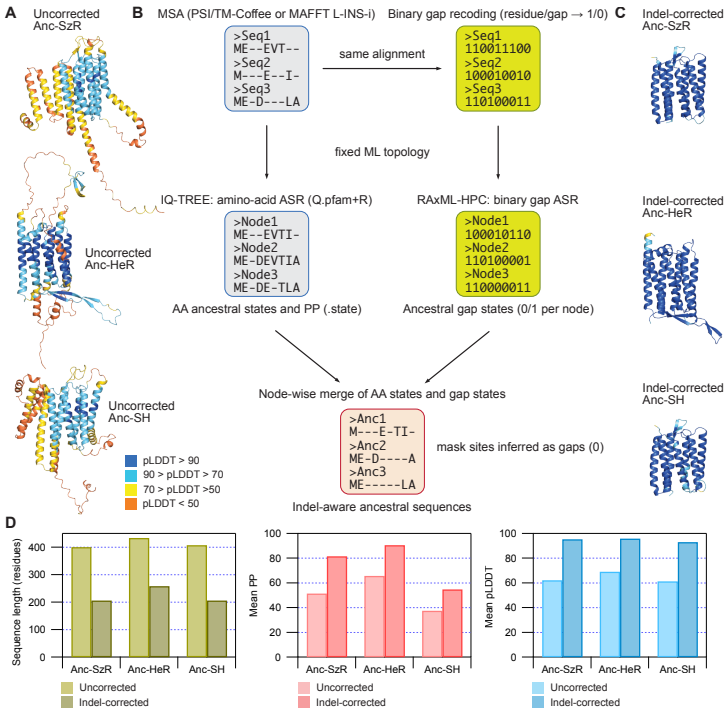

Figure 2: **Indel-aware refinement restores realistic sequence lengths and improves statistical and structural confidence.** (A) AlphaFold3 monomeric models for uncorrected ancestral sequences inferred directly from the full-length PSI/TM-Coffee alignment under Q.pfam+R7, without explicit indel treatment. Structures are colored by per-residue pLDDT using the standard AlphaFold scale. All three ancestors display elongated extra-membrane segments and low-confidence tails despite forming overall 7TM-like folds. (B) Schematic overview of the indel-aware ASR pipeline. The same full-length amino-acid alignment is used for amino-acid ASR in IQ-TREE and for binary gap coding followed by ancestral gap-state inference in RAxML on the fixed IQ-TREE topology. Node-specific binary gap reconstructions are then merged with the IQ-TREE `.state` files to mask sites inferred as gaps at each internal node, yielding indel-corrected ancestral sequences. (C) AlphaFold3 monomeric models for the indel-corrected ancestors obtained after node-specific gap masking, colored as in panel (A). Indel-aware correction contracts over-extended extra-membrane regions, removes low-confidence tails, and yields compact 7TM folds with uniformly high pLDDT. (D) Effects of indel-aware refinement on sequence length and per-site confidence for Anc-SzR, Anc-HeR and Anc-SH. Bar plots show sequence length, mean posterior probability (PP) and mean pLDDT for uncorrected and indel-corrected ancestors inferred from the PSI/TM-Coffee alignment under Q.pfam+R7. Indel-aware refinement simultaneously increases both mean PP and mean pLDDT.

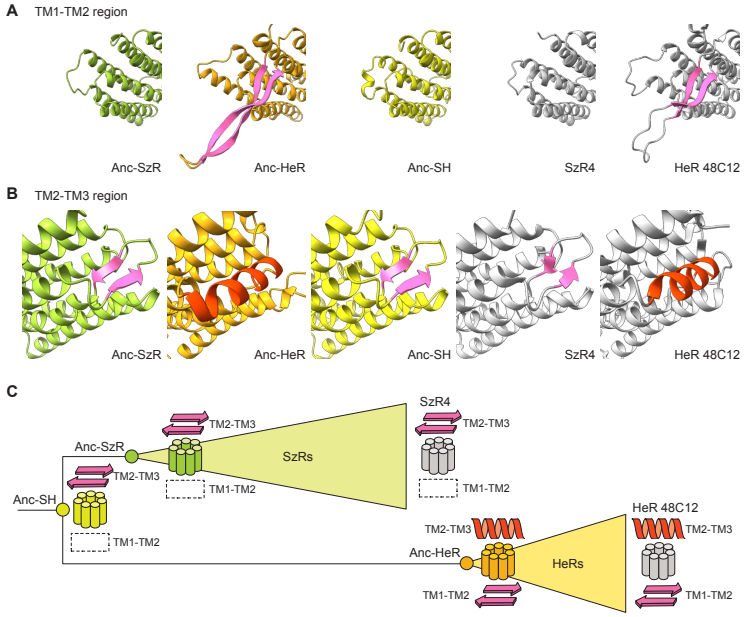

Figure 3: **Evolutionary gain and loss of extra-membrane secondary-structure motifs between TM1–TM2 and TM2–TM3.** (A) Close-up views of the extra-membrane region between TM1 and TM2 for indel-corrected Anc-SzR (green), Anc-HeR (orange), Anc-SH (light green), and the extant structures SzR4 (PDB ID 7e4g) and HeR 48C12 (PDB ID 6su3; both light grey), aligned to their respective templates. β-strands in this region are highlighted in pink. Anc-HeR and HeR 48C12 share long β-strands between TM1 and TM2, whereas Anc-SzR, Anc-SH and SzR4 lack defined secondary structure at this position and form a loop. (B) Corresponding close-up views of the extra-membrane segment between TM2 and TM3, β-strands are highlighted in pink and short α-helices in red. In Anc-SH, Anc-SzR and SzR4, this region forms short β-strands, whereas in Anc-HeR and HeR 48C12 it is replaced by a short α-helix, indicating a topological shift of secondary-structure motifs along the HeR lineage. (C) Schematic summary of motif gain and loss mapped onto a simplified HeR–SzR phylogeny. Each ancestor or extant protein is represented by a seven-cylinder cartoon indicating the TM helices, with the TM2–TM3 and TM1–TM2 regions annotated above and below. The pink arrows denote β-strands, the red ribbons represent short α-helices, and the dashed boxes mark loop-like regions lacking defined secondary structure. Together, these panels show that Anc-SzR and SzR4 largely retain the Anc-SH pattern (short β-strands between TM2 and TM3 and a loop between TM1 and TM2), whereas the HeR lineage loses the TM2–TM3 β-strands, gains a short α-helix in its place, and acquires a new β-strands between TM1 and TM2.

Anc-SzR and Anc-HeR reached mean PP values of approximately 82–94% and mean pLDDT values above 95, regardless of whether PSI/TM-Coffee or MAFFT L-INS-i alignments were used. As the PP×LDDT approached its upper range in the TM regions, there was a strong agreement between the sequence-based and structure-based confidence in the membrane-embedded helix cores.

Second, EM regions retain unexpectedly high confidence. Although EM regions were slightly lower than TM regions on average, mean PP and pLDDT remained high. EM segments generally exhibited mean PP values around 76–88% and mean pLDDT values around 89–94, and PP×pLDDT values supported the view that many extra-membrane segments are reliably reconstructed rather than being dominated by alignment noise (Figure 4B and Table 4). It is noteworthy that EM segments corresponding to characteristic β-strand motifs in HeR 48C12 and SzR4 also retained high PP and pLDDT in Anc-HeR and Anc-SzR. This finding indicates that these EM architectures are recovered as structured elements rather than being lost or degraded by indel-aware correction.

Third, local hotspots of uncertainty coincide with TM1 and EM segments adjacent to TM5. Across ancestors and alignments, PP tended to be locally reduced in TM1 and in EM segments proximal to TM5, consistent with the slightly lower PP and PP×pLDDT values observed in these regions in Tables S4 and S6. These segments likely represent evolutionarily labile regions where fold-level constraints are maintained but residue-level identities vary more freely.

The PP−pLDDT difference provided additional insight at the local scale. Region-averaged PP−pLDDT values were similar between TM and EM segments (Figure 4B and Table 4), consistent with the overall high confidence of both compartments. However, site-wise PP−pLDDT profiles revealed distinct patterns: sites with relatively low PP but high pLDDT mark positions where the three-dimensional fold is stable even though the specific residue is ambiguous. This indicates that structural constraints are stronger than sequence conservation. Conversely, sites with high PP but lower pLDDT tend to occur in flexible EM regions where the alignment is confident but conformational variability is intrinsically high. These metrics may also facilitate more transparent

manual trimming decisions. For instance, regions exhibiting both low PP and low pLDDT are indicative of alignment regions that are poorly supported at both the sequence and structural levels. Conversely, sites with low PP but high pLDDT correspond to structurally well-packed but sequence-wise ambiguous positions. These positions might be more appropriately classified as designable mutational hotspots rather than as candidates for deletion.

Together, these region-level analyses demonstrate that OPM-guided TM/EM partitioning provides a more realistic picture than a simple TM-versus-loop dichotomy and that full-length ancestors preserve meaningful extra-membrane architecture.

Node-level reliability: divergence between evolutionary and structural support

We next examined node-level reliability at three key ancestors: the SzR ancestor (Anc-SzR), the HeR ancestor (Anc-HeR), and their common ancestor (Anc-SH). For each node, we summarized three metrics: (i) ultrafast bootstrap support (BS) from the UFBoot2 consensus tree for the corresponding split, (ii) mean PP across sites in the indel-corrected ancestors, and (iii) mean pLDDT of the AlphaFold models (Table 2). Branches that were not recovered in the UFBoot2 consensus (i.e. collapsed into a polytomy) were assigned BS = 0 to reflect unresolved topology. Unless otherwise noted, the BS values reported below refer to UFBoot2 support extracted from the consensus tree; these UFBoot2 values together with SH-aLRT support on the ML topology are shown for major branches in Figure 1. Conventional ultrafast bootstrap values from the ML tree are summarized in Table 2.

Anc-SzR exhibited uniformly high support across metrics. On the UFBoot2 consensus trees, BS at the SzR ancestral node reached 99% (PSI/TM-Coffee) and 96% (L-INS-i). Furthermore, the indel-corrected Anc-SzR sequences exhibited high mean PP (80–82%) and high mean pLDDT (94–96). TM segments, in particular, displayed near-ceiling pLDDT values and high PP, consistent with a well-resolved split and a structurally coherent 7TM ancestor. Taken together, these metrics identify Anc-SzR as a robust candidate for experimental

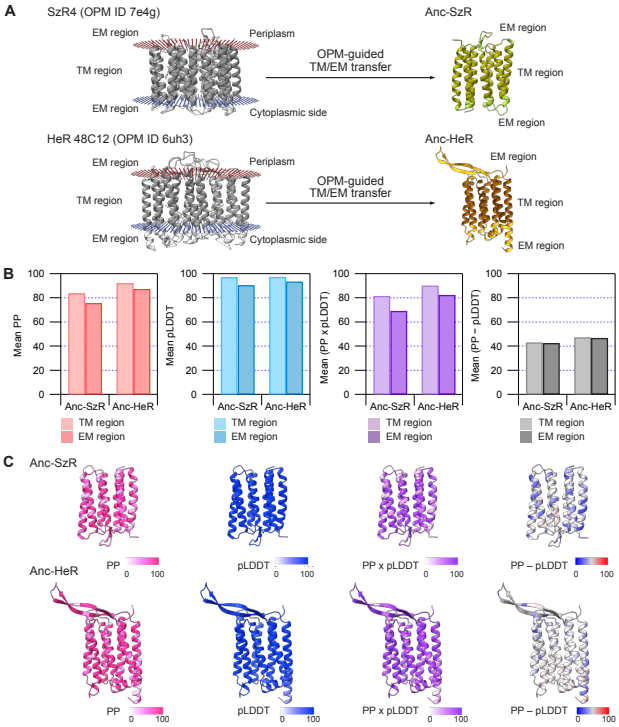

Figure 4: **Region-level reliability of TM and EM segments in Anc-SzR and Anc-HeR.** (A) OPM-guided transfer of transmembrane (TM) and extra-membrane (EM) annotations from extant structures to indel-corrected ancestors. OPM entries for SzR4 and HeR 48C12 are shown on the left, and their TM/EM boundaries are projected onto Anc-SzR and Anc-HeR via the PSI/TM-Coffee alignment under Q.pfam+R7. In the ancestral models, TM residues are colored dark green (Anc-SzR) or brown (Anc-HeR), and EM residues light green (Anc-SzR) or orange (Anc-HeR). (B) Region-level summary statistics for Anc-SzR and Anc-HeR. For each ancestor, bar plots report the mean PP, mean pLDDT, mean (PP×pLDDT) and mean (PP−pLDDT) separately for TM and EM regions defined in panel (A). All metrics are shown on a 0–100 scale. TM segments display uniformly high confidence, whereas EM regions show only slightly reduced but still high values, indicating that many extra-membrane segments are reliably reconstructed. (C) Per-residue reliability mapped onto Anc-SzR and Anc-HeR structures. For each ancestor, four views are shown with Cα atoms colored by PP (white to red), pLDDT (white to blue), PP×pLDDT (white to purple), or PP−pLDDT (blue through grey to red), using the same 0–100 scale. These maps highlight that TM cores are consistently high-confidence, whereas local pockets of reduced PP with high pLDDT in EM segments mark positions where the fold is well defined but residue identity remains more ambiguous. Per-residue reliability mapped onto Anc-SH are shown in Figure S3.

resurrection.

Anc-HeR displayed a more nuanced profile. Under the PSI/TM-Coffee alignment, the HeR ancestral split was recovered in the UFBoot2 consensus with only moderate support (BS = 54%), despite the indel-corrected Anc-HeR sequence showed very high mean PP (≈91%) and mean pLDDT (≈96). Under the MAFFT L-INS-i alignment, the corresponding internal branch was collapsed into a near-polytomy in the UFBoot2 consensus, and we therefore treated its BS as 0 in our summaries, despite similarly high PP and pLDDT for the Anc-HeR sequence. This discrepancy—high residue-level and structural confidence versus limited or absent UFBoot2 support—indicates that alternative near-optimal arrangements of closely related HeR lineages are frequently sampled in the bootstrap replicates. In other words, the HeR ancestor is well supported once a particular local HeR subtree topology is chosen, but the local arrangement of HeR sequences remains sensitive to alignment details.

Anc-SH presented the opposite pattern. The deeper split separating HeRs and SzRs from the outgroup received maximal UFBoot2 support (BS = 100%) under all conditions, indicating a robust topology at this depth. However, the indel-corrected Anc-SH sequences showed substantially lower mean PP (≈55–57%) than the terminal ancestors, even though mean pLDDT remained high (≈93–94) and the overall 7TM fold appeared well formed. Thus, the deep ancestral split is stable, but residue-level identities at Anc-SH are more diffuse, with many sites supported by multiple plausible amino-acid states.

These contrasting cases illustrate that no single metric—BS, PP or pLDDT—is sufficient to summarize ASR reliability. Topological stability does not guarantee residue-level certainty, and high PP or structural confidence does not automatically imply a unique or unambiguous evolutionary history. In the following section, we therefore introduce a composite index that integrates these three dimensions into an explicit node-level reliability score.

Composite reliability metric integrating PP, pLDDT, and BS

In order to synthesize information from evolutionary, statistical, and structural sources, a composite node-level metric was introduced. The sequence-structure consistency index was initially defined as the mean of the per-site product PP×pLDDT for a given ancestor. This quantity is high only when both the posterior probability of the inferred residue and the structural confidence at that position are high. We then combined this with bootstrap support for the corresponding split, estimated by UFBoot2 on the bootstrap consensus tree, to obtain

$$\text{CompositeScore}_{0\text{–}100} = BS_{\text{UFBoot2}} \times \overline{\text{PP} \times \text{pLDDT}}, \tag{1}$$

where BS_{UFBoot2} is the UFBoot2 frequency of the relevant split (0–1), PP is expressed as a fraction (0–1), and pLDDT is scaled to 0–1 for the computation. In practice, we used the mean PP×pLDDT values (0–100) derived from the indel-corrected Cα atoms, converted them to 0–1, and multiplied by the UFBoot2 support extracted from the `.contree` files. For internal branches that were collapsed in the consensus tree (i.e. unresolved polytomies), we set $BS_{\text{UFBoot2}} = 0$. The resulting composite scores were reported on a 0–100 scale for interpretability (Table 2).

Applied to the PSI/TM-Coffee reconstructions, the mean PP×pLDDT values were approximately 78, 87 and 52 for Anc-SzR, Anc-HeR and Anc-SH, respectively, while the corresponding UFBoot2 supports were 0.99, 0.54 and 1.0 (Figure 5A and Table 2). The combination of these quantities resulted in composite scores of approximately 77 for Anc-SzR, 47 for Anc-HeR, and 52 for Anc-SH. Under MAFFT L-INS-i, the mean PP×pLDDT values were found to be comparable (approximately 76, 86 and 54), but the UFBoot2 supports were 0.96, 0 and 1.0, giving composite scores of approximately 73 for Anc-SzR, 0 for Anc-HeR (reflecting complete collapse of the corresponding branch in the bootstrap consensus), and 54 for Anc-SH (Figure 5B and Table 2). Consequently, across both alignments, Anc-SzR consistently exhibits high composite reliability, Anc-SH attains moderate reliability, and Anc-HeR falls between the two when

the HeR split is resolved but drops to zero when it is not.

This behavior is informative. Anc-HeR combines very high PP×pLDDT with alignment-sensitive BS_{UFBoot2}; its composite score is therefore high under PSI/TM-Coffee, but strongly suppressed under L-INS-i, consistent with the local topological instability described above. By contrast, Anc-SzR combines high PP×pLDDT with robust UFBoot2 support under both alignments, making it the most conservative target for experimental resurrection. Anc-SH, finally, has consistently high UFBoot2 support but lower PP×pLDDT, emphasizing that deep nodes can be topologically stable yet residue-wise diffuse (Figure 5C).

We emphasize that the composite metric is intended as an integrative summary and a decision-support tool, not as a replacement for examining each component separately. In particular, nodes whose branches are collapsed or strongly rearranged in UFBoot2 consensus trees, or whose bootstrap support is clearly alignment-dependent, should be explicitly flagged, and composite scores for such nodes treated as indicative rather than definitive. In this sense, Anc-HeR exemplifies a "conditionally reliable" ancestor whose structural and site-wise statistics are strong, but whose evolutionary context remains sensitive to alignment choices.

Oligomeric-state prediction supports dimeric HeR and trimeric SzR ancestors

Finally, we asked whether the reconstructed ancestors support oligomeric states consistent with extant rhodopsins. AlphaFold-Multimer was utilized to predict oligomeric complexes (dimer, trimer and pentamer) for Anc-SzR, Anc-HeR and Anc-SH, starting from indel-corrected sequences inferred from both PSI/TM-Coffee and MAFFT L-INS-i alignments (Figure 6). The focus of this study is the stoichiometries of 2, 3 and 5 because microbial rhodopsins are experimentally observed as dimers, trimers or pentamers, whereas tetrameric assemblies have not been reported for this family [27, 15]. For each target and stoichiometry, we summarized the predicted TM-score (pTM) of the overall complex and the interface TM-score (ipTM), which together quantify fold accuracy and interface quality [28].

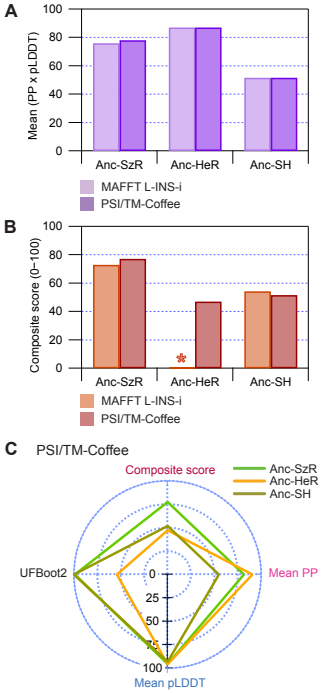

Figure 5: **Node-level sequence–structure consistency and composite reliability across alignments.** (A) Mean PP× pLDDT for Anc-SzR, Anc-HeR and Anc-SH under the PSI/TM-Coffee and MAFFT L-INS-i alignments. Values are averaged over Cα sites and shown on a 0–100 scale. For all three nodes, PP× pLDDT is high and broadly similar between alignments, indicating that residue probabilities and AlphaFold confidence are jointly strong in both TM and extra-membrane regions. (B) Composite node-level reliability scores (0–100) combining UFBoot2 branch support and mean PP× pLDDT for the same nodes and alignments. The composite score is defined from UFBoot2 support for the corresponding split and the mean PP× pLDDT (see Methods). For the MAFFT L-INS-i alignment, the Anc-HeR branch is collapsed in the UFBoot2 consensus; its composite score is designated as 0 and it is denoted by an asterisk. Across both alignments, Anc-SzR attains consistently high composite scores, Anc-SH reaches intermediate values, and Anc-HeR is conditionally reliable, with a high composite score only when the HeR split is recovered in the consensus. (C) Radar plot summarizing node-level metrics for the PSI/TM-Coffee alignment. Anc-SzR (green), Anc-HeR (orange) and Anc-SH (dark yellow) are shown with axes corresponding to mean PP, mean pLDDT, mean PP× pLDDT and the composite score (all on a 0–100 scale). The plot highlights that Anc-SzR is uniformly strong across all metrics, Anc-HeR combines very high sequence–structure consistency with alignment-sensitive branch support, and Anc-SH exhibits a robust topology but lower residue-level certainty.

Anc-SzR was consistently predicted to form a trimer (Figure 6A, C and Table S8). For PSI/TM-Coffee, trimeric models achieved high ipTM (0.88) and pTM (0.90), substantially exceeding the corresponding scores for dimers (ipTM $\approx$ 0.23, pTM $\approx$ 0.56) or pentamers (ipTM $\approx$ 0.51, pTM $\approx$ 0.59). L-INS-i yielded a very similar ranking, with trimeric Anc-SzR again showing the highest interface scores (ipTM $\approx$ 0.85, pTM $\approx$ 0.87) and dimeric and pentameric complexes scoring poorly. These results are consistent with the trimeric assemblies observed for extant SzRs, suggesting that both the 7TM cores and the reconstructed EM elements are compatible with a trimeric ancestral interface.

In contrast, dimeric models were strongly favoured over higher-order assemblies for Anc-HeR (Figure 6B, C and Table S8). Under PSI/TM-Coffee, the best dimeric prediction showed a high ipTM of 0.81 and pTM of 0.86, whereas trimeric and pentameric models displayed markedly lower interface scores (ipTM $\approx$ 0.3–0.4, pTM $\approx$ 0.5). The same pattern was observed for L-INS-i (ipTM $\approx$ 0.79, pTM $\approx$ 0.84 for the dimer, versus ipTM $\lesssim$ 0.3 for trimer and pentamer), indicating that the preference for a dimeric Anc-HeR is robust to alignment strategy. Visual inspection of the dimeric models further revealed that the EM β-strand region between TM1 and TM2 contributes to a plausible dimer interface, consistent with the proposed function of this motif in extant HeRs.

The deeper Anc-SH exhibited intermediate behavior. Trimeric models consistently scored better than dimers or pentamers, with ipTM values of approximately 0.67 (PSI/TM-Coffee) and 0.77 (L-INS-i), whereas dimeric and pentameric assemblies had substantially lower ipTM and pTM scores (Figure 6C and Table S8). However, the interface scores for Anc-SH remained lower than those of the terminal ancestors, consistent with its deeper phylogenetic position, lower mean PP and the mixed HeR/SzR character of its sequence. In summary, the multimeric predictions reinforce the conclusion that our indel-corrected ancestors not only form plausible monomeric 7TM folds with high pLDDT, but also encode oligomerization interfaces compatible with the known dimeric and trimeric states of HeR and SzR, respectively.

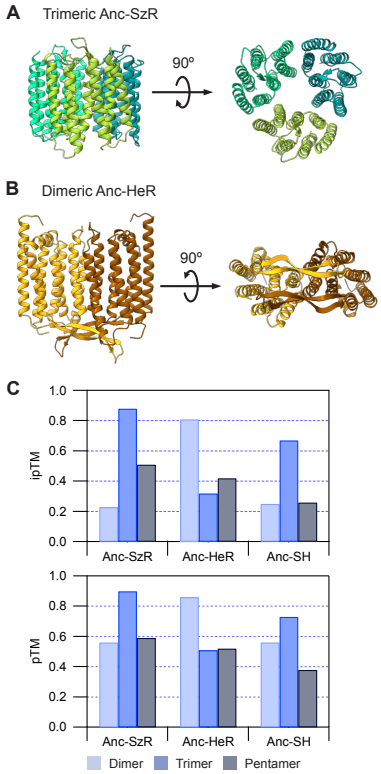

Figure 6: **Oligomeric-state predictions support trimeric Anc-SzR and dimeric Anc-HeR.** (A) AlphaFold-Multimer model of trimeric Anc-SzR viewed parallel to the membrane (left) and from the cytoplasmic side (right). Each protomer is shown in cartoon representation and colored by chain. The trimeric arrangement and packing of the seven-helix bundles closely resemble those observed for extant SzR trimers. (B) AlphaFold-Multimer model of dimeric Anc-HeR viewed parallel to the membrane (left) and from the extracellular side (right), with protomers again colored by chain. The dimer interface involves extra-membrane elements between TM1 and TM2, consistent with proposed dimerization motifs in extant HeRs. (C) Predicted interface TM-scores (ipTM) and overall complex TM-scores (pTM) for dimeric, trimeric and pentameric assemblies of Anc-SzR, Anc-HeR and Anc-SH, based on indel-corrected sequences inferred from the PSI/TM-Coffee alignment. Bars show the top-ranked AlphaFold-Multimer models for each stoichiometry. Anc-SzR attains the highest ipTM and pTM as a trimer, Anc-HeR as a dimer, and Anc-SH shows intermediate ipTM values with a modest preference for trimeric assemblies. Tetrameric models were not evaluated, as microbial rhodopsins are experimentally observed to form dimers, trimers or pentamers rather than tetramers.

Expression and spectral characterization of ancestral rhodopsins

In order to assess the proper expression and binding of retinal by the reconstructed ancestral rhodopsins, the focus was placed on Anc-SzR and Anc-HeR, which were expressed them heterologously in *E. coli* under several induction conditions. The expression constructs were designed to use the ASR-derived sequences essentially "as reconstructed". For each protein, the inferred ancestral amino acid sequence under the PSI/TM-Coffee alignment was inserted as a single open reading frame into the multiple cloning site of a standard pET expression vector, retaining only the vector-encoded His_6 affinity tag cassette. Anc-SzR was cloned into pET-21a(+) such that the ASR-derived sequence is followed by a C-terminal His_6 tag, whereas Anc-HeR was cloned into pET-15b so that an N-terminal His_6 tag and thrombin recognition sequence precede the ASR-derived Anc-HeR open reading frame. No additional fusion partners, solubility tags, truncations, or stabilizing point mutations were introduced. Therefore, the observation of rhodopsin-like pigmentation in *E. coli* directly reflects the properties of the sequences produced by our ASR pipeline.

For both Anc-SzR and Anc-HeR, we compared three combinations of temperature and induction time after the addition of IPTG and all-*trans*-retinal: 37 °C for 4 h, 37 °C for 20 h, and 25 °C for 20 h. In all cases, the cell pellets exhibited visible reddish or purple pigmentation, indicating successful expression of retinal-binding proteins. Pellets from the 25 °C, 20 h and 37 °C, 20 h conditions showed more intense coloration than those from 37 °C, 4 h, and the 25 °C, 20 h condition was chosen for subsequent preparations of both Anc-SzR and Anc-HeR. Under this condition, both ancestral constructs yielded clearly pigmented pellets (Figure 7A), consistent with efficient folding and retinal binding in vivo without engineering beyond a simple His_6 tag.

Detergent-solubilized Anc-SzR and Anc-HeR were further characterized by UV–visible absorption spectroscopy (Figure 7B). Anc-SzR showed a major absorption maximum at 549 nm, whereas Anc-HeR showed a λ_{max} at 543 nm. These values fall within the ranges previously observed for extant schizorhodopsins and heliorhodopsins, respectively, indicating that the ancestral reconstructions retain

the characteristic spectral features of each clade. The line shapes of the spectra were also consistent with a retinal chromophore embedded in a well-ordered protein environment, corroborating the structural plausibility suggested by our AlphaFold models.

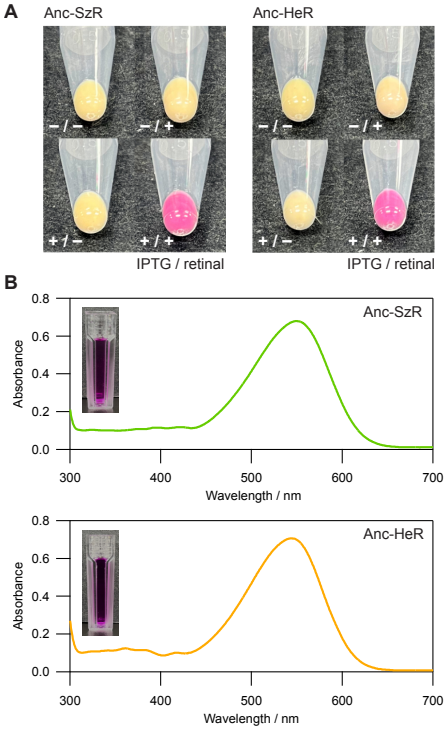

Figure 7: **Expression and spectral properties of Anc-SzR and Anc-HeR.** (A) Whole-cell pigmentation of *E. coli* expressing Anc-SzR or Anc-HeR under four combinations of IPTG induction and all-*trans*-retinal supplementation (−/−, −/+, +/−, +/+). Pronounced coloration is observed only in induced cells supplied with retinal (+/+), consistent with successful expression and chromophore binding of both ancestral proteins. (B) UV–visible absorption spectra of the purified, detergent-solubilized Anc-SzR and Anc-HeR recorded at room temperature after affinity purification. Insets show photographs of the purified samples in cuvettes, illustrating the characteristic visible colors of the ancestral rhodopsins.

Discussion

This study establishes a structure-guided, indel-aware framework for ASR of seven-transmembrane microbial rhodopsins and highlights general lessons for membrane-protein ASR. By combining structure-consistent alignments, profile-based evolutionary models, explicit binary indel modeling and AlphaFold-based structural evaluation, we show that full-length 7TM ancestors can be reconstructed with high statistical and structural confidence while retaining extra-membrane architecture that is usually trimmed away.

Our analyses demonstrate that alignment quality often dominates model choice in determining ancestral stability, especially for HeR clades. Across full and subset datasets, profile models of the Q.pfam+R family were consistently preferred over single-matrix models, with ΔBIC values >200–250 relative to the best non-profile empirical models (e.g. LG+F+R7/8) (Table 1). More complex profile-mixture models such as LG+C60+F+R8 further improved likelihood and BIC, as expected for highly parameterized mixtures, but did not qualitatively change the placement or support of key ancestral nodes: in particular, the Anc-HeR branch remained alignment-sensitive and showed only moderate or unstable UFBoot2 support (Table S2). In contrast, the choice between PSI/TM-Coffee and MAFFT L-INS-i had a strong impact on the local topology of HeR internal nodes, underscoring that alignment quality, rather than model complexity per se, is the dominant factor for ancestral stability in these 7TM datasets. Under PSI/TM-Coffee, the HeR ancestor was recovered with a well-resolved surrounding subtree and retained in the UFBoot2 consensus with moderate support (approximately 50–60%), whereas under L-INS-i the same branch was shortened or collapsed into a near-polytomy despite maximal ultrafast bootstrap values (UFBoot = 100%) on the ML tree. These behaviors imply that the apparent instabilities observed in 7TM phylogenies are largely attributable to subtle alignment artifacts and near-equivalent local rearrangements, rather than being indicative of irreducible evolutionary ambiguity. This highlights the significance of structure-aware alignments for membrane proteins.

The employment of explicit, indel-aware ASR is imperative for the acquisition of realistic full-length ancestors. Without indel modelling, ancestral sequences for both HeRs and SzRs became artificially extended in EM segments, depressing mean PP and degrading structural predictions by introducing elongated, low-confidence tails (Figure 2A). Binary gap–coding strategies similar in spirit have been applied previously to microbial rhodopsins to probe ancient photic niches [29], but those studies focused primarily on TM cores and spectral tuning and did not evaluate EM architecture. By integrating a binary gap model in RAxML with IQ-TREE ASR outputs, we restored sequence lengths that closely match extant crystal structures (e.g. 206 aa for Anc-SzR versus 202 aa for SzR4, and 259 aa for Anc-HeR versus 256 aa for HeR 48C12) and simultaneously improved both PP and pLDDT in the retained residues (Figure 2C). For Anc-SzR in particular, indel-aware refinement transforms an over-extended, low-confidence 401-residue ancestor into a compact 206-residue 7TM model that is nearly indistinguishable in length and overall topology from extant SzR4. This qualitative "before/after" contrast highlights how gap-unaware ASR can generate unrealistically long, structurally implausible ancestors, and how an explicit indel model can recover full-length sequences that are both statistically well supported and structurally coherent. This is particularly important for 7TM proteins, where termini and EM elements contribute to folding, oligomerization and function. Classical ASR studies on G protein-coupled receptors (GPCRs) and other 7TM families have often circumvented these issues by trimming N- and C-termini and most loop regions, effectively restricting reconstruction to the TM core and treating indels implicitly through manual curation of the alignment [30, 31, 32]. Our results argue against such routine aggressive trimming and instead support an indel-aware, full-length strategy in which EM architecture is modeled explicitly and evaluated alongside the TM helices.

More broadly, several authors have noted that many ASR studies provide only limited information about how alignments were curated or how insertions and deletions were handled, even though indel errors can be a major source of reconstruction error [7, 13, 8]. In practice, gaps are often managed implicitly

through manual trimming or the choice of a single alignment method, which makes it difficult to reproduce or systematically compare ASR workflows for proteins with long, heterogeneous EM segments. Our framework addresses this concern by making indel treatment explicit, scriptable and node specific: gaps are first reconstructed under a binary-state indel model, and these inferred ancestral indel patterns are then used to generate indel-aware ancestral sequences that can be re-analyzed, re-aligned or re-scored by any laboratory.

The OPM-guided region-level analyses show that EM segments are not uniformly unreliable. Across ancestors and alignments, TM regions display very high confidence (mean PP typically 82–94 and mean pLDDT 96–98), but EM regions also retain unexpectedly strong support (mean PP 76–88 and mean pLDDT 89–94), with PP×pLDDT values indicating that many residues outside the bilayer are jointly well supported by both sequence and structure (Figure 4B and Table 4). Key architectural features are preserved: in Anc-HeR, the EM segments corresponding to the short α-helix between TM2 and TM3 and the long β-strands between TM1 and TM2 in HeR 48C12 are recovered with strand/helix morphology and high PP and pLDDT; in Anc-SzR, the short β-strands between TM2 and TM3 reported for SzR4 are likewise recapitulated as structured EM elements rather than erased by indel correction (Figure 3A and B). Anc-SH preserves only the SzR-like short β-strands between TM2 and TM3 and lacks the HeR-specific helix/long-β combination, consistent with a gain of these EM motifs along the HeR branch after divergence from Anc-SH. These patterns argue against the common assumption that "loops are noise" in 7TM ASR and instead indicate that, when TM/EM boundaries are defined from experimentally determined orientations, a substantial fraction of EM architecture remains reliably reconstructable.

An important consequence of combining indel-aware ASR with OPM-guided region partitioning is that extra-membrane motifs become traceable evolutionary characters rather than unstructured noise. In the HeR–SzR system, we can follow a specific pattern of gain, loss and relocation of β-strands and short helices between TM1–TM2 and TM2–TM3: Anc-SH carries a single β-strand between

TM2 and TM3, which is retained in Anc-SzR and extant SzR4, whereas the HeR lineage loses this strand, gains a short α-helix between TM2 and TM3 and acquires a new β-strand between TM1 and TM2 in Anc-HeR and HeR 48C12 (Figure 3C). Because these motifs are supported by both high PP and high pLDDT, they represent concrete examples where full-length, indel-aware reconstructions allow the evolutionary history of extra-membrane architecture to be read directly from the tree. This is in contrast to traditional workflows, in which such regions are trimmed away or treated as alignment artefacts. This suggests that loop and peri-membrane elements in other 7TM families may also encode informative evolutionary transitions when reconstructed under a similar framework.

Combining PP and pLDDT elucidates that residue-level and structure-level uncertainties need not coincide, particularly at deep nodes. Across Anc-SzR, Anc-HeR and Anc-SH, PP–pLDDT profiles highlight sites where the three-dimensional fold is predicted with high confidence but the exact residue identity remains ambiguous, and conversely sites where PP is high despite intrinsically flexible local structure (Figure 4B and Table 4). For example, Anc-SH retains a well-formed 7TM fold with high mean pLDDT ($\sim$93–94) even though mean PP is substantially lower than in Anc-SzR and Anc-HeR, especially in TM cores, indicating that structural constraints delimit the fold more tightly than they specify individual side chains (Table 2 and Figure S3). In contrast, some EM segments exhibit high PP but slightly reduced pLDDT, consistent with confidently aligned but conformationally heterogeneous regions. Thus, inspecting PP and pLDDT (and their combinations) is essential for interpreting ASR results: topology and residue probabilities alone can underestimate the recoverability of the fold, whereas structure alone can mask evolutionary ambiguity at the sequence level.

The composite BS–PP–pLDDT metric offers a pragmatic approach to encapsulating node-level reliability without obfuscating the underlying components. By construction, the score is high only when (i) the split is well supported in the UFBoot2 consensus (BS), (ii) residue identities are statistically well de-

termined (PP) and (iii) the resulting structure is confident (pLDDT). In our dataset, Anc-SzR consistently attains the highest composite values, reflecting simultaneously strong UFBoot2 support, high mean PP and very high mean pLDDT (Figure 5B). Anc-HeR and Anc-SH occupy intermediate ranges for different reasons: Anc-HeR combines high PP and pLDDT with composite scores that are highly sensitive to alignment, taking relatively high values under the PSI/TM-Coffee consensus (where the HeR split is retained with moderate frequency) but collapsing to zero under L-INS-i (where the corresponding branch is unresolved), whereas Anc-SH shows consistently maximal UFBoot2 support and good structural coherence but lower mean PP (Figure 5C). These distinctions offer a principled basis for prioritizing ancestors for experimental resurrection (e.g. Anc-SzR as a primary target, Anc-HeR as conditionally reliable, Anc-SH as structurally informative but sequence-wise diffuse) and illustrate how an explicit integrative index can complement, rather than replace, the inspection of BS, PP and pLDDT individually. General-purpose ASR pipelines such as ancseq [33] already emphasize automation and reproducibility for indel-aware ASR; our contribution is to demonstrate how an indel-aware, structure-guided strategy can be tuned specifically for full-length 7TM proteins and coupled to explicit structural reliability metrics. Conceptually, our BS–PP–pLDDT metrics are not intended as a new score in isolation, but as a way to make explicit and repeatable the otherwise qualitative question of which ancestral nodes are trustworthy enough to warrant experimental resurrection.

Multimeric-state predictions reinforce that the reconstructed ancestors are compatible with known oligomerization modes of microbial rhodopsins (Figure 6). Anc-HeR is robustly predicted to form dimers with high ipTM and pTM scores and a plausible interface involving the EM β-strand region between TM1 and TM2, consistent with dimeric assemblies observed for extant HeRs [15, 17]. Anc-SzR is consistently predicted to form trimers with high interface scores, matching the trimeric packing of extant SzRs [18, 19], whereas Anc-SH shows intermediate, trimer-biased interfaces in line with its deeper and more diffuse sequence state. These results indicate that the indel-corrected ancestors not only

form plausible monomeric 7TM folds with high pLDDT but also encode EM elements that support biologically realistic oligomeric interactions.

Importantly, Anc-SzR and Anc-HeR were expressed essentially "as reconstructed": the inferred ancestral sequences were used directly as open reading frames, with only minimal vector-encoded His_6 tags (C-terminal for Anc-SzR in pET-21a(+) and N-terminal for Anc-HeR in pET-15b) and a thrombin site in the case of Anc-HeR. We did not introduce any additional fusion partners or stabilizing mutations that are often required for membrane proteins. The robust production of clearly pigmented cells and rhodopsin-like absorption spectra thus demonstrates that the present ASR pipeline readily yields ancestral seven-transmembrane proteins that are compatible with standard bacterial expression systems.

A comparison with a classical RAxML–PAML workflow suggests that pipeline choice matters most at exactly those nodes that are already difficult. For relatively stable clades such as the SzR ancestor, indel-aware RAxML–PAML reconstructions yielded mean PP and pLDDT values, as well as sequence identities, that closely matched those obtained with the IQ-TREE-based framework. This finding indicates that many existing ASR studies are likely robust in well-supported parts of the tree. In contrast, for the intrinsically unstable HeR ancestor, PAML-based sequences diverged substantially from all IQ-TREE-based ancestors, even though the latter agreed well with each other across alignment and model choices (Table S3). This pattern implies that modern profile-based models and FreeRate heterogeneity, together with structure-guided alignments, can mitigate exactly the cases where traditional pipelines disagree most strongly. The practical implications of our results argue that there is little reason to insist on legacy RAxML–PAML pipelines in new ASR studies: for stable nodes they offer no clear advantage over IQ-TREE, and for fragile nodes their behavior can be more erratic.

Several limitations point to directions for future improvement. The SzR dataset is predominantly composed of metagenome-derived sequences, while most HeRs originate from annotated genomes. Consequently, the apparent EM diver-

sity in SzRs may be attributable to both sampling biases and genuine biological variability. Although AlphaFold predictions (including oligomeric complexes) provide strong evidence of structural plausibility, experimental characterization of resurrected ancestors remains the definitive benchmark for fold and function. The composite BS–PP–pLDDT metric inherits the assumptions and model dependencies of each component and should be interpreted alongside detailed inspection of trees, alignments and structures. In a broader sense, our research has been concentrated on a single 7TM system. However, it will be essential to test the generalizability of these conclusions by applying the same framework to independent 7TM families, including GPCRs, channels, and transporters. Future work will also explore additional structural metrics (e.g. predicted TM-scores or interfacial confidences) in the composite index and integrate our indel-aware workflow into more automated ASR pipelines. This will enable the systematic evaluation of full-length membrane-protein reconstructions, rather than their treatment as special cases.

In practice, many published ASR studies provide only partial descriptions of their alignment curation, trimming rules or reconstruction settings, which makes it difficult to reproduce or systematically compare workflows, especially for complex membrane proteins. One practical contribution of this work is therefore to provide an explicitly documented, scriptable pipeline (ConsistASR: Consistency-aware ASR), including full control files for PAML-based reconstructions, so that each step from alignment to indel-aware ASR and structural evaluation can be replicated or adapted in other systems. Although we focused on seven-transmembrane microbial rhodopsins, nothing in the ConsistASR framework is specific to this family. Any system in which long, heterogeneous regions coexist with more conserved cores—for example, multi-domain enzymes, modular signaling proteins or other membrane proteins with substantial extra-membrane architecture—faces similar challenges of alignment uncertainty, indel placement and uneven structural confidence. A combination of explicit binary indel modeling with structure-based evaluation and reproducible, scriptable workflows has been demonstrated to provide a general template for extending full-length ASR

beyond single-domain soluble proteins and into the multi-domain, multi-state space that many biologically important proteins occupy.

Conclusion

In this study, we combined an indel-aware ASR workflow with AlphaFold-based confidence mapping to reconstruct ancestral 7TM rhodopsins, including not only the transmembrane helices but also the extra-membrane regions. In the case of microbial rhodopsins, functional diversification is often dominated by local changes around the retinal-binding pocket within the TM core, and the extracellular loops may appear secondary at first glance. Nevertheless, our reconstructions suggest that these extra-membrane segments can be inferred with comparable confidence and provide a structurally consistent context for interpreting ancestral photochemistry, oligomerization, and potential protein–protein interactions. More broadly, many eukaryotic 7TM receptors, including GPCRs, rely on extracellular and cytoplasmic loops and C-terminal tails as primary determinants of ligand recognition and G-protein/arrestin coupling. The ability to obtain gap-aware ancestral models that explicitly include these regions therefore opens a route to applying similar ASR strategies to signaling 7TM receptors, where the extra-membrane architecture is a key part of the functional interface. We anticipate that the ConsistASR workflow introduced here will serve as a general starting point for reconstructing and engineering ancestral 7TM proteins in which both the transmembrane core and extra-membrane segments are treated in a physically and statistically consistent manner.

Materials and Methods

Sequence datasets and curation

The sequences of heliorhodopsin (HeR) and schizorhodopsin (SzR) were compiled by combining previously published datasets with additional database searches. Previously characterized HeR and SzR sequences were taken from Inoue et al. and Bulzu et al. [16, 34] and supplemented with homologs identified by

BLAST searches against NCBI protein databases and by profile-based searches of MGnify metagenome assemblies [35]. Because SzR-like sequences are sparsely represented in standard protein databases, most additional SzR candidates were obtained from metagenomic assemblies. Several HeR sequences were directly taken from the metagenome-derived dataset of Bulzu et al. [34]; for these entries, we retained the original contig/gene identifiers when a unique public accession could not be assigned. All amino-acid sequences used in this study are provided as supplementary FASTA files.

Candidate sequences were filtered in four stages. (i) *Length filtering*: sequences shorter than 200 amino acids for HeRs or 150 amino acids for SzRs were discarded to remove obvious fragments and mispredicted multi-pass membrane proteins. (ii) *Functional-residue filtering*: sequences lacking the conserved Schiff-base lysine in TM7 were excluded. (iii) *Redundancy reduction*: highly similar sequences (>95% pairwise identity) were clustered with CD-HIT [36], and a single representative per cluster was retained. (iv) *Phylogenetic diversity filtering*: Treemmer [37] was applied to a preliminary tree to downsample near-identical sequences while preserving long internal branches.

After these automated filters, predicted topologies and hydropathy profiles were visually inspected to confirm the presence of seven transmembrane (7TM) segments, and sequences lacking a clear 7TM architecture or showing obvious misannotation were removed. The final non-redundant dataset comprised 228 sequences, including 115 HeRs, 110 SzRs, and 3 outgroup microbial rhodopsins. Two reduced datasets, HeR+outgroup and SzR+outgroup, were derived from this set, sharing the same outgroup panel. Sequence lengths ranged from approximately 195 to 305 residues, consistent with the known 7TM microbial rhodopsin architectures.

Multiple sequence alignment

Multiple sequence alignments were generated for the full HeR+SzR+outgroup dataset and for the two subset datasets (HeR+outgroup and SzR+outgroup) using PSI/TM-Coffee and MAFFT [20, 21]. Unless otherwise stated, default

parameters of the respective programs or web servers were used.

Structure-aware alignments were obtained with PSI/TM-Coffee in transmembrane mode using UniRef100 as the homologous sequence database. In this mode, predicted transmembrane segments and additional homologous sequences are explicitly used in the consistency-based alignment procedure, thereby stabilizing TM/EM boundaries. The resulting alignments were exported in full-length FASTA format without manual editing of internal regions.

As a non–structure-guided baseline, we generated MAFFT alignments using the L-INS-i strategy, which incorporates local pairwise alignment information and is recommended for accuracy on medium-sized protein families [21]. L-INS-i alignments were produced via the MAFFT web server with default settings and served as a widely used reference for comparison to PSI/TM-Coffee.

For exploratory analyses and for a "classical pipeline" comparison, additional alignments were generated with the faster FFT-NS-1 strategy in MAFFT. These FFT-NS-1 alignments exhibited discernible mismatches at the TM/EM boundaries and frequently resulted in inflated ancestral sequence lengths and depressed PPs in certain HeR reconstructions. Consequently, they were employed exclusively in sensitivity analyses and were not incorporated into the primary indel-aware ASR pipeline.

To examine the influence of mild gap-based trimming, we also produced trimmed versions of each alignment using trimAl with the `gappyout` option [38]. These trimmed alignments were used solely to assess robustness of model selection and topology. All ASR, indel-aware reconstruction and AlphaFold-based structural evaluations were performed on the untrimmed, full-length alignments, and no aggressive manual trimming of loop or terminal regions was applied. The alignment-level handling of EM segments was instead delegated to the indel-aware reconstruction and OPM-guided region partitioning described below.

Phylogenetic inference and model selection

Maximum-likelihood (ML) phylogenies were inferred with IQ-TREE v3.0.1 [23] for the HeR+SzR+outgroup dataset and for the HeR+outgroup and SzR+outgroup

subsets. The PSI/TM-Coffee and MAFFT L-INS-i alignments were analyzed as single-partition protein datasets.

For each alignment, we used ModelFinder [22] (`-m MFP`) to compare empirical matrices (e.g. LG, WAG, JTT) with various rate-heterogeneity schemes and profile families including Q.pfam and related models. Model fit was assessed using the Bayesian information criterion (BIC), and the model with the lowest BIC was selected. Across the full and subset datasets, Q.pfam models with FreeRate heterogeneity (e.g. Q.pfam+R7) were consistently preferred over LG-based models by substantial BIC margins (see Results and Table 1). In addition, ModelFinder was employed as a sensitivity check for moderately trimmed alignments (trimAl `gappyout`). Unless indicated otherwise, ML trees utilized for ASR and structural analyses were re-estimated on untrimmed alignments under the best-fit Q.pfam+R model.

ML tree searches were run in IQ-TREE with ultrafast bootstrap approximation (1,000 replicates) and SH-aLRT support (1,000 replicates) [24]. Three outgroup rhodopsins were specified to root the tree. The best-scoring ML tree (`.treefile`) with SH-aLRT and ultrafast bootstrap support values was used for visualization and as the fixed topology in downstream analyses, while the UF-Boot2 consensus tree (`.contree`) was used to determine whether focal branches were resolved or collapsed and to obtain the branch support values entering the composite reliability metric.

For comparison with more parameter-rich empirical profile-mixture models, additional trees were inferred under LG+C60+F+R8 using otherwise identical settings. These runs served as a benchmark for the impact of model complexity on internal-node stability relative to Q.pfam+R. However, Q.pfam+R trees were adopted as the primary evolutionary framework for ASR.

Ancestral sequence reconstruction in IQ-TREE

Ancestral sequence reconstruction (ASR) was performed in IQ-TREE v3.0.1 on the ML topologies inferred under the best-fit Q.pfam+R models. For each dataset (HeR+SzR+outgroup, HeR+outgroup, SzR+outgroup) and each align-

ment (PSI/TM-Coffee, MAFFT L-INS-i), marginal ASR carried out at all internal nodes using the `--ancestral` option, treating the topology as fixed.

IQ-TREE produces, for each internal node and alignment column, the posterior probability (PP) for all 20 amino-acid states and the corresponding maximum-posterior state. These values were extracted from the `.state` files. For each focal ancestor, the most probable residue and its PP were recorded at every column and used for (i) numerical summaries (mean PP, PP×pLDDT, PP−pLDDT) and (ii) mapping onto AlphaFold models. Extant sequences (tips) were not altered by ASR and served as references for sequence identity and length comparisons.

We focused on three internal nodes in the HeR+SzR+outgroup tree: the SzR ancestor (Anc-SzR), the HeR ancestor (Anc-HeR) and their common ancestor (Anc-SH). ASR for these nodes was performed under the Q.pfam+R model selected for each alignment. Sensitivity analyses under LG+C60+F+R8 were also run on the same alignments and topologies to assess model-dependent variation in residue identities and PP distributions at these nodes (see Results).

The unmasked IQ-TREE ancestral sequences and PPs were then passed to the indel-aware refinement pipeline described below. It should be noted that all PP-based statistics and AlphaFold evaluations reported in this study refer to the indel-corrected ancestral sequences obtained after binary gap-based masking.

Indel-aware reconstruction using binary gap models

To make indel treatment explicit and node specific, we implemented a two-stage indel-aware refinement pipeline inspired by previous binary gap coding approaches [25]. In this workflow, IQ-TREE is used exclusively for amino-acid ASR, and RAxML-HPC v8.2.12 [39] is used exclusively for binary gap inference on the same fixed topology.

First, for each protein alignment (PSI/TM-Coffee or MAFFT L-INS-i), we generated a binary gap matrix by recoding each column as 1 (residue present) or 0 (gap). This binary alignment retains the per-taxon gap pattern of the original multiple sequence alignment while ignoring amino-acid identity. RAxML-HPC

was then employed to infer ancestral gap states under a two-state (gap/non-gap) model on the fixed IQ-TREE ML topology. The resulting output provides, for each internal node and alignment column, a binary ancestral gap call indicating whether a residue is present or absent at that position.

Second, we mapped these binary gap reconstructions back onto the IQ-TREE amino-acid ASR output. Because IQ-TREE and RAxML assign different internal node labels, we implemented a bipartition-based node-mapping procedure to ensure a one-to-one correspondence between IQ-TREE nodes (including Anc-SzR, Anc-HeR and Anc-SH) and RAxML reconstructions. For each ancestor and column, if the binary reconstruction indicated a gap state (0), the corresponding position in the IQ-TREE ASR sequence was masked and removed; if a residue state (1) was inferred, the IQ-TREE maximum-posterior amino acid and its PP were retained. Indel-corrected ancestral sequences were then assembled by concatenating all unmasked positions in their original alignment order.

This node-specific masking contracts over-extended EM segments and termini and removes sites where the presence or absence of a residue is not supported by the indel model. All ancestral sequences used for downstream analyses (sequence identity comparisons, region-level PP and pLDDT summaries, composite reliability scores, and structural modeling) correspond to these indel-corrected sequences. Gap recoding, RAxML parsing and node mapping were implemented in custom Python scripts, which together with example command lines are provided in the ConsistASR pipeline (refer to the Data availability section).

AlphaFold structure prediction and confidence metrics

Predicted structures for ancestral and extant rhodopsins were obtained using an AlphaFold 3-based web server [40]. Both uncorrected and indel-corrected amino-acid sequences were submitted in monomer mode; five models were generated per sequence, and the top-ranked model by mean predicted Local Distance Difference Test (pLDDT) was used in downstream analyses.

Monomeric models were calculated for Anc-SzR, Anc-HeR and Anc-SH. Uncorrected ancestral sequences were utilized to evaluate the impact of indel

treatment on overall fold quality, whereas indel-corrected sequences were employed for all quantitative summaries and comparisons to crystal structures. Per-residue pLDDT values were extracted from the B-factor field of the output coordinate files, and analyses were restricted to C_α atoms, yielding a single pLDDT value per residue. These pLDDT values were aligned to the corresponding PPs from IQ-TREE ASR using alignment coordinates, resulting in paired PP and pLDDT values for each residue retained in the indel-corrected ancestor.

To evaluate oligomeric states, Anc-SzR, Anc-HeR and Anc-SH were also modeled as homomeric complexes with stoichiometries of 1 (monomer control), 2 (dimer), 3 (trimer) and 5 (pentamer) using the same AlphaFold 3-based server in multimer mode. The stoichiometries of 1, 2, 3, and 5 were the focal point of this study due to the experimental observations of microbial rhodopsins forming dimers, trimers, or pentamers. However, there is a paucity of reports on the formation of tetramers. For each stoichiometry and ancestor, we recorded the predicted TM-score (pTM) of the overall complex and the interface TM-score (ipTM) for the top-ranked model, which together quantify fold accuracy and interface quality. Structural inspection (e.g. TM packing, EM β-strands and short helices) was performed in UCSF ChimeraX and PyMOL and qualitatively compared with the SzR4 and HeR 48C12 crystal structures.

Definition of TM and EM regions using OPM

To quantify region-specific reliability, the ancestral model was partitioned into transmembrane (TM) and extra-membrane (EM) regions using membrane embeddings from the Orientations of Proteins in Membranes (OPM) database [26]. For SzR, we used the OPM entry for SzR4 (PDB 7e4g). For HeR, we used the OPM entry corresponding to HeR 48C12 (PDB ID 6uh3) to define TM/EM boundaries. A higher-resolution structure of the same protein (PDB ID 6su3) was used only for illustrative structural overlays in Figure 3; all TM/EM assignments and region-level statistics are based on the OPM entry for 6uh3. For each template, residue-wise membrane location (bilayer-embedded versus solvent-exposed) and the boundaries of the seven canonical TM segments were extracted.

Residues lying within the hydrophobic core of the bilayer were classified as TM, and all remaining residues were classified as EM. In this convention, EM includes N- and C-terminal tails, peri-membrane helix termini, β-strands, short coils and amphipathic helices outside the membrane core.

These annotations were transferred to the reconstructed ancestors by aligning Anc-SzR to SzR4 and Anc-HeR to HeR 48C12 using the same structure-consistent alignments that underlie the ASR analyses (PSI/TM-Coffee or MAFFT L-INS-i). The TM/EM labels were propagated column-wise from the extant sequence to the ancestral sequence. Sites that aligned to gaps in the ancestor or were removed by indel-aware masking were excluded from regional statistics. Utilizing these labels, we proceeded to compute region-level means for PP, pLDDT, PP$\times$pLDDT and PP$-$pLDDT for TM and EM residues, and, where indicated, for individual TM helices and flanking EM segments.

Composite reliability metric

To integrate evolutionary, statistical and structural information at the node level, we defined a composite reliability metric for Anc-SzR, Anc-HeR and Anc-SH. For each ancestor, site-wise marginal PPs were extracted from IQ-TREE `.state` files, mapped onto the indel-corrected sequence using the same node-mapping and masking procedure as for the binary gap model, and paired with pLDDT values from the corresponding AlphaFold model.

Let PP_i and pLDDT_i denote the posterior probability and normalized pLDDT (0–1) for residue i, and let N be the number of C_α sites considered. Node-level summary statistics were defined as

$$\overline{\mathrm{PP}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{PP}_i, \quad \overline{\mathrm{pLDDT}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{pLDDT}_i,$$

$$\overline{\mathrm{PP}\times\mathrm{pLDDT}} = \frac{1}{N}\sum_{i=1}^{N}\big(\mathrm{PP}_i\times\mathrm{pLDDT}_i\big), \quad \overline{\mathrm{PP}-\mathrm{pLDDT}} = \frac{1}{N}\sum_{i=1}^{N}\big(\mathrm{PP}_i-\mathrm{pLDDT}_i\big).$$

In the Results, these quantities are reported on a 0–100 scale by multiplying by 100.

Bootstrap support (BS) for each ancestor was taken from the IQ-TREE UFBoot2 consensus tree (`.contree`) under the corresponding alignment and model. BS was defined as the fraction of bootstrap trees containing the relevant split, expressed as $b = \mathrm{BS}/100$ on a 0–1 scale. When a branch was collapsed in the consensus (i.e. the split was not recovered above the consensus threshold), b was set to 0.

The composite node-level reliability score was then defined as

$$\mathrm{CompositeScore}_{0\text{–}100} = 100 \times b \times \overline{\mathrm{PP} \times \mathrm{pLDDT}}.$$

This formulation first merges sequence-based and structure-based confidence into a single consistency index $\overline{\mathrm{PP} \times \mathrm{pLDDT}}$ and then modulates it by the stability of the corresponding split (b). Composite scores were computed separately for each alignment/model combination; values in the main text correspond to the baseline condition (PSI/TM-Coffee with Q.pfam+R7), and additional combinations are summarized in Supplementary tables.

Expression, purification and spectroscopy of ancestral rhodopsins

Genes encoding Anc-SzR and Anc-HeR were synthesized as contiguous open reading frames corresponding exactly to the inferred amino acid sequences (GenScript), without additional sequence optimization or stabilizing mutations. For Anc-SzR, the gene was cloned into the NdeI and XhoI sites of pET-21a(+), placing the ASR-derived open reading frame in frame with the vector-encoded C-terminal His_6 tag. For Anc-HeR, the gene was cloned into the NdeI and BamHI sites of pET-15b, such that the vector-encoded N-terminal His_6 tag and thrombin recognition sequence precede the ASR-derived open reading frame and the C terminus corresponds exactly to the inferred ancestral sequence. For both Anc-SzR and Anc-HeR, the initiator methionine is encoded by the ATG within the NdeI site (CATATG) and is counted as the first residue of the ASR-derived sequence.

Recombinant Anc-SzR and Anc-HeR were expressed in *E. coli* C43(DE3) harboring the corresponding plasmids. Cultures were grown in LB medium

954 supplemented with ampicillin at 37 °C, and expression was induced at an OD_{600}
955 of approximately 1.0 by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside
956 (IPTG) together with 10 μM all-*trans*-retinal from an ethanol stock. After
957 induction, cultures were incubated for 20 h at 25 °C with shaking, a condition
958 that yielded strongly pigmented cell pellets for both ancestors (see Results). For
959 qualitative comparison of expression conditions, additional cultures were induced
960 at 37 °C for 4 or 20 h.

961 For purification, cells were harvested by centrifugation, resuspended in buffer,
962 and disrupted by sonication. Membrane fractions were collected by ultracen-
963 trifugation, resuspended in detergent-containing buffer, and solubilized with
964 n-dodecyl-β-D-maltoside (DDM). Solubilized pigments were purified by Co^{2+}-
965 affinity chromatography and eluted in imidazole-containing buffer with DDM.
966 Eluted fractions were exchanged into a Tris/NaCl buffer containing DDM for
967 spectroscopic measurements. The purity of the ancestral proteins were confirmed
968 using SDS-PAGE analysis. As is frequently observed for highly hydrophobic
969 multi-pass membrane proteins, both Anc-SzR and Anc-HeR exhibited anoma-
970 lously low apparent molecular weights by SDS–PAGE in comparison with their
971 calculated masses, despite migrating as single, well-resolved bands (Figure S4).
972 Compositions of buffers, conditions for centrifugation, and chromatography
973 gradients are provided in the Supplementary information.

974 UV–visible absorption spectra of detergent-solubilized Anc-SzR and Anc-
975 HeR were recorded at room temperature using a quartz cuvette with a 10 mm
976 path length. Spectra were collected over the range of 300–700 nm, with the
977 corresponding buffer used for baseline correction.

978 *Classical RAxML+PAML workflow for comparison*

979 To establish a classical point of reference, we also performed ancestral re-
980 construction using a RAxML+PAML pipeline on the same 7TM alignments.
981 ML trees were first inferred with RAxML-NG v1.2.2 [41] under an empirical
982 amino-acid model with among-site rate heterogeneity (typically LG+F+G4).
983 These topologies were then supplied to codeml (PAML v4.10.7) [42] for marginal

ancestral reconstruction under a fixed LG rate matrix with gamma-distributed rate variation (four categories). Tree topologies were held fixed, and marginal ancestral states were written to the `rst` file. Full codeml control files, including all parameter settings, are provided as Supplementary data to facilitate exact reproducibility.

Software and reproducibility

Sequence curation used CD-HIT [36] and Treemmer [37]. Multiple sequence alignments were generated with MAFFT (L-INS-i and FFT-NS-1 strategies) [21] and PSI/TM-Coffee in transmembrane mode via the T-Coffee web server [20]. Alignment trimming used trimAl [38]. ML phylogenies, model selection and ASR were performed with IQ-TREE v3.0.1 [23]; binary gap reconstructions used RAxML-HPC v8.2.12 [39]; classical ASR comparisons used RAxML-NG v1.2.2 [41] and PAML v4.10.7 [42]. Structural predictions were obtained from an AlphaFold 3 web interface [40], and structural visualization and figure preparation were carried out in UCSF ChimeraX [43] and PyMOL (Schrödinger, LLC).

Custom Python 3 scripts were written to (i) convert alignments to binary gap matrices, (ii) map RAxML ancestral gap states onto IQ-TREE `.state` files, (iii) generate indel-corrected ancestral sequences, and (iv) compute node-level summary statistics (PP, pLDDT, PP$\times$pLDDT, PP$-$pLDDT, composite scores). These scripts, together with configuration files and example command lines, are distributed as the ConsistASR pipeline in a version-controlled repository and as an archival snapshot (links provided in the Data availability section). An archive of the full workflow (alignment, tree inference, binary indel reconstruction, ASR and structural evaluation) are provided as Supplementary data.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

The authors used a generative AI assistant (ChatGPT, GPT-5.1 Thinking; OpenAI) to assist with scriptwriting and to edit parts of the manuscript. (e.g.

language polishing and restructuring of paragraphs). All scientific ideas, analyses, and conclusions are the authors' own, and all AI-assisted text was carefully checked and, where necessary, revised by the authors.

Acknowledgments

The authors would like to thank Taito Urui and Yu Iritani for helpful discussions and technical assistance. This research was supported by a Grants-in-Aid for Scientific Research to Y.M. (No. JP23H00285) .

Data availability

Supplementary data contains the sequence, alignment, phylogenetic, ancestral sequence, and structural data underlying the analyses in this manuscript, which are also available at Zenodo (DOI: doi.org/10.5281/zenodo.18080157). ConsistASR v1.0.1 is available at GitHub (https://github.com/harutois/ConsistASR) and archived at Zenodo (DOI: doi.org/10.5281/zenodo.18080327).

Tables

Table 1: Summary of datasets, alignments, and best-fit substitution models used for phylogenetic inference and ancestral sequence reconstruction.

Dataset	Alignment strategy	N sequences	Alignment length (residues)	Best-fit model (BIC)	Notes
HeR+SzR+outgroup	PSI/TM-Coffee	228	449	Q.pfam+R7	Baseline dataset used for joint HeR–SzR phylogeny and ASR. Outgroup sequences were constrained as the external clade.
HeR+SzR+outgroup	MAFFT L-INS-i	228	421	Q.pfam+R7	Alternative alignment used to assess alignment sensitivity under the same model family.
HeR+outgroup	PSI/TM-Coffee	118	533	Q.pfam+R6	HeR-only subtree used for node-level analyses of Anc-HeR.
HeR+outgroup	MAFFT L-INS-i	118	401	LG+F+R6	Alignment/model combination that yielded a near-polytomy around the HeR ancestor in UFBoot2 consensus trees.
SzR+outgroup	PSI/TM-Coffee	113	320	Q.pfam+R5	SzR-only subtree; Anc-SzR showed high and stable bootstrap support.
SzR+outgroup	MAFFT L-INS-i	113	287	Q.pfam+R6	Alignment used to confirm robustness of Anc-SzR to alignment strategy.

Best-fit models are those selected by ModelFinder under the Bayesian information criterion (BIC).

Table 2: Composite node-level reliability metrics integrating posterior probabilities (PP), AlphaFold confidence (pLDDT), and ultrafast bootstrap support (UFBoot) for key ancestral nodes.

Alignment	Ancestor	ML UFBoot (%)	Consensus UFBoot2 (%)	Mean PP	Mean pLDDT	Mean (PP×pLDDT)	Composite score (0–100)
PSI/TM-Coffee	Anc-SzR	99	99	81.6	95.5	78.1	77.3
	Anc-HeR	100	54	90.6	95.9	87.2	47.1
	Anc-SH	100	100	54.9	93.2	51.6	51.6
MAFFT L-INS-i	Anc-SzR	96	96	80.4	94.3	76.1	73.0
	Anc-HeR	100	0[a]	91.4	93.6	86.0	0.0[a]
	Anc-SH	100	100	57.5	93.6	54.3	54.3

Mean PP, mean pLDDT, and mean PP×pLDDT are reported on a 0–100 scale. Composite scores were computed as $100 \times b \times \overline{\mathrm{PP} \times \mathrm{pLDDT}}$, where b is the consensus UFBoot2 support expressed as a fraction ($b = \mathrm{UFBoot2}/100$) and $\overline{\mathrm{PP} \times \mathrm{pLDDT}}$ is the mean product on a 0–1 scale.

[a] For MAFFT L-INS-i, the HeR ancestral branch collapses in the UFBoot2 consensus tree (polytomy), so UFBoot2 support is effectively 0 and the composite score is 0 by definition for this alignment.

Table 3: Global impact of indel-aware reconstruction on ancestral node statistics. Mean posterior probabilities (PP), AlphaFold confidence (pLDDT), and sequence length before and after binary gap-based masking.

Alignment	Ancestor	Before indel-aware masking			After indel-aware masking			Extant length[a]
		Mean PP	Mean pLDDT	Length (residues)	Mean PP	Mean pLDDT	Length (residues)	(aa)
PSI/TM-Coffee	Anc-SzR	51.5	62.3	401	81.6	95.5	206	202
	Anc-HeR	65.8	69.2	434	90.6	95.9	259	256
	Anc-SH	37.6	61.3	408	54.9	93.2	206	–
MAFFT L-INS-i	Anc-SzR	52.4	63.6	386	80.4	94.3	203	202
	Anc-HeR	70.4	75.5	414	91.4	93.6	260	256
	Anc-SH	39.8	63.1	396	57.5	93.6	203	–

Mean PP and mean pLDDT are reported on a 0–100 scale and computed over all Cα positions retained (or predicted) in the corresponding full-length ancestral models, before and after indel-aware masking. [a] Extant lengths correspond to SzR4 (202 residues) for Anc-SzR and HeR 48C12 (256 residues) for Anc-HeR. No direct extant structural reference is available for Anc-SH.

Table 4: Region-level reliability of ancestral 7TM rhodopsins. Mean posterior probabilities (PP), AlphaFold confidence (pLDDT), and derived indices for transmembrane (TM) and extra-membrane (EM) segments.

Alignment	Ancestor	Region	Mean PP	Mean pLDDT	Mean (PP×pLDDT)	Mean (PP−pLDDT)
PSI/TM-Coffee	Anc-SzR	TM	83.9	97.4	81.7	43.2
		EM	75.9	90.7	69.3	42.6
	Anc-HeR	TM	92.5	97.6	90.3	47.4
		EM	87.5	93.7	82.5	46.9
MAFFT L-INS-i	Anc-SzR	TM	82.1	96.2	79.2	42.9
		EM	76.2	89.3	68.0	43.5
	Anc-HeR	TM	93.6	95.5	89.5	49.1
		EM	88.0	91.3	81.0	48.3

TM and EM regions were defined by mapping OPM-derived membrane boundaries for SzR4 and HeR 48C12 onto Anc-SzR and Anc-HeR (see Methods). Values are means over Cα positions retained after indel-aware masking.

References

[1] J. W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules, Nat. Rev. Genet. 5 (2004) 366–375. doi:https://doi.org/10.1038/nrg1324.

[2] G. K. Hochberg, J. W. Thornton, Reconstructing ancient proteins to understand the causes of structure and function, Annu. Rev. Biophys. 46 (2017) 247–269. doi:https://doi.org/10.1146/annurev-biophys-070816-033631.

[3] F. F. Nascimento, M. d. Reis, Z. Yang, A biologist's guide to bayesian phylogenetic analysis, Nat. Ecol. Evol. 1 (2017) 1446–1454. doi:https://doi.org/10.1038/s41559-017-0280-x.

[4] H. Ashkenazy, I. Sela, E. Levy Karin, G. Landan, T. Pupko, Multiple sequence alignment averaging improves phylogeny reconstruction, Syst. Biol. 68 (2019) 117–130. doi:https://doi.org/10.1093/sysbio/syy036.

[5] P. D. Williams, D. D. Pollock, B. P. Blackburne, R. A. Goldstein, Assessing the accuracy of ancestral protein reconstruction methods, PLoS Comput. Biol. 2 (2006) e69. doi:https://doi.org/10.1371/journal.pcbi.0020069.

[6] R. Del Amparo, M. Arenas, Consequences of substitution model selection on protein ancestral sequence reconstruction, Mol. Biol. Evol. 39 (2022) msac144. doi:https://doi.org/10.1093/molbev/msac144.

[7] A. Loytynoja, N. Goldman, Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis, Science 320 (2008) 1632–1635. doi:https://doi.org/10.1126/science.1158395.

[8] C. Iglhaut, J. Pečerska, M. Gil, M. Anisimova, Please mind the gap: indel-aware parsimony for fast and accurate ancestral sequence reconstruction and multiple sequence alignment including long indels, Mol. Biol. Evol. 41 (2024) msae109. doi:https://doi.org/10.1093/molbev/msae109.

[9] K. L. Pierce, R. T. Premont, R. J. Lefkowitz, Seven-transmembrane receptors, Nat. Rev. Mol. Cell Biol. 3 (2002) 639–650. doi:https://doi.org/10.1038/nrm908.

[10] S. Lu, X. He, Z. Yang, Z. Chai, S. Zhou, J. Wang, A. U. Rehman, D. Ni, J. Pu, J. Sun, et al., Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design, Nature Commun. 12 (2021) 4721. doi:https://doi.org/10.1038/s41467-021-25020-9.

[11] M. Punta, L. R. Forrest, H. Bigelow, A. Kernytsky, J. Liu, B. Rost, Membrane protein prediction methods, Methods 41 (2007) 460–474. doi:https://doi.org/10.1016/j.ymeth.2006.07.026.

[12] A. J. Situ, T. S. Ulmer, Universal principles of membrane protein assembly, composition and evolution, PLoS One 14 (2019) e0221372. doi:https://doi.org/10.1371/journal.pone.0221372.

[13] K. Aadland, B. Kolaczkowski, Alignment-integrated reconstruction of ancestral sequences improves accuracy, Genome Biol. Evol. 12 (2020) 1549–1565. doi:https://doi.org/10.1093/gbe/evaa164.

[14] A. Pushkarev, K. Inoue, S. Larom, J. Flores-Uribe, M. Singh, M. Konno, S. Tomida, S. Ito, R. Nakamura, S. P. Tsunoda, et al., A distinct abundant group of microbial rhodopsins discovered using functional metagenomics, Nature 558 (2018) 595–599. doi:https://doi.org/10.1038/s41586-018-0225-9.

[15] W. Shihoya, K. Inoue, M. Singh, M. Konno, S. Hososhima, K. Yamashita, K. Ikeda, A. Higuchi, T. Izume, S. Okazaki, et al., Crystal structure of heliorhodopsin, Nature 574 (2019) 132–136. doi:https://doi.org/10.1038/s41586-019-1604-6.

[16] K. Inoue, S. P. Tsunoda, M. Singh, S. Tomida, S. Hososhima, M. Konno, R. Nakamura, H. Watanabe, P.-A. Bulzu, H. L. Banciu, et al., Schizorhodopsins: A family of rhodopsins from Asgard archaea that

function as light-driven inward H^+ pumps, Sci. Adv. 6 (2020) eaaz2441. doi:https://doi.org/10.1126/sciadv.aaz2441.

[17] K. Kovalev, D. Volkov, R. Astashkin, A. Alekseev, I. Gushchin, J. M. Haro-Moreno, I. Chizhov, S. Siletsky, M. Mamedov, A. Rogachev, et al., High-resolution structural insights into the heliorhodopsin family, Proc. Natl. Acad. Sci. 117 (2020) 4131–4141. doi:https://doi.org/10.1073/pnas.1915888117.

[18] A. Higuchi, W. Shihoya, M. Konno, T. Ikuta, H. Kandori, K. Inoue, O. Nureki, Crystal structure of schizorhodopsin reveals mechanism of inward proton pumping, Proc. Natl. Acad. Sci. 118 (2021) e2016328118. doi:https://doi.org/10.1073/pnas.2016328118.

[19] J. E. Besaw, S. Peng, A. Kuo, J. Reichenwallner, R. D. Miller, L. S. Brown, O. P. Ernst, Structural insights into an inward proton pumping rhodopsin, Biophys. J. (2025). doi:https://doi.org/10.1016/j.bpj.2025.11.011.

[20] P. Di Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J.-M. Chang, J.-F. Taly, C. Notredame, T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, Nucleic Acids Res. 39 (2011) W13–W17. doi:https://doi.org/10.1093/nar/gkr245.

[21] K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780. doi:https://doi.org/10.1093/molbev/mst010.

[22] S. Kalyaanamoorthy, B. Q. Minh, T. K. Wong, A. Von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (2017) 587–589. doi:https://doi.org/10.1038/nmeth.4285.

[23] T. K. Wong, N. Ly-Trong, H. Ren, H. Baños, A. J. Roger, E. Susko, C. Bielow, N. De Maio, N. Goldman, M. W. Hahn, et al., IQ-TREE

3: Phylogenomic inference software using complex evolutionary models, Submitted (2025). doi:https://doi.org/10.32942/X2P62N.

[24] D. T. Hoang, O. Chernomor, A. Von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2018) 518–522. doi:https://doi.org/10.1093/molbev/msx281.

[25] K. Aadland, C. Pugh, B. Kolaczkowski, High-throughput reconstruction of ancestral protein sequence, structure, and molecular function, in: Computational Methods in Protein Evolution, Springer, 2018, pp. 135–170. doi:https://doi.org/10.1007/978-1-4939-8736-8_8.

[26] M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes, Nucleic Acids Res. 40 (2012) D370–D376. doi:https://doi.org/10.1093/nar/gkr703.

[27] M. Shibata, K. Inoue, K. Ikeda, M. Konno, M. Singh, C. Kataoka, R. Abe-Yoshizumi, H. Kandori, T. Uchihashi, Oligomeric states of microbial rhodopsins determined by high-speed atomic force microscopy and circular dichroic spectroscopy, Sci. Rep. 8 (2018) 8262. doi:https://doi.org/10.1038/s41598-018-26606-y.

[28] L. J. Galindo, S. Takaramoto, T. Nagata, A. Rozenberg, H. Takahashi, O. Béjà, K. Inoue, Apusomonad rhodopsins: A new family of ultraviolet to blue light–absorbing rhodopsin channels, Proc. Natl. Acad. Sci. 122 (2025) e2510619122. doi:https://doi.org/10.1073/pnas.2510619122.

[29] C. D. Sephus, E. Fer, A. K. Garcia, Z. R. Adam, E. W. Schwieterman, B. Kacar, Earliest photic zone niches probed by ancestral microbial rhodopsins, Mol. Biol. Evol. 39 (2022) msac100. doi:https://doi.org/10.1093/molbev/msac100.

[30] B. S. Chang, K. Jönsson, M. A. Kazmi, M. J. Donoghue, T. P. Sakmar, Recreating a functional ancestral archosaur

visual pigment, Mol. Biol. Evol. 19 (2002) 1483–1489. doi:https://doi.org/10.1093/oxfordjournals.molbev.a004211.

[31] D. Kuang, Y. Yao, D. MacLean, M. Wang, D. R. Hampson, B. S. Chang, Ancestral reconstruction of the ligand-binding pocket of Family CG protein-coupled receptors, Proc. Natl. Acad. Sci. 103 (2006) 14050–14055. doi:https://doi.org/10.1073/pnas.0604717103.

[32] G. Blanchet, D. Alili, A. Protte, G. Upert, N. Gilles, L. Tepshi, E. A. Stura, G. Mourier, D. Servent, Ancestral protein resurrection and engineering opportunities of the mamba aminergic toxins, Sci. Rep. 7 (2017) 2701. doi:https://doi.org/10.1038/s41598-017-02953-0.

[33] Y. Sugihara, J. Kourelis, M. P. Contreras, H. Pai, A. Harant, M. Selvaraj, A. Toghani, C. Martínez-Anaya, S. Kamoun, Helper NLR immune protein NRC3 evolved to evade inhibition by a cyst nematode virulence effector, PLoS Genet. 21 (2025) e1011653. doi:https://doi.org/10.1371/journal.pgen.1011653.

[34] P.-A. Bulzu, V. S. Kavagutti, A.-S. Andrei, R. Ghai, The evolutionary kaleidoscope of rhodopsins, mSystems 7 (2022) e00405–22. doi:https://doi.org/10.1128/msystems.00405-22.

[35] L. Richardson, B. Allen, G. Baldi, M. Beracochea, M. L. Bileschi, T. Burdett, J. Burgin, J. Caballero-Pérez, G. Cochrane, L. J. Colwell, et al., MGnify: the microbiome sequence data analysis resource in 2023, Nucleic Acids Res. 51 (2023) D753–D759. doi:https://doi.org/10.1093/nar/gkac1080.

[36] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics 22 (2006) 1658–1659. doi:https://doi.org/10.1093/bioinformatics/btl158.

[37] F. Menardo, C. Loiseau, D. Brites, M. Coscolla, S. M. Gygli, L. K. Rutaihwa, A. Trauner, C. Beisel, S. Borrell, S. Gagneux, Treemmer: a tool to reduce

large phylogenetic datasets with minimal loss of diversity, BMC Bioinform. 19 (2018) 164. doi:https://doi.org/10.1186/s12859-018-2164-8.

[38] S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinformatics 25 (2009) 1972–1973. doi:https://doi.org/10.1093/bioinformatics/btp348.

[39] A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, Bioinformatics 22 (2006) 2688–2690. doi:https://doi.org/10.1093/bioinformatics/btl446.

[40] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature 630 (2024) 493–500. doi:https://doi.org/10.1038/s41586-024-07487-w.

[41] A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference, Bioinformatics 35 (2019) 4453–4455. doi:https://doi.org/10.1093/bioinformatics/btz305.

[42] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591. doi:https://doi.org/10.1093/molbev/msm088.

[43] E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Tools for structure building and analysis, Protein Sci. 32 (2023) e4792. doi:https://doi.org/10.1002/pro.4792.