Structure-based insights into evolution of rhodopsins

Dmitrii Zabelskii\textsuperscript{1,2,3,15}, Natalia Dmitrieva\textsuperscript{4,15}, Oleksandr Volkov\textsuperscript{1,2,15}, Vitaly Shevchenko\textsuperscript{3,5,15}, Kirill Kovalev\textsuperscript{3,6,7}, Taras Balandin\textsuperscript{1,2}, Dmytro Soloviov\textsuperscript{3,8,9}, Roman Astashkin\textsuperscript{3,6}, Egor Zinovev\textsuperscript{1,2,3,5,6}, Alexey Alekseev\textsuperscript{1,2,3,5}, Ekaterina Round\textsuperscript{1,10}, Vitaly Polovinkin\textsuperscript{1,6,11}, Igor Chizhov\textsuperscript{12}, Andrey Rogachev\textsuperscript{3,8}, Ivan Okhrimenko\textsuperscript{3}, Valentin Borshchevskiy\textsuperscript{1,2,3}, Vladimir Chupin\textsuperscript{3}, Georg Büldt\textsuperscript{3}, Natalia Yutin\textsuperscript{13}, Ernst Bamberg\textsuperscript{14}, Eugene Koonin\textsuperscript{13} & Valentin Gordeliy\textsuperscript{1,2,3,5,6}.

Rhodopsins, most of which are proton pumps generating transmembrane electrochemical proton gradients, span all three domains of life, are abundant in the biosphere, and could play a crucial role in the early evolution of life on earth. Whereas archaeal and bacterial proton pumps are among the best structurally characterized proteins, rhodopsins from unicellular eukaryotes have not been well characterized. To fill this gap in the current understanding of the proton pumps and to gain insight into the evolution of rhodopsins using a structure-based approach, we performed a structural and functional analysis of the light-driven proton pump LR (Mac) from the pathogenic fungus \textit{Leptosphaeria maculans}. The first high-resolution structure of fungi rhodopsin and its functional properties reveal the striking similarity of its membrane part to archaeal but not to bacterial rhodopsins. We show that an unusually long N-terminal region stabilizes the protein through direct interaction with its extracellular loop (ECL2). We compare to our knowledge all available structures and sequences of outward light-driven proton pumps and show that eukaryotic and archaeal proton pumps, most likely, share a common ancestor.
Microbial (type 1) rhodopsins are the most abundant family of light-harvesting proteins. Type 1 rhodopsins are heptahelical transmembrane (7TM) proteins that covalently bind the retinal chromophore and use the energy of light to perform different biological functions, such as ion pumping\(^{1,6}\), ion channeling\(^{1,5}\), sensing\(^{10-12}\), and kinase activity\(^{13}\). The explosion of research on microbial rhodopsins, in large part, owes to the key role of these proteins in optogenetics, a methodology that caused a revolution in neuroscience\(^{14,15}\). Recently, discoveries of genomics and metagenomics show that rhodopsins are highly abundant, perform extremely diverse functions, and are present in all kingdoms of life as well as many large viruses\(^{6,17}\). Rhodopsins are considered to be the most abundant light-harvesting proteins on earth and the major light capturers in the oceans\(^{18}\). Given the ubiquity of rhodopsins and their crucial ecological role, there is little doubt that these proteins played a crucial role in the evolution of life on earth. Recently, rhodopsins have been identified in Asgard archaea\(^{19}\), the archaeal superphylum that includes the likely ancestors of eukaryotes\(^{20,21}\). Furthermore, whole-genome analysis of microbial rhodopsins has led to the hypothesis that retinal-based phototrophy emerged early in the evolution of life on Earth, predating the rise and profoundly impacting the evolution of photosynthesis\(^{22}\). Thus, the study of the functions and evolution of rhodopsins could yield valuable information on the origin and early evolution of life.

Among the factors critical in the evolution of early life was the ability to convert the energy of sunlight into a transmembrane proton gradient that provides for chemiosmotic coupling\(^{23}\). Light-driven proton pumps generating transmembrane gradients are the most abundant among the rhodopsins\(^{24}\). The structure and function of these proteins are exceptionally well studied, providing a rare opportunity to use a structure-based approach to explore evolutionary relationships among proteins.

Whereas structures of archaeal and bacterial proton pumps have been thoroughly characterized, this is not the case for rhodopsins of unicellular eukaryotes. To fill this gap and enable a structure-based analysis of the evolutionary relationships among the rhodopsins, we determined a high-resolution structure and performed an in-depth functional study of a light-driven proton pump LR (Mac) from the fungus <i>Leptosphaeria maculans</i>. <i>Leptosphaeria maculans</i> is a major pathogen of <i>Brassica napus</i>, an agricultural plant that is used as a feed source for livestock and the production of rapeseed oil. A dramatic epidemic of <i>L. maculans</i> occurred in Wisconsin on cabbage. The fungus destroys around 5–20% of canola yields in France\(^{25}\). The disease is also harmful in England\(^{26}\). Rapeseed oil is the preferred European oil source for biofuel due to its high yield. <i>B. napus</i> produces more oil per land area than other sources like soybeans. Thus, apparent from the fundamental importance of a thorough characterization of rhodopsin from a unicellular eukaryote, the study of LR (Mac) also could help understand the role of the rhodopsin in the fungus pathogenicity.

To date, high-resolution structures of multiple proton-pumping archaeal and bacterial rhodopsins have been solved\(^{27-30}\). For instance, more than 100 structures of the most-studied microbial rhodopsin, bacteriorhodopsin from archaean <i>Halobacterium salinarum</i> (HsBR), and its mutants, were deposited to the Protein Data Bank (PDB) since 1997\(^{31}\). By contrast, structural characterization of eukaryotic type-1 rhodopsins lags far behind. Two high-resolution structures of an H\(^+\) pumping rhodopsin from unicellular eukaryotes are currently available, <i>Acetabularia</i> rhodopsin (AR) from the marine alga <i>Acetabularia acetabulum</i>\(^{32,33}\) and Coccomyxa rhodopsin from <i>Coccomyxa subellipsoida</i>\(^{34}\). The AR has been deemed to be closely similar to archaeal HsBR although this protein has a 200 ms long photocycle, compared to the 20 ms photocycle in HsBR\(^{35}\).

Results and discussion

Overall LR structure and function. We performed a comparative functional characterization of the full-length (residues 1–313) LR and its previously partially characterized N-terminally truncated version (residues 49–313). These proteins did not show any detectable expression in <i>E. coli</i>, and therefore, we expressed them in <i>Leishmania tarentolae</i> (LEXSY) as previously described for channelrhodopsin 2 (ChR2)\(^{36}\) (see “Methods” for details).

First, we describe the structure of the full-length protein, which will be the structural basis for understanding the differences in the properties of rhodopsins and their evolution. LR was crystallized with an in meso approach similar to that used previously\(^{5,6}\). The structure was solved with crystals grown at pH 7.0 by the molecular replacement method, using the coordinates of HsBR (PDB ID: 1C3W), and refined to 2.2 Å resolution (Table 1). The crystals belong to the P2₁ space group and comprise two protomers in an asymmetric unit. LR (residues 42–286), all-trans-retinal (ATR), 19 lipid molecules, and 65 water molecules inside the protein are clearly resolved in the electron-density map, whereas the N- and C-terminal domains lack resolution for 41 and 31 residues, respectively (Supplementary Figures 1–2).

The LR bundle consists of seven transmembrane helices: TM1–TM7, connected by three intracellular (ICL 1–3) and three extracellular (ECL 1–3) loops, with retinal chromophore covalently attached to the K270 on the TM7 (Fig. 1). The overall structure of LR aligns well with the archaeal pump HsBR\(^{37}\) (PDB ID: 1C3W) and eukaryotic pumps ARII\(^{32}\) (PDB ID: 5AWZ) and...
Table 1 Crystallographic data collection and refinement statistics.

| Data collection  | LR 1-313 |
|------------------|----------|
| Space group      | P 21 21 21 |
| **Cell dimensions** |          |
| a, b, c (Å)      | 63.54, 70.78, 148.02 |
| α, β, γ (°)      | 90, 90, 90 |
| Wavelength (Å)   | 0.976 |
| Resolution (Å)   | 48.21-2.2 (2.26-2.20) |
| R<sub>merge</sub> (%) | 19.4 (139.8) |
| I/α               | 5.7 (1.1) |
| CC<sub>1/2</sub> (%) | 99.7 (80.4) |
| Completeness (%)  | 99.9 (99.6) |
| Unique reflections| 34,622 (2526) |
| Multiplicity      | 6.4 (6.2) |
| **Refinement**    |          |
| Resolution (Å)   | 19.92-2.20 |
| No. reflections   | 34,340 |
| R<sub>work</sub>/R<sub>free</sub> (%) | 23.8/28.5 |
| No. atoms         |        |
| Protein           | 3650 |
| Retinal cofactor  | 40 |
| Water             | 104 |
| Lipids            | 419 |
| B-factors (Å<sup>2</sup>) | 35 |
| Protein           | 27 |
| Retinal cofactor  | 40 |
| Water             | 44 |

| **R.m.s. deviations** | Bond lengths (Å) | 0.007 |
|                       | Bond angles (°)  | 0.935 |

CsR<sup>44</sup> (PDB ID: 6GYH), with root mean square deviations (RMSD) of 0.65, 0.72, and 0.76 Å respectively (Supplementary Figure 3). The interior portions of LR and HsBR are virtually identical structurally, and the only major difference is observed at the extracellular sides of these rhodopsins (Fig. 1). Specifically, ECL1 of LR is much longer, consisting of 30 amino acids, whereas the corresponding loop in HsBR contains only 16 residues, which results in a major difference in length, 44 Å in LR vs 21 Å in HsBR. Interestingly, ECL1 of LR is reminiscent of the highly conservative ICL1 beta-sheet domain of heliorhodopsins, which was previously shown to play an important role in dimer formation<sup>41,42</sup> (Supplementary Figure 4).

Structural features of LR important for proton transport. In the LR structure, the retinal chromophore is fixed in an all-trans conformation. The protonated retinal Schiff base (RSB) points toward the extracellular side and donates a hydrogen bond to the w8 molecule (Fig. 2a). This arrangement is stabilized by the RSB counterions, D266 and D139, which are counterparts to D212 and D85 in HsBR. The counterions and water molecules (w5, w8, and w10) form a hydrogen bond pentamer, which stabilizes the RSB and the extracellular part of the retinal binding pocket. The counterions take the same positions as the corresponding amino acids of HsBR, and as in HsBR, are additionally stabilized by T143 and S97 residues. By contrast, Gloeobacter rhodopsin (GR) and other proteorhodopsins (PRs) exhibit a different conformation of the RSB region<sup>43</sup>. In the GR structure, D121 (counterpart to D85 in HsBR, the primary proton acceptor) forms a hydrogen bond with H57<sup>44</sup>. This configuration of the counterions was shown to cause pH dependence of the photocycle due to deprotonation of the H57 residue at pH higher than physiological<sup>45</sup> (Supplementary Figure 5). In archaeal pumping rhodopsins, the proton release is mediated by a glutamate pair (E194–E204 in HsBR). In LR and other eukaryotic proton pumps, the release group has a similar construction. The E258 residue interacts with D248 that equivalently substitutes E204 of HsBR in a configuration similar to that of HsBR. By contrast, the bacterial proton-release group consists of only one negatively charged residue, the sole glutamate in TMS5 or TM6 (Supplementary Figures 6 and 7). A common feature of bacterial proton pumps is that they contain a water-accessible cavity that is connected with the extracellular bulk and protrudes up to R136 (Fig. 2a). By contrast, in both archaeal and eukaryotic proton-pumping rhodopsins, extracellular cavities are separated by a negatively charged pair of residues (E194–E204 in HsBR and D248–E258 in LR) into two parts, one of which is open to the bulk (Fig. 2a). The proton-releasing moiety of PRs consists of the sole glutamate residue in TMS5. The configuration of this part is conserved in the entire family of bacterial proton-pumping rhodopsins (Supplementary Figure 7). The extracellular part of the protein is fully accessible from the bulk for the water molecules up to the arginine residue (R76 in Med12BPR), similarly to LR.

The intracellular portion of LR is structurally nearly identical to that of HsBR (Fig. 2b). In particular, protonation occurs through the D150 residue, presumably, followed by a synchronized movement of L147 and T143 residues<sup>46</sup>. In HsBR, the corresponding residues are D96, L93, and T90, respectively. In the GR structure, the residues responsible for the proton uptake are E132, Q129, and T125. This configuration results in nearly complete accessibility of the proton donor E132 to the cytoplasmic bulk (Supplementary Figure 8). In other proteorhodopsins, the glutamate is often replaced by lysine, causing an absorbance shift effect<sup>47</sup>.

Photocycle of LR proton pump. The photocycle kinetics of HsBR and LR are closely similar as suggested by the high structural similarity and are likely to undergo similar structural rearrangements under light illumination<sup>48</sup> (Fig. 3). Both full-length and truncated constructs of LR in DDM micelles exhibit almost similar relaxation time that suggests that N-terminus has little impact on protein photoactivation (Supplementary Figure 9). After photoactivation, the LR photocycle starts with a short-living K state followed by blue-shifted L- and M states associated with proton release in archaeal proton pumps<sup>49</sup>. Next, the protein relaxation into the ground state occurs through N- and O states that complete the photocycle<sup>50-53</sup>. Consistent with RSB configuration difference, PR photocycle, and in the particular photocycle of green-absorbing proteorhodopsin (GPR) has 4 photointermediate states under neutral pH, lacking an apparent L state<sup>54</sup>. Consistent with structural similarity, the photocycle of LR and HsBR demonstrates nearly similar photocycle composition that further supports a close relation between LR and archaeal proton pump HsBR.

LR oligomeric state. Almost all microbial rhodopsins exist as oligomers in the native lipid environment. The oligomeric state was previously shown to contribute to protein stability<sup>55,56</sup> and in some cases to function<sup>57</sup>. Besides that, the stoichiometry of the fundamental oligomer of microbial rhodopsins was recently found to correlate with the phylogenetic origin of rhodopsins<sup>58</sup>. Even though LR forms dimers in the crystal lattice (Supplementary Figure 1), the monomer–monomer interface was surprisingly low (187 Å<sup>2</sup>)<sup>59</sup>. To assess oligomeric states of LR rhodopsin in a solution, we performed glutaraldehyde cross-linking of LR followed by SDS-PAGE analysis (Fig. 4a). This method was previously used in combination with size-exclusion chromatography (SEC) to confirm pentameric assembly of viral.
rhodopsin OLPVRII17, so we believe that combination of cross-linking with SEC provides sufficient reliability for the determination of LR protein oligomerization. For both LR 1–313 and LR 49–313 we observed bands corresponding to oligomers with stoichiometry 3 and lower, which suggests trimeric state as predominant in DDM micelles (Fig. 4a, Extended Fig. 10). Moreover, size-exclusion chromatography (SEC) profiles of LR 1–313 and LR 49–313 in DDM micelles demonstrate the presence of only one oligomeric state, which is likely trimeric. At pH 8.0, the calculated molar masses from the SEC retention volume of LR 1–313 and LR 49–313 peaks were 99 kDa/267 kDa and 73 kDa/216 kDa, respectively (Fig. 4b). Given that the empty DDM micelle size is ~50 kDa, pentameric KR2 and trimeric NsXeR elutes at 321 and 192 kDa respectively (Fig. 4c). Therefore, it is likely that LR in DDM micelles exist in monomeric and trimeric forms only. It is also noteworthy that truncation of N-terminus changes the ratio between monomers and trimers from ~2:5 for LR 1–313 to ~40:1 for LR 49–313 and thus impacts on detergent resistance of the protein.

Using template-based modeling GalaxyHomomer server, we proposed the trimeric state of LR using monomer template-based on archaeorhodopsin-2 model (PDB: 3WQJ)60. LR trimer has a higher monomer–monomer interface area (1799 Å2) and therefore is a likely major oligomeric form of LR rhodopsin (Fig. 4d). Besides that, protomers in both LR and HsBR trimers interact within TM2–TM4 helices (Fig. 4e), that further highlights the similarity between LR and HsBR. It should also be noted that despite the similarity of ECL1 of LR and ICL1 of heliorhodopsins, heliorhodopsins employ different dimer configuration with a much higher interface surface area (Supplementary Figure 11).

**Fig. 1 The overall architecture of LR. a** Crystal structure of LR at pH 7.0 at two different projections. Hydrophobic/hydrophilic interface was calculated with PPM server83 and is shown as gray lines, all-trans-retinal (ATR) cofactor is colored orange. **b** Structural superimposing of LR and HsBR (PDB code: 1C3W) structures. **c** Schematic representation of the ribbon diagram. **d** Structural comparison of extracellular loop 1 (ECL1) of LR and HsBR proteins.

**Structural roles of ECL1 and the N-terminal domain.** We showed that the LR core (membrane part) of the protein is strikingly similar to the core of HsBR and, in general, to the corresponding structures of other archaeal rhodopsin proton pumps as expected of core parts of highly conserved homologous proteins. By contrast, substantial structural variation could be expected to exist in the N- and C-terminus regions. To assess the role of the N-terminal domain in the structural stability of the protein, we searched protein sequence databases for homologs of LR using BLASTP61 and identified multiple opsins with high sequence similarity to LR. Despite the overall high similarity, we identified at least two subgroups among fungal rhodopsins that follow different patterns in the extracellular side of the protein. The first subgroup contained an elongated ECL1 domain, and a short beta-strand bears the N-terminus, whereas the second subgroup lacked both these features (Fig. 5a, Supplementary Figure 12). Secondary-structure prediction with RAPTORX62 suggests the presence of an additional beta-strand near the N-terminus of LR, but not in subgroup 2 of fungal rhodopsins or other light-driven proton pumps. Therefore, the elongated ECL1 domain and the N-terminal beta-strand appear to be signatures of the subfamily of fungal rhodopsins that includes LR rhodopsin. Consistent with the prediction, the LR crystal structure contained 8 residues from the N-terminal region (residues 42–49), that aligned with ECL1 and belonged to a 3-stranded antiparallel β-sheet. This N-terminal β-strand strongly interacts with the ECL1 loop and is directly involved in its stabilization through at least 7 hydrogen bonds. In addition, the N-terminus of LR is connected with the TM α-helices via the V47–S52 and G48–D256 interactions (Fig. 5b). To verify the effect of the N-terminal truncation,
we measured the thermal stability of the LR protein constructs with a nano-DSF method under different pH and detergent concentrations (Supplementary Figure 13). The F350/330 ratio of truncated LR showed two inflection points (corresponding to protein-unfolding temperatures) at 49.3 and 64.8 °C, whereas the full-length LR demonstrated a single inflection point at 66.5 °C (Fig. 5c).

To assess the contribution of the N-terminal domain to protein stabilization we used CABS-flex 2.0 coarse-grain protein modeling server. We used two models, full-length LR structure (resolved residues 42–286) and LR structure without the N-terminal region (residues 49–286), that represent LR constructs used for functional tests (Fig. 5e). Consistent with previous observations, the average root mean square fluctuation (RMSF) of LR 49–286 (0.74 Å) was higher than the average RMSF of LR 42–286 (0.65 Å). In particular, the highest destabilization was predicted for the ECL1 and ECL3 domains that directly interact with the N-terminal domain in the crystal structure. Given that the F350/330 ratio represents a change in tryptophan fluorescence, the additional melting point at 49.3 °C is likely to be caused by tryptophans on the extracellular side of the protein (W56, W190, W192, and W244). Increased flexibility of TM1, ECL1, and ECL3 domains can impact hydration and orientation of tryptophans on the extracellular side, with surprisingly little influence on protein kinetics.

**Proton translocation experiments.** To estimate the proton-pumping activity of the full-length and truncated proteins, we performed ion translocation experiments using POPC:POPS proteoliposomes (see “Methods” for details). Under weak acidic conditions (pH 6.5, 100 mM NaCl), the full-length protein demonstrated a substantially higher proton-pumping activity than the truncated derivative, 0.3 pH units for LR 1–313 against 0.1 for LR 49–313 (Fig. 5d, g).

A possible explanation of such a difference in proton pumping activity of two proteins with nearly similar photocycle is protein orientation in liposomes. Membrane protein orientation generally follows the “positive-inside” rule. Truncation of a slightly negatively charged N-terminus (D4, E7, E8, and D40) substantially altered the charge distribution of the extracellular part of the LR protein. Because extracellular parts of membrane proteins are more negatively charged than intracellular parts, such removal should influence protein orientation in both cell membranes and liposomes. Besides, protein orientation might be also influenced by decreased stability of the extracellular portion of the truncated protein.

**Structure-based phylogenetic analysis of light-driven proton pumps.** Different aspects of the evolution of microbial rhodopsins have been addressed in detail, featuring both large datasets and rhodopsins with putative functions. However, all the conclusions in these studies are derived from sequence-based phylogenetic analysis. To address different perspectives of the discussion, we performed a structure-based clustering of the light-driven outward proton pumps using 16 structures available from the PDB (see Methods for details). To this end, we superimposed all structures of proton pumps and calculated the pairwise RMSD values. We used these values to cluster the rhodopsins with a hierarchical clustering algorithm (Fig. 6). Notably, and in agreement with both the detailed structural comparisons described here and the previous phylogenetic analyses, the proton-pumping rhodopsins from archaea and eukaryote confidently clustered together, to the exclusion of the structurally distinct bacterial rhodopsins (Fig. 6b). Rhodopsin genes appear to be
subject to extensive horizontal gene transfer, resulting in mixed branches in phylogenetic trees and complicating inferences of common ancestry. Nevertheless, the high structural similarity between archaeal and eukaryotic proton pumps including the conformation of key functional regions (Fig. 6c), such as the retinal binding pocket and the regions involved in proton release and proton uptake, along with functional similarities (pH dependence of the photocycle, the intermediate states), presents strong evidence of an archaeal origin of eukaryotic proton-pumping rhodopsins (and, most likely, all other eukaryotic rhodopsins as well). Thus, rhodopsins appear to belong to a small set of genes, apart from the core components of the translation, transcription, and replication systems, that have been inherited by the proteoeukaryotes from their archaeal ancestors.

Methods

Phylogenetic analysis. In total, 28 sequences of rhodopsins with H+-pumping activity were aligned using MUSCLE. Phylogenetic reconstruction was conducted by maximum likelihood (ML) using PhyML with the following parameters: Jones–Thornton–Thornton model, SH-like approximate likelihood-ratio test, and estimated gamma-distribution parameter. The following rhodopsins were used for phylogenetic tree construction and sequence alignment:

- HsBR, Bacteriorhodopsin from Archaee (Halobacterium salinarum; P02945);
- HwBR/MR, Midfrohodopsin from Eubacteria (Halobacterium halobium DSM 16790, Q18D38); ar-1, Archaerhodopsin-1 from Archaee (Halorubrum chiovater, P90351); ar-2, Archaerhodopsin-2 from Archaee (Halobacterium sp. AUS-2, P29563), ar-3, Archaerhodopsin-3 from Archaee (Halobacterium salinarum DSM 17290, B1YF8); Med12, proteorhodopsin from uncultured bacterium (Med12, Q4P594); H775, Blue-light absorbing proteorhodopsin from uncultured bacterium (HOT 75m4, Q4AFF7); Xe, Xanthorhodopsin from Eubacteria (Salinibacter ruber DSM 13855/M31, Q5S2F8); TR, Thermophilic Rhodopsin from Eubacteria (Thermus thermophilus J1-18, H9ZSC3); gPR, Green-light absorbing proteorhodopsin from Eubacteria (Gamma-proteobacterium EBA31A08, Q97P74); Mac, Mac Rhodopsin from Eubacteria (Canidithas Actinomarina minuta, S6D71); PR from O.marina, Oxyrhynchus marina rhodopsin (Oxyrhynchus marina (Dinoflagellate), A7WQE3); GR, Gloeobacter Rhodopsin from Eubacteria (Gloeobacter violaceus ATCC 29082/PGC 7421, Q7NP59); GPR1, Green-absorbing Proteorhodopsin from Eubacteria (Dokdonia donghaensis MED134, EAQ40507); GPR2, Green-absorbing Proteorhodopsin from Eubacteria (Vibrio sp. AND4, ZP_02194911.1); GPR3, Green-absorbing Proteorhodopsin from Eubacteria (Canidithas Pelagibacter ubique HTCC1062, “SAR1” group, Q4FMZ3); BPR, Blue-absorbing Proteorhodopsin from Eubacteria (Photobacterium sp. LCI-200, BAL68143); NM-R1, Nonlabens marinus proteorhodopsin from Eubacteria (Nonlabens marinus S1-08, W8VZ92).

Structure alignment. For structure alignment, 17 known atomic structure models of H+-pumping rhodopsins, including LR, were compared pairwise by RMSD values calculated with the PyMOL align function with parameter cycles = 2 and cutoff = 1.5. After hierarchical clustering of the RMSD correlation values, the analyzed structures were divided into two clusters - “archaeal” and “bacterial”. Then the structure of the reference archaeal pump HsBR was aligned to each one from the “archaeal” cluster and the structure of the reference Med12BPR was aligned to each one from the “bacterial” cluster. To represent the conservation of Ca atom positions within the clusters, for every Ca atom of the reference structure, distance to the nearest Ca atom of the respective aligned structure was measured. The measurement was repeated for every aligned structure. RMSD of the distances corresponding to each Ca atom of each reference structure, CαRMSD, was used as a measure of the Ca position conservation.
Fig. 4 LR oligomerization analysis. a Glutaraldehyde cross-linking of LR 1–313 and LR 49–313. Gradient SDS-PAGE 18.5–8%. Lanes 1 and 1* contain LR 1–313 and LR 49–313 samples treated with glutaraldehyde vapor for 30 min correspondingly. LR samples not treated with glutaraldehyde were used as a control.

b Size-exclusion chromatography profiles of LR 1–313 and LR 49–313 protein used for crystallization trials. Elution profiles of NsXeR and KR2 proteins are shown as examples of trimeric and pentameric proteins with similar SEC experiments. Full details on the expression and purification of those proteins can be found in 5,6.

c The estimated molecular weight of different rhodopsin fractions obtained during the SEC experiment using standard calibration proteins. Full details on cross-linking and SEC calibration can be found in Supplementary Figure 3.

d Extended comparison of multimeric states of LR and HsBR. LR dimer and HsBR trimer correspond to the crystal-packing multimeric state. LR trimeric state was calculated using Homomer server60 with archaeorhodopsin-2 reference model (PDB: 3WQJ).

e Magnified view of an interprotein interaction between monomers in the multimeric state. The average interaction surface area is indicated for all structures.
Cloning. The genes encoding LR (1–313 aa and 49–313 aa) from *Leptosphaeria maculans* (UniProt Q9HGT7) were synthesized de novo. The nucleotide sequences were optimized for *Leishmania tarentolae* expression with GeneOptimizer software (Thermo Fisher Scientific). Both genes in fusion with a C-terminal polyhistidine tags (H6 and H9) were introduced into the integrative inducible expression vector pLEXSY_I-blecherry3 (Jena Bioscience, Germany) through the BglII and NotI restriction sites. Plasmid vectors, optimized genes, and primers used for PCR amplification can be found in Supplementary Tables 1–3.

LEXSY expression, solubilization, and purification. The proteins were expressed as described previously. The *Leishmania tarentolae* cells of the strain LEXSY host TT–TR (Jena Bioscience) were transformed with the LR1-313 and LR49-313 expression plasmids linearized by the SwaI restriction enzyme. After the clonal selection, the transformed cells were grown at 26 °C in the dark in shaking flasks in the Brain-Heart-Infusion Broth (Carl Roth, Germany) supplemented with 5 mg ml\(^{-1}\) Hemin, 50 U ml\(^{-1}\) penicillin, and 50 mg ml\(^{-1}\) streptomycin (AppliChem). When OD\(_{600}\) = 1.0 was reached, 10 mg ml\(^{-1}\) tetracycline was added, for LR1–313 also, 10 \(\mu\)M all-trans-retinal (Sigma-Aldrich) was added, and incubation continued for a further 24 h. The collected cells were disrupted in an M-110P Lab Homogenizer (Microfluidics) at 10,000 psi in a buffer containing 50 mM NaH\(_2\)PO\(_4\)/Na\(_2\)HPO\(_4\), pH 7.6, 0.1 M NaCl, 10% glycerol, 1 mM EDTA, 2 mM 6-aminohexanoic acid (AppliChem), 50 mg l\(^{-1}\) DNase I (Sigma-Aldrich), and complete protease inhibitor cocktail (Roche). The membrane fraction of the cell
Lysate was isolated by ultracentrifugation at 120,000 g for 1 h at 4 °C. The pellet was resuspended in the same buffer but without DNase I and stirred for 30 min at 4 °C. The ultracentrifugation step was repeated. Finally, the membranes were resuspended in the solubilization buffer containing 20 mM HEPES, pH 8.0, 0.2 M NaCl, Complete, 1% n-dodecyl-β-D-maltoside (Cube Biotech), and 20 μM for LR 1–313 and 5 μM for LR 49–313 all-trans-retinal and stirred overnight for solubilization. The insoluble fraction was removed by ultracentrifugation at 120,000 g for 1 h at 4 °C. The supernatant was loaded on a Ni-NTA resin (Cube Biotech), and LR was eluted in a buffer containing 20 mM HEPES, pH 7.5, 0.2 M NaCl, 0.25 M L-Histidine (AppliChem), 0.1 mM phenylmethanesulfonyl fluoride (PMSF), 2 mM 6-aminohexanoic acid, Complete, and 0.1% n-dodecyl β-D-maltoside. The eluate was subjected to size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences) in a buffer containing 10 mM NaH2PO4/Na2HPO4, pH 6.5, 0.2 M NaCl, 1 mM EDTA, 2 mM 6-aminohexanoic acid, Complete, 0.1 mM phenylmethanesulfonyl fluoride (PMSF), 2 mM 6-aminohexanoic acid, Complete, and 0.1% n-dodecyl β-D-maltoside. Protein-containing fractions with an A280/A540 absorbance ratio of about 1.5 were pooled and concentrated to 30–40 mg ml−1 for crystallization. An average yield for LEXSY-optimized constructs of LR 1–313 and LR 49–313 was 20 mg and 10 mg from 1 liter of culture, respectively.

Incorporation of the protein into liposomes. Phospholipids 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine at a ratio 4:1 (wt wt−1) (POPC:POPS) were dissolved in CHCl3 (chloroform ultrapure, Applichem Panreac) and dried under a stream of N2 in a glass vial. The solvent was removed by overnight incubation under vacuum. The dried lipids were resuspended in 4% (w v−1) sodium cholate. The mixture was clarified by sonication at 4 °C and LR was added at a protein:lipid ratio of 1:20 (w w−1). The detergent was removed by the addition of detergent-absorbing beads (Amberlite XAD 2, Supelco) and incubation at 4 °C for 2 days. The 1.5 ml of liposome mixture was dialyzed against 100 mM NaCl (pH 8.0) buffer at 4 °C for 18 h (two 1.5 L changes) to adjust for the desired pH.

Measurement of proton translocation activity in proteoliposomes. The measurements were performed on 1.5 ml of stirred liposome suspension at 4 °C. LR-containing liposomes were prepared following the protocol described above. Liposomes were illuminated for 10 minutes with a halogen lamp (Intralux 5000-1, VOLPI) and then were kept in the dark for another 10 min. Changes in pH were monitored with a pH meter (LAB 850, Schott Instruments). Measurements were
repeated in the presence of 30 μM of carbonyl cyanide m-chlorophenyl hydratine (CCCP, Sigma-Aldrich) under similar conditions.

Nanodsf measurements. Nano-DSF measurements were performed using a Promethuss NT-48 instrument (NanoTemper Technologies, Germany). In total, 6 samples for both full-length and truncated LR were made, having pH 6.0, 7.0, or 8.0 and DDM concentration of 0.05% or 1%. First, samples with purified LR at a concentration of 1 mg mL⁻¹ were dilaayed against 200 μM NaCl, 0.05% DDM buffer for 18 h at 4 °C. After that, the protein was introduced into 50 mM Na₂HPO₄-NaH₂PO₄, 200 mM NaCl buffer with the desired pH, and incubated at 4 °C for another 18 h. About 10 μL of each sample was loaded into UV capillaries (NanoTemper Technologies). The temperature gradient was set at 1 °C min⁻¹ in a range from 15 to 98 °C. Protein unfolding was observed by following the change in tryptophan fluorescence at emission wavelengths of 330 and 350 nm. The ratio between the emission intensities at 350 and 330 nm (F₃₅₀/F₃₃₀) was used to track the structural changes with increasing temperature. Melting-point temperatures (Tₘ) were calculated using the peaks in the first derivative of the signal data.

Time-resolved absorption spectroscopy. The laser-flash photolysis setup was similar to that described by Chizhov and co-workers50. The excitation/detection spectrum of the probe, red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. New Nat. Neurosci. 16, 1499–1508 (2013).

Data availability. The protein coordinates and atomic structure factors have been deposited in the Protein Data Bank (PDB) under accession number 7BMH (LR). Source data are provided as Supplementary Data 1. All other data are available from the corresponding author upon reasonable request.

Received: 7 October 2020; Accepted: 7 June 2021; Published online: 30 June 2021

References
1. Lozier, R. H., Bogomolni, R. A. & Stockenius, W. Bacteriorhodopsin: a light-driven proton pump in Halobacterium Halobium. Biophys. J. 15, 955–962 (1975).
2. Schobert, B. & Lanyi, J. K. Halorhodopsin is a light-driven chloride pump. J. Biol. Chem. 257, 10306–10313 (1982).
3. Hoffmann, A., Hildebrandt, V., Heberle, I., J. & Buldt, G. Phototoxic mitochondria: in vivo transfer of a light-driven proton pump into the inner mitochondrial membrane of Schizosaccharomyces pombe. Proc. Natl. Acad. Sci. 91, 9367–9371 (1994).
4. Inoue, K. et al. A light-driven sodium ion pump in marine bacteria. Nature. Commun. 4, 1678 (2013).
5. Gushchin, I. et al. Crystal structure of a light-driven sodium pump. Nat. Struct. Mol. Biol. 22, 390–396 (2015).
6. Shevchenko, V. et al. Inward H⁺-pump xenorhodopsin: Mechanism and alternative optogenetic approach. Sci. Adv. 3, 1–11 (2017).
7. Nagel, G. et al. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. Proc. Natl. Acad. Sci. 100, 13940–13945 (2003).
8. Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. & Tsien, R. Y. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. Nat. Neurosci. 16, 1499–1508 (2013).
9. Oda, K. et al. Crystal structure of the red light-activated channelrhodopsin Chrimson. Nat. Commun. 9, 3949 (2018).
10. Gordeliy, V. I. et al. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducin complex. Nature 419, 484–487 (2002).
11. Mukhametzianov, R. et al. Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. Nature 440, 115–119 (2006).
12. Boyden, E. S. Optogenetics and the future of neuroscience. Nat. Neurosci. 18, 1200–1201 (2015).
13. Yutin, N. & Koonin, E. V. Proteorhodopsin genes in giant viruses. Biol. Direct 7, 3 (2012).
14. Bratanov, D. et al. Unique structure and function of viral rhodopsins. Nat. Commun. 10, 4339 (2019).
15. Gómez-Consarnau, L. et al. Microbial rhodopsins are major contributors to the solar energy captured in the sea. Sci. Adv. 6, eaaw8855 (2019).
16. Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. & Tsien, R. Y. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. Nat. Neurosci. 16, 1499–1508 (2013).
17. Boyden, E. S. Optogenetics and the future of neuroscience. Nat. Neurosci. 18, 1200–1201 (2015).
18. Zhang, F. et al. Multimodal fast optical interrogation of neural circuitry. Nature 466, 633–639 (2007).
19. Bratanov, D. et al. Unique structure and function of viral rhodopsins. Nat. Commun. 10, 4339 (2019).
20. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. Nat. Rev. Microbiol. 15, 711–723 (2017).
21. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. Nat. Rev. Microbiol. 15, 711–723 (2017).
81. Adams, P. D. et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).
82. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
83. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 40, D370–376 (2012).

Acknowledgements
We acknowledge the Structural Biology Group of the European Synchrotron Radiation Facility (ESRF) for granting access to the synchrotron beamlines. This work was supported by the common program of Agence Nationale de la Recherche (ANR), France and Deutsche Forschungsgemeinschaft (DFG), Germany (ANR-15-CE11-0029-02/FA 301/11-1), by the DFG Research Unit FOR 2518 (Dyson, project P4 to JPM, MA 7525/1-1). VG greatly acknowledge support of Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF) within Helmholtz professorship. DZ was supported by the Russian Foundation for Basic Research project number 20-34-90009. VB, AR, and IO acknowledge the Ministry of Science and Higher Education of the Russian Federation (agreement # 075-00337-20-03, project FSMG-2020-0003). Crystallographic experiments and functional tests were supported by Russian Science Foundation (RSF) Project 19-44-06302. RA was supported by the Russian Foundation for Basic Research project number 17-00-00164. The development and validation of bioinformatics algorithms was supported by Russian Science Foundation (RSF) Project 21-64-00018.

Author contributions
D.Z., N.D., O.V., and V.C. contributed equally and either has the right to list himself first in bibliographic documents. N.D. and O.V. expressed and purified LR in LEXSY expression system. V.S. and I.O. tested LR expression in other expression systems and optimized genetic constructs. E.R, V.P., and R.A. crystallized protein, harvested crystals, and collected diffraction data. K.K. processed the diffraction data and refined the structure. E.Z. performed the search of initial phases. N.D., D.Z., and T.B. performed functional tests. D.S. and I.C. measured the photocycle kinetics of the protein. D.Z., N.D., A.A., N.Y., and E.K. performed the bioinformatics analysis. V.G. designed and supervised the project. E.B., G.B., V.C., V.B., A.A., and V.G. analyzed the data. V.G. with D.Z. and N.D. wrote the manuscript with input from all the authors. All authors have reviewed the final version of the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02326-4.

Correspondence and requests for materials should be addressed to V.G.

Peer review information Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Ross Bathgate and Luke R. Grinham.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021