

THE ORIGINS AND EVOLUTION OF AN EARLY
MICROBIAL RHODOPSIN PROTEIN

by

Cathryn Dawn Sephus

Copyright © Cathryn Dawn Sephus 2021

A Thesis Submitted to the Faculty of the

DEPARTMENT OF MOLECULAR AND CELLULAR BIOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2021

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by: Cathryn Dawn Sephus
titled: THE ORIGINS AND EVOLUTION OF AN EARLY MICROBIAL RHODOPSIN PROTEIN

and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

Betul Kacar

Betul Kacar

Date: Apr 18, 2021

Ryan Gutenkunst

Ryan Gutenkunst


Date: Apr 27, 2021

S. Duhamel

Solange Duhamel

Date: Apr 26, 2021

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement. 

Betul Kacar

Betul Kacar

Thesis Committee Chair
Molecular and Cellular Biology

Date: Apr 18, 2021

ARIZONA

ACKNOWLEDGEMENTS

To begin, I would like to thank my mentor and advisor, Dr. Betul Kacar. She has carved a place in Arizona where biologists, geologists, computer scientists, and astronomers can work together towards answering some of life's biggest questions. Somehow, while fostering this interdisciplinary environment and heading a NASA center, she not only manages to keep every project on course but guides each venture forward to be better than it was the day before. I am lucky to have been given the space and opportunity to grow as a researcher in her lab over the last three years. I sincerely appreciate my ideas always being embraced with enthusiasm and all of the discussions that have led to the synthesis of work presented here in this thesis.

I would also like to thank the members of the Kacar Ancient Biology Lab. Whether you've listened to my practice presentations, taught me how to use equipment, or were around to grab a tea when times were tough, each of you has played an integral part in my development and achievements as a scientist. A special thanks goes to Dr. Amanda Garcia. Thank you for showing me the ropes of ancestral sequence reconstruction, tirelessly answering my questions, and always being there to lend a hand with any ambitious experiments conducted in the lab. Your drive, work ethic, and intelligence have served as a constant inspiration over the years. You're an A+ scientist and a great mentor; I am honored to have had the chance to work with and learn from you.

I thank my committee members Dr. Ryan Gutenkunst and Dr. Solange Duhamel for their reviews, questions, and suggestions during the final stages of my master's program.

I would like to further extend my gratitude to the faculty and staff in the Department of Molecular & Cellular Biology. Thank you for the endless support and encouragement throughout my academic career at the University of Arizona.

Last but not least, I'd like to thank my mother for her unwavering support and incontestable strength. Thank you for never making me put down a book and for always allowing me to pursue my aspirations- even when they seemed completely "out of this world." I couldn't have done it without you and your advice to...

... *Keep Moving Forward.*

TABLE OF CONTENTS

LIST OF TABLES	6
LIST OF FIGURES	7
ABSTRACT	8
1 INTRODUCTION TO MICROBIAL RHODOPSINS	9
1.1 Phototrophic Systems	9
1.2 Rhodopsins	11
1.3 Microbial Rhodopsins.....	12
1.4 Bioenergetics	14
1.5 Considerations of Gene Transfer.....	16
2 ANCIENT GENE RECONSTRUCTION AND ANALYSIS	18
Summary.....	18
2.1 Ancestral Sequence Reconstruction	18
2.1.1 Reconstructing phylogenetic trees.....	20
2.1.2 Reconstructing ancestral states.....	21
2.2 Predicting Biophysical Attributes of Proteins	22
2.2.1 Physicochemical properties	22
2.2.2 Subcellular location	23
2.2.3 Tertiary structure	23
2.2.4 Functionality.....	24
3 MICROBIAL RHODOPSINS RECONSTRUCTED: IMPLICATIONS ON EARLY MEMBRANES & BIOENERGETICS	25
Summary.....	25
3.1 Results	25
3.1.1 Phylogeny shows an organization of rhodopsins based on functionality	25
3.1.2 Ancestral retinal binding site residues yield high statistical support.....	27
3.1.3 Oldest ancestral rhodopsin functional motifs resemble proton pumps.....	30
3.1.4 Protein architecture is consistent over time.....	31
3.1.5 Ancestral rhodopsins were membrane-spanning proteins	34
3.2 Discussion.....	36
3.3 Methods	39
3.4 References	42
3.5 Supplementary Data	51
4 COPYRIGHT PERMISSIONS	60

LIST OF TABLES

Table 1-1	Functional motifs of microbial rhodopsins	14
Table 3-1	Mean posterior probabilities of maximum likelihood ancestral sequences	27
Table 3-2	Results of homology modeling using Phyre2	31
Table 3-3	Results of homology modeling using SWISS-MODEL	32
Table 3-4	SOSUI prediction of ancestral rhodopsin subcellular localizations	35
Table 3-5	Query sequences used to collect homologous sequences from NCBI	39

LIST OF FIGURES

Figure 1-1	Comparative diagram of phototrophic metabolisms.	10
Figure 1-2	Photoisomerization of the retinal chromophore in type-1 and type-2 rhodopsins	11
Figure 1-3	Diagram illustrating functional diversity of microbial rhodopsin proteins	13
Figure 2-1	Schematic of steps used to complete ancestral sequence reconstruction.	20
Figure 3-1	Maximum likelihood phylogeny of type-1 (microbial) rhodopsins with heliorhodopsin outgroup sequences.	26
Figure 3-2	Statistical support for ancestral rhodopsin motif and active site residues	29
Figure 3-3	Homology models of ancestral rhodopsin tertiary structure.	33
Figure 3-4	Hydropathy plots of ancestral rhodopsin proteins	35

ABSTRACT

The advent of cellular organisms took place sometime between the start of prebiotic chemosynthesis on Earth and the evolution of the last universal common ancestor. Cellularity is now a fundamental organizational principle shared by all life on Earth and represents a key transition in evolutionary history. The emergence of cellular organization necessitated organisms to evolve a means to permit and regulate the exchange of material between the intracellular compartment and the extracellular environment. This capability implies both the ability to embed proteins into their membranes and to translocate molecules across their membranes. Other factors integral to the progression of early cellular life were the maintenance of transmembrane potential and chemiosmotic coupling for generating and conserving energy, and pigments to absorb light energy for photosynthetic and phototrophic metabolisms. Microbial rhodopsins, a superfamily of photoactive membrane proteins, have been suggested to be the simplest and possibly most ancient form of a phototrophic metabolism, likely providing a mechanism for microbial energy capture in Earth's early shallow marine ecosystems. The retinal-based photosystem (rhodopsin) is composed of one retinal chromophore and one opsin protein. In this system, light absorption directly drives a conformational change in the protein via the isomerization of the retinal moiety to carry out biological functions such as ion pumping and ATP synthesis. Here, computational approaches were used to investigate the evolutionary history of rhodopsin proteins, combining phylogenetic reconstruction and ancestral sequence inferences. Additionally, protein structure modeling and biophysical predictions were used to reveal ancestral rhodopsin functionality. Together, these results may shed light on the evolution of pigment-based metabolisms and prove beneficial for understanding the characteristics of early cellular membranes.

CHAPTER 1: INTRODUCTION TO MICROBIAL RHODOPSINS

1.1 PHOTOTROPHIC SYSTEMS

Solar radiation is the most abundant and virtually inexhaustible source of energy available on Earth's surface. Life has had to cope with the presence of solar light energy since it first began on the planet. Over time, organisms developed a cache of biomolecules to exploit the potential benefits and detriments of light as an energy source (Pinhassi et al. 2016). Among the earliest light-absorbing molecules to evolve were pigments, which absorbed light in the mid-ultraviolet range and efficiently converted light into energy in aqueous environments (Michaelian, 2015).

Although a variety of proteins such as phytochromes, phototropins, and photoactive yellow proteins have evolved to take advantage of light energy available on the Earth's surface (Briggs, 2001; Hellingwerf, 2003; Pham, 2018), only two biological mechanisms for light-driven energy production are known to exist: chlorophyll-based phototrophy and retinal-based phototrophy (Burnetti and Ratcliff 2020, Bryant 2019, **Figure 1-1**). While both phototrophic systems convert light into chemiosmotic free energy via the generation of a proton gradient, the mechanisms by which they achieve this differ greatly (Hellingwerf et al. 1993). Interestingly, previous studies comparing the bioenergetic systems have found that the two maintain a complementary division of phototrophic niche spaces. For example, their pigments absorb light at reciprocal wavelengths, and they monopolize opposite ends of key trade-offs such as efficiency per unit resource versus efficiency per unit protein and complexity versus simplicity (Burnetti and Ratcliff 2020, DasSarma and Schwieterman 2018, Larkum et al. 2018, Hampp 2000, Hellingwerf et al. 1993).

The chlorophyll-based photosystem is the more efficient and complex mechanism of phototrophy (Bryant and Frigaard 2006). This system uses photons to initiate electron transfers and redox reactions of chlorophyll and an electron acceptor (Hellingwerf 1993). Secondary electron transfer reactions that are independent of light subsequently produce a proton-motive force that is coupled to ATP synthesis (Bryant 2019). A single chlorophyll system necessitates at least 30 different genes, coding for multiple proteins and pigments to form viable reaction centers, antenna structures, and photosystems (Bryant and Frigaard 2006, Pinhassi, 2016). Currently, the

biosynthesis of chlorophyll has yet to be detected in *Archaea* but has been shown to exist in *Bacteria* and *Eukaryotes* (Burnetti, 2020; Pinhassi, 2016). This has led some to suggest photosynthesis evolved after the divergence of bacterial and archaeal-eukaryotic lineages (Bryant and Frigaard 2006).

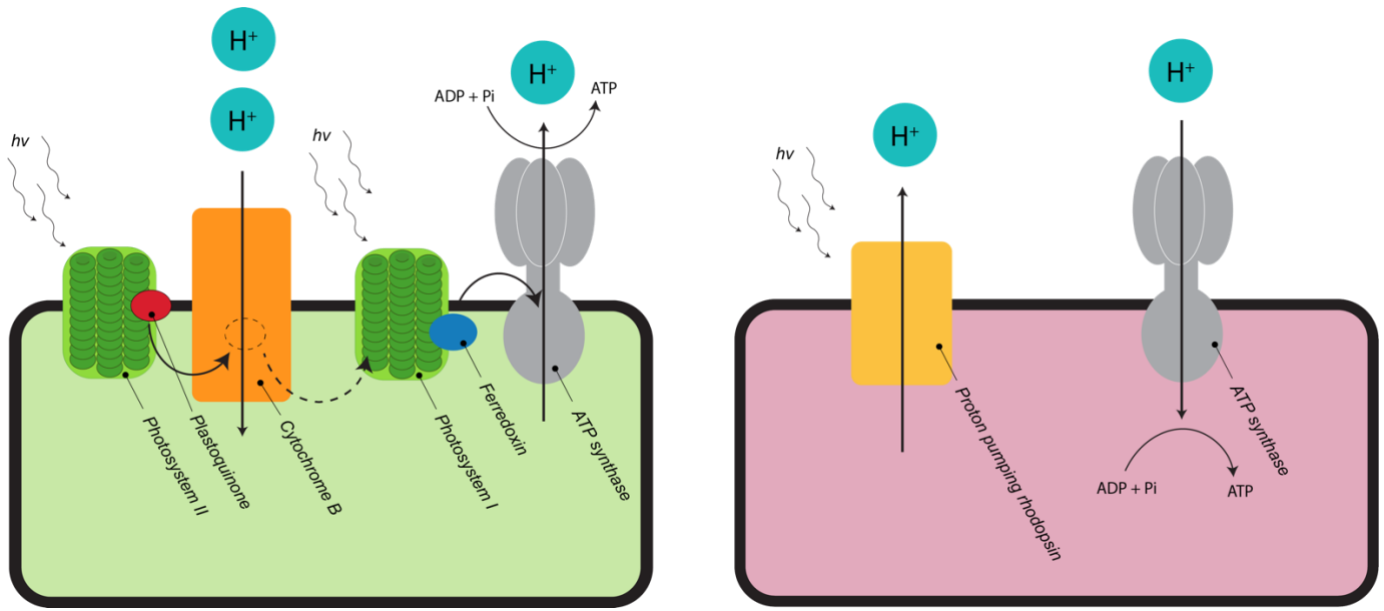


Figure 1-1. Comparative diagram of phototrophic metabolisms. *Left* – simplified schematic of a chlorophyll-based phototrophic system. *Right* – simplified schematic of a retinal-based (rhodopsin) phototrophic system.

Retinal-based phototrophy utilizes a system that is comparably simpler than its chlorophyll-based counterpart (DasSarma and Schwieterman 2018; Nowicka 2016). The retinal-based photosystem (rhodopsin) is composed of one retinal chromophore and one opsin protein (Spudich, 1986, 2014). In this system, light absorption directly drives proton transport via the isomerization of the retinal group to synthesize ATP or carry out other biological functions such as ion pumping, light sensing, or phototaxis (Kurihara and Sudo 2015). The singular chromophore necessitates no cofactors and the organic pigment proteins have been found to be common to the three domains of life (*Bacteria, Archaea, and Eukaryota*) (Gushchin and Gordeliy 2018). The simplicity and widespread abundance of rhodopsin photosystems have led many to speculate a precedent advent to photosynthesis (Raven and Smith 1981, Bryant and Frigaard 2006, Sparks et al. 2007, DasSarma and Schwieterman 2018, Kwon et al. 2020).

1.2 RHODOPSINS

Photoreceptor proteins are a class of light-sensitive proteins harnessed by biology to regulate a range of cellular functions including energy production and signal transduction (Allorent and Petroustos 2017, van der Horst and Hellingwerf 2004). One of the largest and arguably best-characterized groups of photoreceptor proteins are rhodopsins (van der Horst and Hellingwerf 2004, Porter 2016, Bratanov et al. 2019, Kovalev et al. 2020). Rhodopsin proteins constitute a superfamily of photoactive membrane-spanning retinylidene proteins. Members of this family are widely distributed in organisms across the three domains of life and have been identified in some large viruses (Bratanov et al. 2019). All rhodopsins assume a seven-transmembrane α -helical fold which forms an interior cavity known as the ‘retinal binding pocket’ (Mackin et al 2014). This hydrophobic pocket houses a retinal chromophore that’s transiently bound to a lysine located on the seventh helix (Govorunova et al. 2017).

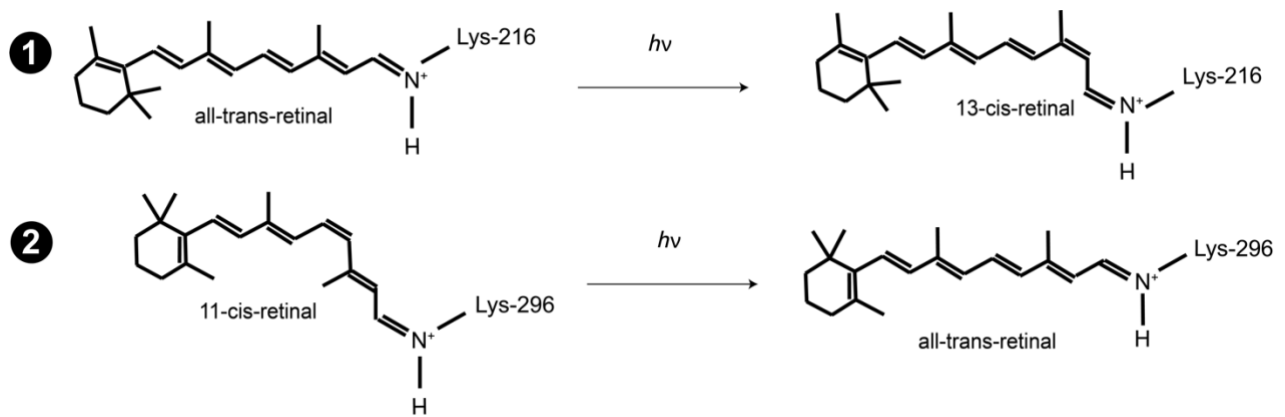


Figure 1-2. Photoisomerization of the retinal chromophore in type-1 and type-2 rhodopsins.

(1) All-trans-retinal isomerizes to 13-cis retinal upon light absorption in type-1 rhodopsins. (2) 11-cis retinal isomerizes to all-trans-retinal upon light absorption in type-2 rhodopsins.

The broad family of photoactive proteins may be partitioned into two distinct groups, microbial (type-1) and animal (type-2) rhodopsins. Type-1 rhodopsins are the most taxonomically diverse of the two, found in many species across the three domains of life. Their retinal

chromophore adopts the stable all-*trans* configuration when at rest then isomerizes to the 13-*cis* configuration when photoactivated (Nogly et al. 2018, Feldman et al. 2016, **Figure 1-2**). In contrast, type-2 rhodopsins are exclusive to eukaryotes and undergo a cis-to-trans conformation change upon photoactivation (Nogly et al. 2018, Inoue 2015, Feldman et al. 2016). Although the two groups share some structural and functional similarities, thus far, little to no sequence homology has been detected. This has called into question whether the two types of proteins represent a case of convergent evolution or extensive divergence from a common ancestor. While the former hypothesis has garnered the consensus of the field (Porter 2016, Oakley and Speiser 2015, Inoue 2015, Larusso et al. 2008), arguments in favor of the latter have left the question of rhodopsin's evolutionary relationships irresolute. (Shen et al 2013, Devine et al. 2013, Mackin et al. 2014).

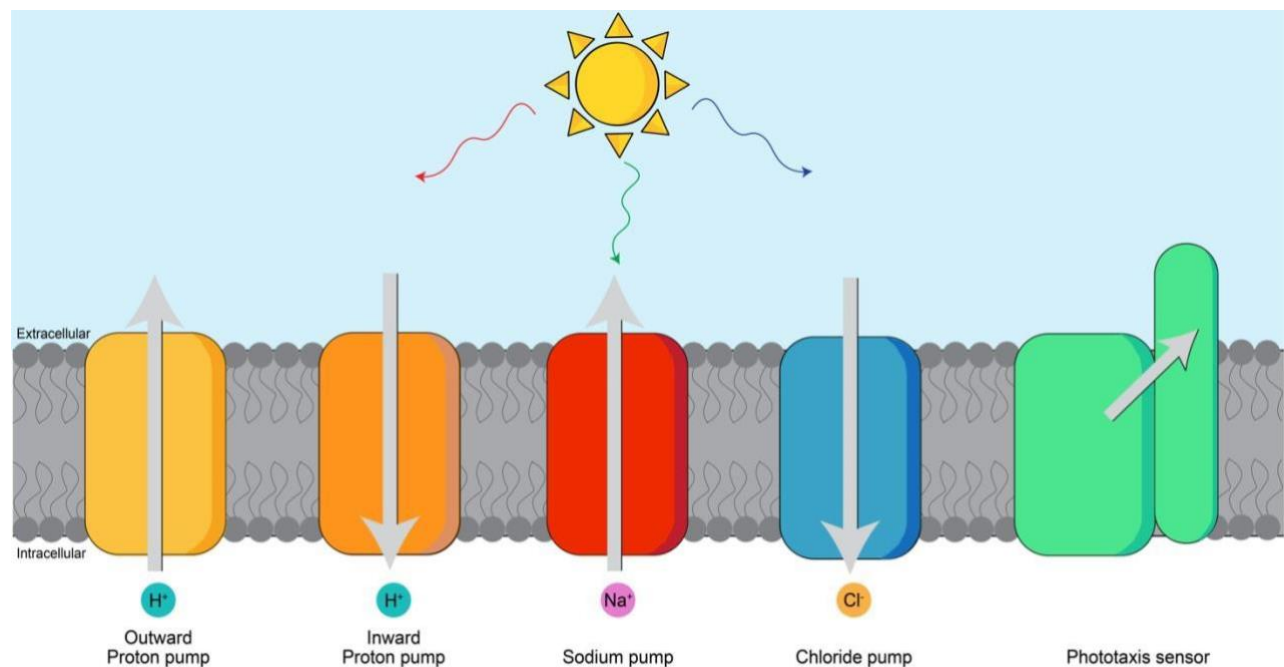
1.3 MICROBIAL RHODOPSINS

The first discovered microbial rhodopsin, bacteriorhodopsin, was identified as a photoactivated outward proton (H^+) pump within the membrane of *Halobacterium halobium* (Oesterhelt and Stoeckenius 1971, 1973). Halorhodopsin, a chloride (Cl^-) pump, and sensory rhodopsins I & II were detected in *Halobacterium*'s membrane soon after (Matsuno-Yagi and Mukohata 1977, Schobert and Lanyi 1982, Bogomolni and Spudich 1982). At first, microbial rhodopsins were thought to be exclusive to halophilic archaea, however, by the early 2000s, proteorhodopsin and xanthorhodopsin, eubacterial outward H^+ pumps, had been discovered in saline environments (Béjà et al. 2000, Balashov et al. 2005, Luecke et al. 2008). Moreover, advancements in metagenomic technologies have led to the discovery of outward sodium (Na^+) pumping rhodopsins and inward Cl^- pumping rhodopsins among marine bacterial species (Inoue et al. 2013, Nakajima et al. 2018, **Figure 1-3**).

Despite the distinct functionalities and primary sequences of microbial rhodopsins, members of the protein family largely retain the same structure – seven transmembrane helices and an internal lysine-bound retinal chromophore (Besaw et al. 2020, Govorunova et al. 2017). Functional variation is the result of small substitutions in the protein sequence, differences in charges of ion substrates, and distribution of internal water molecules (Besaw et al. 2020, Kaneko

et al. 2017, Kandori 2015, Bondar 2008). Ion-pumping rhodopsins in particular epitomize this concept- existing as a group of virtually identical protein structures with different ion affinities.

Currently, there are four known types of ion-pumping rhodopsins, inward H^+ pumps, outward H^+ pumps, Cl^- pumps, and Na^+ pumps (Inoue 2020). The functions of the proteins have been found to correspond to conserved three-residue motifs on the third transmembrane helix (Besaw et al. 2020, Inoue 2020, Kandori 2015, **Table 1-1**). Fungal and archaeal outward H^+ pumps may be distinguished with a characteristic Asp-Thr-Asp (DTD) motif, whereas bacterial outward H^+ pumps bear a DTE/DTK motif. Archaeal Cl^- pumping rhodopsins have a TSA motif at the corresponding sites and bacterial Cl^- pumping rhodopsins have an NTQ motif at the positions. Finally, Na^+ pumping rhodopsins can be distinguished by their hallmark NDQ motif. Recently, using a third helix motif as a functional predictor has proven effective for characterizing novel microbial rhodopsin proteins (Hasegawa et al. 2020).



Microbial rhodopsin functional types

Figure 1-3. Diagram illustrating functional diversity of microbial rhodopsin proteins. Some rhodopsins may act as ion pumps, transporting H^+ , Na^+ , or Cl^- across the membrane. Other

rhodopsins may act as photosensors. Adapted from Kandori, H., *Frontiers in Molecular Biosciences*, 2(52), 2015.

H ⁺ pump			Na ⁺ pump			Cl ⁻ pump		
D	T	D	N	D	Q	T	S	A
D	T	E				N	T	Q
D	T	K				T	S	D
D	T	G						
*D	S	A						

Table 1-1. Functional motifs of ion-pumping rhodopsins.

Information in the table was collected from Inoue et al. 2021, Inoue et al. 2020, Besaw et al. 2020, and Shevchenko et al. 2017.

*DSA motif denotes inward proton pumping functionality.

1.4 BIOENERGETICS

Chemiosmotic coupling, the ability to transport ions across a cellular membrane to facilitate a metabolism is ubiquitous across biological systems (Mitchell 1961, Alberts et al. 2002, Ferreira and Bashford 2006). This deep conservation of functionality suggests that the generation of ion gradients may have been an early invention of evolution, possibly playing a role in life's origins (Lane 2017, Lane et al. 2010). It is largely accepted that the last universal common ancestor (LUCA) hosted a membrane-bound ATP synthase-like complex and was able to maintain a chemiosmotic gradient (Weiss et al. 2018, Lane 2017, Lane and Martin 2012, Mulkidjanian et al. 2007). Still, the question of *how* remains unsolved.

Peter Mitchell's chemiosmotic coupling hypothesis was based on the premise that protein systems could harness available energy sources to actively transport ions (e.g., protons) across a membrane, thereby generating a chemical gradient and promoting the production of ATP via ATP synthase (Mitchell 1961, Ferreira and Bashford 2006). One of the most abundant energy sources available both now and 3.5+ Gya near the time of life's origin was light (Deamer and Weber 2010). Today, sunlight drives virtually all life on Earth, with ~50% of primary activity occurring in

marine environments. (Gómez-Consarnau et al. 2019, Camacho et al. 2017). Notably, light is only able to be transduced through the use of biological pigments (Deamer and Weber 2010).

Chlorophylls, the Earth's current most abundant pigments, enable light to be transformed into chemical energy through a series of steps known as the "light reactions" (Fernandez et al. 2007, Croce and van Amerongen 2014). When a photon is absorbed by chlorophyll, it becomes excited and transfers this excitation energy through a large protein complex to eventually produce an electrochemical gradient capable of manufacturing ATP with ATP synthase (Croce and van Amerongen 2014). It is by this mechanism that life on Earth continues to survive.

This begs the question that if light-driven chemical reactions are so crucial to biology now, could light have also fueled metabolisms earlier in life's history? It is possible, but the pigment receptor likely would not have been chlorophyll as it is today. The smallest chlorophyll-based photosystems necessitate thirty or more distinct genes, and the synthesis of the chlorophyll molecule itself is a complicated process requiring several enzymatic steps (Bryant and Frigaard 2006, Woodward 2009, Deamer and Weber 2010). This would necessitate a fairly sophisticated level of biosynthetic machinery and assembly. In contrast, ion-pumping rhodopsins have shown themselves to be comparatively simpler in terms of both photosystem composition and molecular synthesis (see sections 1.1 and 1.5). While the theory that ion-pumping rhodopsin systems arose before chlorophyll-based systems is favored by many (Burnetti and Ratcliff 2020, Kwon et al. 2020, DasSarma and Schwieterman 2018, Sparks et al. 2007, Hellingwerf et al. 1993, Goldsworthy 1991, Lake et al. 1985, Raven and Smith 1981), the exact time of evolution remains unknown due to a lack of preserved fossil data. It has been put forth that this mystery may never be fully solved due to both light-based photosystems evolving earlier than our current molecular clocks can decipher (Larkum et al. 2018, Lake et al. 1985).

Regardless of which system came first, understanding an early pigment-based metabolism offers some insight into early microbial bioenergetics and membranes. First, the presence and working functionality of a primitive pigment-based metabolism imply the existence of some form of a membrane. Second, we would know this membrane had the capacity to encapsulate genetic information and the catalytic macromolecules whose biosynthesis was encoded by the information (Deamer 1986, DasSarma and Schwieterman 2018). Third, the presence of a membrane protein suggests that the early membrane would not have been rigid but at least semi-permeable (Pohorille

and Deamer 2009). Finally, this membrane must have been cohesive enough to form a permeability barrier able to prevent the free diffusion of ions into or out of the cell- this would have been crucial for the generation and maintenance of an asymmetric H^+ chemiosmotic gradient (DasSarma and Schwieterman 2018, Gunner et al. 2013).

1.5 Considerations of Gene Transfer

In 1971, Oesterhelt and Stoeckenius identified the novel proton-pumping integral protein, bacteriorhodopsin, within the purple membrane of *Halobacterium halobium* (Oesterhelt and Stoeckenius 1971, 1973). Now, fifty years later, advancements in the field of genomics have permitted the discovery of over 7,000 photoactive microbial rhodopsin proteins (Hasegawa et al. 2020, Kwon et al. 2020, Govorunova et al. 2017) across the tree of life. The broad distribution of rhodopsins has exposed a complex evolutionary history that may be described in terms of independent gene loss and horizontal gene transfer (HGT). It has been suggested that a common ancestor may have possessed a light-driven rhodopsin that was able to pump ions (putatively H^+) across its membrane, then, through independent events, this gene may have been lost to some lineages while simultaneously spread to others via HGT (Sharma et al. 2007, Spudich et al. 2014, Sharma et al. 2006, Ihara et al. 1999, Kaneko et al. 2017).

Microbial rhodopsins have been labeled as ‘cosmopolitan genes’, specialty genes that appear or disappear as environmental conditions change (Bryant and Frigaard, 2006, Sharma et al. 2006). Cosmopolitan genes can be thought of as lifestyle genes- capable of aiding adaptations to new or diverse environments (Woese 2004). Rhodopsins may have earned this moniker due to their relative simplicity. As previously discussed, the rhodopsin-based photosystem is composed of one opsin protein and a single retinal chromophore. If an organism already has the ability to synthesize β -carotene, only two genes, coding for β -carotene 15,15'-monooxygenase and an opsin protein, are required to produce a functioning photosystem (Bryant 2019). If an organism does not already produce β -carotene, only six genes, five of which code for the proteins necessary to manufacture retinal, are needed (Martinez et al. 2007). Furthermore, rhodopsin and retinal biosynthetic genes have been described as genetically linked – sharing a common operon – in haloarchaeal, marine bacterial, and fungal species (Sharma et al. 2006, Martinez et al. 2007, McCarren and DeLong

2007, Prado et al. 2004). This genetic organization may further facilitate HGT amongst microbial communities.

The inheritance of a rhodopsin complex translates to the inheritance of an ion pump able to generate a chemiosmotic gradient. This gradient may then be harnessed to produce energy for the cell (ATP), enable substrate uptake, support survival, and sustain a metabolism (Gómez-Consarnau et al. 2019, Steindler et al. 2011, Gómez-Consarnau et al 2016, Gómez-Consarnau et al. 2010). It is no surprise that rhodopsin photosystems have shown themselves to be beneficial assets to many microorganisms residing in photic, nutrient-poor oligotrophic environments (Gómez-Consarnau et al. 2019, Bohorquez et al. 2012).

Given the advantageous and rather straightforward transfer of rhodopsin-based photosystems among organisms, one may wonder why microbes harboring chlorophyll-based photosystems seem to dominate most phototrophic niches. This may in part, be due to the relatively inefficient way photons are harnessed by rhodopsin-based systems as compared to chlorophylls. It could also be due to the smaller rhodopsin protein yielding a smaller absorption cross-section for its singular chromophore as compared to the larger chlorophyll reaction centers- which may hold thousands of chromophores at once (Bryant and Frigaard 2006). Interestingly, the characteristic that makes chlorophyll-based photosystems so efficient and effective, its size, is likely the reason this system is not able to readily transfer horizontally to new organisms. Even the smallest chlorophyll-based photosystems would require the successful transfer of thirty+ different genes between microbes.

CHAPTER 2: ANCIENT GENE RECONSTRUCTION AND ANALYSIS

This chapter includes work to be published in *Methods in Molecular Biology*.

Garcia, A.K., Fer, E., **Sephus, C.D.**, Kacar, B. (2021). An integrated method to reconstruct ancient proteins. *Environmental Microbial Evolution: Methods and Protocols*, Springer, New York. *Accepted.*

SUMMARY

Although it may be one step behind a TARDIS (Time and Relative Dimension in Space) and 1.21 Gigawatts less powerful than a DeLorean, implementing a combination of *in silico* analysis tools can effectively be used to travel through molecular time. Ancestral Sequence Reconstruction (ASR) approaches can be used to infer the amino acid sequences of ancient enzymes. In the first step of the reconstruction process, genetic information provided by extant homologous proteins is used to construct a multiple sequence alignment. Next, using the alignment in conjunction with an evolutionary model, a phylogenetic tree can be reconstructed using a maximum likelihood algorithm. Finally, this phylogeny may subsequently be used to deduce the peptide sequences of interior ancestral nodes.

The reconstructed sequence of the root node is often of great interest, as it represents the common ancestor of all sequences used in the study (Straub and Merkl 2019). Given that ASR reconstructs all interior nodes, it provides a means to study the evolutionary trajectory of a protein or protein family—tracing a lineage from the common ancestor to an extant organism. Additionally, when combined with *in silico*, *in vitro*, or *in vivo* analyses, ASR methodology provides a way to identify key residues that have modified a protein's properties such as structure or function through evolutionary time.

2.1 ANCESTRAL SEQUENCE RECONSTRUCTION

Proteins have played a fundamental role throughout life's history on Earth, acting as enzymatic catalysts that increase the rate of almost all cellular chemical reactions (Garcia and Kaçar 2019; Borgaonkar 2020). Despite their biological abundance, the properties and evolution of ancient proteins are seldom able to be directly studied due to a scarcity of these macromolecules that have remained intact and well-preserved over long spans of time (Thornton 2004; Hanson-Smith 2010). Ancestral Sequence Reconstruction (ASR) provides a method to infer amino acid sequences and determine the evolutionary precursors of modern-day proteins. An ASR approach allows ancient sequences to be deduced from genetic information available in extant organisms and then resurrected *in silico*, *in vitro*, or *in vivo* to elucidate ancestral structures and biochemical attributes (Thornton 2004; Kaçar et al. 2017; Garcia and Kaçar 2019).

Evolutionary biologists have long had a central goal of realizing how the functional biodiversity of modern-day genes and proteins emerged over time (Jukes and Cantor 1969, Kimura 1968, Pauling and Zuckerkandl 1963, Eck and Dayhoff 1966, Doolittle 1981, Felsenstein 2004). The typical approach to address this goal uses statistical methods to infer evolutionary histories from the genetic information present in extant organisms (Felsenstein 2004, Yang and Rannala 2012). Whereas these methods have granted some insight into the relationship between molecular evolution and function, they lack the means to empirically verify, dispute, or challenge any evolutionary inferences suggested by associated models. This limitation hinders the direct inference of ancestral molecular functionality, which otherwise might reveal the adaptive events that have shaped its evolution.

ASR coupled with the subsequent analysis of ancient proteins provides a means to test evolutionary hypotheses. ASR is a method by which ancestral molecular sequences are computationally inferred from the genetic information of extant descendants, given a model describing molecular evolution. A gene coding for an inferred ancestral protein can then be resurrected and analyzed for functionality via two main approaches: 1) synthesize and express the ancestral gene within an organism to then be studied physically at the biochemical or systems level, or 2) computationally resurrect the ancestral protein and investigate using *in silico* methods.

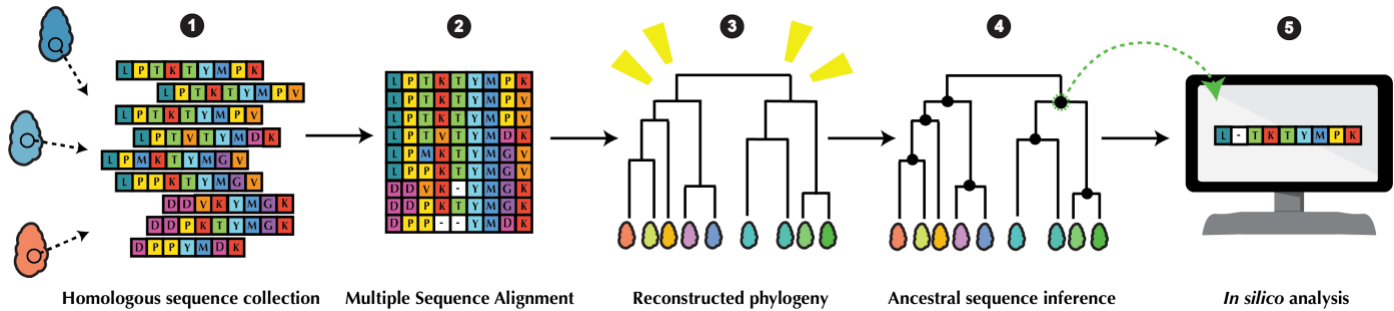


Figure 2-1. Schematic of steps used to complete ancestral sequence reconstruction. (1)

Homologous sequences of extant proteins of interest are collected. (2) Sequences are aligned in a multiple sequence alignment. (3) A phylogenetic tree is reconstructed using the multiple sequence alignment in conjunction with a best-fit evolutionary model. (4) Ancestral sequences are inferred for each interior node of the phylogeny. (5) Inferred ancestral sequences are resurrected and analyzed computationally.

2.1.1 RECONSTRUCTING PHYLOGENETIC TREES

In 1963, Linus Pauling and Emile Zuckerkandl proposed the idea that genetic information provided by extant proteins could one day be used to reconstruct ancestral amino acid sequences and infer the structures and functions of ancient proteins (Pauling and Zuckerkandl 1963). Nearly three decades later, seminal ASR experiments were conducted (Stackhouse et al. 1990, Malcom et al. 1990, Jermann et al. 1995) utilizing statistical methods based on maximum parsimony (Fitch 1971). However, the parsimony approach was restricted to sequence datasets with high similarity (Zhang and Nei 1997), limiting the maximum age of targeted molecular ancestors. Moreover, maximum parsimony methods could not incorporate more complex evolutionary models that described, for example, biases in amino acid substitution probabilities and evolutionary rate variation across branches of a phylogeny.

The statistical foundations of ASR were later refined by implementing more sophisticated models such as maximum likelihood (ML) (Felsenstein 1981, Yang 2007, Nguyen et al. 2015, Yang et al. 1995) and Bayesian inference (Rannala and Yang 1996, Pagel et al. 2004, Huelsenbeck et al 2001, Ronquist et al. 2012, Lartillot et al. 2012) in phylogenetic and ancestral sequence

reconstructions. ML algorithms aim to find the tree topology that confers the highest likelihood (probability) with respect to topology and branch lengths (Svennblad et al. 2006). In contrast, Bayesian inference methods seek to maximize posterior probabilities of putative topologies to resolve the best-fit topology (Svennblad et al. 2006).

While ML and Bayesian methods can both be used to infer tree topology, in this thesis I used and will therefore focus on, a ML methodology for phylogenetic reconstruction (**Figure 2-1**). The reconstruction process begins with the retrieval of extant homologous sequences from the protein group of interest. One of the simplest and most widely used methods to identify putative homologous sequences is performing a BLAST search through a sequence database such as NCBI or Uniprot (O’Leary et al. 2016, The UniProt Consortium 2021). Once the desirable BLAST hits have been collected and unwanted sequence data such as duplicates or partial sequences have been expelled, the extant sequences are then used to generate a multiple sequence alignment. Next, the evolutionary model that the tree will be reconstructed with must be selected. Prottest, a program that evaluates many combinations of amino acid substitution models, rate variation site models, and empirically estimated state frequencies (Darriba et al. 2011), can be used to select the best-fitting model for the given dataset. Finally, using the alignment and the best-fit evolutionary model, the tree may be reconstructed. Tree reconstruction may be executed by a number of programs, including but not limited to RAxML or IQ-TREE (Stamatakis 2014, Nguyen et al. 2015).

2.1.2 RECONSTRUCTING ANCESTRAL STATES

Much like phylogenetic reconstructions, ancestral sequence inferences can be carried out using ML or Bayesian algorithms. In the context of ancestral reconstruction, ML methods assess the likelihood that each character state at each site in an ancestral protein sequence would produce the observed extant sequence dataset, given a user-specified phylogeny and statistical evolutionary model. The Bayesian approach analyzes the posterior probabilities of each ancestral character state at each interior node on a user-specified phylogeny (Joy et al. 2016). For consistency, ancestral states and sequences in this work were reconstructed using a ML framework.

There are two main approaches to ancestral reconstruction, joint and marginal. The first,

joint reconstruction, considers all interior nodes simultaneously and calculates the character states at each node that collectively maximize the posterior probability of the data. The second, marginal reconstruction, considers interior nodes independently and calculates character states that maximize their individual posterior probabilities. This means that a marginally inferred ancestral sequence can be locally optimized but may shift the joint distribution of ancestral states away from the global optimal (Joy et al. 2016, Pupko et al. 2000). In most cases, both joint and marginal reconstructions produce consistent results.

2.2 PREDICTING BIOPHYSICAL ATTRIBUTES OF PROTEINS

Computational methods for studying membrane proteins have become increasingly sophisticated over time, solidly hitting their stride at the dawn of the 21st century (Liang et al. 2012, Nugent et al. 2017, Stansfeld 2017). Algorithms are now able to complement experimental work and offer a more comprehensive analysis of molecular structures, functions, and interactions (Almeida et al. 2017). Several *in silico* analysis tools have been developed based on machine learning software, e.g., Hidden Markov models and support vector machines (Liang et al. 2012, Tusnády and Simon 2001, Käll et al. 2004, Nugent and Jones 2009, Latek et al. 2018). These programs are able to accurately infer protein topology, structures, and protein-protein/protein-substrate interactions. Here, the suite of bioinformatic tools used to analyze, characterize, and predict the attributes of ancestral rhodopsin proteins is described.

2.2.1 PHYSICOCHEMICAL PROPERTIES

The peptide sequence of a protein serves as a blueprint for its structure, function, solubility, stability, dynamics, and interactions with other enzymes. These traits are produced through specific physical and chemical properties of the formative amino acids and their interactions with one another in the sequence space. All amino acids are separate entities that possess different chemical properties. Therefore, residue substitutions in a sequence may be presumed to affect the associated protein in some capacity (Abriata et al. 2015). Understanding the physicochemical properties of the amino acids and their interactions across the sequence can elucidate fundamental protein characteristics and benefit analyses of protein evolution (Abriata et al. 2015, Tokuriki and Tawfik 2009).

The Expert Protein Analysis System (ExPASy) ProtParam tool was used to assess the physical and chemical properties of ancestral rhodopsins (<https://www.expasy.org/resources/protparam>, Wilkins et al. 1999). Using the FASTA-formatted sequences as an input, the theoretical isoelectric points (pI), molecular weights (MW), extinction coefficients ($E_{x_{co}}$), instability indexes (II), and grand average of hydropathy (GRAVY) scores were calculated (**Supplementary-6**).

2.2.2 SUBCELLULAR LOCATION

The SOSUI server (<https://harrier.nagahama-i-bio.ac.jp/sosui/>) was used to calculate the hydropathy score and estimate the probable localization of the ancestral proteins. SOSUI is a program able to discern whether a protein is likely to be cytosolic or membrane-spanning. This web-based prediction tool is highly accurate- correctly predicting the localization of proteins 99% of the time in a benchmark test (Hirokawa et al. 1998). Predicting the localization is dependent upon the physicochemical properties of the peptide sequence submitted, especially hydrophobicity values.

2.2.3 TERTIARY STRUCTURE

A protein's three-dimensional (3D) structure can provide invaluable information regarding its biochemical function and molecular-level interactions (Bordoli et al. 2009). Elucidating the 3D structures of ancestral proteins is a complex task that would generally require the use of X-ray crystallography or Nuclear Magnetic Resonance (NMR) methodologies, both of which may be tedious, time-consuming, and expensive ventures (Satyanarayana et al. 2018). Fortunately, advancements in computational sciences have enabled the 3D structures of proteins to be reliably predicted *in silico* using homology modeling techniques (Kelley et al. 2015, Bordoli et al. 2009). Homology modeling is extensively used for bioinformatic analyses and is the method of choice for determining the 3D structure of uncharacterized protein sequences where an alignment with template structures (known structures) has a greater than a 35% sequence similarity (Bordoli et al. 2009, Satyanarayana et al. 2018).

The tertiary structure modeling of the ancestral rhodopsins was performed by Phyre2. The Phyre2 server (<http://www.sbg.bio.ic.ac.uk/phyre2>, Kelley et al. 2015) used advanced remote homology detection methods to model all 3D structures and analyze amino acid-specific effects on the protein. Homology modeling was also performed by SWISS-MODEL (<https://swissmodel.expasy.org>, Waterhouse et al. 2018) to further survey structural predictions. These results assisted our determination of ancestral rhodopsin's atomic distribution and corroborated our protein architecture inferences.

2.2.4 FUNCTIONALITY

Discerning the biochemical functionality of a protein provides a means to identify important molecular function details of an active site like substrate binding or reaction mechanisms (Fetrow and Babbitt 2018). Over the last two decades, high throughput *in silico* techniques to predict protein functions via algorithmic assignments of functional annotations have become increasingly refined and reliable. (Makrodimitris et al. 2020, Fetrow and Babbitt 2018). The technological advancements in the bioinformatic field allow us to predict the putative functionality of ancestral proteins.

Structure-based functional predictions were determined using the I-TASSER software pipeline (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>, Yang et al. 2015, Roy et al. 2010). The suite of software infers the function of the protein of interest by matching the predicted 3D model structure to previously annotated proteins in PDB. The structural analogs of the protein of interest in the Gene Ontology (GO) library are then matched based on a global similarity score using TM-align (Ashburner et al. 2000, Zhang and Skolnick 2005). The consensus of GO terms was curated based on the frequency of GO term occurrences.

CHAPTER 3: MICROBIAL RHODOPSINS RECONSTRUCTED: IMPLICATIONS ON EARLY MEMBRANES & BIOENERGETICS

This chapter includes unpublished data in preparation for journal submission.

SUMMARY

In the previous chapter, I described ancestral sequence reconstruction and other computational methods that have been developed to analyze proteins *in silico*. In this chapter, the tools previously described have been applied to study the evolutionary history and functionality of ancient microbial rhodopsin proteins. Using a combination of molecular phylogeny, ancestral sequence reconstruction, homology modeling, sequence-based, and structure-based prediction techniques, I show that the common ancestor likely had seven transmembrane helices, was able to form a protein-chromophore linkage in the retinal binding pocket, and would have transported protons across the membrane similar to many extant rhodopsins.

3.1 RESULTS

3.1.1 PHYLOGENY SHOWS AN ORGANIZATION OF RHODOPSINS BASED ON FUNCTIONALITY

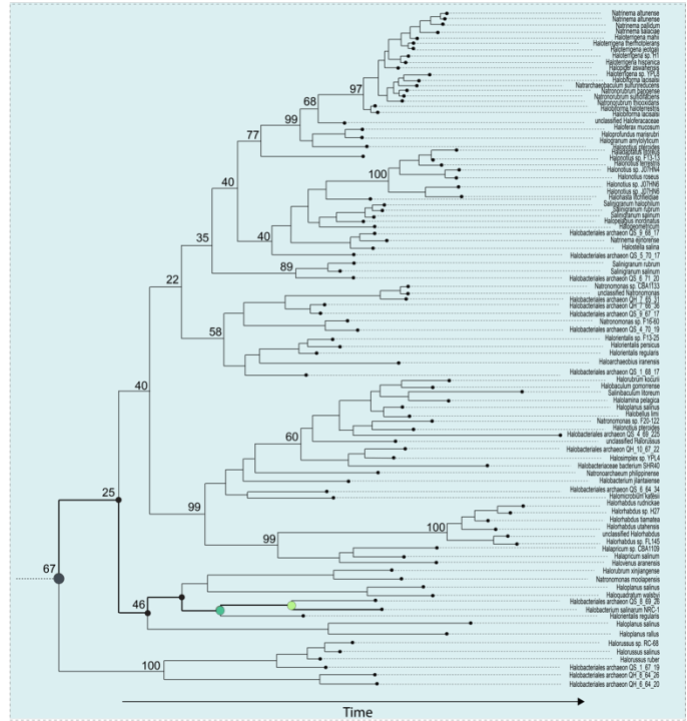
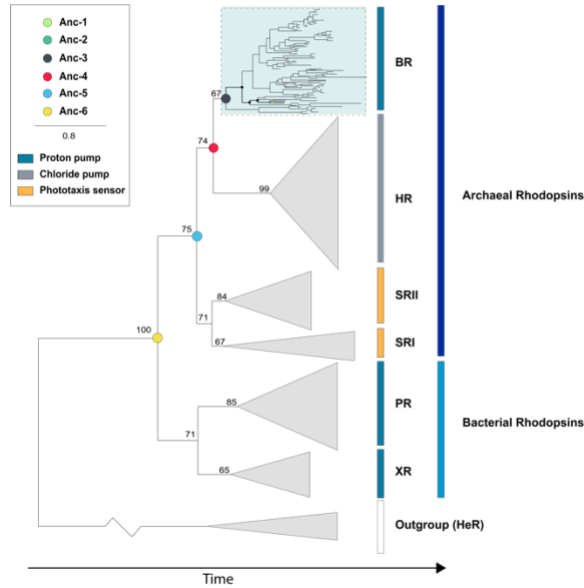


Figure 3-1. Maximum likelihood phylogeny of type-1 (microbial) rhodopsins with heliorhodopsin outgroup sequences. Abbreviations: BR, bacteriorhodopsin; HR, halorhodopsin; SR II, sensory rhodopsin II; SR I, sensory rhodopsin I; PR, proteorhodopsin; XR, xanthorhodopsin; HeR, heliorhodopsin. Scale bar indicates average amino acid substitutions per site. Outgroup branch break used to conserve space. Branch support values are derived from 100 bootstrap replicates. Amino acid sequences were aligned using MAFFT, and evolutionary distances were estimated using the LG model.

Homologous sequences for type-1 rhodopsins and heliorhodopsin proteins were identified and collected from the National Center for Biotechnology Information (NCBI) and Uniprot protein databases (O’Leary et al. 2016). The final set of homologous sequences used for ancestral reconstruction consisted of 452 unique sequences of type-1 rhodopsins and 116 unique sequences of heliorhodopsins. Four major archaeal rhodopsin groups (bacteriorhodopsin, halorhodopsin, sensory rhodopsin I, and sensory rhodopsin II) and two major bacterial groups (proteorhodopsin, xanthorhodopsin) were selected to constitute the ingroup of our analyses. Sequence selection was based on the availability of complete protein sequences and previous experimental data establishing protein functionality (Béja et al. 2001; Vogeley et al. 2004).

Heliorhodopsin proteins were selected as the outgroup to root our type-1 rhodopsin sequences. This recently discovered group of proteins constitutes a third class of proteins reported to share homology with type-1 rhodopsins (Pushkarev et al. 2018; Kovalev et al. 2020).

An alignment of the 568 protein sequences was used to generate a maximum likelihood phylogeny of type-1 rhodopsins (**Figure 3-1**). Our phylogeny displays two major lineages of type-1 rhodopsins. The first clade consists of archaeal rhodopsins, and the second consists of bacterial rhodopsins. While modern type-1 rhodopsins share a relatively low sequence identity across functional groups (**Supplementary-3**), the phylogeny supports the segregation of these groups into distinct clades, similar to those previously published (Shibata et al 2018; Pinhassi et al. 2016; Sharma et al. 2006; Ruiz-González and Marín 2004; Ihara et al. 1999; Adamian et al. 2006). The sequences within the archaeal group further separate into 4 subgroups known to differ in functionalities. (1) bacteriorhodopsin, outward proton pumps; (2) halorhodopsin, inward chloride pumps; (3) sensory rhodopsin I, positive phototaxis mediator; and (4) sensory rhodopsin II, negative phototaxis mediator (Ernst et al. 2014). Monophyly of all archaeal rhodopsin groups is supported with $\geq 67\%$ bootstrap values. Sequences within the bacterial group further segregate into 2 distinct subgroups, proteorhodopsin and xanthorhodopsin. Both groups of bacterial rhodopsins have been demonstrated to behave as proton pumps (Claassens et al. 2013; Olson et al. 2018). Experimental evidence and characterization of biochemistry and ion specificity have been obtained in a limited number of rhodopsin proteins (Pinhassi et al. 2016). The remaining sequences are assumed to have the same functionality as others within the same clade. Future biochemical and physiological experiments are needed to definitively assign a function to these uncharacterized rhodopsins. As previously reported, type-1 rhodopsin amino acid sequence alignments and phylogenetic inferences are sensitive to what sequences are included in the dataset, resulting in instances of low bootstrap support values (Pinhassi 2016), as seen in the bacteriorhodopsin clade (**Figure 3-1 - right**).

3.1.2 ANCESTRAL RETINAL BINDING SITE RESIDUES YIELD HIGH STATISTICAL SUPPORT

Ancestor	Mean ancestral site posterior probability ($\pm 1\sigma$)	Mean ancestral active-site posterior probability ($\pm 1\sigma$)
Anc-1	0.91 (± 0.16)	1.00 (± 0.00)

Anc-2	0.90 (\pm 0.18)	1.00 (\pm 0.00)
Anc-3	0.83 (\pm 0.21)	1.00 (\pm 0.01)
Anc-4	0.76 (\pm 0.24)	0.97 (\pm 0.09)
Anc-5	0.71 (\pm 0.26)	0.97 (\pm 0.08)
Anc-6	0.55 (\pm 0.29)	0.91 (\pm 0.15)

Table 3-1. Mean posterior probabilities of maximum likelihood ancestral sequences. Posterior probabilities of maximum likelihood rhodopsins are averaged across full sequences and 23 binding-site residues.

From the reconstructed phylogeny, ancestral sequences from six internal nodes were chosen as targets for protein analysis. The six targeted nodes (**highlighted in Figure 3-1**) consist of the common ancestor to extant bacterial and archaeal rhodopsins (Anc-6), the common ancestor to archaeal rhodopsins (Anc-5), the common ancestor of halorhodopsin and bacteriorhodopsin groups (Anc-4), the common ancestor to bacteriorhodopsin sequences (Anc-3), and recent ancestors to the modern bacteriorhodopsin protein isolated from *H. salinarum* *NRC-1* (Anc-2 and Anc-1). The maximum likelihood inferred reconstructed ancestral rhodopsin sequences yielded mean site posterior probabilities between ~0.55 and 0.91 (**Table 3-1**). Probability values exhibited a trend whereby ancestral sequence support decreased as phylogenetic node age increased. For example, the youngest ancestor to modern-day *H. salinarum* *NRC-1* (Anc-1) had the highest mean posterior probability, (0.91 \pm 0.16), and the oldest (Anc-6) had the lowest mean posterior probability 0.55 (\pm 0.29).

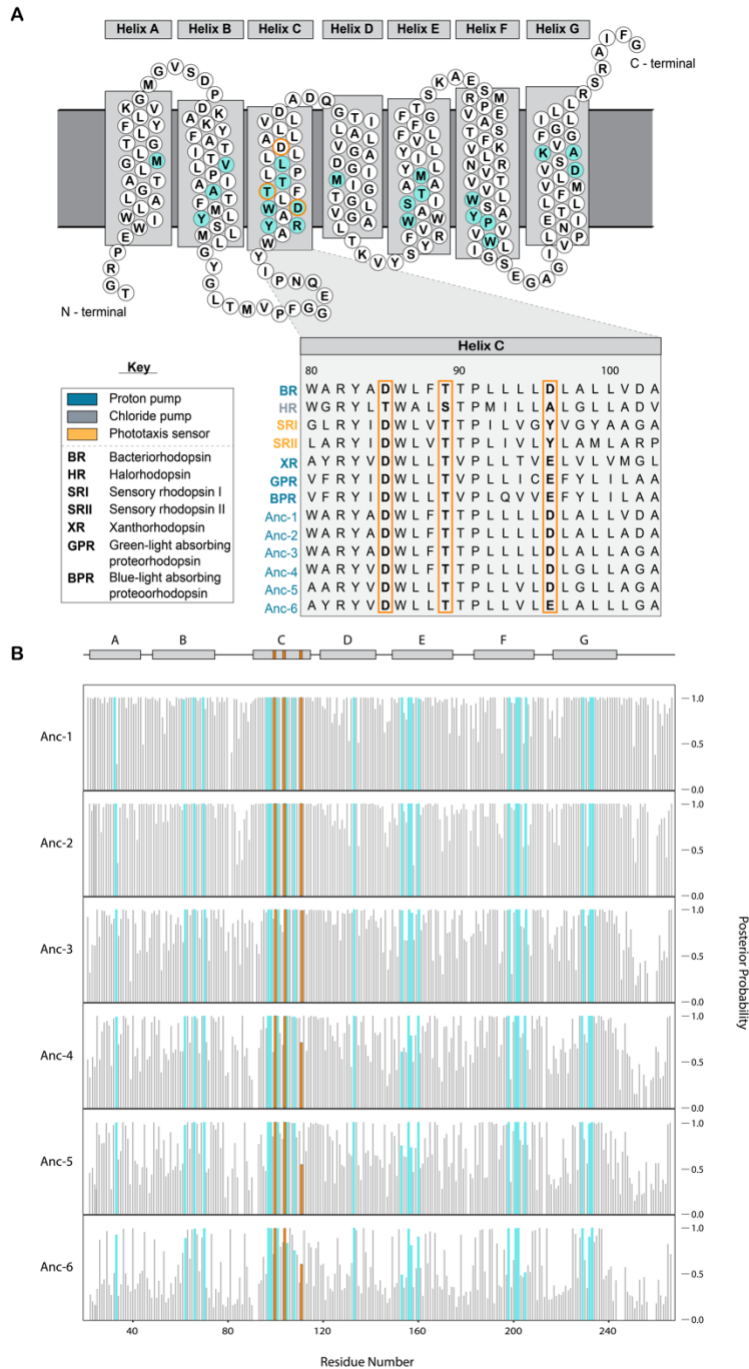


Figure 3-2. Statistical support for ancestral rhodopsin motif and active site residues.

A) Secondary structure of Bacteriorhodopsin. Single-letter residues of the BR amino acid sequence are used (PDB: 1C3W). The seven helices are labeled A-G, starting from the amino terminus. Key residues involved in retinal chromophore binding are marked in cyan. Residues denoting functional motifs are outlined in orange. **Inset: Protein alignment of rhodopsin sequences.** Multiple sequence alignment of helix C residues between extant and ancestral type-1 rhodopsins. Positions corresponding to D-85, T-89, and D-96 in bacteriorhodopsin are highlighted by orange bounding boxes.

B) Posterior probabilities of ancestral sequences. Posterior probabilities of residues involved with retinal binding are highlighted in cyan, posterior probabilities of residues corresponding to D-85, T-89, and D-96 are highlighted in orange.

In addition to analyzing the statistical support for full ancestral sequences, we analyzed support for 23 residues previously reported to form the retinal-binding pocket of bacteriorhodopsin. These residues included Met-20, Arg-82, Tyr-83 (Babitzki et al. 2009); Trp-86, Trp-138, Trp-182, Trp-189 (Greenhalgh et al. 1993; Mogi et al. 1989); Tyr-57, Tyr-185 (Mogi et al. 1987); Asp-85, Asp-212 (Greenhalgh et al. 1993; Mogi et al. 1988); Thr-89, Thr-90, Ser-141 (Marti et al. 1991); Pro-186 (Ahl et al. 1988; DeLange et al. 1998); Leu-93 (Subramaniam et al. 1991); Val-49, Ala-53, Met-118 (Greenhalgh 1993); Thr-142, Met-145, Ala-215 (Kusnetzow et al. 1999); Lys-216 (Lemke and Oesterhelt 1981; Krebs and Khorana 1993; Pushkarev et al. 2018). (Site numbering here and hereafter is based on *H. salinarum* *NRC-1* bacteriorhodopsin). These residues are not adjacent to each other- they are spread across all 7 helical structures (**Figure 3-2A - top**). Mean posterior probabilities of ancestral retinal-binding site residues ranged between ~0.91 – 1.00, invariably greater than the whole reconstructed ancestral rhodopsin sequences (**Figure 3-2B; Table 3-1**). Only 3 of the 23 retinal-binding site residues had probable alternative reconstructions with a posterior probability greater than 0.30. These included Val-49, Trp-138, and Met-145.

3.1.3 OLDEST ANCESTRAL RHODOPSIN FUNCTIONAL MOTIFS RESEMBLE PROTON PUMPS

In order to predict ancestral rhodopsin functionality, we analyzed features of extant and ancestral rhodopsin protein sequences. Specifically, we focused our attention on the third transmembrane helix (helix C). It has widely been reported that rhodopsin protein functionality can be conferred by three residues located in the third transmembrane helix (Inoue et al. 2016; Kandori 2015; Kataoka et al. 2019; Needham 2019). The archaeal proton pump bacteriorhodopsin contains a DTD motif (D-85, T-89, and D-96) (Kataoka et al. 2019). Similarly, bacterial proton pumps such as xanthorhodopsin and proteorhodopsin contain a DTE motif at these positions (Kataoka et al. 2019; Yoshizawa et al. 2014). The characterizing motifs of proton-pumping rhodopsins contain a carboxylate at the first and third residue position which act during proton transfer (Kandori 2015; Yamauchi et al. 2017). In contrast, a chloride anion pump such as halorhodopsin possesses the TSA motif which contains neutral residues only (Yamauchi et al. 2017). **Figure 3-2A - bottom** shows

inferred ancestral sequences aligned to a set of representative type-1 rhodopsin sequences which have been experimentally tested for functionality (Lozier et al. 1975; Schobert and Lanyi 1982; Spudich and Bogomolni 1984; B  j   et al. 2000; Balashov et al. 2011). The amino acid alignment of residues known to determine function in previously characterized type-1 rhodopsins is inconsistent with chloride transportation or photosensory functionality for ancestral rhodopsins. Anc-1, 2-5 contain the DTD motif. Anc-6 (the common ancestor to bacterial and archaeal type-1 rhodopsins) contains the DTE motif. The presence of these motifs suggests that early rhodopsins may have functioned similarly to that of a light-driven proton pump such as bacteriorhodopsin or proteorhodopsin.

Ancestor	Phyre Template ID (PDB ID)	Confidence	% ID	Coverage (Range)	Type (Organism)
Anc-1	D1m0ka (1M0K)	100%	85%	0.86 (17-244)	Bacteriorhodopsin (<i>Halobacterium salinarum</i>)
Anc-2	C4qidB (4QID)	100%	63%	0.86 (18-241)	Bacteriorhodopsin (<i>Haloquadratum walsbyi</i>)
Anc-3	D1c8sa (1C8S)	100%	65%	0.88 (6-220)	Bacteriorhodopsin (<i>Halobacterium salinarum</i>)
Anc-4	C4jr8A (4JR8)	100%	61%	0.92 (4-229)	Cruxrhodopsin-3 (<i>Haloarcula vallismortis</i>)
Anc-5	C4fbzA (4FBZ)	100%	52%	0.94 (3-233)	Deltarhodopsin (<i>Haloterrigena thermotolerans</i>)
Anc-6	D1c8sA (1C8S)	100%	43%	0.90 (3-222)	Bacteriorhodopsin (<i>Halobacterium salinarum</i>)

3.1.4 PROTEIN ARCHITECTURE IS CONSISTENT OVER TIME

Table 3-2. Results of homology modeling using Phyre2. Atomic-level structural models of six ancestral rhodopsin proteins (Anc-1, 2-6) were generated by comparison modeling to known rhodopsin structures using the Phyre2 server (Kelley et al. 2015). Template structures were selected based on confidence and % identity values (% ID), where confidence represents the

probability that the match between the target (submitted) and template sequences are homologous and % ID denotes the percentage identity shared between the target and template sequences.

Ancestor	PDB ID	GMQE	QMEAN	Coverage (Range)	Type (Organism)
Anc-1	2ZZL	0.75	-1.85	1.00 (17-245)	Bacteriorhodopsin (<i>Halobacterium salinarum</i>)
Anc-2	6GUY	0.77	-1.77	0.92 (10-248)	Archaerhodopsin-3 (<i>Halorubrum sodomense</i>)
Anc-3	4PXX	0.75	-2.52	1.00 (3-232)	Bacteriorhodopsin (<i>Haloarcula marismortui</i>)
Anc-4	4QI1	0.70	-2.30	0.96 (9-226)	Bacteriorhodopsin (<i>Haloquadratum walsbyi</i>)
Anc-5	3WQJ	0.70	-2.59	0.97 (7-224)	Archaerhodopsin-2 (<i>Halobacterium sp. AUS-2</i>)
Anc-6	1XJI	0.68	-3.54	0.96 (5-230)	Bacteriorhodopsin (<i>Halobacterium salinarum</i>)

Table 3-3. Results of homology modeling using SWISS-MODEL. Atomic-level structural models of six ancestral rhodopsin proteins (Anc-1, 2-6) were generated by comparison modeling to known rhodopsin structures using the SWISS-MODEL server (Waterhouse et al. 2018). Template structures were chosen according to QMEAN (Qualitative Model Energy ANalysis) and GMQE (Global Model Quality Estimation) scores, which denote the degree of nativeness and expected accuracy of the model produced by the given alignment and template respectively.

All modern-day rhodopsins are known to adopt a seven-transmembrane alpha-helical fold, sometimes referred to as the G-protein coupled receptor fold (Mackin et al. 2014). This seven-helix structure forms a pocket wherein the retinal chromophore covalently binds to an opsin protein via a Schiff base linkage to a lysine present on the seventh helix (Lys-216) (Mackin et al. 2014; Zhang et al. 2011; Ernst et al. 2014). We were interested to see if the seven-transmembrane structure or retinal binding environment changed over time.

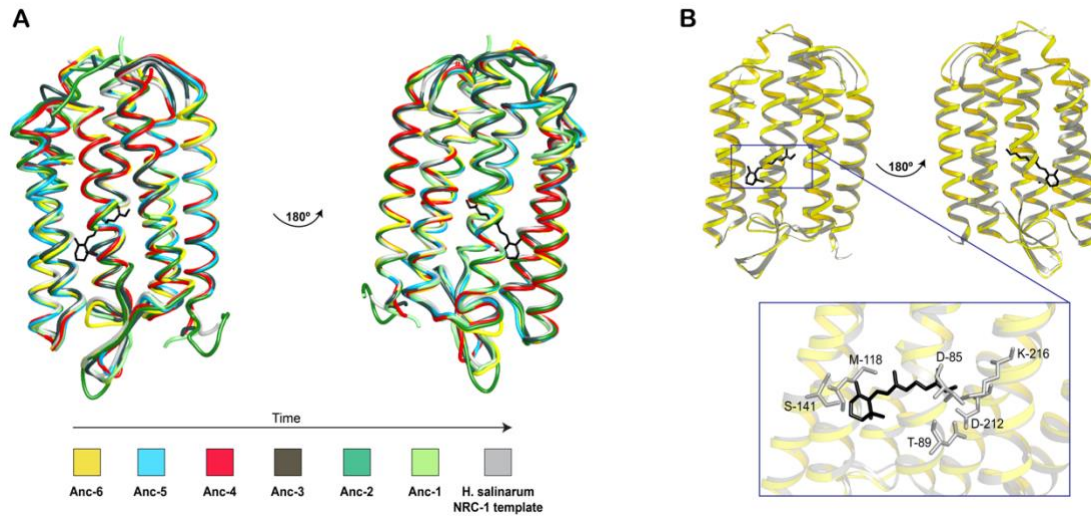


Figure 3-3. Homology models of ancestral rhodopsin tertiary structure.

A) Structural modeling of ancestral rhodopsins. Modeled protein structures of ancestral rhodopsins inferred from maximum-likelihood phylogeny aligned to an *H. salinarum* *NRC-1* structural template (PDB: 1C3W).

B) Retinal-binding pocket modeling. An interhelical view of a modeled retinal-binding pocket in ancestral type-1 rhodopsin Anc-6 aligned to *H. salinarum* *NRC-1* structural template. This view shows several amino acid residues that interact with the retinal chromophore. These include Lys-216, the retinal binding site (Pushkarev et al. 2018); Ser-141, a proximal residue to the β -ionone ring of retinal (Grigorieff et al. 1996), M-118, the binding pocket cap forming residue (Grigorieff et al. 1996); T-89, a residue that has direct contact with retinal near the Schiff base (Kandori et al. 2000); Asp-85 and Asp-212, members of the Schiff base counterion complex (Grigorieff et al. 1996).

In order to visualize the spatial distribution of ancestral structures and characterize the functions of ancestral rhodopsin proteins, we generated homology models using the Phyre2 server (Kelley et al. 2015). The structural predictions of ancestral rhodopsin proteins showed similarities to known structures of bacteriorhodopsin, cruxrhodopsin-3, and deltarhodopsin proteins (PDB ID: 1M0K, 4QID, 1C8S, 4JR8, 4FBZ). Rhodopsin proteins in extant microorganisms with the highest predicted structural homology to the inferred ancestral protein were bacteriorhodopsin from *Halobacterium salinarum* (Schobert et al. 2002; Hsu et al. 2015; Luecke et al. 1999), cruxrhodopsin-3 from *Haloarcula vallismortis* (Chan et al. 2014), and deltarhodopsin from

Haloterrigena thermotolerans (Zhang et al. 2013) (**Table 3-2**). To further explore the potential structure space of ancestral rhodopsin proteins and corroborate our Phyre2 results, we performed additional homology modeling using the SWISS-MODEL server (Waterhouse et al. 2018, **Table 3-3**). The modeled structure results from both servers revealed a consistent seven-transmembrane helical fold and formation of a retinal-binding pocket (**Figure 3-3A**). Modeling also indicated that the positioning of amino acids known to interact with the retinal chromophore such as the hallmark Lys-216 in helix G or stabilizing Thr-89 and Ser-141 in helices C and E have remained largely unchanged over time (**Figure 3-3B**).

To further characterize the functionality of the ancestral rhodopsin proteins, we performed structure-based function predictions on the reconstructed sequences using the I-TASSER server (Buchan and Jones 2019, Roy et al. 2010). Structure-based protein characterization (**Supplementary-5**) predicted that ancestral rhodopsins would have carried out phototransduction (GO: 0007602), been able to form protein-chromophore linkages (GO:0018298), and act as proton transmembrane transporters (GO: 0015992). Additionally, ancestral rhodopsins were predicted to be localized to the plasma membrane (GO:0005886) and contain components that span the hydrophobic region of the membrane (GO: 0016021), just as modern-day rhodopsin proteins do.

3.1.5 ANCESTRAL RHODOPSINS WERE MEMBRANE-SPANNING PROTEINS

To examine how the seven-helix structures may have been oriented, we generated Kyte and Doolittle hydrophathy plots for each of the six reconstructed rhodopsin proteins (Kyte and Doolittle 1982). Topology analysis of all plots reveals seven regions of high hydrophobicity, indicating a putative presence of seven transmembrane domains (**Figure 3-4**). Each hydrophobic region spans a length of 20-25 amino acids, comparable to the helical length for extant rhodopsin proteins (Rao et al. 1983, Lueke et al 2008, Ugalde et al. 2011). Additionally, the ~21-residue-long C-terminal sequence of all ancestral rhodopsins formed a distinct hydrophilic region, indicating a possible orientation towards the cytoplasm as observed with modern-day type-1 rhodopsins (Inoue et al. 2020). The conserved hydrophathy profiles of each reconstructed rhodopsin suggest that the ancestral proteins might all fold in a similar way within the membrane.

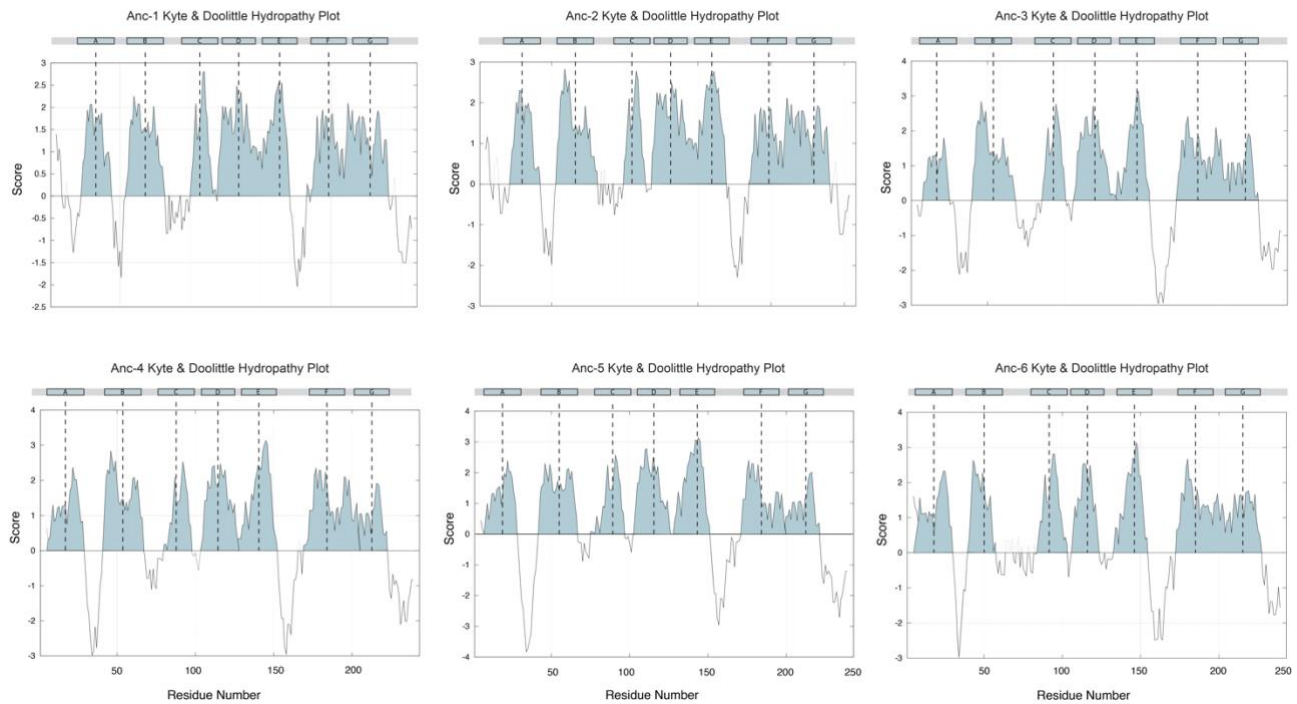


Figure 3-4. Hydropathy plots of ancestral rhodopsin proteins. Hydropathy plots using the Kyte & Doolittle hydrophobicity scale were drawn using ExPASy (Gasteiger et al. 2005, Kyte and Doolittle 1982) with a window size of 9. Amino acid residue number is presented on the x-axis and hydrophobicity values are presented on the y-axis. Hydrophobic domains corresponding to transmembrane helices are shaded. The approximate positions of the putative seven transmembrane helices are illustrated above each plot.

Ancestor	Protein localization	Average Hydrophobicity score
Anc-1	Membrane	0.769349
Anc-2	Membrane	0.814342
Anc-3	Membrane	0.532099
Anc-4	Membrane	0.670371
Anc-5	Membrane	0.562140
Anc-6	Membrane	0.668313

Table 3-4. Prediction of ancestral rhodopsin cellular localization. SOSUI server results for ancestral protein hydrophobicity and subcellular localization.

To further assess the possible subcellular localization of ancestral rhodopsin proteins, we used SOSUI, a sequence-based prediction tool able to discern whether proteins are likely to be located in the cytoplasm or embedded within the membrane based on physicochemical attributes of the amino acid sequence (Hirokawa et al. 1998, Bharat Siva Varma et al. 2015). Our SOSUI results indicated that all six of our reconstructed ancestors would have been mostly hydrophobic and localized in the membrane (Hirokawa et al. 1998, **Table 3-4**). Average sequence hydrophobicity scores for Anc-1, 2-6 were as follows: 0.769, 0.814, 0.532, 0.670, 0.562, and 0.668 respectively.

3.2 DISCUSSION

Type-1 rhodopsins are present in a diverse range of microorganisms spanning all three domains of life as well as some giant viruses (Govorunova 2017, Philosof 2013, Yutin and Koonin 2012). With the rhodopsin's seemingly ubiquitous nature, it is unclear what their original function or role in a prehistoric environment may have been. Therefore, we sought to understand the evolutionary histories of microbial rhodopsin homologs and establish the biochemical functions of the ancestral proteins. This work specifically explored the evolutionary relationships between bacteriorhodopsin, halorhodopsin, sensory rhodopsins I and II, proteorhodopsin, and xanthorhodopsin proteins. Eukaryotic microbial rhodopsin sequences were excluded from this study due to sequences contributing to long-branch attraction in phylogenetic reconstructions.

Here, we used a combination of molecular phylogeny, ancestral sequence reconstruction, homology modeling, sequence-based, and structure-based prediction techniques to investigate the evolutionary history and functionality of microbial rhodopsins. Our phylogenetic results are largely consistent with previous studies targeting microbial rhodopsins (Shibata et al 2018; Pinhassi et al. 2016; Sharma et al. 2006; Ruiz-González and Marín 2004; Ihara et al. 1999; Adamian et al. 2006). The reconstructed maximum likelihood phylogeny suggests an evolutionary history where bacterial and archaeal rhodopsins descended from a common ancestor but diverged into two independent lineages (**Figure 3-1**). Additionally, within each lineage (bacterial and archaeal), rhodopsin proteins form separate clades based on functionality. The common ancestor to the sequences used in this study indicates an early H⁺ pumping functionality, a finding that supports previous postulations (Spudich et al. 2014, Sharma et al. 2006, Inoue et al. 2016, Kaneko et al. 2017) of a precursory H⁺ pump.

In recent years, the identification of conserved motifs within the third transmembrane helix (helix C) has served as a functional litmus test for microbial rhodopsin functionalities. Outward H⁺ pumps possess a DTX motif (where the X represents D, E, K, or G), Cl⁻ pumps possess either a TSA/D or NTQ motif, and Na⁺ pumping rhodopsins feature the NDQ motif. Differences in the biochemical properties of the three-residue motifs contribute to functional variation among rhodopsins. For example, the Asp-Thr-Asp (DTD) motif of proton pumping rhodopsins contains a carboxylate at the first and third residue position which facilitates proton transfer (Kandori 2015, Yamauchi et al. 2017). In contrast, the Asn-Asp-Gln (NDQ) motif of Na⁺ pumping rhodopsins contains a single carboxylate residue (in the second residue position), and the three amino acids aid Na⁺ binding and transport kinetics (Yamauchi et al. 2017, Besaw et al. 2020, Inoue et al. 2015).

In 2016, Inoue et al. conducted a phylogenetic study focused on understanding the evolutionary history of eubacterial rhodopsins (Inoue et al. 2016). Their findings strongly suggested that bacterial rhodopsins originated as H⁺ pumps, followed by the evolution of Cl⁻ pumps, then Na⁺ pumps. Their reconstructed phylogeny delineated a basal placement of proteorhodopsin and xanthorhodopsin groups bearing the DTE motif. While our study had an expanded goal- to discern the evolutionary history of bacterial and archaeal rhodopsins- we obtained a similar result. The earliest reconstructed ancestor (Anc-6) had the DTE motif encoded within its primary sequence (**Figure 3-2**). The presence of this motif coupled with our phylogenetic inference indicates H⁺ pumping may have predated other functionalities in the archaeal lineage as well.

When an experimentally determined structure of a protein of interest is unavailable, as was the case for our ancestral sequences, predictive modeling can be used to understand the molecular architecture of a protein (Khor et al. 2015, Garcia et al. 2020). Homology modeling produces highly accurate membrane protein structures on the condition that the template (known structure) and target (unknown structure) proteins share above 30-40% sequence identity, avoiding what is known as the “twilight zone of sequence homology” (Forrest et al. 2006, Schmidt et al. 2014, Khor et al. 2015, AlQuraishi 2019). Homology models of our proteins were principally generated using the Phyre2 server (Kelley et al. 2015). Protein templates selected for modeling shared between an 85% to 43% sequence identity with our ancestral sequences. The predicted structures of all sequences folded into seven α -helices, retained a retinal-binding pocket, and given the atomic

configuration of the carboxyl groups, Asp-85 & Glu-96 (Anc-6) or Asp-85 & Asp-96 (Anc-1, 2-5), indicated a capability to transport H⁺ ions.

Mindful that our oldest ancestor's identity to the template structure had a borderline twilight zone value (43%), homology modeling was also performed with the SWISS-MODEL server (Waterhouse et al. 2018) to confirm our results. Though SWISS-MODEL selected a different template structure (PDB: 1XJI) to model Anc-6, the results were largely the same. The predicted structure had seven α -helices, an interior retinal-binding pocket, and an atomic configuration facilitative of a proton-pumping ability. Although these types of analyses are fundamental predictions, the fact that two separate algorithms predicted seven transmembrane domains, a binding pocket, and architecture conducive to proton pumping, gives us confidence in the inferred structures of ancestral rhodopsins.

Finally, to characterize the functions of ancestral rhodopsins, we performed structure-based function predictions with I-TASSER (Yang et al. 2015). The structure-based predictions strongly support our sequence-based functional predictions. The highest scoring biological functions indicated that ancestral rhodopsins would have bound a chromophore and transported protons across the cell membrane (**Supplementary-5**). Additionally, ancestral proteins were inferred to be photoreceptive integral membrane proteins (**Supplementary-5**).

Understanding the evolutionary origins of rhodopsins may provide insight into early microbial bioenergetics. Modern microorganisms have demonstrated an ability to harness proton pumping type-1 rhodopsin's simplistic infrastructure to extend viability in low nutrient environments, maintain a basal metabolism, produce ATP under anaerobic conditions, and drive electrochemical transmembrane gradients (Song et al. 2019, Gómez-Consarnau et al. 2019, Brock and Petersen 1976, Danon and Stoeckenius 1973, Waschuk et al. 2005). Our results have evidenced an ancestral integral membrane protein composed of seven helices, that was able to bind a chromophore molecule, and transport protons across a cellular membrane. It is possible that an ancestral rhodopsin may have facilitated beneficial bioenergetic processes in early photic zone microorganisms in a manner similar to today's proteorhodopsins or bacteriorhodopsins (Damer and Deamer 2020). Furthermore, the presence of a functional ancestral protein that was able to transform light to an electrochemical gradient may illuminate details of an ancient cellular

membrane. For instance, in order for an asymmetric proton gradient to be maintained, the membrane must have been cohesive enough to form a permeability barrier capable of preventing the free diffusion of protons into or out of the cell (Deamer and Nichols 1989). Additionally, the transmembrane nature of rhodopsins would have necessitated a membrane environment that was not rigid, but at least semi-permeable (Pohorille and Deamer 2009). Future laboratory experiments to assess synthetically resurrected ancestral rhodopsins may continue to improve our understanding of rhodopsin proteins and early membrane bioenergetics.

3.3 METHODS

EXTANT RHODOPSIN SEQUENCE COLLECTION

A dataset of homologous rhodopsin proteins was created by collecting type-1 rhodopsin amino acid sequences from the National Center for Biotechnology Information non-redundant protein database in December 2020 (O'Leary et al. 2016). Homologous sequences were identified and obtained via BLASTp (Camacho et al. 2009) using query amino acid sequences acquired from Uniprot (The UniProt Consortium 2019) (**Table 3-5**) and an expect value cutoff of $<1e-5$. The final dataset consisted of 568 sequences- 96 BR sequences, 43 BPR sequences, 40 GPR sequences, 144 HR sequences, 29 SRI sequences, 56 SR II sequences, 44 XR sequences, and 116 outgroup HeR sequences. The dataset was aligned using the multiple sequence alignment MAFFT algorithm (Katoh et al. 2002) and iteratively curated to remove duplicate and partial sequences. Fungal and algal type-1 rhodopsin sequences were excluded from the phylogeny because their presence contributed to long-branch attraction within the bacterial rhodopsin clade.

Rhodopsin Type	Accession Number
Bacteriorhodopsin	P02945 (BACR_HALSA)
Halorhodopsin	B0R2U4 (BACH_HALS3)
Sensory rhodopsin I	P0DMH8 (BACS1_HALSA)
Sensory rhodopsin II	P71411 (BACS2_HALSA)
Blue-light absorbing proteorhodopsin	Q9AFF7 (PRRB_PRB02)
Green-light absorbing proteorhodopsin	Q9F7P4 (PRRG_PRB01)
Xanthorhodopsin	Q2S2F8 (Q2S2F8_SALRD)

Table 3-5. Query sequences that were used to collect homologous sequences from NCBI.

PHYLOGENETIC RECONSTRUCTION AND ANCESTRAL SEQUENCE INFERENCE

The full set of 568 rhodopsin sequence MAFFT alignment was used to select the best-fit evolutionary model and amino acid substitution using IQ-TREE (Nguyen et al. 2015). The alignment data was best fit with the LG evolutionary model and gamma rate distribution according to the Akaike (Akaike 1973) and Bayesian Information Criterion (Schwartz 1978) and was used to infer a maximum likelihood phylogenetic reconstruction and ancestral amino acid states. Specifically, RAxML version 8 was used to infer the maximum likelihood tree topology, branch lengths, and evolutionary rates (Stamatakis 2014). Branch support was evaluated using 100 rapid bootstrap replicates. The tree was rooted with heliorhodopsin sequences, a protein group known to share a distant homology with type-1 rhodopsins (Pushkarev et al. 2018; Kovalev et al. 2020). The codeml program of the Phylogenetic Analysis by Maximum Likelihood (PAML) package version 4 (Yang 2007) was used to reconstruct the ancestral sequences at each node of the RAxML phylogeny. Ancestral insertion/deletion sites were reconstructed following Aadland 2019.

The full MAFFT protein alignment contained large gaps concentrated at the N-terminal. Gaps at the N-terminus were present largely due to some rhodopsin peptides possessing a signal anchor sequence of ~10-13 amino acids in this region (such as bacteriorhodopsin), while other rhodopsin groups lacked such a sequence. To verify that retaining this gapped region would not significantly alter the results of our maximum likelihood tree topology, we generated a phylogeny using a multiple sequence alignment wherein the gap-rich regions were removed. A trimmed alignment was prepared by eliminating poorly conserved regions from the alignment using TrimAl's 'gappyout' selection. The resulting phylogenetic reconstruction produced a topology equivalent to that generated by the untrimmed alignment.

HOMOLOGY MODELING OF ANCESTRAL RHODOPSIN SEQUENCES

The tertiary modeling of the ancestral rhodopsin 3D structures was performed using two different homology modeling programs, Phyre2 and SWISS-MODEL.

The reconstructed sequences inferred from six selected ancestral nodes were submitted to the Protein Homology/Analogy Recognition Engine V2.0 Phyre2 (Kelley et al. 2015) to identify the

best-suited protein template in PDB. The best template for each ancestor was selected based on sequence identity, domain coverage, E-value, and confidence scores. Based on protein searches and fold recognitions, the chosen PDB codes selected for homology modeling are as follows: 1M0K, 4QID, 1C8S, 4JR8, 4FBZ.

The SWISS-MODEL online server was also used to predict the tertiary structures of the same six ancestral rhodopsin proteins (Waterhouse et al. 2018). The best templates for each ancestor were selected based on QMQE, sequence identity, and coverage scores. Ten 3D models were generated per ancestor and the final protein model was selected based on GMQE, GMEAN, sequence identity, and coverage scores. The chosen PDB codes selected for homology modeling are as follows: 2ZZL, 6GUY, 4P XK, 4QI1, 3WQJ, 1XJI.

The homologous structures were all rendered and analyzed using UCSF Chimera (Pettersen et al. 2004).

FUNCTIONAL PREDICTION OF ANCESTRAL RHODOPSIN SEQUENCES

To understand the possible biophysical characteristics ancestral rhodopsins may have had, we used the I-TASSER to predict our sequences' most probable gene ontology (GO) terms (Yang and Zhang 2015). Structural analogs of our six selected ancestors were matched to similar proteins in the GO library based on their root mean square deviation.

PROTEIN LOCALIZATION PREDICTIONS

The sequence-based prediction tool SOSUI was used as a means to determine whether our ancestral proteins were likely to be cytoplasmic or integrated into the membrane. This theoretical approach is dependent on the inferred physicochemical properties of the amino acid sequence (Hirokawa et al. 1998).

SERVER-BASED BIOINFORMATIC TOOLS USED IN THIS STUDY:

<http://www.sbg.bio.ic.ac.uk/phyre2>

<https://swissmodel.expasy.org>

<http://zhanglab.ccmb.med.umich.edu/I-TASSER>

https://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html

3.4 REFERENCES

- Aadland, K., Pugh, C., Kolaczowski, B. (2019). High-Throughput Reconstruction of Ancestral Protein Sequence, Structure, and Molecular Function. *Methods in molecular biology*, 1851, 135–170. https://doi.org/10.1007/978-1-4939-8736-8_8
- Adamian, L., Ouyang, Z., Tseng, Y.Y., Liang, J. (2006). Evolutionary patterns of retinal-binding pockets of type I rhodopsins and their functions. *Photochemistry and photobiology*, 82, 1426–1435. <https://doi.org/10.1111/j.1751-1097.2006.tb09795.x>
- Ahl, P. L., Stern, L.J., Mogi, T., Khorana, H.G., Rothschild, K.J. (1988). Effects of amino acid substitutions in the F helix of bacteriorhodopsin. Low temperature ultraviolet/visible difference spectroscopy. *Journal of Biological Chemistry*, 263, 13594-13601. [https://doi.org/10.1016/S0021-9258\(18\)68283-3](https://doi.org/10.1016/S0021-9258(18)68283-3)
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory. *Akademia Kiado*. 267-281
- AlQuraishi, M. (2019). ProteinNet: a standardized data set for machine learning of protein structure. *BMC bioinformatics*, 20, 1-10. <https://doi.org/10.1186/s12859-019-2932-0>
- Babitzki, G., Denschlag, R., Tavan, P. (2009). Polarization effects stabilize bacteriorhodopsin's chromophore binding pocket: a molecular dynamics study. *The journal of physical chemistry B*, 113, 10483–10495. <https://doi.org/10.1021/jp902428x>
- Balashov, S. P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., Lanyi, J.K. (2005). Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science*, 309, 2061–2064. <https://doi.org/10.1126/science.1118046>
- Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S.B., Gates, C.M., Feldman, R. A., Spudich, J.L., Spudich, E.N., DeLong, E.F. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, 289, 1902–1906. <https://doi.org/10.1126/science.289.5486.1902>
- Béjà, O., Spudich, E.N., Spudich, J.L., Leclerc, M., DeLong, E.F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, 411, 786-789. <https://doi.org/10.1038/35081051>
- Besaw, J. E., Ou, W.L., Morizumi, T., Eger, B.T., Sanchez Vasquez, J.D., Chu, J., Harris, A., Brown, L.S., Miller, R., Ernst, O.P. (2020). The crystal structures of a chloride-pumping microbial rhodopsin and its proton-pumping mutant illuminate proton transfer determinants. *The Journal of biological chemistry*, 295, 14793–14804. <https://doi.org/10.1074/jbc.RA120.014118>

- Bharat Siva Varma, P., Adimulam, Y.B., Kodukula, S. (2015). *In silico* functional annotation of a hypothetical protein from *Staphylococcus aureus*. *Journal of infection and public health*, 8, 526–532. <https://doi.org/10.1016/j.jiph.2015.03.007>
- Brock, T.D., Petersen, S. (1976). Some effects of light on the viability of rhodopsin-containing halobacteria. *Archives of microbiology*, 109, 199–200. <https://doi.org/10.1007/BF00425136>
- Buchan, D., Jones, D.T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*, 47, W402–W407. <https://doi.org/10.1093/nar/gkz297>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, 10, 1-9. <https://doi.org/10.1186/1471-2105-10-421>
- Chan, S. K., Kitajima-Ihara, T., Fujii, R., Gotoh, T., Murakami, M., Ihara, K., Kouyama, T. (2014). Crystal structure of Cruxrhodopsin-3 from *Haloarcula vallismortis*. *PloS one*, 9, e108362. <https://doi.org/10.1371/journal.pone.0108362>
- Claassens, N.J., Volpers, M., dos Santos, V.A., van der Oost, J., de Vos, W.M. (2013). Potential of proton-pumping rhodopsins: engineering photosystems into microorganisms. *Trends in biotechnology*, 31, 633–642. <https://doi.org/10.1016/j.tibtech.2013.08.006>
- Consortium, The Uniprot (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47, D506–D515.
- Damer, B., Deamer, D. (2020). The Hot Spring Hypothesis for an Origin of Life. *Astrobiology*, 20, 429-452. <https://doi.org/10.1089/ast.2019.2045>
- Danon, A., Stoeckenius, W. (1974). Photophosphorylation in *Halobacterium halobium*. *Proceedings of the National Academy of Sciences*, 71, 1234–1238. <https://doi.org/10.1073/pnas.71.4.1234>
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 27, 1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>
- Deamer, D.W., Nichols, W. (1989). Proton Flux Mechanisms in Model and Biological Membranes. *The Journal of Membrane Biology*, 107, 91-103. <https://doi.org/10.1007/BF01871715>
- DeLange, F., Klaassen, C.H.W., Wallace-Williams, S.E., Bovee-Geurts, P.H.M., Liu, X.-M., DeGrip, W.J., Rothschild, K.J. (1998). Tyrosine Structural Changes Detected during the Photoactivation of

- Rhodopsin. *Journal of Biological Chemistry*, 273, 23735-23739.
<https://doi.org/10.1074/jbc.273.37.23735>
- Ernst, O. P., Lodowski, D.T., Elstner, M., Hegemann, P., Brown, L.S., Kandori, H. (2014). Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical reviews*, 114, 126–163. <https://doi.org/10.1021/cr4003769>
- Forrest, L. R., Tang, C.L., Honig, B. (2006). On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophysical journal*, 91, 508–517.
<https://doi.org/10.1529/biophysj.106.082313>
- Garcia, A. K., McShea, H., Kolaczowski, B., Kaçar, B. (2020). Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology*, 18, 394-411.
<https://doi.org/10.1111/gbi.12381>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. Walker J.M. (eds) *The Proteomics Protocols Handbook, Springer Protocols Handbooks*. <https://doi.org/10.1385/1-59259-890-0:571>
- Gómez-Conarnau, L., Raven, J.A., Levine, N.M., Cutter, L.S., Wang, D., Seegers, B., Arístegui, J., Fuhrman, J.A., Gasol, J.M., Sañudo-Wilhelmy, S.A. (2019). Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science Advances*, 5, eaaw8855.
<https://doi.org/10.1126/sciadv.aaw8855>
- Govorunova, E., Sineshchekov, O.A., Li, H., Spudich, J.L. (2017). Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications. *Annual Review of Biochemistry*, 86, 845-872.
<https://doi.org/10.1146/annurev-biochem-101910-144233>
- Greenhalgh, D. A., Farrens, D.L., Subramaniam, S., Khorana, H.G. (1993). Hydrophobic Amino Acids in the Retinal-binding Pocket of Bacteriorhodopsin. *The Journal of biological chemistry*, 268, 20305-20311. [https://doi.org/10.1016/S0021-9258\(20\)80729-7](https://doi.org/10.1016/S0021-9258(20)80729-7)
- Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., Henderson, R. (1996). Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *Journal of Molecular Biology*, 259, 393-421. <https://doi.org/10.1006/jmbi.1996.0328>
- Hirokawa, T., Boon-Chieng, S., Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14, 378–379.
<https://doi.org/10.1093/bioinformatics/14.4.378>

- Hsu, M. F., Fu, H.Y., Cai, C.J., Yi, H.P., Yang, C.S., Wang, A.H. (2015). Structural and Functional Studies of a Newly Grouped Haloquadratum walsbyi Bacteriorhodopsin Reveal the Acid-resistant Light-driven Proton Pumping Activity. *The Journal of biological chemistry*, 290, 29567–29577. <https://doi.org/10.1074/jbc.M115.685065>
- Ihara, K., Umemura, T., Katagiri, I., Kitajima-Ihara, T., Sugiyama, Y., Kimura, Y., Mukohata, Y. (1999). Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation. *Journal of Molecular Biology*, 285, 163–174. <https://doi.org/10.1006/jmbi.1998.2286>
- Inoue, K., Konno, M., Abe-Yoshizumi, R., Kandori, H. (2015). The Role of the NDQ Motif in Sodium-Pumping Rhodopsins. *Angewandte Chemie*, 127, 11698–11701. <https://doi.org/10.1002/ange.201504549>
- Inoue, K., Nomura, Y., Kandori, H. (2016). Asymmetric Functional Conversion of Eubacterial Light-driven Ion Pumps. *The Journal of biological chemistry*, 291, 9883–9893. <https://doi.org/10.1074/jbc.M116.716498>
- Inoue, K., Tsunoda, S.P., Singh, M., Tomida, S., Hososhima, S., Konno, M., Nakamura, R., Wantanabe, H., Bulzu, P.-A., Banciu, H.L., Andrei, A.-S., Uchihashi, T., Ghai, R., Bèjà, O., Kandori, H. (2020). Schizorhodopsins: A family of rhodopsins from Asgard archaea that function as light-driven inward H⁺ pumps. *Science Advances*, 6, eaaz2441. <https://doi.org/10.1126/sciadv.aaz2441>
- Kandori, H., Kinoshita, N., Yamazaki, Y., Maeda, A., Shichida, Y., Needleman, R., Lanyi, J.K., Bizounok, M., Herzfeld, J., Raap, J., Lugtenburg, J. (2000). Local and distant protein structural changes on photoisomerization of the retinal in bacteriorhodopsin. *Proceedings of the National Academy of Sciences*, 97, 4643–4648. <https://doi.org/10.1073/pnas.080064797>
- Kandori, H. (2015). Ion-pumping microbial rhodopsins. *Frontiers in molecular biosciences*, 2, 1–11. <https://doi.org/10.3389/fmolb.2015.00052>
- Kaneko, A., Inoue, K., Kojima, K., Kandori, H., Sudo, Y. (2017). Conversion of microbial rhodopsins: insights into functionally essential elements and rational protein engineering. *Biophysical reviews*, 9, 861–876. <https://doi.org/10.1007/s12551-017-0335-x>
- Kataoka, C., Inoue, K., Katayama, K., Bèjà, O., Kandori, H. (2019). Unique Photochemistry Observed in a New Microbial Rhodopsin. *Journal of Physical Chemistry Letters*, 10, 5117–5121. <https://doi.org/10.1021/acs.jpcllett.9b01957>
- Katoh, K., Misawa, K., Kuma, K., Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 259, 3059–3066.

<https://doi.org/10.1093/nar/gkf436>

- Kelley, L. A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*, 845-885.
<https://doi.org/10.1038/nprot.2015.053>
- Khor, B. Y., Tye, G.J., Lim, T.S., Choong, Y. S. (2015). General overview on structure prediction of twilight-zone proteins. *Theoretical biology & medical modelling*, *12*, 1-11.
<https://doi.org/10.1186/s12976-015-0014-1>
- Kovalev, K., Volkov, D., Astashkin, R., Alekseev, A., Gushchin, I., Haro-Moreno, J.M., Chizhov, I., Siletsky, S., Mamedov, M., Rogachev, A., Balandin, V., Borshchevskiy, V., Popov, A., Bourenkov, G., Bamberg, E., Rodriguez-Valera, F., Büldt, G., Gordeliy, V. (2020). High-resolution structural insights into the heliorhodopsin family. *Proceedings of the National Academy of Sciences*, *117*, 4131-4141. <https://doi.org/10.1073/pnas.1915888117>
- Krebs, M.P., Khorana, H.G. (1993). Mechanism of light-dependent proton translocation by bacteriorhodopsin. *Journal of Bacteriology*, *175*, 1555–1560.
<https://doi.org/10.1128/jb.175.6.1555-1560.1993>
- Kusnetzow, A., Singh, D.L., Martin, C.H., Barani, I.J., Birge, R.R. (1999). Nature of the chromophore binding site of bacteriorhodopsin: the potential role of Arg82 as a principal counterion. *Biophysical journal*, *76*, 2370–2389. [https://doi.org/10.1016/S0006-3495\(99\)77394-7](https://doi.org/10.1016/S0006-3495(99)77394-7)
- Kyte, J., Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*, 105-132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Lemke, H.-D., Oesterhelt, D. (1981). Lysine 216 is a binding site of the retinyl moiety in bacteriorhodopsin. *FEBS Letters*, *128*, 255-260. [https://doi.org/10.1016/0014-5793\(81\)80093-2](https://doi.org/10.1016/0014-5793(81)80093-2)
- Lozier, R.H., Bogomolni, R.A., Stoeckenius, W. (1975). Bacteriorhodopsin: a light-driven proton pump in Halobacterium Halobium. *Biophysical journal*, *15*, 955–962. [https://doi.org/10.1016/S0006-3495\(75\)85875-9](https://doi.org/10.1016/S0006-3495(75)85875-9)
- Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P., Lanyi, J.K. (1999). Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution. *Science*, *286*, 255–261.
<https://doi.org/10.1126/science.286.5438.255>
- Luecke, H., Schobert, B., Stagno, J., Imasheva, E.S., Wang, J.M., Balashov, S.P., Lanyi, J.K. (2008). Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proceedings of the National Academy of Sciences*, *105*, 16561-16565.
<https://doi.org/10.1073/pnas.0807162105>

- Mackin, K.A., Roy, R.A., Theobald, D.L. (2014). An empirical test of convergent evolution in rhodopsins. *Molecular Biology and Evolution*, 31, 85–95. <https://doi.org/10.1093/molbev/mst171>
- Marti, T., Otto, H., Mogi, T., Rösselet, S.J., Heyn, M.P., Khorana, H.G. (1991). Bacteriorhodopsin mutants containing single substitutions of serine or threonine residues are all active in proton translocation. *Journal of Biological Chemistry*, 266, 6919-6927. [https://doi.org/10.1016/S0021-9258\(20\)89590-8](https://doi.org/10.1016/S0021-9258(20)89590-8)
- Mogi, T., Stern, L.J. Hackett, N.R., Khorana, H.G. (1987). Bacteriorhodopsin mutants containing single tyrosine to phenylalanine substitutions are all active in proton translocation. *Proceedings of the National Academy of Sciences*, 84, 5595-5599. <https://doi.org/10.1073/pnas.84.16.5595>
- Mogi, T., Stern, L.J., Marti, T., Chao, B.H., Khorana, H.G. (1988). Aspartic acid substitutions affect proton translocation by bacteriorhodopsin. *Proceedings of the National Academy of Sciences*, 85, 4148-4152. <https://doi.org/10.1073/pnas.85.12.4148>
- Mogi, T., Marti, T., Khorana, H.G. (1989). Structure-Function Studies on Bacteriorhodopsin: IX. Substitutions of Tryptophan Residues Affect Protein-retinal Interactions in Bacteriorhodopsin. *Journal of Biological Chemistry*, 264, 14197-14201. [https://doi.org/10.1016/S0021-9258\(18\)71662-1](https://doi.org/10.1016/S0021-9258(18)71662-1)
- Mohana Rao, J. K., Hargrave, P.A., Argos, P. (1983). Will the seven-helix bundle be a common structure for integral membrane proteins? *FEBS Letters*, 156, 165-169. [https://doi.org/10.1016/0014-5793\(83\)80270-1](https://doi.org/10.1016/0014-5793(83)80270-1)
- Needham, D. M., Yoshizawa, S., Hosaka, T., Poirier, C., Choi, C.J., Hehenberger, E., Irwin, N.A.T., Wilken, S., Yung, C.-M., Bachy, C., Kurihara, R., Nakajima, Y., Kojima, K., Kimura-Someya, T., Leonard, G., Malmstrom, R.R., Mende, D.R., Olson, D.K., Sudo, Y., Sudek, S., Richards, T.A., DeLong, E.F., Keeling, P.J., Santoro, A.E., Shirouzu, M., Iwasaki, W., Worden, A.Z. (2019). A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proceedings of the National Academy of Sciences*, 116, 20574-20583. <https://doi.org/10.1073/pnas.1907517116>
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Pruitt, K.D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44, D733-745. <https://doi.org/10.1093/nar/gkv1189>
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W. ... Pruitt, K.D. (2016). Reference sequence (RefSeq) database at NCBI:

- current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*, 733-745. <https://doi.org/10.1093/nar/gkv1189>
- Olson, D.K., Yoshizawa, S., Boeuf, D., Iwasaki, W., DeLong, E.F. (2018). Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. *The ISME journal*, *12*, 1047–1060. <https://doi.org/10.1038/s41396-018-0074-4>
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*, 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Philosof, A., Bèjà, O. (2013). Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environmental Microbiology Reports*, *5*, 475-482. <https://doi.org/10.1111/1758-2229.12037>
- Pinhassi, J., DeLong, E.F., Bèjà, O., González, J.M., Pedrós-Alió, C. (2016). Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiology and Molecular Biology Reviews*, *80*, 929–954. <https://doi.org/10.1128/MMBR.00003-16>
- Pohorille, A., Deamer, D. (2009). Self-assembly and function of primitive cell membranes. *Research in Microbiology*, *160*, 449-456. <https://doi.org/10.1016/j.resmic.2009.06.004>
- Pushkarev, A., Inoue, K., Larom, S., Flores-Uribe, J., Singh, M., Konno, M., Tomida, S., Ito, S., Nakamura, R., Tsunoda, S. P., Philosof, A., Sharon, I., Yutin, N., Koonin, E. V., Kandori, H., Bèjà, O. (2018). A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature*, *558*, 595–599. <https://doi.org/10.1038/s41586-018-0225-9>
- Roy, A., Kucukural, A., Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*, 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Ruiz-González, M. X., Marín, I. (2004). New Insights into the Evolutionary History of Type 1 Rhodopsins. *Journal of Molecular Evolution*, *58*, 348–358. <https://doi.org/10.1007/s00239-003-2557-8>
- Schmidt, T., Bergner, A., Schwede, T. (2014). Modelling three-dimensional protein structures for applications in drug design. *Drug discovery today*, *19*, 890–897. <https://doi.org/10.1016/j.drudis.2013.10.027>
- Schobert, B., Lanyi, J.K. (1982). Halorhodopsin is a light-driven chloride pump. *The Journal of biological chemistry*, *257*, 10306–10313. [https://doi.org/10.1016/S0021-9258\(18\)34020-1](https://doi.org/10.1016/S0021-9258(18)34020-1)

- Schober B, C.-V. J., Hornak V, Smith S, Lanyi J. (2002). Crystallographic structure of the K intermediate of bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal. *Journal of Molecular Biology*, 321, 715-726. [https://doi.org/10.1016/s0022-2836\(02\)00681-2](https://doi.org/10.1016/s0022-2836(02)00681-2)
- Schwartz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6, 461 - 464. <https://doi.org/10.1214/aos/1176344136>
- Sharma, A. K., Spudich, J. L., Doolittle, W.F. (2006). Microbial rhodopsins: functional versatility and genetic mobility. *Trends in Microbiology*, 14, 463–469. <https://doi.org/10.1016/j.tim.2006.09.006>
- Shibata, M., Inoue, K., Ikeda, K., Konno, M., Singh, M., Kataoka, C., Abe-Yoshizumi, R., Kandori, H., Uchihashi, T. (2018). Oligomeric states of microbial rhodopsins determined by high-speed atomic force microscopy and circular dichroic spectroscopy. *Scientific reports*, 8, 8262. <https://doi.org/10.1038/s41598-018-26606-y>
- Song, Y., Cartron, M.L., Jackson, P.J., Davison, P.A., Dickman, M.J., Zhu, D., Huang, W.E., Hunter, C. N. (2019). Proteorhodopsin Overproduction Enhances the Long-Term Viability of *Escherichia coli*. *Applied and Environmental Microbiology*, 86, 1-14. <https://doi.org/10.1128/AEM.02087-19>
- Spudich, J.L., Bogomolni, R.A. (1984). Mechanism of colour discrimination by a bacterial sensory rhodopsin. *Nature*, 312, 509–513. <https://doi.org/10.1038/312509a0>
- Spudich, J.L., Sineshchekov, O.A., Govorunova, E.G. (2014). Mechanism divergence in microbial rhodopsins. *Biochimica et Biophysica Acta – Bioenergetics*, 1837, 546-552. <https://doi.org/10.1016/j.bbabi.2013.06.006>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Subramaniam, S., Greenhalgh, D.A., Rath, P., Rothschild, K.J., Khorana, H.G. (1991). Replacement of leucine-93 by alanine or threonine slows down the decay of the N and O intermediates in the photocycle of rhodopsin: implications for proton uptake and 13-cis-retinal --- all-trans-retinal reisomerization. *Proceedings of the National Academy of Sciences*, 88, 6873-6877. <https://doi.org/10.1073/pnas.88.15.6873>
- Ugalde, J.A., Podell, S., Narasingarao, P., Allen, E.E. (2011). Xenorhodopsins, an enigmatic new class of microbial rhodopsins horizontally transferred between archaea and bacteria. *Biology Direct*, 6, 1-8. <https://doi.org/10.1186/1745-6150-6-52>
- Vogele, L., Sineshchekov, O.A., Trivedi, V.D., Sasaki, J., Spudich, J.L., Luecke, H. (2004). Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science Advances*, 306, 1390-1393. <https://doi.org/10.1126/science.1103943>

- Waschuk, S. A., Bezerra, A.G., Jr, Shi, L., Brown, L.S. (2005). Leptosphaeria rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote. *Proceedings of the National Academy of Sciences*, *102*, 6879–6883. <https://doi.org/10.1073/pnas.0409659102>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*, W296–W303. <https://doi.org/10.1093/nar/gky427>
- Yamauchi, Y., Konno, M., Ito, S., Tsunoda, S. P., Inoue, K., Kandori, H. (2017). Molecular properties of a DTD channelrhodopsin from *Guillardia theta*. *Biophysics and physicobiology*, *14*, 57–66. https://doi.org/10.2142/biophysico.14.0_57
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, *12*, 7–8. <https://doi.org/10.1038/nmeth.3213>
- Yang, J., Zhang Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*, *43*, W174–W181. <https://doi.org/10.1093/nar/gkv342>
- Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W., DeLong, E.F., Kogure, K. (2014). Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proceedings of the National Academy of Sciences*, *111*, 6732–6737. <https://doi.org/10.1073/pnas.1403051111>
- Yutin, N., Koonin, E.V. (2012). Proteorhodopsin genes in giant viruses. *Biology Direct*, *7*, 1–6. <https://doi.org/10.1186/1745-6150-7-34>
- Zhang, F., Vierock, J., Yizhar, O., Fenno, L. E., Tsunoda, S., Kianianmomeni, A., Prigge, M., Berndt, A., Cushman, J., Polle, J., Magnuson, J., Hegemann, P., Deisseroth, K. (2011). The microbial opsin family of optogenetic tools. *Cell*, *147*, 1446–1457. <https://doi.org/10.1016/j.cell.2011.12.004>
- Zhang, J., Mizuno, K., Murata, Y., Koide, H., Murakami, M., Ihara, K., Kouyama, T. (2013). Crystal structure of deltarhodopsin-3 from *Haloterrigena thermotolerans*. *Proteins*, *81*, 1585–1592. <https://doi.org/10.1002/prot.24316>

3.5 SUPPLEMENTARY DATA

1 - Ancestral sequences

Anc-1 (youngest ancestor)

MLAPLSAGVWFWIGTMGMAAGTLVFFFLARNQVDEERRTFLLITALIPGIAAVSYFGMR
NVWEATGAVEVDGQPLVDA YRYVDWLLTTPLLVLELALLGADRSTIARLVAADVLM
IVLGYAGELTTGPGLRWLWGAVSTVPFLVILYLLFAELPKQVEEQDDPEVQKLFKTLRN
LVVVLWAVYPIAWLLGPAGLAGALDVAATSVGYTILDVVAKVGFVGFIIYNALDEAADE
ADEGEPAPAD

Anc-2

MLQPVDTTVWFWIGTLGMALGTLVFVYMARNEEDEERREYYVITALIPGIAAVSYLGM
ALGIGVVEVDGQNTVYAARYVDWLLTTPLLVLDLGLLAGADRNTIATLVALDVLMIVT
GFAAALTTGPALRWVWFVAVSTVAFLVILYLLFAELPEQAKEQNDDVQSLFKKLRNLTV
VLWLVPVWVWLLGPEGLGLLDVATTS LGYTVLDVTAKVGFVGFIVLNSRDTLDEAADER
PDEGGEAPPAD

Anc-3

MLQPLPTSIWLWIGTAGMALGTLFFIYMGRNVEDEERREFYVITILIPAIAAASYLGMAL
GIGLVEVEGGQDTVYWAR YVDWLFTTPLL LLDLGLLAGADRNTIATLVGLDVFMI VTG
LAAALTTAPALRWVWFAISTA AFLVILYLLFAELPEQAKEQDDDDVQSLFNTLRNLTVVL
WLVPVWVWLLGTEGLGIVGLGVTS LGYTVLDVTAKVGFVGFILLRSRGTLD EAAQEPDE
GGEAAPAD

Anc-4

MLQPGPESIW LWIGTAGMALGTL YFIYKGRGVEDEEAREFYVITILIPAIAAASYLSMAL
GFGLTEVEVGGQTQDVYWAR YADWLFTTPLL LLDLALLAGADRNTIATLVGLDVFMI V
TGLVGALTTAPAFRYVWVAISTGAFLVILYLLFSSLPEQAKEQDDDDTQSTFNTLRNLIVV

LWTVYPVVWLVGTEGLGIVGLGVETVGYTVLDVTAKVGF GFILLRSRNTLDEASQESE
GEAAPAD

Anc-5

MPLLLQTPAINKAAITGRPESIWLAI GTALMVLGTLYFMAKGWGVEDEEAKEFYVITILI
PAIAAASYLSMFLGFGLTEVELGGGQTLDIYWARYADWLFTTPLL LLDLALLADRST
ILALIGLDAFMIITGLVGALTTVFTFRFVWWAISTVAMLFILYFLFSALTAKAEEMDEDT
QSTFRVLRNLTIVLWTVYPVLWIVGTEGAGIVPLFVETLGF MVLDVTAKVGF GFILLRSR
AILGETSAPEPSGEAAAAD

Anc-6 (oldest ancestor)

MAALLQTAAEVSQAQITGRPEWIWLALGTALMFLGTLYFMVKGMGVEDPEAKKFYVI
TTLIPAIAFASYLSMLLGFGLTRVPFGGEQTLDIYWARYADWLFTTPLL LLDLALLVDAD
RSTILALIGADAFMIITGLVGALTTVFSFRFVWWAISTAAMLYILYV LFFGFTAKAEDMD
EETASTFRVLRNITIVLWSVYPVVWLIGTEGAGIVPLFVETLGF MVLDVSAKVGFGFILL
RSRAIFGEAEAPEPSGDGEAATAD

3 - Sequence Identity Charts

A

	BR	HR	SRI	SRII	BPR	GPR	XR
BR	100	29	24	29	23	22	20
HR	29	100	18	21	14	13	17
SRI	24	18	100	32	16	15	16
SRII	29	21	32	100	23	18	16
BPR	23	14	16	23	100	79	24
GPR	22	13	15	18	79	100	27
XR	20	17	16	16	24	27	100

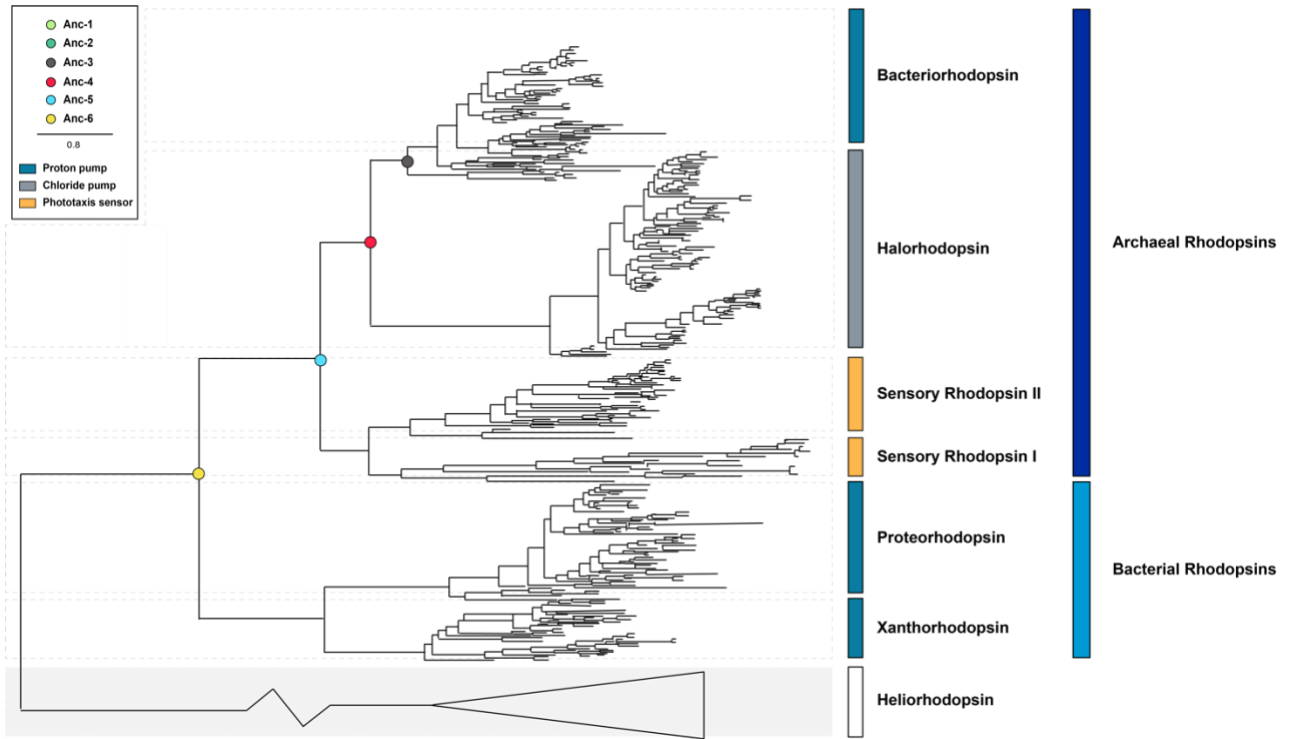
B

	BR	Anc-1	Anc-2	Anc-3	Anc-4	Anc-5	Anc-6
BR	100	83	68	56	53	44	37
Anc-1	83	100	82	63	58	47	40
Anc-2	68	82	100	71	63	52	43
Anc-3	56	63	71	100	85	70	57
Anc-4	53	58	63	85	100	82	61
Anc-5	44	47	52	70	82	100	68
Anc-6	37	40	43	57	61	68	100

A) Percent Sequence Identity Chart of Extant Rhodopsins. Seven representative extant sequences were selected from the maximum likelihood rhodopsin phylogeny: BR (P02945); HR (B0R2U4); SRI (P0DMH8); SRII (P71411); BPR (Q9AFF7); GPR (Q9F7P4); XR (Q2S2F8). The chart displays percent sequence identities shared between all seven rhodopsin sequence pairs.

B) Percent Sequence Identity Chart of Ancestral Rhodopsins. The chart expresses a pairwise sequence identity comparison between the extant bacteriorhodopsin sequence from *Halobacterium salinarum* NRC-1 (P02945) and six reconstructed ancestral rhodopsins.

4 - Expanded Phylogeny



Expanded maximum likelihood phylogeny of microbial rhodopsins and heliorhodopsin outgroup sequences from Figure 3-1.

5 – Structure-based function predictions of ancestral rhodopsins

Ancestor	Biological Process (GO Term GO-Score)	Cellular Component (GO Term GO-Score)	Molecular Function (GO Term GO-Score)
Anc-1	Phototransduction (0007602 0.99)	Plasma membrane (0005886 0.99)	Ion channel activity (0005216 0.99)
	Protein-chromophore linkage (0018298 0.99)	Integral component of membrane (0016021 0.99)	Photoreceptor activity (0009881 0.99)
	Proton transmembrane Transport (0015992 0.99)		
Anc-2	Phototransduction (0007602 0.99)	Plasma membrane (0005886 0.99)	Ion channel activity (0005216 0.99)
	Protein-chromophore linkage (0018298 0.99)	Integral component of membrane (0016021 0.99)	Photoreceptor activity (0009881 0.99)
	Proton transmembrane Transport (0015992 0.99)		
Anc-3	Phototransduction (0007602 1.00)	Plasma membrane (0005886 1.00)	Ion channel activity (0005216 1.00)
	Protein-chromophore linkage (0018298 1.00)	Integral component of membrane (0016021 1.00)	Photoreceptor activity (0009881 1.00)
	Proton transmembrane Transport (0015992 1.00)		
Anc-4	Phototransduction (0007602 0.99)	Plasma membrane (0005886 0.99)	Ion channel activity (0005216 1.00)
	Protein-chromophore linkage (0018298 0.99)	Integral component of membrane (0016021 0.99)	Photoreceptor activity (0009881 0.99)

	Proton transmembrane Transport (0015992 0.97)		
Anc-5	Phototransduction (0007602 0.99)	Plasma membrane (0005886 0.99)	Ion channel activity (0005216 1.00)
	Protein-chromophore linkage (0018298 0.99)	Integral component of membrane (0016021 0.99)	Photoreceptor activity (0009881 0.99)
	Proton transmembrane Transport (0015992 0.97)		
Anc-6	Phototransduction (0007602 1.00)	Plasma membrane (0005886 1.00)	Ion channel activity (0005216 1.00)
	Protein-chromophore linkage (0018298 1.00)	Integral component of membrane (0016021 1.00)	Photoreceptor activity (0009881 1.00)
	Proton transmembrane Transport (0015992 0.99)		

Prediction of biophysical characteristics of ancestral rhodopsins. The table shows consensus GO terms inferred from the top-scoring homologous GO templates in PDB according to I-TASSER.

6 - Physicochemical Properties of Reconstructed Rhodopsins

Calculated using ExPASy's ProtParam Tool

Bacteriorhodopsin (P02945)	
Number of Amino Acids	262
Molecular Weight	28256.26
Theoretical pI	4.58
Instability Index	27.03
Grand Average of Hydropathicity (GRAVY)	0.723
Ancestor	
Anc-1	
Number of Amino Acids	261
Molecular Weight	28549.48
Theoretical pI	4.35
Instability Index	30.28
Grand Average of Hydropathicity (GRAVY)	0.769
Extinction Coefficient (Ex _{co})	55920
Ancestor	
Anc-2	
Number of Amino Acids	258
Molecular Weight	28125.09
Theoretical pI	4.29
Instability Index	35.40
Grand Average of Hydropathicity (GRAVY)	0.814
Extinction Coefficient (Ex _{co})	54430
Ancestor	
Anc-3	
Number of Amino Acids	243
Molecular Weight	26431.45
Theoretical pI	4.12
Instability Index	32.26
Grand Average of Hydropathicity (GRAVY)	0.532
Extinction Coefficient (Ex _{co})	58900
Ancestor	
Anc-4	

Number of Amino Acids	243
Molecular Weight	26356.70
Theoretical pI	4.10
Instability Index	38.91
Grand Average of Hydropathicity (GRAVY)	0.670
Extinction Coefficient (Ex _{co})	55920

Ancestor	Anc-5
----------	-------

Number of Amino Acids	243
Molecular Weight	26486.71
Theoretical pI	4.13
Instability Index	30.78
Grand Average of Hydropathicity (GRAVY)	0.562
Extinction Coefficient (Ex _{co})	51910

Ancestor	Anc-6
----------	-------

Number of Amino Acids	243
Molecular Weight	26303.69
Theoretical pI	4.20
Instability Index	25.60
Grand Average of Hydropathicity (GRAVY)	0.668
Extinction Coefficient (Ex _{co})	55920

Instability Index values less than 40 denote a stable protein.

Extinction Coefficient (Ex_{co}) is in units of M⁻¹ • cm⁻¹ at 280 nm measured in water.

CHAPTER 4: COPYRIGHT PERMISSION

As the author of the work as it is written in chapter 3, and co-author of the work presented in chapter 2, I retain certain copyrights which allow it to be published in this thesis. The works have been referenced in chapter 2, section 1 (2.1) and chapter 3, section 1 (3.1) of this thesis.

My retained copyrights are defined on the publisher's permission request policy webpage.

Springer Nature Policy:

<https://www.nature.com/nature-research/reprints-and-permissions/permissions-requests>

Relevant copyrights include but are not limited to:

1. I retain the right to reproduce the contribution in whole or in part in any printed volume (book or thesis) of which I am the author.
2. I retain the right to reuse figures or tables I created and contained within the *Contribution in oral presentations and other works.
3. I retain the right to post a copy of the contribution as accepted for publication after peer review in my institutional repository, six months after publication of the printed or online edition of the Journal, provided it is also linked to the contribution on the publisher's website.

* The term contribution refers to my own version, not the final version as it is published in the Journal

CHAPTER 5: REFERENCES

- Abriata, L.A., Palzkill, T., Dal Peraro, M. (2015). How structural and physicochemical determinants shape sequence constraints in a functional enzyme. *PloS one*, *10*, e0118684. <https://doi.org/10.1371/journal.pone.0118684>
- Alberts, B., Johnson, A., Lewis, J., et al. (2002). The Universal Features of Cells on Earth. *Molecular Biology of the Cell*, 4th edition. <https://www.ncbi.nlm.nih.gov/books/NBK26864/>
- Allorent, G., Petroustos, D. (2017). Photoreceptor-dependent regulation of photoprotection. *Current Opinion in Plant Biology*, *37*, 102-108. <https://doi.org/10.1016/j.pbi.2017.03.016>
- Almeida, J. G., Preto, A.J., Koukos, P.I., Bonvin, A., Moreira, I.S. (2017). Membrane proteins structures: A review on computational modeling tools. *Biochimica et biophysica acta-Biomembranes*, *1859*, 2021–2039. <https://doi.org/10.1016/j.bbamem.2017.07.008>
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, *25*, 25–29. <https://doi.org/10.1038/75556>
- Balashov, S. P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., Lanyi, J.K. (2005). Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science*, *309*, 2061–2064. <https://doi.org/10.1126/science.1118046>
- Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S.B., Gates, C.M., Feldman, R. A., Spudich, J.L., Spudich, E.N., DeLong, E.F. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, *289*, 1902–1906. <https://doi.org/10.1126/science.289.5486.1902>
- Besaw, J. E., Ou, W.L., Morizumi, T., Eger, B.T., Sanchez Vasquez, J.D., Chu, J., Harris, A., Brown, L.S., Miller, R., Ernst, O.P. (2020). The crystal structures of a chloride-pumping microbial rhodopsin and its proton-pumping mutant illuminate proton transfer determinants. *The Journal of biological chemistry*, *295*, 14793–14804. <https://doi.org/10.1074/jbc.RA120.014118>
- Bogomolni, R. A., Spudich, J.L. (1982). Identification of a third rhodopsin-like pigment in phototactic *Halobacterium halobium*. *Proceedings of the National Academy of Sciences*, *79*, 6250–6254. <https://doi.org/10.1073/pnas.79.20.6250>

- Bohorquez, L. C., Ruiz-Pérez, C.A., Zambrano, M.M. (2012). Proteorhodopsin-Like Genes Present in Thermoacidophilic High-Mountain Microbial Communities. *Applied and Environmental Microbiology*, 78, 7813–7817. <https://doi.org/10.1128/AEM.01683-12>
- Bondar, A.-N., Baudry, J., Suhai, S., Fischer, S., Smith, J.C. (2008). Key Role of Active-Site Water Molecules in Bacteriorhodopsin Proton-Transfer Reactions. *Journal of Physical Chemistry B*, 112, 14729-14741. <https://doi.org/10.1021/jp801916f>
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols*, 4, 1–13. <https://doi.org/10.1038/nprot.2008.197>
- Borgaonkar, K., Patil, R. (2020). RNA as ENZYMES and comparison of its properties with PROTEINS as ENZYMES. *MedPulse International Journal of Biochemistry*, 13, 12-19. <https://doi.org/10.26611/10021314>
- Bratanov, D., Kovalev, K., Machtens, J.P., Astashkin, R., Chizhov, I., Soloviov, D., Volkov, D., Polovinkin, V., Zabelskii, D., Mager, T., Gushchin, I., Rokitskaya, T., Antonenko, Y., Alekseev, A., Shevchenko, V., Yutin, N., Rosselli, R., Baeken, C., Borshchevskiy, V., Bourenkov, G., ... Gordeliy, V. (2019). Unique structure and function of viral rhodopsins. *Nature Communications*, 10, 4939. <https://doi.org/10.1038/s41467-019-12718-0>
- Briggs, W. R., Beck, C.F., Cashmore, A.R., Christie, J.M., Hughes, J., Jarillo, J.A., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A., Okada, K., Salomon, M., Rüdiger, W., Sakai, T., Takano, M., Wada, M., Watson, J.C. (2001). The phototropin family of photoreceptors. *The Plant cell*, 13, 993–997. <https://doi.org/10.1105/tpc.13.5.993>
- Bryant, D. A., Frigaard, N.-U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology*, 14, 488-496. <https://doi.org/10.1016/j.tim.2006.09.001>
- Bryant, D. A. (2019). Phototrophy and Phototrophs. *Encyclopedia of Microbiology Fourth Edition*, 527-537.
- Burnetti, A., Ratcliff, W. (2020). The Origin of Phototrophy Reveals the Importance of Priority Effects for Evolutionary Innovation. *Preprints: 2020110700*.
- Camacho, A., Walter, X.A., Picazo, A., Zopfi, J. (2017). Photoferrotrophy: Remains of an Ancient Photosynthesis in Modern Environments. *Frontiers in microbiology*, 8, 1-17. <https://doi.org/10.3389/fmicb.2017.00323>
- Consortium, The UniProt (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49, D480-D489. <https://doi.org/10.1093/nar/gkaa1100>

- Croce, R., van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. *Nature Chemical Biology*, 10, 492-501. <https://doi.org/10.1038/nchembio.1555>
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 27, 1164-1165. <https://doi.org/10.1093/bioinformatics/btr088>
- DasSarma, S., Schwieterman, E.W. (2018). Early evolution of purple retinal pigments on Earth and implications for exoplanet biosignatures. *International Journal of Astrobiology*, 1, 1-10. <https://doi.org/10.1017/S1473550418000423>
- Deamer, D., Weber, A.L. (2010). Bioenergetics and life's origins. *Cold Spring Harbor perspectives in biology*, 2, a004929. <https://doi.org/10.1101/cshperspect.a004929>
- Deamer, D. W. (1986). Role of amphiphilic compounds in the evolution of membrane structure on the early earth. *Origins of life and evolution of the biosphere: the journal of the International Society for the Study of the Origin of Life*, 17, 3–25. <https://doi.org/10.1007/BF01809809>
- Devine, E. L., Oprian, D.D., Theobald, D.L. (2013). Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proceedings of the National Academy of Sciences*, 110, 13351–13355. <https://doi.org/10.1073/pnas.1306826110>
- Doolittle, R. (1981). Similar amino acid sequences: chance or common ancestry? . *Science*, 214, 149-159. <https://doi.org/10.1126/science.7280687>
- Eck, R. V., Dayhoff, M.O. (1966). Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. *Science*, 152, 363-366. <https://doi.org/10.1126/science.152.3720.363>
- Feldman, T. B., Smitienko, O.A., Shelaev, I.V., Gostev, F.E., Nekrasova, O.V., Dolgikh, D.A., Nadtochenko, V.A., Kirpichnikov, M.P., Ostrovsky, M.A. (2016). Femtosecond spectroscopic study of photochromic reactions of bacteriorhodopsin and visual rhodopsin. *Journal of Photochemistry and Photobiology B: Biology*, 164, 296-305. <https://doi.org/10.1016/j.jphotobiol.2016.09.041>
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17, 368-376. <https://doi.org/10.1007/bf01734359>
- Felsenstein, J. (2004). Inferring phylogenies. *Sinauer Associates, Sunderland, Mass.*

- Fernandes, T. M., Gomes, B.B., Lanfer-Marquez, U.M. (2007). Apparent absorption of chlorophyll from spinach in an assay with dogs. *Innovative Food Science & Emerging Technologies*, 8, 426-432. <https://doi.org/10.1016/j.ifset.2007.03.019>
- Ferreira, A. M., Bashford, D. (2006). Model for proton transport coupled to protein conformational change: application to proton pumping in the bacteriorhodopsin photocycle. *Journal of the American Chemical Society*, 128, 16778–16790. <https://doi.org/10.1021/ja060742d>
- Fetrow, J. S., Babbitt, P.C. (2018). New computational approaches to understanding molecular protein function. *PLoS Computational Biology*, 14, e1005756. <https://doi.org/10.1371/journal.pcbi.1005756>
- Fitch, W. M. (1971). Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Zoology*, 20, 406-416. <https://doi.org/10.2307/2412116>
- Garcia, A. K., Kaçar, B. (2019). How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radical Biology and Medicine*, 140, 260-269. <https://doi.org/10.1016/j.freeradbiomed.2019.03.033>
- Goldsworthy, A. (1991). Phycobilins. *Photoreceptor Evolution and Function*, M.G. Holmes, ed., Academic Press, London.
- Gómez-Conarnau, L., Raven, J.A., Levine, N.M., Cutter, L.S., Wang, D., Seegers, B., Arístegui, J., Fuhrman, J.A., Gasol, J.M., Sañudo-Wilhelmy, S.A. (2019). Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science Advances*, 5, eaaw8855. <https://doi.org/10.1126/sciadv.aaw8855>
- Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L., González, J. M., Pinhassi, J. (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biology*, 8, e1000358. <https://doi.org/10.1371/journal.pbio.1000358>
- Gómez-Consarnau, L., González, J.M., Riedel, T., Jaenicke, S., Wagner-Döbler, I., Sañudo-Wilhelmy, S.A., Fuhrman, J.A. (2016). Proteorhodopsin light-enhanced growth linked to vitamin-B1 acquisition in marine Flavobacteria. *The ISME journal*, 10, 1102–1112. <https://doi.org/10.1038/ismej.2015.196>
- Govorunova, E., Sineshchekov, O.A., Li, H., Spudich, J.L. (2017). Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications. *Annual Review of Biochemistry*, 86, 845-872. <https://doi.org/10.1146/annurev-biochem-101910-144233>

- Gunner, M. R., Amin, M., Zhu, X., Lu, J. (2013). Molecular mechanisms for generating transmembrane proton gradients. *Biochimica et Biophysica Acta- Bioenergetics*, 1827, 892-913.
<https://doi.org/10.1016/j.bbabi.2013.03.001>
- Gushchin, I., Gordeliy, V. (2018). Microbial Rhodopsins. *Sub-cellular biochemistry*, 87, 19–56.
https://doi.org/10.1007/978-981-10-7757-9_2
- Hampp, N. (2000). Bacteriorhodopsin as a Photochromic Retinal Protein for Optical Memories. *Chemical reviews*, 100, 1755-1775. <https://doi.org/10.1021/cr980072x>
- Hanson-Smith, V., Kolaczowski, B., Thornton, J.W. (2010). Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty. *Molecular Biology and Evolution*, 27, 1988-1999.
<https://doi.org/10.1093/molbev/msq081>
- Hasegawa, M., Hosaka, T., Kojima, K., Nishimura, Y., Nakajima, Y., Kimura-Someya, T., Shirouzu, M., Sudo, Y., Yoshizawa, S. (2020). A unique clade of light-driven proton-pumping rhodopsins evolved in the cyanobacterial lineage. *Scientific reports*, 10, 16752.
<https://doi.org/10.1038/s41598-020-73606-y>
- Hellingwerf, K. J., Crielgaard, W., Westerhoff, H.V. (1993). Comparison of Retinal-Based and Chlorophyll-Based Photosynthesis: A Biothermokinetic Description of Photochemical Reaction Centers. *Schuster S., Rigoulet M., Ouhabi R., Mazat JP. (eds) Modern Trends in Biothermokinetics. Springer, Boston, MA.* https://doi.org/10.1007/978-1-4615-2962-0_9
- Hellingwerf, K. J., Hendriks, J., Gensch, T. (2003). Photoactive Yellow Protein, A New Type of Photoreceptor Protein: Will This “Yellow Lab” Bing Us Where We Want to Go? *Journal of Physical Chemistry A*, 107, 1082-1094. <https://doi.org/10.1021/jp027005y>
- Hirokawa, T., Boon-Chieng, S., Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14, 378–379.
<https://doi.org/10.1093/bioinformatics/14.4.378>
- Huelsenbeck, J. P., Bollback, J.P., Olmstead, R. (2001). Empirical and Hierarchical Bayesian Estimation of Ancestral States. *Systematic Biology*, 50, 351-366.
<https://doi.org/10.1080/106351501300317978>
- Ihara, K., Umemura, T., Katagiri, I., Kitajima-Ihara, T., Sugiyama, Y., Kimura, Y., Mukohata, Y. (1999). Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation. *Journal of Molecular Biology*, 285, 163–174.
<https://doi.org/10.1006/jmbi.1998.2286>

- Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., Kandori, H. (2013). A light-driven sodium ion pump in marine bacteria. *Nature Communications*, 4, 1678. <https://doi.org/10.1038/ncomms2689>
- Inoue, K. (2016). The Study and Application of Photoreceptive Membrane Protein, Rhodopsin. *Bulletin of the Chemical Society of Japan*, 89, 1416-1424. <https://doi.org/10.1246/bcsj.20160235>
- Inoue, K. (2020). Shining light on rhodopsin selectivity: How do proteins decide whether to transport H⁺ or Cl⁻?. *The Journal of biological chemistry*, 295, 14805–14806. <https://doi.org/10.1074/jbc.H120.016032>
- Inoue, K., Karasuyama, M., Nakamura, R., Konno, M., Yamada, D., Mannen, K., Nagata, T., Inatsu, Y., Yawo, H., Yura, K., Béjà, O., Kandori, H., Takeuchi, I. (2021). Exploration of natural red-shifted rhodopsins using a machine learning-based Bayesian experimental design. *Communications biology*, 4, 362. <https://doi.org/10.1038/s42003-021-01878-9>
- Jermann, T. M., Opitz, J.G., Stackhouse, J., Benner, S.A. (1995). Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature*, 374, 57-59. <https://doi.org/10.1038/374057a0>
- Joy, J. B., Liang, R.H., McCloskey, R.M., Nguyen, T., Poon, A.F.Y. (2016). Ancestral Reconstruction. *PLoS Computational Biology*, 12, e1004763. <https://doi.org/10.1371/journal.pcbi.1004763>
- Jukes, T. H., Cantor, C.R. (1969). Evolution of Protein Molecules. *Munro H (ed) Mammalian Protein Metabolism, Volume III*, 21-132.
- Kacar, B., Guy, L., Smith, E., Baross, J. (2017). Resurrecting ancestral genes in bacteria to interpret ancient biosignatures. *Philosophical Transactions of the Royal Society A*, 375, 20160352. <http://dx.doi.org/10.1098/rsta.2016.0352>
- Käll, L., Krogh, A., Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338, 1027–1036. <https://doi.org/10.1016/j.jmb.2004.03.016>
- Kandori, H. (2015). Ion-pumping Microbial Rhodopsins. *Frontiers in molecular biosciences*, 2, 1-11. <https://doi.org/10.3389/fmolb.2015.00052>
- Kaneko, A., Inoue, K., Kojima, K., Kandori, H., Sudo, Y. (2017). Conversion of microbial rhodopsins: insights into functionally essential elements and rational protein engineering. *Biophysical reviews*, 9, 861–876. <https://doi.org/10.1007/s12551-017-0335-x>

- Kelley, L. A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*, 845–858.
<https://doi.org/10.1038/nprot.2015.053>
- Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, *217*, 624–626.
<https://doi.org/10.1038/217624a0>
- Kovalev, K., Volkov, D., Astashkin, R., Alekseev, A., Gushchin, I., Haro-Moreno, J.M., Chizhov, I., Siletsky, S., Mamedov, M., Rogachev, A., Balandin, V., Borshchevskiy, V., Popov, A., Bourenkov, G., Bamberg, E., Rodriguez-Valera, F., Büldt, G., Gordeliy, V. (2020). High-resolution structural insights into the heliorhodopsin family. *Proceedings of the National Academy of Sciences*, *117*, 4131–4141. <https://doi.org/10.1073/pnas.1915888117>
- Kurihara, M., Sudo, Y. (2015). Microbial rhodopsins: wide distribution, rich diversity and great potential. *Biophysics and physicochemistry*, *12*, 121–129. https://doi.org/10.2142/biophysico.12.0_121
- Kwon, S.-K., Jun, S.-H., Kim, J.F. (2020). Omega Rhodopsins: A Versatile Class of Microbial Rhodopsins. *Journal of Microbiology and Biotechnology*, *30*, 633–641.
<https://doi.org/10.4014/jmb.1912.12010>
- Lake, J. A., Clark, M.W., Henderson, E., Fay, S.P., Oakes, M., Scheinman, A., Thornber, J.P., Mah, R.A. (1985). Eubacteria, halobacteria, and the origin of photosynthesis: the photocytes. *Proceedings of the National Academy of Sciences*, *82*, 3716–3720. <https://doi.org/10.1073/pnas.82.11.3716>
- Lane, N., Allen, J.F., Martin, W. (2010). How did LUCA make a living? Chemiosmosis in the origin of life. *BioEssays*, *32*, 271–280. <https://doi.org/10.1002/bies.200900131>
- Lane, N., Martin, W.F. (2012). The Origin of Membrane Bioenergetics. *Cell*, *151*, 1406–1416.
<https://doi.org/10.1016/j.cell.2012.11.050>
- Lane, N. (2017). Proton gradients at the origin of life. *BioEssays*, *39*, 1–8.
<https://doi.org/10.1002/bies.201600217>
- Larkum, A. W. D., Ritchie, R.J., Raven, J.A. (2018). Living off the Sun: chlorophylls, bacteriochlorophylls and rhodopsins. *Photosynthetica*, *56*, 11–43. <https://doi.org/10.1007/s11099-018-0792-x>
- Lartillot, N., Lepage, T., Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, *25*, 2286–2288.
<https://doi.org/10.1093/bioinformatics/btp368>

- Larusso, N. D., Ruttenberg, B.E., Singh, A.K., Oakley, T.H. (2008). Type II Opsins: Evolutionary Origin by Internal Domain Duplication? *Journal of Molecular Evolution*, 66, 417–423. <https://doi.org/10.1007/s00239-008-9076-6>
- Latek, D., Trzaskowski, B., Niewieczerał, S., Miszta, P., Młynarczyk, K., Dębiński, A., Puławski, W., Yuan, S., Sztyler, A., Orzeł, U., Jakowiecki, J., Filipek, S. (2018). Modeling of Membrane Proteins. Liwo A. (eds) *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*, 8 (Springer Series on Bio- and Neurosystems). https://doi.org/10.1007/978-3-319-95843-9_12
- Liang, J., Naveed, H., Jimenez-Morales, D., Adamian, L., Lin, M. (2012). Computational studies of membrane proteins: Models and predictions for biological understanding. *Biochimica et Biophysica Acta – Biomembranes*, 1818, 927-941. <https://doi.org/10.1016/j.bbamem.2011.09.026>
- Luecke, H., Schobert, B., Stagno, J., Imasheva, E.S., Wang, J.M., Balashov, S.P., Lanyi, J.K. (2008). Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proceedings of the National Academy of Sciences*, 105, 16561-16565. <https://doi.org/10.1073/pnas.0807162105>
- Mackin, K. A., Roy, R.A., Theobald, D.L. (2014). An empirical test of convergent evolution in rhodopsins. *Molecular Biology and Evolution*, 31, 85–95. <https://doi.org/10.1093/molbev/mst17>
- Makrodimitris, S., van Ham, R., Reinders, M. (2020). Automatic Gene Function Prediction in the 2020's. *Genes*, 11, 1-18. <https://doi.org/10.3390/genes11111264>
- Malcolm, B. A., Wilson, K.P., Matthews, B.W., Kirsch, J.F., Wilson, A.C. (1990). Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature*, 345, 86–89. <https://doi.org/10.1038/345086a0>
- Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., DeLong, E.F. (2007). Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proceedings of the National Academy of Sciences*, 104, 5590-5595. <https://doi.org/10.1073/pnas.0611470104>
- Matsuno-Yagi, A., Mukohata, Y. (1977). Two possible roles of bacteriorhodopsin; a comparative study of strains of *Halobacterium halobium* differing in pigmentation. *Biochemical and biophysical research communications*, 78, 237–243. [https://doi.org/10.1016/0006-291x\(77\)91245-1](https://doi.org/10.1016/0006-291x(77)91245-1)
- McCarren, J., DeLong, E.F. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environmental microbiology*, 9, 846-858. <https://doi.org/10.1111/j.1462-2920.2006.01203.x>

- Michaelian, K., Simeonov, A. (2015). Fundamental molecules of life are pigments which arose and co-evolved as a response to the thermodynamic imperative of dissipating the prevailing solar spectrum. *Biogeosciences*, 12, 4913-4937. <https://doi.org/10.5194/bg-12-4913-2015>
- Mitchell, P. (1961). Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. *Nature*, 191, 144–148. <https://doi.org/10.1038/191144a0>
- Mulkidjanian, A. Y., Makarova, K.S., Galperin, M.Y., Koonin, E.V. (2007). Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nature reviews microbiology*, 5, 892–899. <https://doi.org/10.1038/nrmicro1767>
- Nakajima, Y., Tsukamoto, T., Kumagai, Y., Ogura, Y., Hayashi, T., Song, J., Kikukawa, T., Demura, M., Kogure, K., Sudo, Y., Yoshizawa, S. (2018). Presence of a Haloarchaeal Halorhodopsin-Like Cl- Pump in Marine Bacteria. *Microbes and environments*, 33, 89–97. <https://doi.org/10.1264/jsme2.ME17197>
- Nguyen, L. T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268-274. <https://doi.org/10.1093/molbev/msu300>
- Nogly, P., Weinert, T., James, D., Carbajo, S., Ozerov, D., Furrer, A., Gashi, D., Borin, V., Skopintsev, P., Jaeger, K., Nass, K., Båth, P., Bosman, R., Koglin, J., Seaberg, M., Lane, T., Kekilli, D., Brünle, S., Tanaka, T., Wu, W., ... Standfuss, J. (2018). Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science*, 361, eaat0094. <https://doi.org/10.1126/science.aat0094>
- Nowicka, B., Kruk, J. (2016). Powered by light: Phototrophy and photosynthesis in prokaryotes and its evolution. *Microbiological Research*, 186-187, 99-118. <https://doi.org/10.1016/j.micres.2016.04.001>
- Nugent, T., Jones, D.T. (2009). Transmembrane protein topology prediction using support vector machines. *BMC bioinformatics*, 10, 1-11. <https://doi.org/10.1186/1471-2105-10-159>
- Nugent, T., Jones, D., Hayat, S. (2017). Advances in Computational Methods for Transmembrane Protein Structure Prediction. *J. Rigden D. (eds) From Protein Structure to Function with Bioinformatics*, Springer, Dordrecht. https://doi.org/10.1007/978-94-024-1069-3_5
- O'Leary, N. A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W. ... Pruitt, K.D. (2016). Reference sequence (RefSeq) database at

- NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44, 733-745. <https://doi.org/10.1093/nar/gkv1189>
- Oakley, T. H., Speiser, D.I. (2015). How Complexity Originates: The evolution of Animal Eyes. *Annual Review of Ecology, Evolution, and Systematics*, 46, 237-260. <https://doi.org/10.1146/annurev-ecolsys-110512-135907>
- Oesterhelt, D., Stoeckenius, W. (1971). Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature New Biology*, 233, 149-152. <https://doi.org/10.1038/newbio233149a0>
- Oesterhelt, D., Stoeckenius, W. (1973). Functions of a new photoreceptor membrane. *Proceedings of the National Academy of Sciences*, 70, 2853-2857. <https://doi.org/10.1073/pnas.70.10.2853>
- Pagel, M., Meade, A., Barker, D., Thorne, J. (2004). Bayesian Estimation of Ancestral Character States on Phylogenies. *Systematic Biology*, 53, 673-684. <https://doi.org/10.1080/10635150490522232>
- Pauling, L., Zuckerkandl, E. (1963). Chemical Paleogenetics. Molecular “Restoration Studies” of Extinct Forms of Life. *Acta Chemica Scandinavia*, 17, 9-16. <https://doi.org/10.3891/acta.chem.scand.17s-0009>
- Pham, V. N., Kathare, P.K., Huq, E. (2018). Phytochromes and Phytochrome Interacting Factors. *Plant physiology*, 176, 1025–1038. <https://doi.org/10.1104/pp.17.01384>
- Pinhassi, J., DeLong, E.F., Béjà, O., González, J.M., Pedrós-Alió, C. (2016). Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiology and Molecular Biology Reviews*, 80, 929-954. 10.1128/MMBR.00003-16
- Pohorille, A., Deamer, D. (2009). Self-assembly and function of primitive cell membranes. *Research in Microbiology*, 160, 449-456. <https://doi.org/10.1016/j.resmic.2009.06.004>
- Porter, M. L. (2016). Beyond the Eye: Molecular Evolution of Extraocular Photoreception. *Integrative and Comparative Biology*, 56, 842–852. <https://doi.org/10.1093/icb/icw052>
- Prado, M.M., Prado-Cabrero, A., Fernández-Martín, R., Avalos, J. (2004). A gene of the opsin family in the carotenoid gene cluster of *Fusarium fujikuroi*. *Current genetics*, 46, 47-58. <https://doi.org/10.1007/s00294-004-0508-6>
- Pupko, T., Pe, I., Shamir, R., Graur, D. (2000). A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Molecular Biology and Evolution*, 17, 890-896. <https://doi.org/10.1093/oxfordjournals.molbev.a026369>

- Rannala, B., Yang, Z. (1996). Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution*, 43, 304-311.
<https://doi.org/10.1007/bf02338839>
- Raven, J. A., Smith, F.A. (1981). H⁺ transport in the evolution of photosynthesis. *Biosystems*, 14, 95-111.
[https://doi.org/10.1016/0303-2647\(81\)90025-3](https://doi.org/10.1016/0303-2647(81)90025-3)
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 6, 539-542.
<https://doi.org/10.1093/sysbio/sys029>
- Roy, A., Kucukural, A., Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5, 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Satyanarayana, S. D. V., Krishna, M.S.R., Kumar, P.P., Jeeredy, S. (2018). *In silico* structural homology modeling of nif A protein of rhizobial strains in selective legume plants. *Journal of Genetic Engineering and Biotechnology*, 16, 731-737. <https://doi.org/10.1016/j.jgeb.2018.06.006>
- Schobert, B., Lanyi, J.K. (1982). Halorhodopsin is a light-driven chloride pump. *The Journal of biological chemistry*, 257, 10306–10313. [https://doi.org/10.1016/S0021-9258\(18\)34020-1](https://doi.org/10.1016/S0021-9258(18)34020-1)
- Sharma, A. K., Spudich, J. L., Doolittle, W.F. (2006). Microbial rhodopsins: functional versatility and genetic mobility. *Trends in Microbiology*, 14, 463–469. <https://doi.org/10.1016/j.tim.2006.09.006>
- Sharma, A. K., Walsh, D.A., Baptiste, E., Rodriguez-Valera, F., Ford Doolittle, W., Papke, R.T. (2007). Evolution of rhodopsin ion pumps in haloarchaea. *BMC Evolutionary Biology*, 7, 1-13.
<https://doi.org/10.1186/1471-2148-7-79>
- Shen, L., Chen, C., Zheng, H., Jin, L. (2013). The evolutionary relationship between microbial rhodopsins and metazoan rhodopsins. *The Scientific World Journal*, 2013, 435651.
<https://doi.org/10.1155/2013/435651>
- Shevchenko, V., Mager, T., Kovalev, K., Polovinkin, V., Alekseev, A., Juettner, J., Chizhov, I., Bamann, C., Vavourakis, C., Ghai, R., Gushchin, I., Borshchevskiy, V., Rogachev, A., Melnikov, I., Popov, A., Balandin, T., Rodriguez-Valera, F., Manstein, D. J., Bueltdt, G., Bamberg, E., ... Gordeliy, V. (2017). Inward H⁺ pump xenorhodopsin: Mechanism and alternative optogenetic approach. *Science Advances*, 3, e1603187. <https://doi.org/10.1126/sciadv.1603187>
- Sparks, W. B., DasSarma, S., Reid, I.N. (2007). Evolutionary competition between primitive systems: existence of an early purple Earth? *Bulletin of the American Astronomical Society*, 38, 901.

- Spudich, J. L., McCain, D.A., Nakanishi, K., Okabe, M., Shimizu, N., Rodman, H., Honig, B., Bogomolni, R.A. (1986). Chromophore/protein interaction in bacterial sensory rhodopsin and bacteriorhodopsin. *Biophysical journal*, 49, 479–483. [https://doi.org/10.1016/S0006-3495\(86\)83657-8](https://doi.org/10.1016/S0006-3495(86)83657-8)
- Spudich, J. L., Sineshchekov, O.A., Govorunova, E.G. (2014). Mechanism divergence in microbial rhodopsins. *Biochimica et Biophysica Acta – Bioenergetics*, 1837, 546-552. <https://doi.org/10.1016/j.bbabi.2013.06.006>
- Stackhouse, J., Presnell, S.R., McGeehan, G.M., Nambiar, K.P., Benner, S.A. (1990). The ribonuclease from an extinct bovid ruminant. *FEBS Letters*, 262, 104-106. [https://doi.org/10.1016/0014-5793\(90\)80164-e](https://doi.org/10.1016/0014-5793(90)80164-e)
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stansfeld, P. J. (2017). Computational studies of membrane proteins: from sequence to structure to simulation. *Current Opinion in Structural Biology*, 45, 133–141. <https://doi.org/10.1016/j.sbi.2017.04.004>
- Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., Giovannoni, S.J. (2011). Energy starved *Candidatus Pelagibacter ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PloS one*, 6, e19725. <https://doi.org/10.1371/journal.pone.0019725>
- Straub, K., Merkl, R. (2019). Ancestral Sequence Reconstruction as a Tool for the Elucidation of a Stepwise Evolutionary Adaptation. *Methods in molecular biology*, 1851, 171–182. https://doi.org/10.1007/978-1-4939-8736-8_9
- Svennblad, B., Erixon, P., Oxelman, B., Britton, T. (2006). Fundamental Differences Between the Methods of Maximum Likelihood and Maximum Posterior Probability in Phylogenetics. *Systematic Biology*, 55, 116-121. <https://doi.org/10.1080/10635150500481648>
- Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Reviews Genetics*, 5, 366-375. <https://doi.org/10.1038/nrg1324>
- Tokuriki, N., Tawfik, D.S. (2009). Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, 19, 596–604. <https://doi.org/10.1016/j.sbi.2009.08.003>
- Tusnady, G. E., Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *17*, 849–850. <https://doi.org/10.1093/bioinformatics/17.9.849>

- van der Horst, M. A., Hellingwerf, K.J. (2004). Photoreceptor Proteins, “Star Actors of Modern Times”: A Review of the Functional Dynamics in the Structure of Representative Members of Six Different Photoreceptor Families. *Accounts of chemical research*, 37, 13-20.
<https://doi.org/10.1021/ar020219d>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46, W296–W303.
<https://doi.org/10.1093/nar/gky427>
- Weiss, M. C., Preiner, M., Xavier, J.C., Zimorski, V., Martin, W.F. (2018). The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS genetics*, 14, e1007518.
<https://doi.org/10.1371/journal.pgen.1007518>
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology*, 112, 531–552. <https://doi.org/10.1385/1-59259-584-7:531>
- Woese, C.R. (2004). A New Biology for a New Century. *Microbiology and Molecular Biology Reviews*, 68, 173-186.
- Woodward, R.B. (1961). The total synthesis of chlorophyll. *Pure and Applied Chemistry*, 2, 383-404.
<https://doi.org/10.1351/pac196102030383>
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, 12, 7–8. <https://doi.org/10.1038/nmeth.3213>
- Yang, Z., Kumar, S., Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, 141, 1641-1650. <https://doi.org/10.1093/genetics/141.4.1641>
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586-1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z., Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, 13, 303-314. <https://doi.org/10.1038/nrg3186>
- Zhang, J., Nei, M. (1997). Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *Journal of Molecular Evolution*, 44, S139-S146.
<https://doi.org/10.1007/pl00000067>
- Zhang, Y., Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33, 2302–2309. <https://doi.org/10.1093/nar/gki524>