Functional divergence and spectral tuning of microbial rhodopsins from an ancestral proton pump

Cathryn D. Sephus[1], Evrim Fer[2,3], Amanda K. Garcia[3], Zachary R. Adam[4,5], Edward W. Schwieterman[5,6], Betül Kaçar[3*]

1. Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA
2. Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA
3. Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
4. Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA
5. Blue Marble Space Institute of Science, Seattle, WA, USA
6. Department of Earth and Planetary Sciences, University of California, Riverside, CA, USA

*To whom correspondence should be addressed: bkacar@wisc.edu

Running title: Functional divergence and spectral tuning of microbial rhodopsins from an ancestral proton pump

Keywords: early Earth; phototrophy; rhodopsin; radiative transfer; origins of life; biosignature; habitability

ABSTRACT

For billions of years, life has continuously adapted to dynamic physical conditions near the Earth's surface. Fossils and other preserved biosignatures in the paleontological record are the most direct evidence for reconstructing the broad historical contours of this adaptive interplay. However, biosignatures dating to Earth's earliest history are exceedingly rare. Here, we combine phylogenetic inference of primordial rhodopsin proteins with modeled spectral features of the Precambrian Earth environment to reconstruct the paleobiological history of this essential family of photoactive transmembrane proteins. Our results suggest that ancestral microbial rhodopsins likely acted as light-driven proton pumps and were spectrally tuned toward the absorption of green light, which would have enabled their hosts to occupy depths in a water column or biofilm where UV wavelengths were attenuated. Subsequent diversification of rhodopsin functions and peak absorption frequencies track the diversification of surface ecological niches induced by the accumulation of atmospheric oxygen. Inferred ancestors retain distinct associations between extant functions and peak absorption frequencies. Our findings suggest that novel information encoded by biomolecules can be used as paleosensors for conditions of ancient, inhabited niches of host organisms not represented elsewhere in the paleontological record. The coupling of functional diversification and spectral tuning of this pervasive protein family underscores the utility of rhodopsins as universal testbeds for inferring remotely detectable biosignatures on inhabited planetary bodies.

INTRODUCTION

Decoding the complex relationships between life and the environments it inhabits is central to reconstructing the contours of our planet's history of habitability. These contours are critical for constraining the physical limits of biological functionality, identifying possibly 'universal' biological behavioral modes, and aiding the search for life elsewhere in our universe[1,2]. Planetary habitability is strongly influenced by solar insolation and photon radiance flux, which couple planetary and stellar evolution to the pervasiveness of life[3,4]. The Earth and sun, however, each possess idiosyncratic internal structures, stability points, and observational limitations that obfuscate the reconstruction of long-term habitability parameters.

The information encoded in life itself may provide novel insights into how our planet has maintained planetary habitability where geologic and stellar inferences fall short. The earliest

paleontological record lacks abundant, direct evidence of major metabolic and morphogenetic features[5,6], but the contemporary bioinformatic record is incredibly rich and includes several environment-sensitive biomolecular components. Reconstructing the ancestral sequences of biomolecules that are strongly coupled to habitability parameters may provide a means of extrapolating ancient biological phenotypes—and the environmental conditions that shaped them—back to the Last Universal Common Ancestor (LUCA)[7-9].

Photosensitive enzymes are key intermediaries that connect intracellular chemical states, extracellular substrate availability, and solar irradiance. These biomolecules constitute a promising system for tracking ancient physical parameters not recorded in the geologic record. All known phototrophic metabolisms on Earth rely on one of three energy-converting pigments which transform light energy into chemical energy. These pigments include chlorophylls, bacteriochlorophylls, and retinal[10]. Retinal-based pigment proteins, known as rhodopsins, have been found in Archaea, Bacteria, Eukarya, and giant viruses[11,12]. Rhodopsin pervasiveness was initially attributed to high rates of gene transfer between microbial organisms[13], but subsequent analyses indicate that a rhodopsin-like functionality may have been present in LUCA[14].

Retinal-based phototrophy utilizes a light-capture system that is simpler than its chlorophyll-based counterparts[15]. Rhodopsin pigments are composed of one seven-transmembrane opsin protein and one retinal chromophore[16]. Photoisomerization of the retinal chromophore within the opsin active site (also referred to as the retinal binding pocket) triggers a series of cyclical conformational changes that can directly drive ion transport across the cellular membrane, generating a proton-motive force that may be harnessed by ATP synthase to generate ATP[17]. The simple bipartite architecture of rhodopsins has also been harnessed by biology for uses beyond ion transport, including both phototaxis and photophobia, substrate uptake, and starvation prevention[18-21]. Rhodopsins can be tuned to absorb photons across much of the visible light spectrum[22] and have been sampled from exposed terrestrial surfaces to the sub-photic marine zone[23]. The structural simplicity and functional diversity of microbial (type I) rhodopsins may have promoted their spread to global econiches[24]. Today, type I rhodopsins are among the dominant energy-transducing agents that harvest solar energy in the surface ocean[10]. The combination of cladistic pervasiveness, phylogenetic tractability, and niche ubiquity indicates that ancestral type I rhodopsins are excellent candidate phylogenetic systems for characterizing long-term conditions of habitability and paleoecology where geologic indicators are unavailable[25-27].

We reconstructed the evolutionary history of microbial type I rhodopsins to explore ancestral properties that might serve as a guide for understanding early Earth paleoecology and habitability. Protein sequence analyses of both extant and phylogenetically inferred ancestral rhodopsins were used to constrain possible ancestral rhodopsin functions. In addition, maximum absorbance wavelength of ancestral rhodopsins was predicted by a machine-learning algorithm that models the relationship between spectral absorbance peaks and amino acid sequence[28]. Our analyses suggest proton-pump functionality and green wavelength absorption for the last common ancestor of ancestral type I rhodopsins. Green wavelength absorption is parsimonious with our modeled spectral features of the Precambrian Earth environment, providing an adaptive advantage to shallow coastal marine niches with abundant nutrient and metabolically useful light availability but attenuated phototoxic UV irradiation. The proliferation of specialized functions and the spectral tuning of rhodopsins tracks ecological upheaval wrought by the rise of atmospheric oxygen, with significant accumulation marked by the interval known as the Great Oxidation Event (GOE) approximately 2.3 billion years ago[29,30]. The pattern of niche proliferation imprinted upon the rhodopsin protein network is broadly consistent with trends in oxygen-sensitive stable isotope signatures and mineral distributions throughout the Precambrian[29,31]. The convergence of our results from reconstructed, ancestral biomolecular characteristics and indirect physical models constrains the manner by which phototrophic organisms have adapted to changing habitability parameters throughout Earth's macroevolutionary history.

RESULTS AND DISCUSSION

Conservation and statistical support for ancestral rhodopsin residues important for function and spectral tuning

We assembled a dataset of representative protein sequences with which to reconstruct the evolutionary history and ancestral traits of microbial rhodopsins. BLASTp[32] was used to identify 452 homologous type I rhodopsin sequences from the National Center for Biotechnology Information (NCBI) non-redundant and UniProt protein databases. We additionally identified 116 homologs belonging to heliorhodopsins, a recently discovered rhodopsin class that exhibits an inverted membrane topology relative to type I rhodopsins and constitutes a distinct outgroup[33-35]. Animal, or type II, rhodopsins, though exhibiting a similar seven transmembrane architecture, possess little sequence homology with microbial type I rhodopsins and were thus excluded from the dataset[13,35]. Our type I rhodopsin sequence dataset comprises four major archaeal rhodopsin

125 groups: 1) bacteriorhodopsin, outward proton pumps; 2) halorhodopsin, inward chloride pumps;
126 3) sensory rhodopsin I, positive phototaxis mediators; and 4) sensory rhodopsin II, negative
127 phototaxis mediators[36]. The dataset also includes two major bacterial rhodopsin groups: 1)
128 proteorhodopsin and 2) xanthorhodopsin, both outward proton pumps, with the latter
129 distinguished by the ability to bind a second carotenoid chromophore[36]. These taxonomically and
130 functionally distinct lineages were resolved into their respective clades by maximum likelihood
131 phylogenetic reconstruction (Fig. 1), producing a tree topology largely in agreement with previous
132 analyses[11,37,38]. Alternate phylogenetic trees, generated by incorporating next-best-fit evolutionary
133 models, reproduced this general topology (see Methods; Supplementary Fig. 1). Ancestral amino
134 acid sequences were inferred by maximum likelihood from internal nodes of our rhodopsin
135 phylogeny. Five of the oldest internal nodes were selected as targets for further phenotypic
136 analysis: 1) Anc1, ancestral to bacteriorhodopsins and halorhodopsins, 2) Anc2, ancestral to
137 sensory rhodopsins, 3) Anc3, ancestral to archaeal rhodopsins, 4) Anc4, ancestral to bacterial
138 rhodopsins, and 5) Anc5, ancestral to all type I rhodopsins (Fig. 1, Supplementary Fig. 2).
139
140 Due to their deep position in the phylogenetic tree, mean site posterior probabilities for ancestors
141 Anc1-Anc5 range between 0.55 and 0.76 (Fig. 2a; Supplementary Table 1). However, statistical
142 support is enriched for 23 ancestral binding pocket residues (<5 Å from the retinal chromophore),
143 constituting the most conserved element of microbial rhodopsin structure[39]. Mean posterior
144 probabilities for binding pocket residues fall between 0.90 and 0.97 (Supplementary Table 1, 2).
145 Nine of these residues have ≥95% mean pairwise identity across all extant sequences included
146 in our dataset (Y83, W86, P91, G122, W182, Y185, W189, D212, and K216 in the
147 bacteriorhodopsin of *Halobacterium salinarum*). Most have significant, experimentally determined
148 contributions to rhodopsin function[39-43], including the strictly conserved K216 residue that is
149 covalently bonded to retinal via a Schiff base linkage[40]. Other functions include maintenance of
150 the internal charge environment for ion-pumping rhodopsins[39,44] or mediation of transducer
151 interactions required for phototactic response in sensory rhodopsins[45]. Finally, binding pocket
152 residues have been implicated as having outsized influence in rhodopsin color tuning[46,47]. Thus,
153 the enriched statistical support for ancestral binding pocket residues permits sequence-based
154 inferences of historical rhodopsin functionality.
155
156
157
158

159 **Ancestral type I rhodopsins likely acted as light-driven proton pumps**
160
161 Comparative analysis of extant variants provides compelling circumstantial evidence that
162 primordial rhodopsins functioned as transporters but the evolutionary progression of functions or
163 of photon absorptivity have not been explicitly demonstrated to date[37]. As an initial, global
164 approach to constraining early rhodopsin functionality, 3D protein structures of maximum
165 likelihood ancestral sequences Anc1-Anc5 were predicted with Phyre2[51] and used for functional
166 annotation by the COFACTOR algorithm within the I-TASSER software pipeline[52]. Modeled
167 ancestral rhodopsin structures preserve the characteristic seven-transmembrane alpha-helical
168 fold, otherwise known as the G-protein coupled receptor fold[16]—of extant microbial type I
169 homologs (Fig. 3). Kyte and Doolittle hydropathy plots for Anc1-Anc5 reveal seven highly
170 hydrophobic regions, consistent with the conservation of seven transmembrane domains
171 (Supplementary Fig. 3). Membrane localization of ancestral rhodopsins was also predicted by the
172 SOSUI tool[53] (Supplementary Table 3). Additionally, the ~21-residue-long C-terminal sequence
173 of these ancestral rhodopsins form a distinct hydrophilic region, indicating an orientation towards
174 the cytoplasm as observed with extant type I rhodopsins[33]. Finally, the structure-based functional
175 predictions are consistent with early type I rhodopsins that carried out phototransduction, formed
176 protein-chromophore linkages, and acted as proton transmembrane transporters (Supplementary
177 Table 4).
178
179 Given the comparatively high statistical support and functional significance of ancestral binding
180 pocket residues, we interrogated these specific sequence features to provide complementary
181 insights into plausible functionality of ancient rhodopsins. For extant, uncharacterized rhodopsins,
182 functional annotation is often guided by a characteristic, three-residue motif that includes sites
183 homologous to the D85, T89, and D96 residues (a.k.a., "DTD" motif) in bacteriorhodopsin (*H.*
184 *salinarum*)[35,39,54,55]. These sites are located within a section of the third transmembrane helix
185 ("Helix C"; Fig. 3c, Fig. 4) that constitutes a portion of the retinal binding pocket. Variation in this
186 motif is strongly correlated with functional differentiation in rhodopsins. For example, "DTD" or
187 "DTE" motifs (glutamic acid replaces the D96 residue in the latter) are characteristic of
188 bacteriorhodopsins, xanthorhodopsins, and proteorhodopsins, which all function as light-driven
189 proton pumps and exhibit similar photosensitive mechanisms[39]. By contrast, chloride-pumping
190 halorhodopsins exhibit a "TSA" motif and sensory rhodopsins instead exhibit a "DTY" of "DTF"
191 motif.
192

193 The functional correlation of the three-residue motif can be explained by the role these sites play
194 during ion binding and transfer. In proton-pumping rhodopsins, the negatively charged residues
195 of the DTD/DTE motifs are protonated during intermediate steps of the bacteriorhodopsin,
196 xanthorhodopsin, and proteorhodopsin photocycles, permitting unidirectional proton transport out
197 of the cytoplasm[39]. By contrast, the TSA motif typical of chloride-pumping rhodopsins lacks a
198 negative charge relative to the DTD/DTE motifs, which is compensated by the binding of the
199 chloride ion[55,56]. The functional significance of these motifs has been underscored by experiments
200 demonstrating the conversion of proton-pumping bacteriorhodopsins into chloride pumps by a
201 D85T mutation at the first site of the DTD motif[57]. In certain cases, functional predictions of extant
202 rhodopsins based on these characteristic motifs have been shown to prove reliable upon
203 experimental characterization[54].
204
205 At the deepest nodes in the rhodopsin phylogeny, the most likely ancestral motifs are: 1) Anc1,
206 DTD, 2) Anc2, DTY, 3) Anc3, DTD, 4) Anc4, DTE, and 5) Anc5, DTE (Fig. 3). These motifs suggest
207 a proton-pumping function for all ancestors bearing the DTD/DTE motifs, except for Anc2, which,
208 as the common ancestor to sensory rhodopsins I and II, exhibits the DTY motif consistent with
209 extant sensory rhodopsins. Most inferred residues at these three sites have posterior probabilities
210 >0.90 for Anc1-Anc5 (Supplementary Table 2). The exceptions are in Anc1, Anc3, and Anc5, for
211 which the third site in the motif has a posterior probability of 0.71, 0.54, or 0.61, respectively.
212 However, the next most likely ancestral states for these sites are either aspartic acid or glutamic
213 acid, maintaining the negatively charged residue observed in extant proton transporters. Thus, a
214 motif consistent with an ancestral proton-pumping function is well supported for Anc1, Anc3,
215 Anc4, and Anc5, which together suggests that other microbial functions such as other forms of
216 ion transport and phototaxis may be derived from an ancestral, light-driven proton pump
217 functionality. This model is compatible with experimental findings that sensory rhodopsins, when
218 expressed in alternate microbial hosts in the absence of associated transducer proteins, retain
219 the ability to transport protons, suggesting that the photosensory function is more derived[49].
220
221 The early evolution of proton transport by rhodopsins or rhodopsin-like proteins would likely have
222 provided an adaptive, simplistic means of maintaining the intracellular environment of primitive
223 cells. Life on early Earth likely faced several hurdles related to membrane permeability that
224 implicate the presence of relatively simple, membrane-spanning motive proteins proximal to
225 LUCA[58-60]. One of these hurdles was circumventing the disruptive cellular tendency to expand
226 and lyse due to the influx of salts from their environment passing across their semi-permeable

227 protocell membrane[61]. The simplistic membrane-spanning ion-pump infrastructure of an early
228 rhodopsin could have provided cells inhabiting photic environments with the means to counteract
229 diffusive flux with a mechanism of active transport[62,63]. Biological preference for transporting
230 cations in lieu of anions is consistent with the proposed acidic conditions of the early Archaean
231 ocean, where a high concentration of chloride-based salts such as H^+, Na^+, K^+, Ca^+, and Mg^+
232 would have been prevalent[62,64]. Acquiring a means of agency over intercellular ion gradient
233 dynamics would have permitted the maintenance of transmembrane potential and chemiosmotic
234 coupling for energy generation, which may have been present in LUCA[14,65,66]. From a bioenergetic
235 perspective, our results support the premise that a rhodopsin-like proton pump could have
236 enabled chemiosmotic coupling and simplistic energy production in primordial cells.
237
238 **The type I rhodopsin ancestor was likely tuned to absorb green light**
239
240 Extant rhodopsin variants exhibit absorption capacity across much of the visible spectrum, and it
241 is unclear which portion of the spectrum the earliest rhodopsins would have attenuated, nor
242 whether changes in absorption correlated with the diversification of new functions. To understand
243 the evolution of rhodopsin spectral tuning, we used the recently developed machine-learning
244 algorithm, GrplassoSeq[28]. This method approximates the maximum absorbance wavelength
245 (λ_{max}) of rhodopsin protein sequences by producing a statistical model that describes the
246 relationship between amino acids within the seven transmembrane helices and absorbance
247 features of experimentally characterized homologs. The model was demonstrated to predict λ_{max}
248 of rhodopsins with an average error of ±7.8 nm. Here, λ_{max} values were predicted for extant
249 sequences in the phylogenetic dataset, inferred maximum likelihood ancestors, and alternative
250 ancestors for nodes Anc1-An5 generated to test the robustness of wavelength prediction to the
251 statistical uncertainty of ancestral sequence inference (see Methods; Fig. 2b,c)
252
253 The first-order spectral patterns predicted for ancestral rhodopsins indicate lineage-, and,
254 potentially function-specific, color tuning. λ_{max} for all ancestors range between ~480 to 585 nm,
255 occupying the cyan-to-yellow range of the visible light spectrum (Fig. 1). Different rhodopsin
256 lineages, in addition to being associated with varying functions, also exhibit characteristic λ_{max}
257 values. For example, our results suggest that halorhodopsins evolved to absorb yellow light
258 (~565-585 nm) early, with secondary shifts toward green light absorption in a specific sublineage.
259 The bacteriorhodopsin, sensory rhodopsin I, and xanthorhodopsin clades largely exhibit green
260 spectral tuning (~500-565 nm), whereas shifts towards cyan (~480-500 nm) absorption can be

observed within the sensory rhodopsin II and proteorhodopsin clades. Notable are the so-called blue-light and green-light absorbing members of the proteorhodopsin clade, whose differential spectral tuning has been suggested to correlate with habitat depth of host microbes in the marine environment[67]. Because green light is attenuated at shallower depths, it has been argued that blue-absorbing rhodopsins are hosted by organisms that occupy comparatively deeper portions of the water column. Our λ_{max} predictions suggest that blue-light absorbing proteorhodopsins are derived from a green-light absorbing ancestor which would thus have been better adapted to a shallow marine environment.

All oldest, maximum likelihood ancestral sequences (Anc1-Anc5) from our rhodopsin phylogeny have predicted λ_{max} values between 537 and 543 nm (Fig. 1). Anc5 is the deepest ancestral node, having both the greatest ancestral sequence uncertainty and, by extension, the greatest variation (~20 nm range) in predicted λ_{max} values across alternate ancestral reconstructions. Even accounting for alternate ancestral sequence reconstructions at these same nodes, the predicted λ_{max} range for early ancestors is only expanded to ~525 to 545 nm. This range is still well within the green portion of the visible light spectrum, even if additionally accounting for the ±7.8 nm average error associated with the wavelength prediction algorithm[28]. Together, our spectral tuning analyses support a model wherein extant type I rhodopsins diversified from an ancestor tuned toward the absorption of green light.

Indicators of early rhodopsin host niche habitability and coevolution with Earth environment

Correlations between tuning of photon absorption and cellular function across modern variants might present a logical basis to infer the most likely niche that hosts of primordial, microbial rhodopsins would have occupied. Paleoecological and paleoenvironmental inferences based on organism-level associations and anatomical attributes are well-established in paleobiology, particularly for ecosystems that existed billions of years ago and for which sedimentological indicators are poor, absent or contested[68-70]. Inferences of 'phylogenetic paleoecology' based on molecule-level attributes constitute a relatively new approach to these problems[25,27].

Color tuning over microbial rhodopsin evolution has likely been shaped by a complex interplay of both functional and environmental constraints[15,22,46,71]. Regarding the former, it is not unexpected that both spectral absorbance patterns and protein function are evolutionarily correlated, given

295 the observation that many of the same rhodopsin binding pocket sites strongly determine both
296 features[22,28,71]. Our analysis likewise indicates that λ_{max} values correlate with cellular function
297 throughout rhodopsin's history. The oldest common ancestors of each clade possess comparable
298 λ_{max} values to modern counterparts, which themselves appear generally binned according to
299 functionally distinct lineages (Fig. 1).
300
301 Spectral tuning may be further segregated by the photon availability of the host environment and
302 by selective pressures to enable simultaneous optical control of distinct functions by different
303 colors of light. Host organisms containing rhodopsin variants are found almost everywhere near
304 Earth's surface, occupying a variety of different niches. For example, chloride-pumping
305 halorhodopsins are mostly limited to halophilic Archaea[72] and maximally absorb yellow light,
306 matching the spectral insolation peak experienced in the terrestrial salt ponds in which they
307 inhabit. By contrast, sensory rhodopsin II and proteorhodopsin absorbances hosted in shallow-
308 water photic zones are shifted towards blue light, where peak spectral irradiance is attenuated by
309 water transmissivity. Sensory rhodopsin I (SRI) tuning is unique because it appears to have been
310 shaped both by its dual roles in phototactic and photophobic responses, as well as by the
311 absorption profiles of other rhodopsin proteins in the same host. In a host halobacterium, orange
312 light interaction with SRI produces electrons that drive flagella motion toward light, but absorption
313 of extra photons in the UV triggers another cascade that stops flagella motion to avoid
314 phototoxicity[73,74], thereby guiding the host to wavelengths where both halorhodopsin and
315 bacteriorhodopsin absorb maximally[21].
316
317 Inferring the environmental constraints on early rhodopsin evolution is complicated by the fact
318 that Earth's atmospheric evolution has drastically altered the surface photon irradiance profile
319 over billions of years. Thus, we reconstructed photon irradiance profiles at various water depths
320 for both modern and haze-free Archean Earth scenarios, informed by an updated prescription for
321 the composition of Earth's early atmosphere with high carbon dioxide and methane, but no
322 oxygen[75] (Fig. 5). These models coupled an advanced atmospheric radiative transfer model
323 SMART[76] and the absorption profile of liquid water[77]. We found that the level of shielding enjoyed
324 at the surface of Earth today in a haze-free Archean atmosphere is only possible at depths of tens
325 of meters or deeper, and that the equivalent protective depth level is a strong function of
326 wavelength. At visible wavelengths, red and infrared photons also become strongly attenuated
327 with depth, leading to a wavelength of maximum transparency at ~485 nm in the blue-green
328 portion of the visible spectrum (assuming no impurities, particulate matter will preferentially scatter

329 blue light over green light). This contrasts markedly with a peak spectral irradiance at Earth's
330 surface without water attenuation, which ranges from roughly 500-530 nm with a corresponding
331 color of green-yellow.
332
333 In addition to reconstructing historical changes in Earth surface irradiation, we explored the impact
334 that the GOE[29], one of the most significant environmental transitions in Earth-life coevolution, may
335 have had in the evolutionary diversification of microbial type I rhodopsins. Light and oxygen both
336 destroy the rhodopsin-associated retinal easily[78], so the various functions of rhodopsins as
337 sensors of these external parameters are intimately tied to sequence changes throughout the
338 phylogenetic family. The last common ancestor of haloarchaea is thought to have possessed the
339 halorhodopsin (and perhaps a bacteriorhodopsin) gene[48]. As haloarchaea are oxygen-consuming
340 heterotrophs, they would likely have originated only after the rise of oxygenic photosynthesis[24],
341 constraining the maximum age of the halorhodopsin clade to at least the GOE at ~2.3 Ga, and
342 perhaps as far back as the first indicators of oxygenation around 3.0 Ga[79,80]. In addition, sensory
343 rhodopsin II further drives a photophobic response under conditions of abundant oxygen to avoid
344 photooxidative damage, a function that likely evolved after the emergence of a well oxygenated
345 post-GOE environment. These observations would indicate that early rhodopsins (i.e., Anc1-
346 Anc5) reflect ancestral functions carried out within anoxic hosts of great (likely early Precambrian)
347 age, regardless of the placement of the root of the rhodopsin phylogenetic network or the exact
348 timing of primordial rhodopsin origination.
349
350 Together, our results indicate that ancestral rhodopsins transported protons across the cellular
351 membrane and exhibited spectral tuning toward green photons over a span of time encompassing
352 much of Earth's early history of inhabitation. An assessment of the pervasiveness of light-driven
353 sodium/proton export pumps among prokaryotes has been interpreted as indicating proximity of
354 rhodopsins to the shared bacterial-archaeal superfamily ancestor[14]. This would place the deepest
355 proton-pumping ancestral nodes of our phylogenetic tree proximal to LUCA, with inferred ages of
356 greater than ~3.5 Ga. Such early, green-absorbing rhodopsins would have been well adapted to
357 the water transmission peak at blue-green wavelengths in the marine early Earth surface
358 environment, even at moderate depths in the ocean such that their host organisms could escape
359 damaging UV rays prior to the accumulation of protective atmospheric ozone (Fig. 5a,b, Fig. 6).
360 These spectral regions notably occupy the window excluded by the absorption peaks of later-
361 evolved chlorophylls and bacteriochlorophylls (Fig. 5d, Fig. 6b). Rhodopsin terrestrial and surface
362 water host niches following the GOE were later expanded by reduced UV irradiance, but shallow

363 water niches were simultaneously more sharply discriminated by the increase in dissolved oxygen
364 concentrations, contributing to the extant pattern of functional and sequence divergence of the
365 rhodopsin protein family. This is consistent with the presumed antiquity of the rhodopsin enzyme,
366 and broadly compatible with trends in mid-to-late Precambrian microfossil abundance, which
367 (questions of taphonomic bias aside) consistently indicate significant microfossil diversity and
368 abundance in near-shore marine deposits[81,82]. Host microorganisms residing in these photic
369 zones would have been able to access available nutrients (e.g., phosphorous and nitrate), and
370 avoid harmful UV radiation while simultaneously harnessing beneficial regions of the visible
371 electromagnetic spectrum via the selective attenuation of light (and selective avoidance of
372 dissolved oxygen) within the water column.
373
374 **Implications for planetary habitability and the search for life beyond Earth**
375
376 The coevolution of environment and life early in Earth's history serves as a model for predicting
377 universal, detectable biosignatures that might be generated on a microbe-dominated planet
378 beyond our solar system. The current approach to constraining the possible suite of remotely
379 detectable biosignatures leverages the aggregate spectral features of photosystems and pigment
380 proteins which are sensitive to different portions of the Sun's emission spectrum (Fig. 5)[24].
381 Photosystems, however, are complicated, specific structures with a relatively limited capacity for
382 functional variability or spectral tunability. The dual roles of active site residues as determinants
383 of both function and peak photon absorptivity demonstrate that opsin pigment proteins generate
384 a suite of biosignatures that are tunable to our Sun's peak irradiance profile, to long-term changes
385 in the Earth surface chemical system, and to ecological niche diversification over billions of years.
386 Our phylogenetic analysis indicates that rhodopsin's structural simplicity, functional variability and
387 spectral tunability make it an ideal testbed for assessing universal properties of biosignature
388 production on candidate microbe-dominated planets that may differ from our referent biological
389 (Earth-Sun) system.
390
391 A generic characteristic of life is that it is 'tuned' to the conditions found on our planet that enable
392 its existence. These conditions refer to the chemical and physical characteristics that circumscribe
393 habitability, namely substrate availability, ambient temperature, pH, atmospheric pressure and
394 composition, salt concentration, and solar insolation, and irradiance, among other geochemical
395 and atmospheric attributes. As demonstrated here, enzymes can be far more tuned to these
396 external conditions than other molecules or anatomical features that compose organisms,

397 precisely because many enzymes serve as intermediaries for channelizing photons and
398 substrates into metabolic activity. Our results show that enzyme families (and to some extent, the
399 niches occupied by their hosts) can be amenable to paleobiological reconstruction in ways that
400 cannot be accomplished using conventional molecular biosignatures alone.
401
402 Utilizing enzymes as paleosensors of geologic and stellar conditions in deep time, leveraging the
403 informatic characteristics of enzyme composition, and the tight coupling between internal (cellular)
404 and external (environmental) conditions, permits future studies exploring circumstances of
405 habitability and biosignature production over Earth's deep and varied history. Light absorbance
406 profiles produced by microbial pigments can be used to prime future disk-integrated observations
407 of exoplanets, provide surficial albedo parameters for complex atmospheric radiative transfer
408 models, and directly aid the development of instrumentation for the detection and characterization
409 of life in the universe both within and outside our Solar System. Forthcoming integrative studies
410 that reconcile experimentally validated biological photon absorbances with more sophisticated
411 planetary system analyses would thus be beneficial, for instance, in applying microbial spectral
412 properties to assess the potential habitability or inhabitation of extrasolar planets[83-85]. Overall,
413 understanding biomolecular adaptations may benefit planetary habitability assessments when
414 planetary and stellar indicators are poorly constrained.
415
416 **METHODS**
417
418 **Phylogenetic reconstruction and ancestral sequence inference**
419
420 A type I microbial rhodopsin protein sequence dataset was created by identifying homologs from
421 the National Center for Biotechnology Information non-redundant protein database in December
422 2020. Homologous sequences were identified and obtained via BLASTp[32] using query amino acid
423 sequences acquired from UniProt[86] (Supplementary Table 5) and an expect value cutoff of <1e-
424 5. The sequence dataset was curated to remove duplicate and partial sequences. The final
425 dataset consisted of 568 sequences: 96 bacteriorhodopsin, 144 halorhodopsin, 83
426 proteorhodopsin, 29 sensory rhodopsin I, 56 sensory rhodopsin II, 44 xanthorhodopsin, and 116
427 outgroup heliorhodopsin sequences. The dataset was aligned by MAFFT (v7.427)[87] with a gap
428 open penalty of 1.53 and offset value of 0.0. Fungal and algal type I rhodopsin sequences were
429 excluded from the dataset due to long-branch-attraction effects observed within the bacterial
430 rhodopsin clade.

431 The rhodopsin protein sequence alignment was used to select the best-fit evolutionary model by
432 ModelFinder[88] implemented in the W-IQ-TREE web server[89]. Models accounting for the proportion
433 of invariant sites ("+I") were not included in model selection due to issues optimizing rate
434 heterogeneity parameters independently[90]. The three top-ranking models scored by the Akaike
435 and Bayesian Information Criterion, LG+G+F, WAG+G+F, and VT+G+F, were used for maximum
436 likelihood phylogenetic inference by RAxML (v8)[91]. Branch support was evaluated using 100 rapid
437 bootstrap replicates. The type I rhodopsin clade was rooted with heliorhodopsin sequences, a
438 protein group known to share a distant homology with type I rhodopsins[33,92]. Because all tested
439 evolutionary models yielded comparable topologies, the best-fit model, LG+G+F, was
440 implemented for ancestral sequence inference by PAML (v4)[93]. Ancestral insertion/deletion sites
441 were reconstructed using a binary likelihood model following Aadland et al.[94].
442
443 **Homology modeling and functional predictions of ancestral rhodopsin sequences**
444
445 Inferred maximum likelihood ancestral sequences for the five deepest phylogenetic nodes (Anc1-
446 Anc5) were submitted to the Phyre2 web server[51] for structural homology modeling. Selected
447 structural templates for modeling and associated statistics are listed in Supplementary Table 6.
448 Predicted ancestral rhodopsin structures were used for structure-based functional annotation by
449 COFACTOR implemented in the I-TASSER software suite[52], which predicts protein function
450 based on global and local similarities to template proteins with known structures and functions.
451
452 **Alternative ancestral sequence library generation**
453
454 For ancestral nodes Anc1-Anc5, 50 alternate protein sequences were generated from site-wise
455 posterior probabilities to test the robustness of downstream analyses to statistical uncertainty of
456 ancestral inference. For each ancestral node, regularly increasing percentages (>0% to 100%) of
457 ambiguously reconstructed sites in the maximum likelihood sequence—defined here as having
458 posterior probabilities <0.7—were replaced with the next most probable residue. Thus, the least
459 probable ancestral sequence explored here by this approach had all ambiguous residues
460 replaced by an alternate residue, likely representing the fringe of plausible ancestral sequence
461 space[95].
462
463 **Wavelength Prediction**
464

Ancestral maximum absorbance wavelength (λ_{max}) values were predicted using the machine learning model, GrplassoSeq[28] implemented in R. This statistical model describes the relationship between primary amino acid sequence and rhodopsin color tuning based on a training rhodopsin sequence dataset with experimentally validated λ_{max} values. Maximum likelihood and alternate ancestral rhodopsin sequences, as well as all extant rhodopsin sequences used for phylogenetic inference, were aligned to the training dataset for λ_{max} prediction.

Radiative transfer and pigment spectra

Surface photon irradiance spectra were calculated with the Spectral Mapping Atmospheric Radiative Transfer (SMART) model developed by D. Crisp[76]. Calculation assumed cloud-free conditions with a solar zenith angle of 60 degrees and summed both direct and diffuse radiation at the surface. The modern Earth atmospheric profile was assumed to be a 1962 U.S. Standard Atmosphere employed by the InterComparison of Radiation Codes in Climate Models (ICRCCM) project[96], whereas the Archean scenario was assumed to be a haze-free anoxic atmosphere from Arney et al.[75] with a CH_4/CO_2 ratio of 0.1. The absorption coefficients of Segelstein[77] and the Beer-Lambert Law were used to calculate the transmissivity and the attenuation of the photon irradiances as a function of depth.

Data availability

All code and phylogenetic datasets can be found at https://github.com/kacarlab/Rhodopsin.

ACKNOWLEDGEMENTS

We thank I. Takeuchi, M. Karasuyama, and K. Inoue for access to the machine learning algorithms used to predict rhodopsin wavelength absorbances used in this work and for the supportive assistance, and S. DasSarma for the insightful conversations. This work was supported by the NASA Arizona Space Grant (C.D.S.), the NASA Astrobiology Postdoctoral Fellowship, (A.K.G.), the NASA Interdisciplinary Astrobiology Research Consortium (ICAR) MUSE, a member of the NASA Early Cell to Multicellularity Research Coordination Network funded via the Astrobiology Research Program No 80NSSC21K0592 (B.K.), the NASA Early Career Faculty Award No. 80NSSC19K1617 (B.K.) and an NSF Emerging Frontiers Program Award No. 1724090 (B.K.), NASA's Alternative Earths team funded by the NASA Astrobiology Institute under Cooperative

Agreement Number NNA15BB03A and the NASA ICAR program, and the Virtual Planetary Laboratory, a member of the NASA Nexus for Exoplanet System Science (NExSS) funded via NASA Astrobiology Program Grant No. 80NSSC18K0829 (E.W.S.).

AUTHOR CONTRIBUTIONS

B.K. conceived, designed, and supervised the overall study. C.D.S. designed and performed the phylogenetic and modeling experiments, collected, and analyzed the data. E.F. generated the ancient protein library and wavelength prediction pipeline with input from A.K.G. E.W.S. conducted the radiative transfer calculations. All authors discussed and interpreted the results. B.K. and C.D.S wrote the manuscript with input and/or text contributions from all authors.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1 McKay, C. P. Requirements and limits for life in the context of exoplanets. *Proceedings of the National Academy of Sciences* **111**, 12628-12633, doi:https://doi.org/10.1073/pnas.1304212111 (2014).

2 Harrison, J. P., Gheeraert, N., Tsigelnitskiy, D., Cockell, C.S. The limits for life under multiple extremes. *Trends in microbiology* **21**, 204-212, doi:https://doi.org/10.1016/j.tim.2013.01.006 (2013).

3 Lunine, J. I. Physical conditions on the early Earth. *Philosophical Transactions of the Royal Society B* **361**, 1721-1731, doi:https://doi.org/10.1098/rstb.2006.1900 (2006).

4 Newman, M. J., Rood, R.T. Implications of Solar Evolution for the Earth's Early Atmosphere. *Science* **198**, 1035-1037, doi:https://doi.org/10.1126/science.198.4321.1035 (1977).

5 Garcia, A. K., Cavanaugh, C. M. & Kacar, B. The curious consistency of carbon biosignatures over billions of years of Earth-life coevolution. *ISME J*, doi:10.1038/s41396-021-00971-5 (2021).

6 Butterfield, N. J. Macroevolution and Macroecology through Deep Time. *Palaeontology* **50**, 41-55, doi:10.1111/j.1475-4983.2006.00613.x (2007).

7 Doolittle, W. F. The nature of the universal ancestor and the evolution of the proteome. *Current opinion in structural biology* **10**, 355–358, doi:https://doi.org/10.1016/s0959-440x(00)00096-8 (2000).

8 Benner, S. A., Caraco, M.D., Thomson, J.M., Gaucher, E.A. Planetary biology—paleontological, geological, and molecular histories of life. *Science* **296**, 864-868, doi:https://doi.org/10.1126/science.1069863 (2002).

9 Garcia, A. K., Kaçar, B. How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radical Biology and Medicine* **140**, 260-269, doi:https://doi.org/10.1016/j.freeradbiomed.2019.03.033 (2019).

10 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. . Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science Advances* **5**, eaaw8855, doi:https://doi.org/10.1126/sciadv.aaw8855 (2019).

11 Pinhassi, J., DeLong, E.F., Béjà, O., González, J.M., Pedrós-Alió, C. Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiology and Molecular Biology Reviews* **80**, 929–954, doi:https://doi.org/10.1128/MMBR.00003-16 (2016).

12 Bryant, D. A. & Frigaard, N.-U. Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology* **14**, 488-496, doi:10.1016/j.tim.2006.09.001 (2006).

13 Sharma, A. K., Spudich, J. L., Doolittle, W.F. Microbial rhodopsins: functional versatility and genetic mobility. *Trends in microbiology* **14**, 463–469, doi:https://doi.org/10.1016/j.tim.2006.09.006 (2006).

14 Shalaeva, D. N., Galperin, M.Y., Mulkidjanian, A.Y. Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biology Direct* **10**, 1-12, doi:https://doi.org/10.1186/s13062-015-0091-4 (2015).

15 DasSarma, S. & Schwieterman, E. W. Early evolution of purple retinal pigments on Earth and implications for exoplanet biosignatures. *International Journal of Astrobiology*, 1-10, doi:10.1017/s1473550418000423 (2018).

16 Mackin, K. A., Roy, R.A., Theobald, D.L. An empirical test of convergent evolution in rhodopsins. *Molecular biology and evolution* **31**, 85–95, doi: https://doi.org/10.1093/molbev/mst171 (2014).

17 Racker, E., Stoeckenius, W. Reconstitution of purple membrane vesicles catalyzing light-driven proton uptake and adenosine triphosphate formation. *The Journal of Biological Chemistry* **249**, 662–663, doi:https://doi.org/10.1016/S0021-9258(19)43080-9 (1974).

18 Akram, N., Palovaara, J., Forsberg, J., Lindh, M.V., Milton, D.L., Luo, H., González, J.M., Pinhassi, J. Regulation of proteorhodopsin gene expression by nutrient limitation in the marine bacterium Vibrio sp. AND4. *Environmental microbiology* **15**, 1400–1415, doi:https://doi.org/10.1111/1462-2920.12085 (2013).

19 Govorunova, E., Sineshchekov, O.A., Li, H., Spudich, J.L. Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications. *Annual Review of Biochemistry* **86**, 845-872, doi: https://doi.org/10.1146/annurev-biochem-101910-144233 (2017).

20 Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., Giovannoni, S.J. . Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PloS one* **6**, e19725, doi:https://doi.org/10.1371/journal.pone.0019725 (2011).

21 Sharma, A. K. *Insights into the ecology and evolution of the light-harvesting microbial rhodopsins*, Dalhousie University, (2009).

22 Engqvist, M. K. M. *et al.* Directed Evolution of Gloeobacter violaceus Rhodopsin Spectral Properties. *Journal of Molecular Biology* **427**, 205-220, doi:https://doi.org/10.1016/j.jmb.2014.06.015 (2015).

23 Inoue, K. *et al.* Spectroscopic Study of Proton-Transfer Mechanism of Inward Proton-Pump Rhodopsin, Parvularcula oceani Xenorhodopsin. *The Journal of Physical Chemistry B* **122**, 6453-6461, doi:https://doi.org/10.1021/acs.jpcb.8b01279 (2018).

24 Schwieterman, E. W. *et al.* Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life. *Astrobiology* **18**, 663-708, doi:https://doi.org/10.1089/ast.2017.1729 (2018).

25 Lamsdell, J. C., Congreve, C. R., Hopkins, M. J., Krug, A. Z. & Patzkowsky, M. E. Phylogenetic Paleoecology: Tree-Thinking and Ecology in Deep Time. *Trends in Ecology & Evolution* **32**, 452-463, doi:10.1016/j.tree.2017.03.002 (2017).

26 Finkel, O. M., Béjà, O. & Belkin, S. Global abundance of microbial rhodopsins. *The ISME Journal* **7**, 448-451, doi:10.1038/ismej.2012.112 (2012).

27 Kacar, B., Guy, L., Smith, E. & Baross, J. Resurrecting ancestral genes in bacteria to interpret ancient biosignatures. *Philos Trans R Soc A* **375**, doi:10.1098/rsta.2016.0352 (2017).

28 Karasuyama, M., Inoue, K., Nakamura, R., Kandori, H. & Takeuchi, I. Understanding Colour Tuning Rules and Predicting Absorption Wavelengths of Microbial Rhodopsins by Data-Driven Machine-Learning Approach. *Sci Rep* **8**, 15580, doi:10.1038/s41598-018-33984-w (2018).

29 Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307-315, doi:10.1038/nature13068 (2014).

30 Sánchez-Baracaldo, P., Bianchini, G., Wilson, J. D. & Knoll, A. H. Cyanobacteria and biogeochemical cycles through Earth history. *Trends in Microbiology*, doi:10.1016/j.tim.2021.05.008 (2021).

31 Hazen, R. M. *et al.* Mercury (Hg) mineral evolution: A mineralogical record of supercontinent assembly, changing ocean geochemistry, and the emerging terrestrial biosphere. *American Mineralogist* **97**, 1013-1042, doi:10.2138/am.2012.3922 (2012).

32 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

33 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., Béjà, O. . A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**, 595–599, doi:https://doi.org/10.1038/s41586-018-0225-9 (2018).

34 Lu, Y. *et al.* Crystal structure of heliorhodopsin 48C12. *Cell Research* **30**, 88-90, doi:10.1038/s41422-019-0266-0 (2019).
35 Rozenberg, A., Inoue, K., Kandori, H. & Béjà, O. Microbial Rhodopsins: The Last Two Decades. *Annual Review of Microbiology* **75**, 427-447, doi:10.1146/annurev-micro-031721-020452 (2021).
36 Ernst, O. P., Lodowski, D.T., Elstner, M., Hegemann, P., Brown, L.S., Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews* **114**, 126–163, doi:https://doi.org/10.1021/cr4003769 (2014).
37 Ihara, K., Umemura, T., Katagiri, I., Kitajima-Ihara, T., Sugiyama, Y., Kimura, Y., & Mukohata, Y. Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation. *Journal of molecular biology* **285**, 163–174, doi:https://doi.org/10.1006/jmbi.1998.2286 (1999).
38 Shibata, M., Inoue, K., Ikeda, K., Konno, M., Singh, M., Kataoka, C., Abe-Yoshizumi, R., Kandori, H., & Uchihashi, T. . Oligomeric states of microbial rhodopsins determined by high-speed atomic force microscopy and circular dichroic spectroscopy. *Scientific reports* **8**, 8262, doi:https://doi.org/10.1038/s41598-018-26606-y (2018).
39 Kandori, H. Ion-pumping microbial rhodopsins. *Front Mol Biosci* **2**, 52, doi:10.3389/fmolb.2015.00052 (2015).
40 Babitzki, G., Denschlag, R., Tavan, P. Polarization effects stabilize bacteriorhodopsin's chromophore
binding pocket: a molecular dynamics study. *The journal of physical chemistry B* **113**, 10483–10495, doi:https://doi.org/10.1021/jp902428x (2009).
41 Greenhalgh, D. A., Farrens, D.L., Subramaniam, S., Khorana, H.G. Hydrophobic Amino Acids in the Retinal-binding Pocket of Bacteriorhodopsin. *The Journal of Biological Chemistry* **268**, 20305-20311, doi:https://doi.org/10.1016/S0021-9258(20)80729-7 (1993).
42 Ding, X., Sun, C., Cui, H., Chen, S., Gao, Y., Yang, Y., Wang, J., He, X., Iuga, D., Tian, F., Watts, A., Zhao, X. Functional roles of tyrosine 185 during the bacteriorhodopsin photocycle as revealed by in situ spectroscopic studies. *Biochimica et biophysica acta: bioenergetics* **1859**, 1006–1014, doi:https://doi.org/10.1016/j.bbabio.2018.05.011 (2018).
43 Kusnetzow, A., Singh, D.L., Martin, C.H., Barani, I.J., Birge, R.R. Nature of the chromophore binding site of bacteriorhodopsin: the potential role of Arg82 as a principal counterion. *Biophysical journal* **76**, 2370–2389, doi:https://doi.org/10.1016/S0006-3495(99)77394-7 (1999).
44 Song, Y. & Gunner, M. R. Halorhodopsin pumps Cl- and bacteriorhodopsin pumps protons by a common mechanism that uses conserved electrostatic interactions. *Proc Natl Acad Sci USA* **111**, 16377-16382, doi:10.1073/pnas.1411119111 (2014).
45 Sudo, Y., Furutani, Y., Kandori, H. & Spudich, J. L. Functional importance of the interhelical hydrogen bond between Thr204 and Tyr174 of sensory rhodopsin II and its alteration during the signaling process. *J Biol Chem* **281**, 34239-34245, doi:10.1074/jbc.M605907200 (2006).
46 Man, D. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J* **22**, 1725-1731, doi:10.1093/emboj/cdg183 (2003).
47 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. . Exploration of natural red-shifted rhodopsins using a machine learning-based Bayesian experimental design. *Communications biology* **4**, 362, doi:https://doi.org/10.1038/s42003-021-01878-9 (2021).
48 Sharma, A. K., Walsh, D.A., Bapteste, E., Rodriguez-Valera, F., Ford Doolittle, W., Papke, R.T. Evolution of rhodopsin ion pumps in haloarchaea. *BMC evolutionary biology* **7**, 1-13, doi:https://doi.org/10.1186/1471-2148-7-79 (2007).

49 Bogomolni, R. A. *et al.* Removal of transducer HtrI allows electrogenic proton translocation by sensory rhodopsin I. *Proc Natl Acad Sci USA* **91**, 10188-10192, doi:10.1073/pnas.91.21.10188 (1994).
50 Spudich, J. L., Sineshchekov, O. A. & Govorunova, E. G. Mechanism divergence in microbial rhodopsins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1837**, 546-552, doi:10.1016/j.bbabio.2013.06.006 (2014).
51 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845-858, doi:10.1038/nprot.2015.053 (2015).
52 Roy, A., Kucukural, A., Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**, 725–738, doi:https://doi.org/10.1038/nprot.2010.5 (2010).
53 Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378-379, doi:10.1093/bioinformatics/14.4.378 (1998).
54 Hasegawa, M. *et al.* A unique clade of light-driven proton-pumping rhodopsins evolved in the cyanobacterial lineage. *Sci Rep* **10**, 16752, doi:10.1038/s41598-020-73606-y (2020).
55 Inoue, K. Shining light on rhodopsin selectivity: How do proteins decide whether to transport H+ or Cl–? *J Biol Chem* **295**, 14805-14806, doi:10.1074/jbc.H120.016032 (2020).
56 Besaw, J. E. *et al.* The crystal structures of a chloride-pumping microbial rhodopsin and its proton-pumping mutant illuminate proton transfer determinants. *J Biol Chem* **295**, 14793-14804, doi:10.1074/jbc.RA120.014118 (2020).
57 Sasaki, J. *et al.* Conversion of Bacteriorhodopsin into a Chloride Ion Pump. *Science* **269**, 73-75, doi:10.1126/science.7604281 (1995).
58 Mulkidjanian, A. Y., Galperin, M. Y. & Koonin, E. V. Co-evolution of primordial membranes and membrane proteins. *Trends Biochem Sci* **34**, 206-215, doi:10.1016/j.tibs.2009.01.005 (2009).
59 Deamer, D. W. & Dworkin, J. P. in *Prebiotic Chemistry Topics in Current Chemistry* Ch. Chapter 136806, 1-27 (2005).
60 Pohorille, A., Wilson, M. A. & Chipot, C. *Origins of Life and Evolution of the Biosphere* **33**, 173-197, doi:10.1023/a:1024627726231 (2003).
61 Pohorille, A., Deamer, D. . Self-assembly and function of primitive cell membranes. *Research in Microbiology* **160**, 449-456, doi:https://doi.org/10.1016/j.resmic.2009.06.004 (2009).
62 Damer, B., Deamer, D. . The Hot Spring Hypothesis for an Origin of Life. *Astrobiology* **20**, 429-452, doi:https://doi.org/10.1089/ast.2019.2045 (2020).
63 Maloney, P. C., Wilson, T.H. The Evolution of Ion Pumps. *BioScience* **35**, 43-48, doi:https://doi.org/10.2307/1310083 (1985).
64 Ueda, H., Shibuya, T. Composition of the primordial ocean just after its formation: Constraints from the reactions between the primitive crust and a strongly acidic, CO2-rich fluid at elevated temperatures and pressures. *Minerals* **11**, doi:https://doi.org/10.3390/min11040389 (2021).
65 Penny, D., Sojo, V., Pomiankowski, A. & Lane, N. A Bioenergetic Basis for Membrane Divergence in Archaea and Bacteria. *PLoS Biology* **12**, doi:10.1371/journal.pbio.1001926 (2014).
66 Lane, N., Allen, J. F. & Martin, W. How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* **32**, 271-280, doi:10.1002/bies.200900131 (2010).
67 Ozaki, Y., Kawashima, T., Abe-Yoshizumi, R. & Kandori, H. A color-determining amino acid residue of proteorhodopsin. *Biochemistry* **53**, 6032-6040, doi:10.1021/bi500842w (2014).

68 Adam, Z. R., Skidmore, M. L., Mogk, D. W. & Butterfield, N. J. A Laurentian record of the earliest fossil eukaryotes. *Geology* **45**, 387-390, doi:10.1130/g38749.1 (2017).

69 Butterfield, N. J. & Rainbird, R. H. Diverse organic-walled fossils, including "possible dinoflagellates," from the early Neoproterozoic of arctic Canada. *Geology* **26**, doi:10.1130/0091-7613(1998)026<0963:Dowfip>2.3.Co;2 (1998).

70 Hofmann, H. J. & Jackson, G. D. Shale-Facies Microfossils from the Proterozoic Bylot Supergroup, Baffin Island, Canada. *J Paleontol* **68**, 1-35, doi:10.1017/s0022336000062314 (1994).

71 Tsujimura, M. & Ishikita, H. Insights into the Protein Functions and Absorption Wavelengths of Microbial Rhodopsins. *The Journal of Physical Chemistry B* **124**, 11819-11826, doi:10.1021/acs.jpcb.0c08910 (2020).

72 Ugalde, J. A., Podell, S., Narasingarao, P. & Allen, E. E. Xenorhodopsins, an enigmatic new class of microbial rhodopsins horizontally transferred between archaea and bacteria. *Biology Direct* **6**, doi:10.1186/1745-6150-6-52 (2011).

73 Wilde, A. & Mullineaux, C. W. Light-controlled motility in prokaryotes and the problem of directional light perception. *FEMS Microbiology Reviews* **41**, 900-922, doi:10.1093/femsre/fux045 (2017).

74 Spudich, J. L. & Bogomolni, R. A. Mechanism of colour discrimination by a bacterial sensory rhodopsin. *Nature* **312**, 509-513, doi:10.1038/312509a0 (1984).

75 Arney, G., Domagal-Goldman, S.D., Meadows, V.S., Wolf, E.T., Schwieterman, E., Charnay, B., Claire, M., Hébrard, E., Trainer, M.G. The Pale Orange Dot: The Spectrum and Habitability of Hazy Arhean Earth. *Astrobiology* **16**, 873-899, doi:https://doi.org/10.1089/ast.2015.1422 (2016).

76 Meadows, V. S., Crisp, D. Ground-based near-infrared observations of the Venus nightside: The thermal structure and water abundance near the surface. *Journal of Geophysical Research: Planets* **101**, 4595-4622, doi:https://doi.org/10.1029/95JE03567 (1996).

77 Segelstein, D. J. The complex refractive index of water. *University of Missouri*, doi:http://hdl.handle.net/10355/11599 (1981).

78 Hampp, N. & Oesterhelt, D. in *Nanobiotechnology* (eds Christof M. Niemeyer & Chad A. Mirkin) 146-167 (2004).

79 Cardona, T. Early Archean origin of heterodimeric Photosystem I. *Heliyon* **4**, doi:10.1016/j.heliyon.2018.e00548 (2018).

80 Olson, S. L., Kump, L. R. & Kasting, J. F. Quantifying the areal extent and dissolved oxygen concentrations of Archean oxygen oases. *Chem Geol* **362**, 35-43, doi:10.1016/j.chemgeo.2013.08.012 (2013).

81 Adam, Z. R., Skidmore, M.L., Mogk, D.M. Paleoenvironmental implications of an expanded microfossil assemblage from the Chamberlain Formation, ower Belt Supergroup, Montanta. *The Geological Society of America Special Paper* **522** (2016).

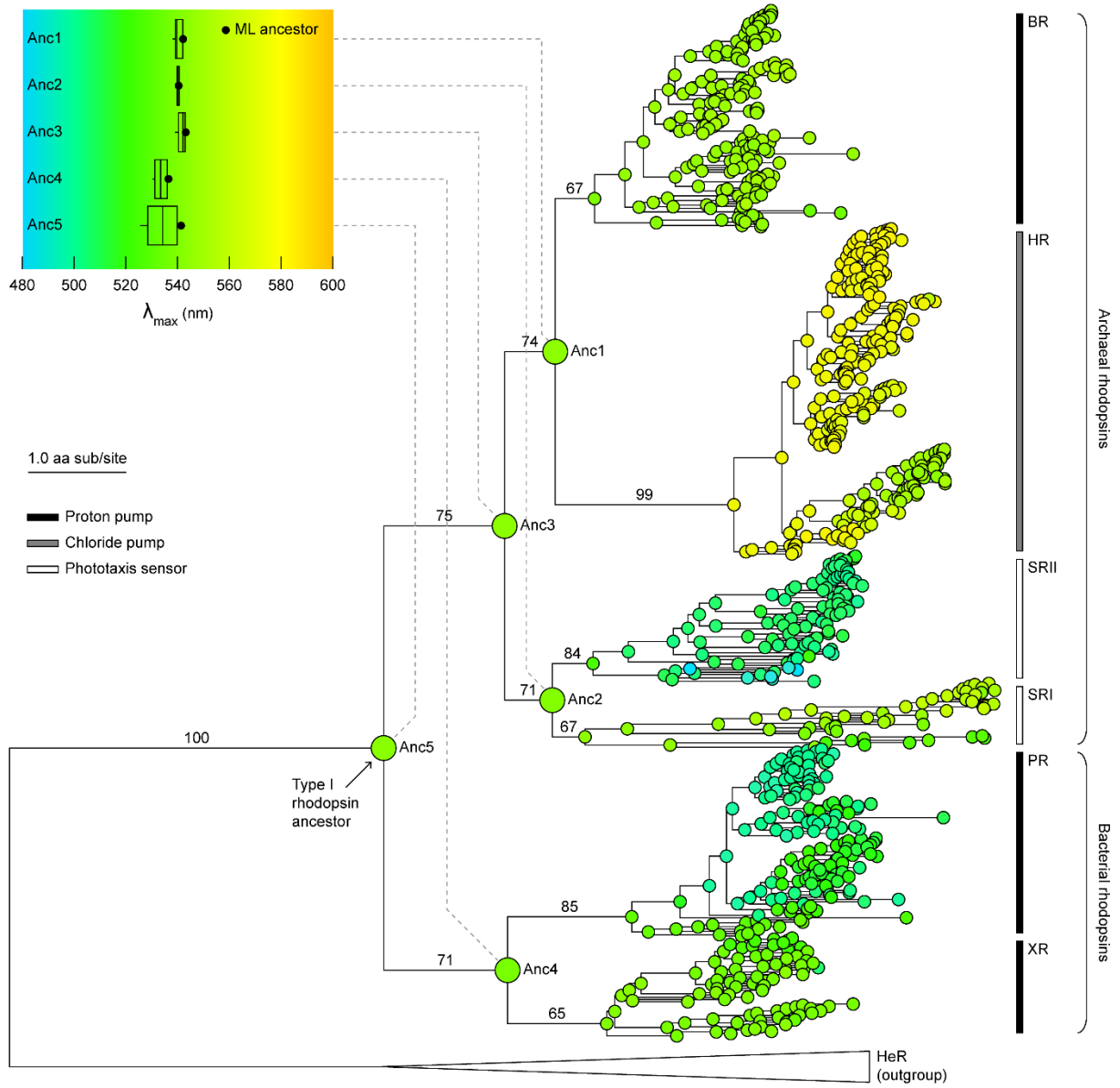
82 Javaux, E. J., Knoll, A.H. Micropaleontology of the lower Mesoproterozoic Roper Group, Australia, and implications for early eukaryotic evolution. *Journal of Paleontology* **91**, 199-229, doi:https://doi.org/10.1017/jpa.2016.124 (2017).

83 DasSarma, S., DasSarma, P., Laye, V.J., Schwieterman, E.W. Extremophilic models for astrobiology: haloarchaeal survival strategies and pigments for remote sensing. *Extremophiles* **24**, 31–41, doi:https://doi.org/10.1007/s00792-019-01126-3 (2020).

84 Hegde, S., Paulino-Lima, I.G., Kent, R., Kaltenegger, L., Rothschild, L. Surface biosignatures of exo-Earths: Remote detection of extraterrestrial life. *Proceedings of the National Academy of Sciences* **112**, 3886–3891, doi:https://doi.org/10.1073/pnas.1421237112 (2015).

85 Schwieterman, E. W., Cockell, C.S., Meadows, V.S. Nonphotosynthetic Pigments as Potential Biosignatures. *Astrobiology* **15**, 341–361, doi:https://doi.org/10.1089/ast.2014.1178 (2015).
86 Uniprot Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).
87 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
88 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587-589, doi:10.1038/nmeth.4285 (2017).
89 Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44**, W232-W235, doi:10.1093/nar/gkw256 (2016).
90 Yang, Z. *Computational Molecular Evolution*. 374 (Oxford University Press, 2006).
91 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).
92 Shibukawa, A., Kojima, K., Nakajima, Y., Nishimura, Y., Yoshizawa, S., Sudo, Y. . Photochemical Characterization of a New Heliorhodopsin from the Gram-Negative Eubacterium Bellilinea caldifistulae (BcHeR) and Comparison with Heliorhodopsin-48C12. *Biochemistry* **58**, 2934–2943, doi:https://doi.org/10.1021/acs.biochem.9b00257 (2019).
93 Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586–1591, doi:https://doi.org/10.1093/molbev/msm088 (2007).
94 Aadland, K., Pugh, C. & Kolaczkowski, B. High-Throughput Reconstruction of Ancestral Protein Sequence, Structure, and Molecular Function. *Methods Mol Biol* **1851**, 135-170, doi:10.1007/978-1-4939-8736-8_8 (2019).
95 Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J. & Thornton, J. W. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol Biol Evol* **34**, 247-261, doi:10.1093/molbev/msw223 (2017).
96 Luther, F. M., Ellingson, R.G., Fouquart, Y., Fels, S., Scott, N.A., Wiscombe, W.J. Intercomparison of Radiation Codes in Climate Models (ICRCCM): Longwave Clear-Sky Results—A Workshop Summary. *Bulletin of the American Meteorological Society* **69**, 40-48, doi:https://doi.org/10.1175/1520-0477-69.1.40 (1988).
97 Chen, M., Schliep, M., Willows, R.D., Cai, Z.-L., Neilan, B.A., Scheer, H. A Red-Shifted Chlorophyll. *Science* **329**, 1318–1319, doi:https://doi.org/10.1126/science.1191127 (2010).
98 Chen, M., Blankenship, R.E. Expanding the solar spectrum used by photosynthesis. *Trends in Plant Science* **16**, 427–431, doi:https://doi.org/10.1016/j.tplants.2011.03.011 (2011).
99 Frigaard, N., Larsen, Kim, L., Cox, R.P. Spectrochromatography of photosynthetic pigments as a fingerprinting technique for microbial phototrophs. *FEMS Microbiology Ecology* **20**, 69–77, doi:https://doi.org/10.1016/0168-6496(96)00005-0 (1996).
100 Jeffrey, S. W. Purification and properties of chlorophyll c from Sargassum flavicans. *Biochemical Journal* **86**, 313-318, doi:https://doi.org/10.1042/bj0860313 (1963).

815 **FIGURES**



816

817 **Fig. 1 Evolution of spectral tuning mapped to a maximum likelihood phylogeny of microbial**
818 **type I rhodopsin protein sequences.** Nodes are colored according to the predicted maximum
819 absorption wavelength (λ_{max}) of the associated maximum likelihood (ML) ancestral sequence
820 (color scale shown in plot, top left). Boxplot shows the predicted λ_{max} distribution among alternate
821 ancestral sequences for nodes Anc1-Anc5 (n = 50 per node; see Methods and Fig. 2b,c). Branch
822 support values are derived from 100 bootstrap replicates. Branch scale indicates 1.0 amino acid
823 substitutions per site. Heliorhodopsin outgroup clade is collapsed. BR: bacteriorhodopsin; HR,
824 halorhodopsin; SR II, sensory rhodopsin II; SR I, sensory rhodopsin I; PR, proteorhodopsin;
825 XR, xanthorhodopsin; HeR, heliorhodopsin.

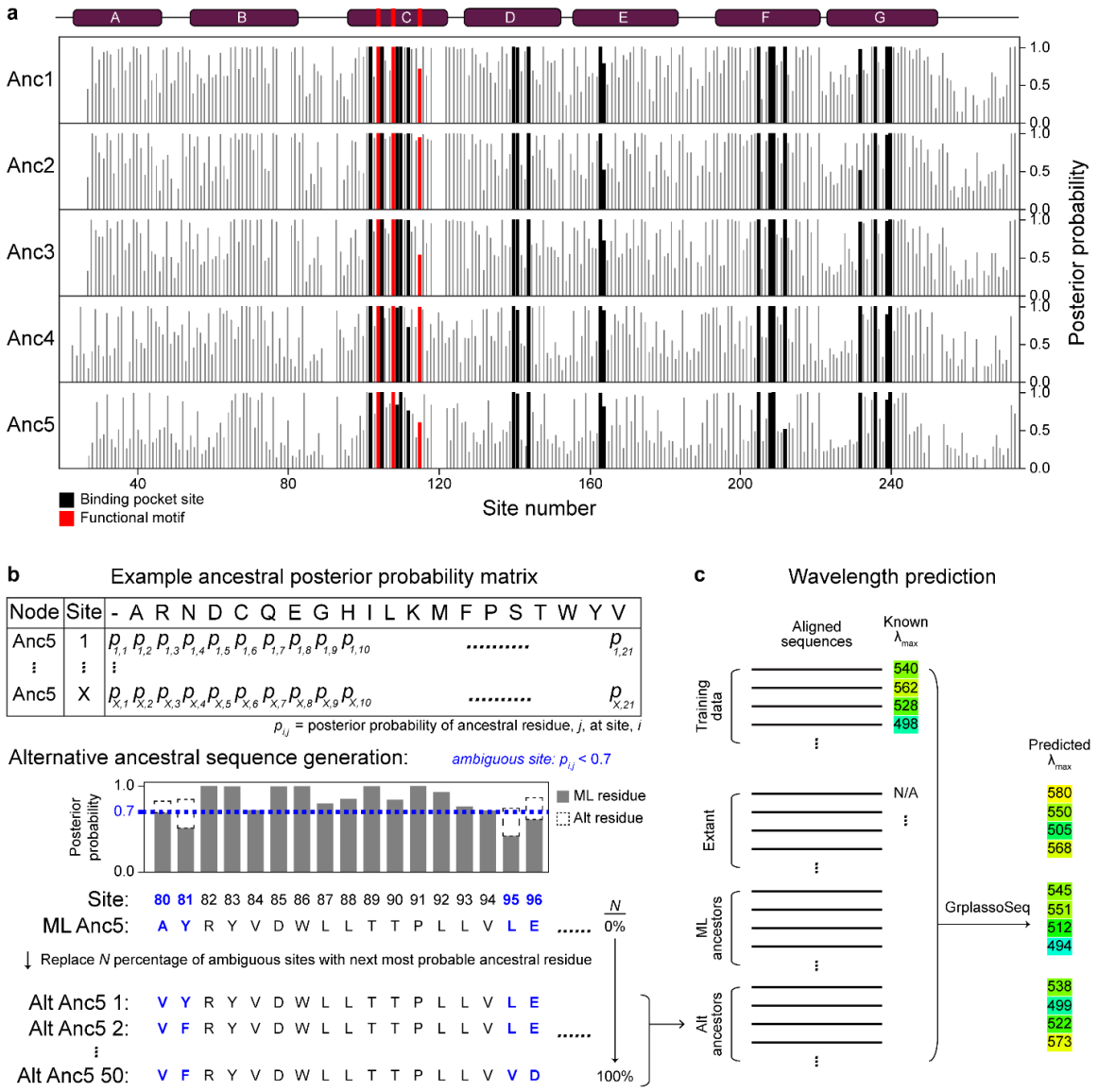
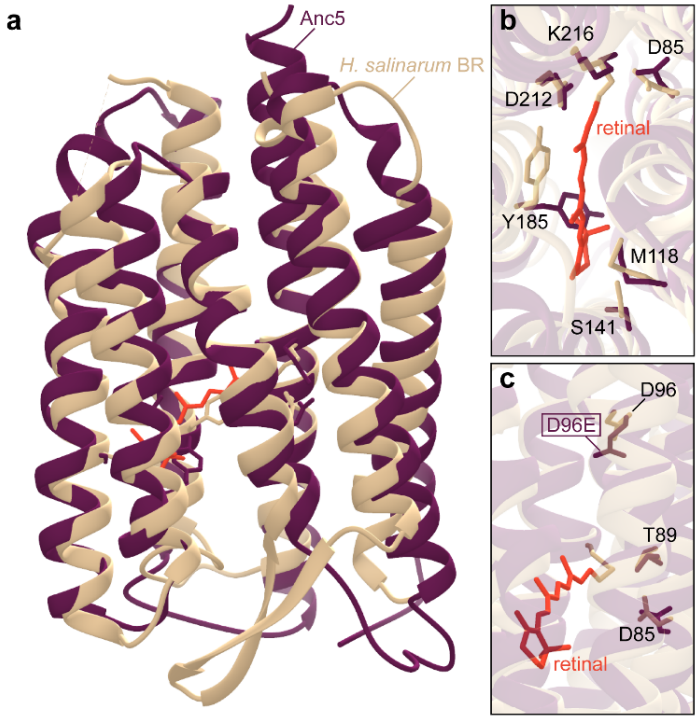


Fig. 2 Incorporation of statistical uncertainty of ancestral rhodopsins inference into wavelength prediction. a Site-wise posterior probabilities of maximum likelihood (ML) ancestral rhodopsin sequences. Sites labeled "functional motif" are homologous to the three-residue DTD motif in extant *H. salinarum* bacteriorhodopsin. Helices are labeled above. **b** Generation of alternative ("alt") ancestral sequences. For each node, alternate sequences were generated by replacing 0% to 100% of ambiguous sites (defined as having a posterior probability < 0.7) with the next most probable residue. **c** Methodology for maximum absorbance wavelength (λ_{max}) prediction. λ_{max} values of extant and ancestral (both ML and alternative) rhodopsin sequences were predicted by a machine-learning model (GrplassoSeq), trained on a dataset of experimentally characterized rhodopsin proteins with known λ_{max} values.



838

839 **Fig. 3 Structural conservation between the inferred, type I rhodopsin last common**
840 **ancestor, Anc5, and extant bacteriorhodopsin**. **a** Homology model of Anc5 aligned to the
841 extant *H. salinarum* bacteriorhodopsin ("BR"; PDB: 1C3W). **b** Retinal binding pocket of ancestral
842 and extant rhodopsins. Conserved residues that interact with the retinal chromophore are labeled.
843 **c** Residues homologous to the *H. salinarum* DTD motif correlative with rhodopsin function. Extant
844 bacteriorhodopsin exhibits a DTD motif, compared to the DTE motif in Anc5. **b,c** All displayed
845 Anc5 residues are identical to those of extant bacteriorhodopsin except for site 96 (D96E, boxed).

846

847

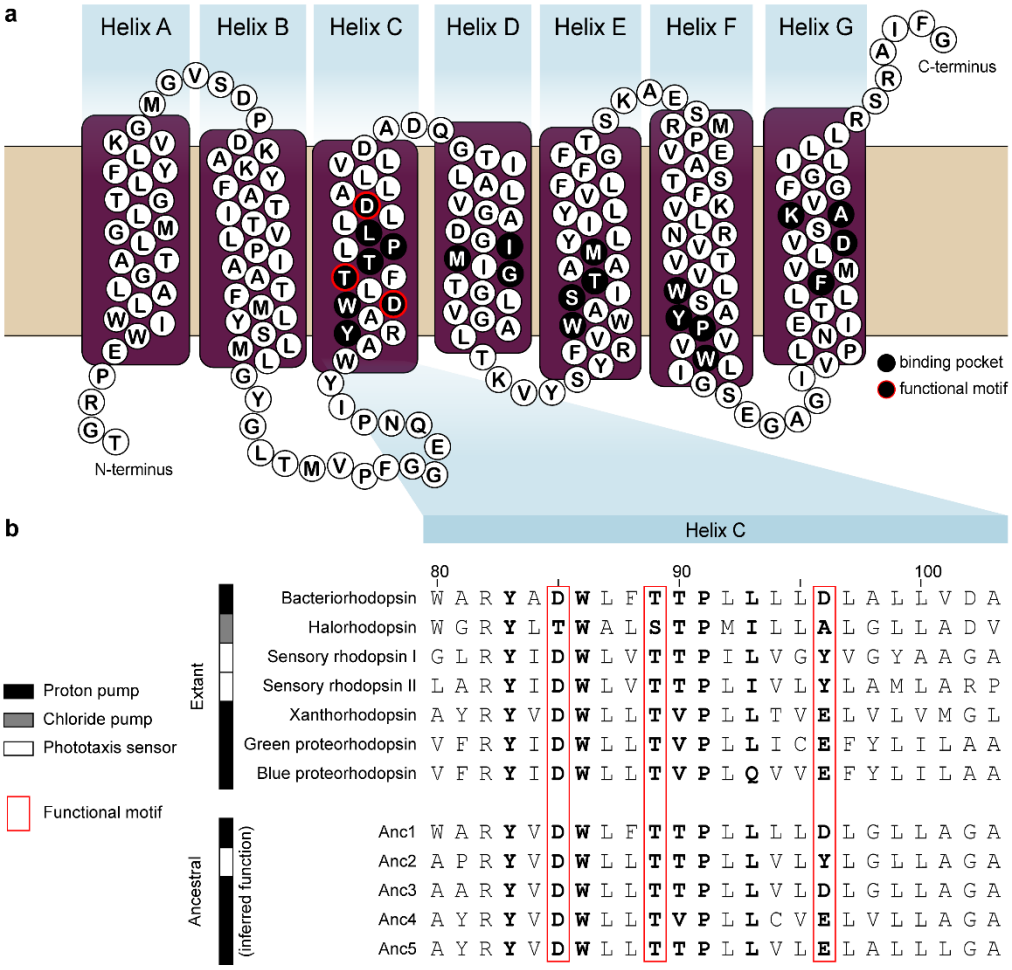


Fig. 4 Rhodopsin binding pocket residues in extant and inferred ancestral sequences. a Secondary structure of extant *H. salinarum* bacteriorhodopsin, indicating the binding pocket residues as well as the functionally correlative DTD motif. **b** Protein sequence alignment of Helix C residues for representative, extant type I rhodopsins and ancestral rhodopsins. Binding pocket residues are indicated by bold font. **a,b** Sites labeled "functional motif" are homologous to the three-residue DTD motif in extant *H. salinarum* bacteriorhodopsin.

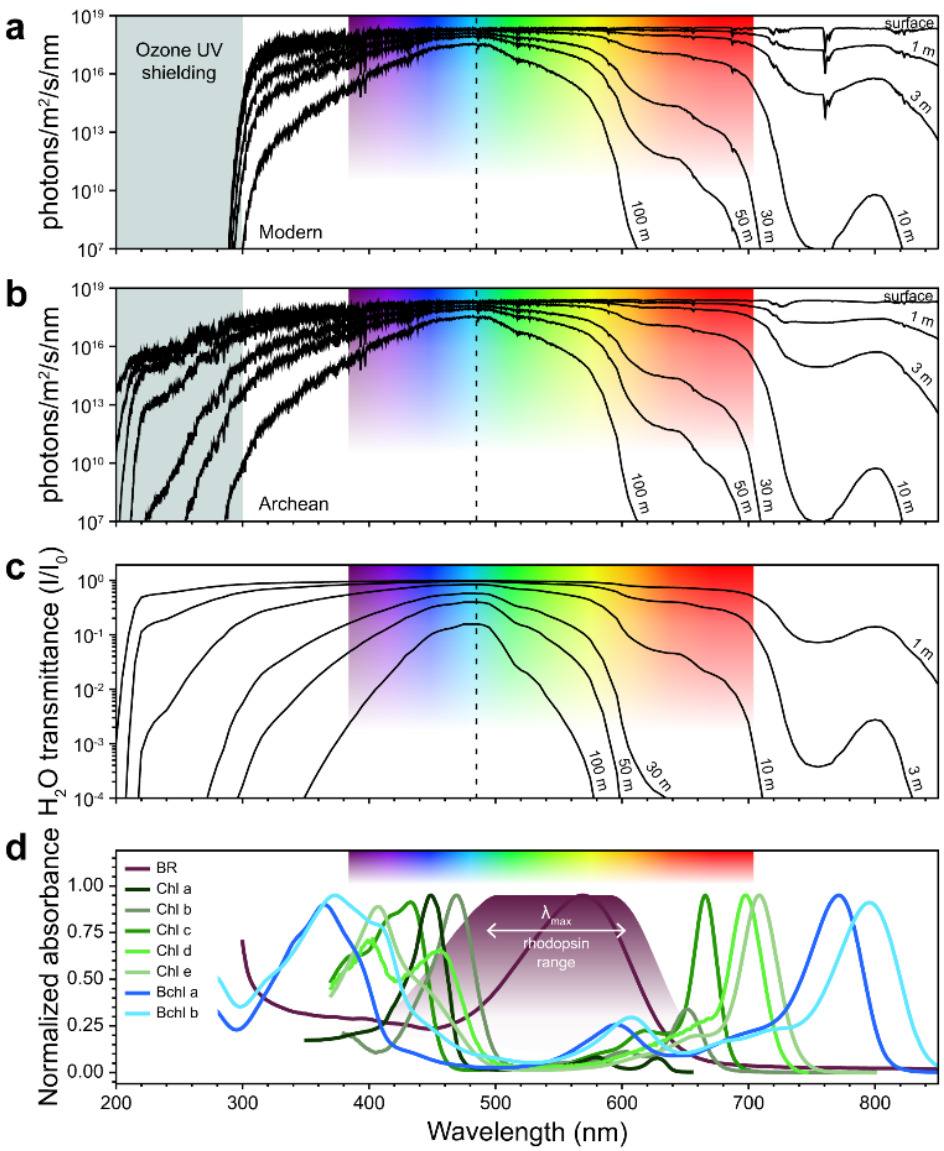


Fig. 5 Diversity of biological pigment absorbance profiles in the context of changing surface photon irradiance over Earth history. **a,b** Photon irradiance profiles at varying water depths, given a modern (**a**) or Archean (**b**) atmospheric composition. **c** Water transmission profile at varying water depths. Transmission was calculated from the coefficients of Segelstein[77] and the Beer-Lambert Law. **a,b,c** Vertical dashed line indicates the maximum water transmission wavelength of 485 nm. **d** Normalized absorbance for bacteriorhodopsin ("BR"), chlorophylls ("Chl"), and bacteriochlorophylls ("Bchl"). Shaded purple region indicates the extant and ancestral range of predicted rhodopsin absorption peaks in Fig. 1. Chlorophyll and bacteriochlorophyll spectra are sourced from the Virtual Planetary Laboratory's Biological Pigments Database (http://vplapps.astro.washington.edu/pigments), with contributions from refs [97-100]. The bacteiorhodopsin spectrum was sourced from DasSarma and Schwieterman[15].

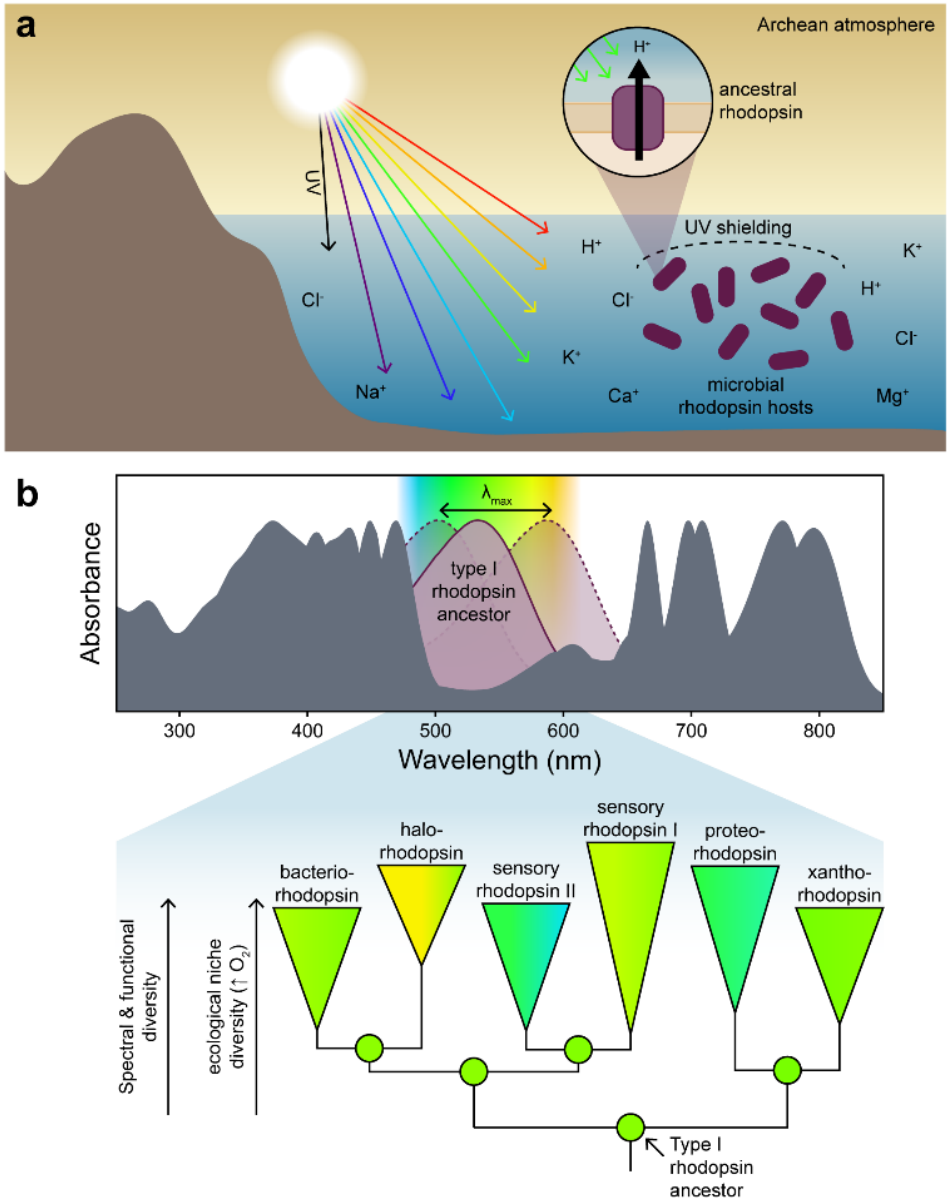


Fig. 6 Coupling between surface irradiance, spectral tuning, and functional diversity over microbial rhodopsin evolution. a Microbial hosts of ancestral type I rhodopsins would likely have occupied the marine environment where high Archean UV irradiation was attenuated and where they could absorb green wavelengths to which they were maximally tuned. These ancestral rhodopsins would likely have functioned as light-driven proton pumps, providing both a simplistic means of energy production and maintenance of the intracellular environment. **b** Color tuning of microbial rhodopsins explored the spectral window not already occupied by other biological pigments such as chlorophylls and bacteriochlorophylls (grey, adapted from Fig. 5d). Spectral diversification would have accompanied functional specialization, which itself was likely responsive to expanding ecological niches driven by the rise of atmospheric oxygen.