An Empirical Test of Convergent Evolution in Rhodopsins by Examination of Novel Folds

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences

Brandeis University

Department of Biochemistry

Douglas L. Theobald, PhD., Advisor

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Kristine Ann Mackin

May 2014
This dissertation, directed and approved by Kristine Ann Mackin’s Committee, has been approved by the Faculty of Brandeis University in partial fulfillment of the requirements for the degree of:

DOCTOR OF PHILOSOPHY

Malcolm Watson, Dean

Graduate School of Arts and Sciences

Dissertation Committee:

Douglas L. Theobald, PhD., Department of Biochemistry

Daniel D. Oprian, PhD., Department of Biochemistry

Timothy O. Street, PhD., Department of Biochemistry

Donald M. Engelman, PhD., Department of Molecular Biophysics and Biochemistry, Yale Univeristy
ACKNOWLEDGEMENTS

My first and most important thanks go to my adviser, Doug Theobald. I almost didn’t rotate in the lab, but I’m very glad I did. Thank you for letting me take on a project that seemed like a totally crazy idea, and allowing me to run with it. I found this work fascinating and the last 5 years have flown by.

I also want to thank all the members of the lab, past and present. Amy Eisenberg and Adam Drake for being two very different but very fun undergraduates to work with, Mackenzie Gallegos for making every day fun, and Brian Beckett for bringing a sense of the absurd to my last eighteen months. I especially want to thank Jeff Boucher for setting the standard I aspire to, even though I often fall short. I have learned so much from you and I will always value the time we spent working together.

Thanks to my friends at Hovey Players and the ACME Improv Company, for giving me a creative outlet and fulfilling my occasional need to be the center of attention. Thanks to my grad school friends, for going through this experience together and always being on the same page as me. Thanks to my Carleton friends, for keeping me from talking about graduate school all the time, and reminding me that there are some many other interesting things to learn and talk about.

Last and most important I want to thank my family, especially my husband Reid. My parents, and my wonderful in-laws, have been so supportive over the last five years. Reid, your support and love mean so much to me, and I look forward to many more years with you.
ABSTRACT

An Empirical Test of Convergent Evolution in Rhodopsins
by Examination of Novel Mutant Folds

A dissertation presented to the Faculty of the
Graduate School of Arts and Sciences of Brandeis University
Waltham, Massachusetts

By Kristine Ann Mackin

Type I and type II rhodopsins share an identical fold, but have no detectable sequence similarity. The evolutionary relationship between these proteins has been debated for decades. Convergent evolution predicts that this shared fold originated due to functional restraints. Homology is typically established when proteins have sequence similarity, although a lack of similarity is a prediction of descent from an ancient common ancestor. Most recent reviews claim these proteins originated through convergent evolution, but current data is not sufficient to definitively conclude that these proteins are not homologous.

In this work, I test the key prediction of convergent evolution. That is, that only one fold is capable of targeting to the membrane, folding correctly, binding retinal, and undergoing the light-activated proton pumping to provide a critical energy source for the archaeon. I designed and characterized seven novel fold mutants and one point mutant that changes the position of the
strictly conserved lysine required for covalent binding to the retinal cofactor. These proteins all target to the cell membrane, fold, and bind retinal, causing a color change upon expression in the E. coli cells. The seven fold mutants all absorb within 2 nm of the wild-type, indicating that the protein is forming the correct helical contacts and binding retinal. When reconstituted into liposomes, the mutants all exhibit light-activated proton pumping activity from 40-170% of the wild-type. Only one of the lysine point mutants, A53K/K216A, has reproducible proton pumping activity at about 11% of the wild-type. The ability of these seven novel folds and the single point mutant indicated that evolution is not constrained to one particular fold, so convergence is unlikely. Therefore I propose that type I and type II rhodopsins share an ancient common ancestor and are homologous.
LIST OF TABLES

3.1 Absorbance of Lysine Swap Mutants ......................................................... 43
3.2 Cm of Mutants .................................................................................. 45

4.1 pH Dependence of Proton Pumping ......................................................... 54

A1 Protein Segments ........................................................................ 73
A2 Mutant Sequences ........................................................................ 74
## LIST OF FIGURES

1.1 Comparison of Rhodopsin Folds .......................................................... 2  
1.2 Retinal Isomerization in Rhodopsins ....................................................... 3  
1.3 Convergent Evolution of Wing Structures .............................................. 6  
1.4 The Structure of Bacteriorhodopsin ....................................................... 12  
1.5 Schiff Base Absorbance Changes ............................................................ 14  
1.6 Proposed bR Folding Mechanism ........................................................... 16  
1.7 The Photocycle ................................................................................. 17  
1.8 Halorhodopsin Compared to Bacteriorhodopsin ........................................ 18  
1.9 Construction of Novel Folds ................................................................. 24  

2.1 Predicted Evolutionary Rates ................................................................. 28  
2.2 Complete Set of Fold Mutants ............................................................... 31  
2.3 Model of the A61K/K225A Mutant .......................................................... 32  
2.4 Colored Cell Pellets .............................................................................. 35  
2.5 Membrane Fractionation ...................................................................... 35  

3.1 Mutant Absorbance .............................................................................. 41  
3.2 Schiff Base Formation and Acid Trapped Spectra .................................... 43  
3.3 Unfolding of Mutants ........................................................................... 44  

4.1 Wild-Type Proton Pumping .................................................................... 53  
4.2 Light Induced Activity of Fold Mutants .................................................. 56  
4.3 Calculated Relative Activity ................................................................... 56  

5.1 Various N-terminal Sequences Target the Membrane ............................. 60  
5.2 Folding Pathway of GBCDEFA .............................................................. 61  
5.3 Proposed Rhodopsin Phylogeny .............................................................. 70
Chapter One

Background and Significance
1.1 Project Overview

The evolutionary relationship between type I and type II rhodopsins is an unresolved debate, starting from the first description of an archaeal rhodopsin nearly fifty years ago (1). The shared seven transmembrane (7TM) α-helical fold (Figure 1.1) and other similarities are evidence that these families may be descended from a common ancestor, that is, the are homologous (2). However, the absence of detectable sequence similarity, as well some structural and functional differences, may indicate that the shared fold emerged due to convergent evolution (2). Despite the lack of conclusive evidence, recent reviews have favored a non-homologous relationship; it is claimed that the similarities are due to physical constraints in the membrane environment (3-7).

Figure 1.1: Comparison of Rhodopsin Folds: The retinal cofactor is shown in magenta. a) The structure of bacteriorhodopsin from *Halobacterium sp. aus-1*, PDB entry 1UAZ. b) The 7TM fold, with helices marked using the convention for Type I rhodopsins. Solid lines are intracellular loops, and dashed lines are extracellular. c) The structure of bovine rhodopsin, PDB 3C9L.
Type I rhodopsins were discovered in the purple membrane fraction of a halophilic archaea (1); this bacteriorhodopsin (bR) from *Halobacterium salinarum* is the canonical member of the type I family. All type I rhodopsins share the 7TM structure, and bind covalently to retinal (5). A Schiff base is formed between a conserved lysine in the middle of the 7th (G) helix and the cofactor (5). When activated by a photon, the retinal in bR undergoes a light-induced isomerization, converting from all-trans to 13-cis (Figure 1.2). This isomerization initiates a series of conformational changes that drive the photocycle. Each photocycle results in the net transfer of a proton from the interior to the exterior of the cell, and the accumulated proton gradient is harnessed to generate ATP, a critical energy source for the cell in low nutrient environments (8).

![Figure 1.2: Retinal Isomerization in Rhodopsins](image)

Figure 1.2: Retinal Isomerization in Rhodopsins: a) The type II isomerization from 11-cis to all-trans. b) The type I isomerization from all-trans to 13-cis.
Type II rhodopsins use the same 7TM fold and form the Schiff base between retinal and a conserved lysine in the 7th helix (Figure 1.1) (5). However, there are crucial differences in function. The canonical type II rhodopsin, bovine rhodopsin, undergoes an isomerization from 11-cis to all-trans retinal (Figure 1.2). The conformational change is coupled to activation of the G-protein transducin, which initiates a signaling cascade (5). Interestingly, bovine rhodopsin also transfers a proton from the positively charged Schiff base to a nearby acidic residue, but does not transfer the proton completely across the membrane (9). After a single isomerization the eukaryotic rhodopsin must be renewed through the removal and replacement of the cofactor with a new 11-cis retinal (10). Type II rhodopsins thereby function as a light-sensing molecule for the organism, rather than an energy source for an individual bacterium.

The work presented in this dissertation is a functional test of the convergent hypothesis in rhodopsins. If one or only a few folds are functional, then evolution is more likely to arrive at a single fold multiple times. No previous work has tested the ability of alternate folds to perform the required function; rather, the literature has relied on structural and sequence based arguments to support convergence or divergence. My aim is to examine the structural constraints on the 7TM fold by determining whether alternative topologies can insert into the cell membrane, fold correctly, and undergo the light-activated conformational changes. This chapter 1) explains the current understanding of evolutionary constraints on protein folds and 2) provides background on the structure, function, and evolution of both type I and type II rhodopsins.
1.2 Evolutionary Overview

1.2.1 Evolution of Protein Structures

When proteins share an identical fold but have little or no detectable sequence similarity, it is difficult to determine whether the fold arose through convergent or divergent evolution. Homology, or descent from a common ancestor, can be established if there is significant sequence identity (11). The SCOP database sets the threshold to establish protein families at 30% identity; the E-values generated using the BLAST algorithm are also a good measure of similarity, with smaller values indicating higher similarity (11-13). Sequence space is practically infinite, so the chance of unrelated proteins converging on a similar sequence is vanishingly small (11, 14, 15). Fold space is much smaller than sequence space; for this work the SCOP definition of a fold is used, which is the secondary structural elements and, critically, their topological connections (12). There are numerous examples of proteins that have identical folds but have little or no detectable sequence similarity (16). The problem arises when folds are identical but the sequence similarity is not statistically significant. A lack of sequence similarity is insufficient to reject homology without additional evidence, leaving the evolutionary pathway ambiguous (17, 18).

Convergent evolution will drive unrelated proteins to share a single fold, if that fold is uniquely well suited to perform a function. Early in the study of protein structure, researchers hypothesized that there were a limited number of stable structures available (19, 20). This observation has born out; even as the number of known protein structures continues to increase, the number of novel folds identified each year is decreasing (21). By induction, there is a finite number of protein folds available, making convergence on some folds almost inevitable (22).
This structural convergence is most easily illustrated by a familiar morphological example. The wing structure of birds, bats, and pterosaurs demonstrates the results of convergent evolution (Figure 1.3). Flappable wings are constrained in terms of aerodynamics, strength-to-weight ratio, surface area, and biomechanics. Due to the physical requirements for flappable wings, these unrelated lineages have all converged on similar wing structures. Convergence in proteins is most easily identified when information in addition to the sequence is available, especially if the ancestral state is known, and differs from the modern form. In the case of flappable wings, the phylogeny contains a non-winged common ancestor which makes it superficially easy to identify the forces of convergent evolution at work. If a certain function can only be supported using a specific structure, this is strong evidence in favor of convergence. However, these constraints are difficult to establish; folds can naturally change during evolution, demonstrating considerable flexibility.

Figure 1.3: Convergent Evolution of Wing Structures: The wing structure of birds, pterosaurs, and bats evolved independently but appears similar due to functional constraints.
Various pathways can lead to novel folds during the course of protein evolution. Circular permutation can be used to generate alternate protein folds, both naturally and artificially. A circular permutation generates a new topology by ligating the native N and C termini and cutting elsewhere to generate new termini (23). Naturally occurring circular permutations have been found in some families of soluble proteins and is one mechanism for fold change (24-26). Artificial permutations have also been created in both soluble and membrane proteins to study folding and the relationship between structure and function (23, 27, 28). Generation of a circular permutation is easiest when the termini are proximal in the native fold, although larger spans can be bridged using a flexible glycine-glycine-serine or similar peptide linker (23). In most cases, the activity of the protein is reduced in the permuted form, although some permutations can result in improvements to function (29-31).

Even more dramatic structural rearrangements have been observed in the evolution of some soluble proteins (15, 25). Intermediate folds are preserved in other family members, and sequence similarity is sufficient to establish the evolutionary relationships between the transitional forms and the endpoints; examples include deletion of secondary structural elements, or \( \alpha \)-helices converting to \( \beta \)-sheets (25). Fold change can also occur by gene duplication and fusion, adding complexity and generating novel folds of increased size (32-34).

Despite known examples of structural convergence in soluble proteins, the discovery of apparently unrelated membrane proteins with identical folds came unexpectedly. Bacteriorhodopsin was just the first in a series where a shared fold was identified despite the absence of sequence similarity (14, 35). The unanswered question remains: do these proteins
share a fold because of functional restraints, or due to divergence from an extremely ancient common ancestor?

1.2.2 Constraints on Membrane Protein Evolution

The number of folds available to membrane proteins is hypothesized to be even smaller than for soluble proteins, leading to frequent structural convergence (14). Membrane proteins interact with more environments than a soluble protein; a greasy lipid phase, two aqueous extramembrane regions, and two transition zones, although the boundaries between these phases are not well-defined (36, 37). Additionally, these proteins must insert into the cell membrane, fold, and function correctly. These constraints hint that fold space for membrane proteins may be smaller than for soluble proteins, making convergence more likely (14). In addition, fold change is less common in membrane proteins than for soluble proteins, further supporting the idea of limited fold space and making it likely that related proteins would maintain a fold even with very long evolutionary distances (38).

Structural permutation can be used to explore the ability of integral membrane proteins to adopt alternative folds, but to date it has been used on only two families. One of those studied is the bacterial glucose transporter family (27, 39). The topology of the glucose transporters was predicted at the time of the first investigation, which a crystal structure later confirmed (40). In this study the predicted topology was used to generate circularly permuted variations through systematic cleavage of each extramembrane loop. This resulted in proteins with decreased transport activity (27). However, the ability of these mutants to fold and insert in the membrane is in itself a significant achievement.
Similar experiments were done in the Omp family of porins (28, 41). OmpX is a β-barrel, and the unstructured loops were chosen as cleavage sites (28, 42). In this study, the circularly permuted mutants were capable of inserting into the membrane, although less efficiently than the wild-type (28). Another experiment performed both circular and non-circular permutations on the related OmpA protein. The cut sites were based on a predicted topology, which was confirmed when the crystal structure was solved (41, 43). In this case, the circular mutants were all capable of inserting into the membrane, though the rate of insertion was reduced. None of the more extreme non-circular permutations were able to insert correctly (41). The activity was not analyzed (41).

Another way folds can change in membrane proteins is through gene duplication and fusion. There are sequence and structural hints about a possible ancestral duplication event in many transmembrane proteins (33), and inverted repeats are common in channel proteins (35, 44). The evolutionary intermediates leading to this structure may be exemplified in some modern protein families; the Fluc family of proteins and EmrE demonstrate two intermediates on this pathway (45, 46). EmrE is the pre-duplication form; it functions as a homodimer in which the monomers insert randomly to form either a parallel or an anti-parallel dimer (46). The Fluc family demonstrates various evolutionary intermediates, including the duplicated but unfused gene in bacteria and the fused state in eukaryotes (45). Additionally, gene duplication is the hypothesized origin of the 7TM and other α-helical folds. One hypothesis predicts that an ancestral 4-helix protein underwent duplication in the first 3 helices to form the modern 7-helix bundle (47).
Another constraint on the evolution of novel membrane protein folds is the ability to target and insert correctly into the membrane. The circular permutation studies show that non-native N-termini are able to facilitate membrane targeting and insertion (27, 28). The expression and purification of membrane proteins in *E. coli* often hinges on correct folding and insertion into the membrane or the ability to refold a protein expressed in inclusion bodies (48). Most integral membrane proteins achieve this by binding to the signal recognition particle (SRP), which then targets the ribosome to the translocon (36). The emerging peptide is inserted through the pore of the translocon, and the protein folds co-translationally and is partitioned into the membrane after folding (49). It is currently unclear what kind of signal is required for SRP binding, but the ability of permuted proteins to correctly incorporate into the membrane demonstrates that many different sequences are permissible (50). Significantly, the *E. coli* SRP is capable of recognizing the signal sequences of non-native proteins, which greatly facilitates the study of membrane proteins and indicates that this may not be an evolutionary bottleneck (51).

The folding pathway for integral membrane proteins may also contribute to structural restrictions. A two-stage model has historically been favored for α-helical membrane proteins; first, each helix achieves its secondary structure and second, the helices condense into the tertiary structure (52). This model is now considered too simple; additional complexity was added to the pathway as more proteins were studied (53, 54). Other steps can include the folding of extramembrane domains or cofactor binding (54). During the folding of some proteins, certain helices are retained at the translocon, adding additional complexity to the folding pathway (55, 56).
The ability to change folds using circular and non-circular permutations, as well as SRP recognition of non-native helices, indicates that structural constraints for membrane proteins may be less stringent than previously thought. However, structural changes will be more difficult if certain sequences are required to fold and condense in order for the protein to access the functional tertiary structure. Convergence is more likely if only one or a very small number of folds are able to perform the function; if multiple folds are functional, there is little evolutionary pressure for unrelated proteins to arrive at a single topology. In this case, divergence from an ancient common ancestor is the more parsimonious explanation for the shared fold, especially given that fold change appears to be uncommon in membrane proteins, so a fold is likely to be preserved even over long evolutionary distances \((38)\). These two opposing hypothesis have both been proposed to explain the shared seven transmembrane \(\alpha\)-helix fold of type I and type II rhodopsin, and this conflict has not yet been satisfactorily resolved.

1.3 The Two Rhodopsin Families

1.3.1 Type I Rhodopsins

The canonical type I rhodopsin is the bacteriorhodopsin from \textit{Halobacterium salinarum} (called \textit{H. halobium} in early papers) \((1)\). This protein was immediately noted to resemble the known visual opsin; it also formed a Schiff base with a retinal cofactor, had a similar molecular weight, and was light-sensitive \((1)\). This paper proposed a photoreceptor function for this new protein; however, the proton pumping activity was soon discovered, and bacteriorhodopsin was shown to be an energy source for the cell when glucose and other metabolites are scarce \((57-59)\). The type I family also contains a light-activated chloride pump, halorhodopsin, as well as two
sensory rhodopsins, but bacteriorhodopsin remains the most studied member of this family (5, 7).

Bacteriorhodopsin shares the seven $\alpha$-helical fold found in type II rhodopsins and other G-protein coupled receptors (GPCRs) (Figure 1.4) (5). The alpha helical structure was determined by electron microscopy; this was the first membrane protein to be structurally characterized (60). The topology was elucidated later, when higher resolution models were determined (61). The structure has now been solved to 2 Å resolution, thanks to repeated x-ray crystallographic and electron microscopy studies (61-64). In fact, bR is often used as a proof of principle for new membrane protein crystallographic techniques (65).

Figure 1.4: The Structure of Bacteriorhodopsin a) The structure of bacteriorhodopsin from Halobacterium sp. aus-I, PDB entry 1UAZ. The retinal cofactor is shown in magenta. b) Residues within 10 Å of retinal are shown as sticks, with the top of the structure rotated towards the viewer by 90 degrees.

Bacteriorhodopsin is expressed at high levels when the native archaebacteria are grown under illumination (58, 66, 67). This is expensive in terms of both time and materials, as high
salt and extended incubations are required, so recombinant expression systems were explored (68). Using the native halophile for expression of mutants is also more difficult than using commercially available competent cells; the native bR must be knocked out, then the mutant transformed into the halophilic cells (69). Expression in E. coli is possible, but several challenges must be overcome. Although SRP activity is observed both in the native H. salinarum and in E. coli (51), bR expressed in E. coli does not achieve the native fold (70, 71). It binds to retinal but forms an inactive conformation, which must be extracted and reconstituted to obtain active protein (72). In spite of these challenges, bR has been extensively studied in both the native archaebacteria and in reconstituted systems.

The retinal cofactor imparts a characteristic absorbance, which is an easy way to determine if the protein is expressed and folded correctly. Free retinal has a λmax of 380 nm; the Schiff base absorbs at 440 nm, and the packing of the hydrophobic core of the protein around retinal red-shifts the absorbance spectrum (Figure 1.5) (73). The absorbance of the folded protein is tuned to different wavelengths in different species; a λmax of 547 nm is observed in bacteriorhodopsin from H. turkmenica, which is blue-shifted from the H. salinarum bR that absorbs around 570 nm (74). Even conservative mutations in the binding pocket shift the absorbance spectrum, providing a visual and facile indication of structural changes to the protein (72). Mutations at the counterion Asp 85 or other residues in the interior of the protein cause large spectral shifts (numbering of bR from H. salinarum is used throughout) (72, 75-77). However, the activity is resistant to many different mutations, and bR can even maintain proton pumping using retinylidene-ethylamine when Lys 216 is mutated to disrupt the Schiff base linkage (72, 75, 78). Most surprisingly, the functional bR molecule can be reconstituted from
various sets of fragments that have been expressed and purified separately, only requiring that all seven native helices are present (77, 79).

![All-Trans Retinal Spectra](image)

**Figure 1.5: Schiff Base Absorbance Changes:** The absorbance of all-trans retinal shifts upon Schiff base formation, and shifts further when it is bound in correctly folded protein.

The folding and stability of bacteriorhodopsin has also been well studied using the characteristic absorbance to detect the folded state. The folding of membrane proteins has been notoriously hard to study; many unfold irreversibly, forming aggregates so that the native state is unrecoverable (80). This prohibits calculation of the free energy of unfolding (81).

Bacteriorhodopsin can be refolded from the denatured state in SDS simply by diluting the solution below the critical micelle concentration (CMC) (52). This reaction is easily observed by monitoring the appearance of the characteristic bR absorption (52). These two advantages have facilitated studies of unfolding in both native and mutant bR (82–86). Recently, several assumptions about the kinetics of unfolding and refolding in bR have been questioned, challenging previous studies (87, 88). The refolding reaction was found to be slower than was
previously reported (88). Most previous studies monitored the transition of \( \text{bR} \rightleftharpoons \text{bRu} \), but this mechanism does not represent the final equilibrium and may not accurately model the unfolding transition (87). The new proposal incorporates the hydrolysis of retinal into the equation, \( \text{bR} \rightleftharpoons \text{bRu} \rightarrow \text{bOu} + \text{RET} \). However, experiments using equations derived from this model yield \( \Delta G \) values very close to those obtained using the simpler assumption, and conclusions from previous experiments are likely to remain valid (87).

Studies of unfolding in vitro as well as the results from in vivo experiments initially described the folding of bR using a two-state model (89). First the helices insert into the membrane and form the secondary structure; second they condense into the native tertiary structure (90). The three-step model currently favored adds an additional step to account for binding of the cofactor (54). Retinal is already present in the membrane, due to its hydrophobicity, and inserts into the protein (54, 91). It then binds covalently to the strictly conserved lysine at position 216 in the middle of the G (7th) helix (92). In the second step of the proposed folding pathway, helices F and G are retained outside the membrane until helices A through E form a stable folding core, which then induces insertion of the final two helices (Figure 1.6) (52, 93). This result has been corroborated by the observation that the G helix is largely unfolded in the transition state, and requires contact with neighboring helices to form the tertiary structure (85).
Figure 1.6: Proposed bR Folding Mechanism: The stable folding core and the FG unit insert separately into the cell membrane, after Booth’s model (52).

The photocycle of bacteriorhodopsin is well characterized (Figure 1.7) (5, 8). The protein passes through various structural states, each having a characteristic absorbance (94). Many states have been trapped and studied crystallographically, enabling detailed models for the movement of individual amino acids (8, 94, 95). These models are controversial, as some researchers believe the crystals consist of mixed states or represent states that are not physiologically relevant, leading to conflicting models of proton pumping, and even the suggestion that bR may be a hydroxide rather than a proton pump (8). However, it is clear that the isomerization of retinal is the primary driving force of the conformational changes, and each state to has an observable $\lambda_{\text{max}}$, which can be used to monitor the structural changes in real time (94).
The resting state of the protein is called the bR state, and has a characteristic absorbance near 570 nm. Critically, the Schiff base nitrogen is positively charged (5, 63). Upon absorption of a photon, retinal isomerizes from the all-trans to 13-cis. The most cited models state that the protein passes rapidly through the K state and enters the L state, and the Schiff base proton is transferred to Asp 85 (Figure 1.7b, step 1). Upon entering the M state, Arg 82 releases its proton to the extracellular side (Figure 1.7b, step 2) (63). The transition to the N state brings Asp 96 into position to reprotonate the Schiff base (step 3) (8). The thermal relaxation of retinal to the all-trans state drives the final steps of the photocycle, allowing reprotonation of Asp 96 from the cytoplasm in the O state (step 4). In step 5, Asp 85 reprotonates Arg 82 and brings the protein back to the resting bR state, primed to restart the photocycle. The result is the net transfer of one proton from the interior to the exterior of the cell (8).
Halorhodopsin (hR) is a related type I rhodopsin that shares the 7TM fold and undergoes a similar photocycle. However, hR transports a chloride ion in the opposite direction of the proton transported by bR (Figure 1.8) (96). Point mutants D85S or D85T can convert bR into a chloride pump (97), but the converse point mutation in hR does not convert it to a proton pump. Varying light conditions or a high concentration of azide, which can serve as the counterion, is needed to achieve proton pumping in hR (98, 99).

![Figure 1.8: Halorhodopsin Compared to Bacteriorhodopsin](Figures from Cartailler & Luecke 2000)

In addition to bR and hR, there are two sensory rhodopsins that are widespread in archaeal genomes; sequence similarity between these four proteins is relatively high, clearly supporting homology (7). The retinal binding pocket in particular has high sequence conservation, likely due to functional constraints (100). The type I rhodopsins have a broad but patchy distribution in prokaryotes and some single-celled eukaryotes (101); this is attributed to
rampant interdomain lateral gene transfer (7). The current hypothesis for the evolution of this family of proteins predicts that the sensory rhodopsins emerged through duplication and differentiation from an ancestral protein possessing transport activity (102). The proton pump is hypothesized to be basal to the other type I rhodopsins (7). Interestingly, under certain conditions sensory rhodopsin I exhibits residual proton pumping ability (103). The phylogeny of type I rhodopsins is an informative example of a new function using a scaffold that originated for a different purpose. Accumulated mutations have altered the function, but the fold is more difficult to change and is therefore preserved.

1.3.2 Type II Rhodopsins

Type II rhodopsins share the 7TM topology of bR (104). The most well-studied member of the type II family is bovine rhodopsin. The first high resolution crystal structure confirmed the topology and provided structural details, further complicating the debate about homology between type I and type II rhodopsins (105).

Although type II rhodopsins bind to 11-\textit{cis} retinal instead of all-\textit{trans} retinal in the resting state, the photo-response is still remarkably similar. The protonated Schiff base is formed with a conserved lysine at position 296 (using the numbering from bovine rhodopsin), in the middle of the 7th helix (5, 104). The retinal isomerizes to all-\textit{trans} upon absorption of a photon, resulting in a conformational change that deprotonates the Schiff base (106). In bovine rhodopsin, the counterion is a glutamic acid at position 113, rather than the aspartic acid present in type II rhodopsins (107). Nonetheless, the substitution is conservative, and proton transfer from the Schiff base to the counterion is required for interaction with transducin, the G-protein coupled to
rhodopsin (9). At this point, bovine rhodopsin has completed its light response; unlike type I rhodopsin there is no thermal relaxation or further conformational changes. In vertebrate rhodopsins, the all-trans retinal is removed and regenerated in a complicated series of enzymatic steps (10), in contrast to type I rhodopsin’s ability to return to the resting state and continue to cycle (108).

Bovine rhodopsin is a member of the G-protein coupled receptor (GPCR) family, which serve various sensory roles instead of a transport function like bacteriorhodopsin. GPCRs are found throughout eukaryotes and take on diverse biological functions (109). All five families of GPCR are thought to have descended from a cyclic AMP receptor (cAMP) present in the common ancestor of all eukaryotes (110). Type II rhodopsins were recently found to be present in fungi, and this phylogenetic overlap is an intriguing hint that the type I and type II rhodopsin families may have evolved from an extremely ancient common ancestor (111). It is not yet established how the 7TM core originated, leaving open both the possibility of convergent evolution to this extremely useful architecture, or divergent evolution from an ancestor present in an ancient organism.

1.3.3 Evolutionary Hypotheses

The significant structural similarities between type I and type II rhodopsins initially led researchers to believe these two families were homologous, and that the more tractable bR could be used as a model for all opsins (1, 112). The shared fold has long been the primary argument in favor of an ancient common origin for these two protein families (2, 113). The lack of sequence similarity is often cited as evidence against homology; however, this observation is also a
prediction of an ancient divergence. Type I and type II rhodopsins are phylogenetically distinct and the evolutionary distance between them is large enough that sequence identity is predicted to be lost due to accumulated mutations \((2, 102)\). The rate of mutation for bacteriorhodopsin in halophilic bacteria is predicted to be 0.5 substitutions per site per \(10^9\) years \((102)\). This calculation predicts one mutation per site since the hypothesized divergence between type I and type II rhodopsins, which is sufficient to explain the lack of sequence similarity \((102, 114)\).

It is possible to align portions of type I and type II sequences, prompting speculation about gene shuffling or an internal duplication in the evolution of these families \((47, 113, 115)\). If the transmembrane helices are excised from the native sequence, some portions of the sequence appear to be conserved between type I and type II rhodopsins \((115, 116)\). Similarities between helices 1-3 of type II rhodopsins and helices 5-7 of type I rhodopsins led to the proposal that an ancient internal duplication event led to the formation of the modern 7-helix bundle. The purported ancestor had only 4 helices, of which the first three duplicated to form helices 5-7. The superficial similarity between helices 1-3 of type II and helices 5-7 of type I rhodopsins could be explained by this ancient duplication event \((47)\).

This ability to align helices that do not correspond in the native sequence is not a conclusive argument for homology, and the weak similarity between some helices is unconvincing. The most compelling analysis aligns not only bR and bovine rhodopsin, but includes many opsins from both families, and concludes there is no ancestral relationship between the two families \((3)\). Further work has probed for evidence of exon shuffling, decoupling the helices from their native order and attempting to align discreet segments of bR and bovine rhodopsin. This analysis has found some identity between non-concurrent helices, but
again, this result is not sufficient to resolve the debate about homology (116). Another analysis looked for evidence of a duplication in type I rhodopsin (17). The lack of sequence similarity between helices 1-3 and 5-7 in type II rhodopsins does not support a duplication event in the evolution of this family (17).

The convergent hypothesis often emphasizes the minor structural differences between the proteins as evidence of non-homology. The helices in type II rhodopsins contain kinks due to proline residues near the center of the membrane (105). These kinks cause the helices in type II rhodopsins to sit at an angle, while the helices in type I rhodopsins lie perpendicular to the membrane surface (5). Proline is commonly found in transmembrane helices, so the presence of kinks in the helices is somewhat unremarkable (117, 118). Additionally, recent work shows it is relatively easy for evolution to insert kinks into transmembrane helices, diminishing the impact of these minor differences as evidence against divergence (119).

The isomerization from 11-cis to all-trans retinal in type II rhodopsins may also constitute evidence against divergence; the proteins bind structurally different forms of retinal in the resting state (3). The recently discovered middle rhodopsin, which isomerizes from all-trans to 11-cis instead of to 13-cis like the classical bR, may be a missing link in the evolutionary path between the two families of rhodopsins, or at least demonstrates that alternative retinal configurations can arise with relative ease (120).

On the basis of sequence and structural evidence, recent reviews all cite convergent evolution as the origin of these two families (3-7). In spite of this consensus, the identical fold must still be explained. This problem has been visited and revisited since the original papers proposing the 7TM topology (3, 17, 121). The fold has been suggested to be a “stable
arrangement” \((121)\) which is “uniquely suited” \((122)\) and “may well represent an extremely favorable framework” for transmembrane proteins \((123)\). The probability of independently evolving proteins converging on an identical fold relies on the membrane being constrained so that only one fold is capable of folding, targeting and inserting into the membrane, binding retinal, and performing the light sensitive function \((14, 124-126)\).

The repeated use of the 7 TM GPCR fold as a scaffold for various functions begs the question whether this fold is functionally unique. The strict definition of a fold, from the SCOP database, is the secondary structural elements and their connectivity, so if the topology changes, then by definition the protein has a different fold \((127, 128)\). A simple thought exercise, illustrated in Figure 1.9, demonstrates that many other topologies are capable of forming a seven helix bundle in the membrane. When the connective loops are stripped away from the GPCR fold the seven directional \(\alpha\)-helices remain. Starting from this stripped down secondary structure, it is possible to construct 144 unique ways to connect the helices as a single polypeptide, which constitutes a novel topology and maintains the directionality of the helices. If multiple different folds are functional, convergent evolution is less likely because independently evolving proteins can use alternative topologies. If multiple unrelated proteins have converged on one fold, it is probable that one fold out of 144 is best able to insert into the membrane, fold, and perform the light-activated conformational change.

Although this functional uniqueness has been the underlying assumption in favor of a convergence of type I and type II rhodopsins, it has never been experimentally tested. Bacteriorhodopsin is incredibly resistant to cleavage, maintaining the proton pumping activity even when fragments are expressed independently \((77, 79)\). Functional bovine rhodopsin can
also be reconstituted from fragments, but these cleaved proteins do not constitute alternative
topologies (129). These results do not answer the question of what will happen when the
connective loops are rearranged, since the cleavage does not put any strain on the protein, but
allows it to form the native contacts.

Figure 1.9: Construction of Novel Folds: An illustration of the method for
counting new topologies of bR. In a) there are four choices for the location of the
N-terminus, marked by stars. In b) there are three choices for the location of this
first connection. Once all connections have been made, there are
\[ 4 \times 3 \times 2 \times 2 \times 1 = 144 \]
possible topologies.

1.4 Outline of Dissertation

I set out to test the prediction made by the convergent hypothesis, that only a limited set
of folds are functional, by generating and characterizing fold mutants of bR. I choose to use
bacteriorhodopsin from *Haloterrigena turkmenica*, which is 53% identical to the canonical *H*. 

24
salinarum bR. The homolog from *H. turkmenica* can insert directly into the membrane of *E. coli* when expressed, making it an easier protein to study although it does not form the highly structured purple membrane of *H. salinarum* (74). When correctly folded and bound to retinal, the *H. turkmenica* bR causes the pelleted cells to take on a distinctive pink color, an easy visual test for successful expression. Since the protein is folded, purification can easily be performed following detergent extraction from the membrane fraction using standard biochemical methods. For these reasons, this member of the type I rhodopsin family was chosen to evaluate the function of novel folds.

First I designed relatively conservative swap mutants, exploiting less conserved portions of the protein. More complex and extreme mutants followed. Once these mutants were successfully expressed and purified, I demonstrated that the helices in the mutants pack similarly to the wild-type, although the stability varies. Finally, each mutant was shown to be functional in a light-dependent proton pumping assay. The ability of multiple folds to function contradicts the main prediction of convergent evolution; if multiple folds are functional, it is less likely that independently evolving proteins will arrive at an identical fold. Based on my evidence, I suggest that type I and type II rhodopsins diverged from a common ancestor in opisthokonts roughly 1-2 billion years ago.
Chapter Two

Design and Expression of Mutants
2.1 Design of Mutants

2.1.1 Analysis of Known Structures

Phylogenetic methods were used to estimate the rate of mutation for each residue position in bR. First, the sequence of bR from *Haloterrigena* sp. arg-4 (NCBI entry 4579714) was used in a BLAST search against the non-redundant protein sequence database (13, 130). This sequence differs at only two positions from bR from *Haloterrigena turkmenica*, which is used for the experimental portion of this work (74). Sequences with BLAST P E-values <10⁻⁵ were downloaded. The collected sequences were aligned using the program MUSCLE (131, 132). Making the sequence alignment smaller decreases the time required to perform the subsequent computations, so the alignment was visually examined using SeaView, and sequences that appeared to be much shorter or much longer than average were removed. PROBCONS was then used on the reduced set of sequences to make a more accurate alignment (133). This alignment was processed by SeqCon (in house, Theobald, unpublished), which removed gaps in the alignment by deleting columns where fewer than 5% of the sequences contained an amino acid.

This alignment was then converted into Phylip format and used to generate a phylogenetic tree using PhyML (134). The calculated phylogeny, sequence alignment, and Chain A of PDB entry 1UAZ were given as inputs to CONSURF (135). This results in a new PDB file where the B-factors are indexed to the calculated rate of evolution.

No crystal structure is available for bR from *Haloterrigena turkmenica*, so the transmembrane helices were annotated using homology modeling. A homology model of HtbR was generated using the known structure of the closely related archaerhodopsin from *Halorubrum* sp. aus-1 (136). The online homology modeling tool SwissProt was used, with the
starting structure 1UAZ and the sequence of \textit{H. turkmenica} bR (137-139). The calculated model retained the 7-helix topology and closely resembles the known structures of several related type I rhodopsins. The sequences of the \textit{H. turkeminca} bacteriorhodopsin and the \textit{H. sp. aus-1} archaerhodopsin were aligned using Clustal W 1.83 (140). Comparison of structures and sequences allowed the confident annotation of the beginning and end point of each helix in the 7TM structure.

2.1.2. Design of Mutants

The predicted evolutionary rates of type I rhodopsins were used to design the first simple swap mutant. The core of the protein is evolving quite slowly, as expected. Positions exposed to the hydrophobic lipid chains have an intermediate rate of mutation. The cytoplasmic side of the
protein appears to be evolving more quickly than the extracellular side (Figure 2.1). In particular, the β-sheet formed between helices B and C is a site of relatively rapid mutation; bacteriorhodopsin cleaved in this loop maintains proton pumping activity (77, 141). However, to maintain the directionality of the seven helices, one of the more conserved extracellular loops must be chosen as the site of cleavage.

The loop between helices A and B was cut to create the new termini; this connection is the least conserved of the extracellular loops and can be cleaved while preserving function in reconstituted bR (141). In order to limit the changes in this first mutant, the rest of the helices were maintained in the native order, with two new connections formed between the G and B, and the F and A helices. This results in a swap of the A and G helices in primary sequence, as illustrated in the ribbon shown in Figure 2.2a. To ensure that the new connection between the F and A helix is long and flexible enough to permit folding, the structural model was examined and an SSG bridge was added to span the distance between the helices. The native N-terminal sequence was retained and assumed to be long enough to span the G and B helices (for the full sequence of the mutant, see Appendix A). The resulting swap mutant was designed with the intention of preserving the formation of the native contacts and maintaining the native packing of the retinal cofactor. Nevertheless, this is a dramatic mutation that creates a novel fold not found in nature. Exploration of this mutant quickly yielded promising experimental data, encouraging the exploration of additional novel topologies.

Additional mutants were designed as research progressed, taking into account successful strategies from previous designs to build up more complicated changes. The primary sequence of each mutant was constructed to ensure that all native helices were preserved, and that the non-
native loops would be sufficient to span the physical space between the ends of the helices (see Appendix A for full sequences). If the cut ends of the native loops were not deemed sufficient, GSG linkers were added. In total, four novel fold mutants were successfully expressed and characterized based on the helix swapping strategy. These mutants have the new N-termini in two different non-native positions, at the C and G helices. Additional mutants were desired to identify how many of the seven native helices were capable of being identified by the signal recognition particle and targeted to the membrane, but a slightly different strategy is required to design mutants that place the N-terminal on the intracellular side.

The odd number of helices in bR puts the N- and C-termini on opposite sides of the membrane. This topology makes it impossible to perform a true circular permutation using the native helices. To enable circular permutations, an 8th helix was added to the native 7-helix fold. The WALP21 sequence spontaneously forms an α-helix and inserts into the membrane (142), and this artificial helix was added to the native C-terminus of bR. The 8 helix structure was used in the generation of three additional circularly permuted mutants which place the N-termini on the prohibited intracellular side of the protein. These use B, D, and F as the first helix in the sequence. Twelve novel folds were designed, and seven mutants supported activity as demonstrate in sections 3 and 4. These mutants use six different N-terminal helices and comprises seven novel folds in addition to the native structure, for a total of eight total topologies (Figure 2.2). As a control for the addition of the WALP helix, this artificial sequence was added to the native C-terminus of bacteriorhodopsin to construct the AW mutant.
Figure 2.2: Set of Active Fold Mutants: The topology of the eight different folds. The solid lines are intracellular loops, and dashed are extracellular. The retinal cofactor is in magenta. a) Wild-type bR, b)GBCDEFA, c) CDEFGBA, d) GFABCDE, e) CDEFABG, f) FW, g) DW, h) BW
An additional set of point mutants was designed to determine if the conserved retinal binding position is required for function. The structure of bR was examined to identify residues within 10 Å of the lysine. With input from Rick Roy and Douglas Theobald, the double mutants A53K/K216A, A53K/K216I, and T89K/K216A were designed. Position 53 is in the center of helix B, and 89 is in the middle of helix C; both mutations were designed with the intention that they would easily access and bind to the retinal cofactor (Figure 2.3).

Figure 2.3: Model of the A53K/K216A Mutant: a) Position 61 is within 7 Å of the retinal cofactor. b) The double mutant moves the Schiff base linkage to helix B.

2.2. Initial Tests of Mutant bRs

2.2.1 Methods

Genes coding for the mutant folds were artificially synthesized and subcloned by GenScript (Piscataway, NJ) into pET-21c vectors, which places a poly-His tag at the C-terminus of the cloned sequence (except the WT and GBCDEFA, which were subcloned into pET-21b and also contain a poly His sequence). Each coding sequence, excepting the wild-type, was generated using GenScript’s OptimumGene codon optimization technique (143). The wild-type gene was
obtained from Takashi Kikukawa at Hokkaido University. Point mutants were generated using the QuickChange Lightning kit from Agilent (Santa Clara, CA) using primers synthesized by IDT (Coralville, IA).

These vectors were transformed into BL21(DE3) pLys cells. For expression, cells were grown in 2 XYT media to an OD$_{600}$ between 0.3-0.6 and induced by adding 1 mM IPTG and 10 μM all-trans retinal directly to the growth media (Sigma-Aldrich, St. Louis, MO) (74). Cells were harvested after 1 to 3 hours by centrifugation at 10000xg for 15 minutes. Pellets were examined at this point for the characteristic pink color that indicates successful expression (Figure 2.4) (74). The cell pellet was resuspended in buffer (50 mM Tris-Cl, pH 8.0, 5 mM MgCl$_2$) after centrifugation, then stored overnight at 4°C, or for longer periods at -80°C. Frozen pellets were thawed overnight at 4°C before purification.

100 units of Pierce Universal Nuclease per 10 g of wet cell pellet was added to the resuspended cells at 4°C and incubated for 5-10 minutes (Thermo Scientific, Rockford, IL). The suspension was sonicated, then the membrane fraction was isolated by centrifugation at 22,000xg for 30 minutes (Figure 2.5). The membrane pellet was resuspended in buffer, using 2 mL for each gram of bacteria harvested (300 mM NaCl, 50 mM MES pH 6.5). Dodecyl maltoside (DDM) was added to 1.0% (w/v). The sample was incubated with gentle shaking for an hour or more at room temperature. The insoluble fraction was pelleted at 22000xg for 45 minutes, and the detergent soluble supernatant was retained.

The resulting supernatant was purified using a Co$^{2+}$-based metal-chelate chromatography resin (HisPur Cobalt Resin, Thermo Scientific) pre-equilibrated with wash buffer (300 mM NaCl, 50 mM MES, pH 6.5). Supernatant was pumped over the column at a rate of 2 mL per
minute, followed by 10 column volumes of wash buffer. Bacteriorhodopsin was eluted with 300 mM NaCl, 300 mM imidazole, 50 mM Tris-HCl, pH 7.0 containing 0.1% DDM. The detergent solubilized protein was further purified using a Superdex 200 column with buffer containing 150 mM NaCl, 20 mM Tris-HCl, pH 7.5 and 0.1% DDM (GE Healthcare Biosciences, Pittsburg, PA). Samples were concentrated to ~10 mg/mL, flash frozen in liquid nitrogen, and stored at -80° C.

2.2.2 Results

All novel fold mutants discussed resulted in a pink color in the bacterial cell pellet (Figure 2.4). Additional mutants were attempted, but those with no color were not pursued further. This pink color simplified the development of the purification protocol, as well as allowing real-time monitoring of the success of each subsequent purification. After sonication and centrifugation, this color migrated to the membrane pellet rather than remaining in the soluble phase. (Figure 2.5). Solubilization with DDM allows this hydrophobic protein to partition into the soluble phase, indicating that bR is inserting into the E. coli membrane and can be extracted using detergent.
Figure 2.4. Colored Cell Pellets: a) The uninduced WT cell pellet is brown. b) Induced WT pellet is pink due to functional rhodopsin. c) The GBCDEFA induced cell pellet is also pink.

Figure 2.5. Membrane Fractionation: a) The pelleted membrane fraction following sonication. b) Insoluble fraction after DDM extraction. c) Detergent solubilized protein, showing the purple color of the protein.
Several alternative detergents at varying concentration were tested for the ability to solubilize bR, but DDM was found to be the most effective. 1% DDM w/v was the minimum amount needed to remove the majority of the protein from the insoluble fraction, as determined by the color of the second pellet and SDS-PAGE analysis. Neither decylmaltoside nor octylglucoside were capable of extracting all of the protein from the membrane fraction, even at high concentrations or with a large excess of buffer. Purification using the cobalt resin resulted in slightly higher purity than using nickel resin.

The ability of the mutants to insert into the *E. coli* membrane confirms that the non-native N-termini are capable of taking advantage of the SRP-dependent insertion mechanism of *E. coli* membrane proteins. Although the exact mechanism of insertion for these mutants has not been confirmed experimentally, the pink color in the pellet is sufficient evidence of membrane targeting, regardless of the pathway used. These results demonstrate that there is considerable flexibility in the choice of N-terminal sequence.

The double point mutants also resulted in colored pellets, although not the characteristic pink of the wild-type bR. Instead the pellets were peach colored, the first indication of a spectral shift. However, these mutants are still capable of folding and inserting into the membrane.
Chapter Three

Structural Characterization of Mutants
3.1 Material and Methods

3.1.1. Spectra and Acid Trapping

Purified protein was added to 200 mM NaPO₄ buffered at either pH 2 or pH 7, then incubated for 10 minutes at room temperature. The absorbance spectrum was measured from 2 uL samples on a NanoDrop 1000-C. Sodium dodecyl sulfate (SDS) was added to 0.5% w/v, and the solution was incubated for 30 minutes at room temperature to allow the denaturation to proceed fully. The absorbance spectrum between 210 and 800 nm was recorded and examined to identify the \( \lambda_{\text{max}} \) of each sample.

3.1.2. SDS Titration

The relative stability of mutants was measured using SDS titration. The protein concentration was calculated based on A280 and the extinction coefficient calculated using ExPASy’s ProtParam tool (http://web.expasy.org/protparam/)\(^{(144)}\). A final concentration of 2 mg/mL protein was used, with SDS concentrations varying from 0 to 1.0% w/v in 0.1% increments in the purification buffer (150 mM NaCl, 20 mM Tris-HCl, pH 7.5). The absorbance was measured after 12 and 36 hour incubations at room temperature.

The decrease in absorbance at 547 nm was graphed relative to the final percent of SDS in the sample. This data was fit in KaleidaGraph using equations derived from the equilibrium equation describing the denaturation of bacteriorhodopsin (the full derivation is shown in Appendix B). The equation can be derived assuming either a linear or a logarithmic relationship between the fraction of protein folded and the mole fraction of SDS \(^{(87)}\). Fitting the data with either the linear or logarithmic equation results in very similar parameter values and an almost
equally good fit. Equation 1 was chosen to fit the data, assuming a logarithmic relationship between SDS and fraction folded, and equal total amounts of protein and retinal. The $C_m$ is the detergent concentration at the midpoint of the unfolding transition, and the calculated value for the various mutants is compared to determine the relative stability. Instead of $\chi_{SDS}$ representing the mole fraction as in Cao et al. 2012 (87), here it is percent SDS.

$$\text{Abs}_{560} = (a\chi_{SDS} + b) \times (1 + (0.25)\left(\frac{\chi_{SDS}}{C_m}\right)^m - \sqrt{0.0625\left(\frac{\chi_{SDS}}{C_m}\right)^{-2m/RT}} - 0.5\left(\frac{\chi_{SDS}}{C_m}\right)^{-m/RT})$$  \hspace{1cm} (1)

Additional denaturation measurements were made using different detergents, to determine if alternative detergents might yield higher stability or activity. In addition to DDM at 2 mM, DM was tested at 3 mM, OG at 30 mM, and Zwittergent (ZW) 3-10 at 50 mM (DDM, DM and OG from Anatrace, Maumee, OH; ZW from Santa Cruz Biotechnology, Dallas, TX). The detergent concentrations chosen are all about 1.5x the CMC of the detergent. All were prepared in 150 mM NaCl, 20 mM Tris-HCl, pH 7.5 buffer. Purified protein was exchanged into these detergents using a PD-10 column (GE Healthcare, Piscataway, NJ). SDS aliquots and additional buffer were added to reach a protein concentration of 2 mg/ml, again varying the final concentration from 0 to 1.0% SDS w/v in 0.1% intervals.

### 3.1.3 Crystallization Attempts

Multiple crystallization trials were conducted. The first trial used the hanging-drop vapor diffusion method with wild-type samples in 150 mM NaCl, 20 mM Tris-HCl, pH 7.5 with 0.1% DDM. Drops were mixed consisting of 2 uL of protein and 2 uL of well solution using Crystal
Screen™, Crystal Screen 2™, and the Membfac™ screen from Hampton Crystallography (Aliso Viejo, CA). The crystal trays were incubated at room temperature. Some initial hits were identified; birefringent purple objects were formed under some conditions but did not have well-defined edges. Additional screening did not yield any protein crystals able to diffract.

Bicelle crystallization was also attempted (64). The bicelle mixture was prepared by dissolving 99.8 mg of CHAPSO in 600 uL of deionized water, then adding 300.2 mg of DMPC. The suspension was repeatedly vortexed, frozen in liquid nitrogen, heated at 55°C for 10 minutes, then cooled on ice for 10 minutes. This cycling results in a clear but very viscous bicelle mixture which is mixed 1:4 with purified protein, to a final concentration of 8% bicelle and 10 mg/mL protein. After a 30 minute incubation on ice, 1 uL of bicelle/protein mixture was mixed with 2.5 uL of well solution, again using the Crystal Screen™, Crystal Screen 2™, and the Membfac™ from Hampton Crystallography, as well as many conditions from the literature. While some initial hits were obtained, no further refinements yielded crystals capable of diffraction.

Finally, the lipid cubic phase method was attempted. Solid monoolein was melted in a 40°C water bath, then mixed with protein to a final ratio of 3:2 lipid:protein (145). This solution was dispensed into the Lipid Cubic Phase Screening Plate (Hampton Crystallography) using a repeating dispenser and a 25 uL syringe loaded with the cubic phase (Hamilton, Reno, NV). Each globule of protein was overlaid with 1 uL of crystallization solution, once again starting with Crystal Screen™, Crystal Screen 2™, and the Membfac™ as well as a selection of conditions from the literature. Plates were scored using a numerical system from 0-9 (145), and hits scoring above 5 (birefringent crystallites and spherulites) were optimized by screening
around those conditions. At the time of writing, several conditions remain to be optimized and may yet yield crystal structures of the wild-type and mutants.

### 3.2 Results

#### 3.2.1 Spectra and Acid Trapping

All fold mutants had a maximum absorbance within 2 nm of the wild-type (Figure 3.1). The absorbance spectra of type I rhodopsins are exquisitely sensitive to changes in the local retinal environment in the protein interior. The similarity of the mutant spectra to the wild-type confirms that even with dramatic structural rearrangements, the retinal is packing correctly in the hydrophobic pocket inside the protein. In the absence of a crystal structure, this is a good indication that the helices are forming contacts very similar to the wild-type instead of being pulled out of place by the artificial loops.

![Figure 3.1: Mutant Absorbance](image)

<table>
<thead>
<tr>
<th>Mutant</th>
<th>$\lambda_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>547</td>
</tr>
<tr>
<td>GBCDEFA</td>
<td>545</td>
</tr>
<tr>
<td>CDEFGBA</td>
<td>548</td>
</tr>
<tr>
<td>GFABCDE</td>
<td>546</td>
</tr>
<tr>
<td>CDEFABG</td>
<td>547</td>
</tr>
<tr>
<td>FW</td>
<td>549</td>
</tr>
<tr>
<td>DW</td>
<td>547</td>
</tr>
<tr>
<td>BW</td>
<td>547</td>
</tr>
</tbody>
</table>

Figure 3.1: Mutant Absorbance: The absorbance spectrum of each fold mutant overlays nicely with the wild-type. The table displays the absorbance maximum for each mutant.
The acid trap experiments confirm the formation of a covalent bond between the protein and the cofactor. When retinal is attached to a positively charged Schiff base, the absorbance shifts from the free retinal $\lambda_{\text{max}}$ at 380 nm to a new $\lambda_{\text{max}}$ at 440 nm. Under acidic conditions the Schiff base does not hydrolyse; when the protein unfolds the Schiff base remains intact, absorbing at 440 nm (73, 79). In the folded state at pH 7, the protein absorbs at 550 nm. When the protein is incubated at neutral pH then denatured using SDS, free retinal is released through hydrolysis and the absorbance at 380 nm is observed. At low pH there is very little free hydroxide available and the counterion Asp85 is protonated, preventing nucleophilic attack at the Schiff base. The rate of hydrolysis is significantly reduced, and absorption at 440 is derived from retinal maintaining the covalent bond to nitrogen (Figure 3.2). Denaturation under acidic conditions can therefore be used to confirm that the protein has formed a covalent bond to the retinal.

In all of the mutants discussed here, the acid trap experiment confirms the existence of the covalent bond. This was particularly significant for the point mutants in which the lysine position was changed. These three mutants all had shifted absorption spectra relative to the wild-type, which is attributed to perturbed packing interactions. The altered lysine position prevents the retinal cofactor from binding correctly in the hydrophobic pocket, pulling it out of place and shifting the absorbance. However, upon denaturation in acid each of the point mutants discussed displayed the characteristic Schiff base absorption at 440 nm, confirming that these new lysine positions were within reach of the cofactor and able to undergo the chemistry required to form the Schiff base.
Figure 3.2: Schiff Base Formation and Acid Trapped Spectra: Acid trapping experiments of the wild-type and double mutant proteins. a) The formation and hydrolysis of the Schiff base; hydrolysis is prevented at low pH. b) The absorption spectra of WT and A53K/K216A acid trap experiments.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>$\lambda_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>547</td>
</tr>
<tr>
<td>A53K/K216A</td>
<td>520</td>
</tr>
<tr>
<td>A53K/K216I</td>
<td>510</td>
</tr>
<tr>
<td>T89K/K216T</td>
<td>600</td>
</tr>
</tbody>
</table>

Table 3.1: Absorbance of Lysine Mutants
3.2.2 SDS Titration

Titration with SDS caused the gradual loss of absorbance at 550. The $C_m$ allows comparison of the relative stability of the mutants, as presented in Table 3.2. $C_m$ is calculated at the mid-point of the unfolding transition, and relies on this transition zone being well resolved. The GBCDEFA fit in Figure 3.3 shows the transition zone clearly; this mutant is more stable than the WT. The calculated $C_m$ for the CDEFABG mutant is not well supported; this experiment has not been replicated nor corroborated by other observations, and the value of the $C_m$ varies depending on the parameters used to fit the equation. The data does not show a clear unfolding transition as required for the curve fit, and this experiment will need to be repeated (Figure 3.3).

Figure 3.3: Unfolding of Mutants: a) The unfolding fit of GBCDEFA shows the unfolding transition while b) the unfolding fit of CDEFABG does not have a clear transition zone and is therefore difficult to fit.
No data was obtained for the DW or AW mutants; these proteins denature quite readily. Due to the low stability and the tendency to aggregate, the absorbance data is not reliable and the \( C_m \) was not calculated. In fact, the DW mutant is even observed to denature over time while frozen at -80\(^\circ\) C, as indicated by a loss of color in the frozen sample. Upon thawing, a white precipitate is observed, revealing that this mutant cannot be retrieved from storage and used for further experiments.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>( C_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBCDEFA</td>
<td>0.54</td>
</tr>
<tr>
<td>CDEFABG</td>
<td>0.50</td>
</tr>
<tr>
<td>WT</td>
<td>0.43</td>
</tr>
<tr>
<td>GFABCDE</td>
<td>0.41</td>
</tr>
<tr>
<td>BW</td>
<td>0.33</td>
</tr>
<tr>
<td>FW</td>
<td>0.08</td>
</tr>
<tr>
<td>CDEFGBA</td>
<td>ND</td>
</tr>
<tr>
<td>DW</td>
<td>ND</td>
</tr>
</tbody>
</table>

Table 3.2: \( C_m \) of Mutants: The \( C_m \) was calculated for each fold mutant in size-exclusion buffer with 0.1\% DDM.

The denaturation experiment was attempted on the three lysine swap point mutants described in Section 2. These mutants were all highly unstable in DDM, as shown by a total loss of native absorbance at the lowest SDS concentration used. This prevented fitting, but the very low stability was firmly established. Additionally, the T89K/K216I mutant precipitated even at 0\% SDS during the 12 hour room temperature incubation.
Testing of various detergents was done on the wild-type and both A53K point mutants. The wild-type is less stable in DDM than in DM, but DDM is the only detergent capable of extracting the protein from the cell membrane. OG caused precipitation of the protein during the overnight incubation. DM and ZW 3-10 both increased resistance to SDS denaturation under these conditions, with ZW 3-10 completely preventing denaturation up to 1% SDS. However, the total mass of the detergent was not constant across the different detergents tested, and these experiments should be repeated while accounting for possible mass effects. Additionally, higher concentrations of SDS were not tested, so the unfolding transition is not well-characterised and prevents calculation of the $C_m$.

All of the lysine point mutants were much less stable than the wild-type, regardless of detergent. A53K/K216A and A53K/K216I were both tested in DM and OG. ZW 3-10 was not used for either of these mutants; in both cases when the protein was exchanged over the PD10 column into buffer containing this detergent, it completely denatured in the time it took to elute. The color change from pink to yellow was clearly visible, and none of the native absorbance remained after elution. For these mutants, DM appeared to increase stability relative to DDM, while OG caused precipitation during the incubation. Unfortunately, not enough data was collected in the unfolding transition zone to allow a good fit to the denaturation equation (1), so these observations are based solely on the higher concentration of SDS needed to cause the loss of absorbance at the native $\lambda_{\text{max}}$. In spite of the improved stability in DM, it is clear that none of the three point mutants achieved stability even close to that of the wild-type.
3.2.3. Crystallography

Although many of the crystallization trials yielded encouraging initial hits, none produced diffraction. At the time of writing, exploration of conditions using the lipid cubic phase method is still underway and may yet result in a structure of the fold mutants.
Chapter Four

Activity of Mutant bR Folds
4.1 Materials & Methods

4.1.1. Preparation of Lipids from Halophilic Cells

Archael lipids were used in some proton pumping experiments as an alternative to soybean lipids for reconstitution. *Halobacterium salinarum* cultures were obtained from Judy Hertzfeld and grown in 5 mL of high salt media at 37°C for 5 to 7 days under constant illumination with shaking (58). Once these cultures had purple color, they were combined with 1 L of media and grown under ambient light for 3 to 4 days at 30°C with continuous shaking. The salty purple media was then spun down at 10000xg for 20 minutes, resulting in a loose purple pellet.

The cell pellet was suspended in 4 M NaCl to a final volume of 80 mL, then 300 mL of methanol-chloroform (2:1) was added (146). The slurry was stirred at room temperature for an hour, then centrifuged at 10000xg for 10 minutes. The pellet was resuspended in 190 mL methanol-chloroform-water (2:1:0.8), shaken, and centrifuged again immediately. The two supernatants were combined and poured into a separatory funnel, and 150 mL each of chloroform and water were added to induce phase separation. This solution was carefully mixed and allowed to separate overnight. The lower organic phase was withdrawn, then placed in a round-bottom flask which was attached to a rotary evaporator and concentrated to a final volume of about 1 mL. The resulting solution is bright orange due to retinal isolated along with the halophilic lipids due to its hydrophobicity. Over-concentration may cause lipids to precipitate, but the addition of chloroform-methanol (1:1) can resuspend the solids. This suspension was stored at -4°C for up to a week.
The total lipids were then separated into polar and nonpolar fractions (146). The total lipid fraction was evaporated in a glass culture tube under a nitrogen stream to 200-300 µL, then 5 mL of acetone was added. This mixture was vortexed and incubated for 1 hour on ice. The solution was centrifuged for 3 minutes at 1200xg in a swinging bucket rotor, then the supernatant was carefully decanted. The slurry was washed twice using 1 mL of ice cold acetone, followed by centrifugation for 1 minute and removal of the supernatant. At the end of the last centrifugation, the remaining white suspension consists of polar archaeal lipids. This lipid slurry was transferred to a weighed scintillation vial and dried under a stream of nitrogen.

Resuspension buffer was added to obtain 2% w/v lipid, with 14 mM OG added. This was sonicated to clarity and the lipids used for reconstitution and proton pumping experiments.

4.1.2 Reconstitution into Liposomes

Purified bacteriorhodopsin was reconstituted into liposomes as follows. Soybean lipids (Fisher Scientific, Pittsburg, PA) were reconstituted at 2% w/v in buffer containing 50 mM KCl, 100 mM KP$_i$, pH 7.0 with 14 mM octyl glucoside (OG; Anatrace, Maumee, OH), and sonicated to clarity, about 30 minutes. The soybean lipids were used for the majority of samples, but some experiments were also performed using alternative lipids. E. coli polar lipids were purchased as 10 mg/mL stocks in chloroform (Avanti Polar Lipids, Alabaster, AL), dried under nitrogen gas, dissolved in pentane, dried again, and resuspended to 40 mg/mL in reconstitution buffer.

POPE:POPG (1-Palmitoyl-2-Oleoyl-sn-Glycero-3-[Phosphoethanolamine]:1-Palmitoyl-2-Oleoyl-sn-Glycero-3-[Phospho-rac-(1-glycerol)]) at a ratio of 3:1 was also used for a set of
experiments; these lipids were stored in chloroform and dried under nitrogen gas before resuspension in buffer and sonicated to clarity (Avanti Polar Lipids).

Protein was added to lipid suspension at a range between 10 and 40 µg protein per mg lipid. The mixture was put into dialysis cassettes with a 10 kDa molecular weight cut off (Slide-A-Lyzer Dialysis Cassette, Pierce) and dialyzed against three changes of buffer. The first two incubations contained 50 mM KCl, 100 mM KP$_i$, pH 7.0 and proceeded for a minimum of 4 hours, and up to 12 hours. The final incubation contained 1.9 mM KCl, 100 mM KP$_i$, pH 7.0 and proceeded for a minimum of 8 hours to overnight. Alternative conditions using 1.9 mM KCl, 100 mM KP$_i$ at pH 6.5 or 7.5 were also examined.

The liposomes were removed from the cassettes and the final volume recorded. Samples were aliquotted between 250-350 µL and flash-frozen at -80°C. On the day of the experiment, the samples were removed from storage and subjected to three freeze-thaw cycles in liquid nitrogen. They were then passed through a 0.4 µm Whatman Nucleopore track-etched polycarbonate membrane 21 times (GE Healthcare Biosciences, Pittsburg, PA), using the liposofast device from Avestin (Ottowa, ON, Canada). Liposomes were then passed over a column containing Sephadex G50 beads suspended in 2M KCl (Fisher Scientific).

4.1.3 Proton Pumping Experiments

The total volume of liposomes collected following the Sephadex G-50 exchange was added to 2 M KCl to the final concentration of 20 µM protein in 2 mL. Experiments were performed with and without the addition of valinomycin dissolved in dimethyl sulfoxide at a final concentration of 2 µg/ml (Fisher Scientific). The pH was corrected to 7.0 using µL additions
of 10 mM HCl or NaOH as needed. The experiment was performed with stirring in a water-cooled chamber at 25°C in a dark room. A pH electrode was positioned so the bulb was completely submerged in the liposome solution.

The sample was illuminated under saturating conditions with a 300 W halogen lamp. The pH change was recorded as relative millivolts using an IonAlyzer analog pH meter with signal digitized with a DataQ (Akron, OH) digitizer. The illumination was continued until the proton pumping reached a steady state, then the lamp was turned off. The experiment was allowed to return to baseline conditions, and this cycle was repeated twice. At the end of the third illumination, 2 µg/ml of the proton ionophore carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP) was added to diffuse the gradient and confirm that the signal was due to proton movement (Fisher Scientific). The system was calibrated with duplicate additions of 1 µL and 2 µL 10 mM HCl.

4.1.4 Data Analysis

Data analysis was performed using Kaleidagraph. The pH curve was fit using the heterogeneous rate equation (2) derived by Douglas Theobald.

\[ p = \frac{k}{\hat{y}} \left( 1 - \frac{1}{\beta} \left( 1 + \frac{\hat{y}}{\beta} \right)^{-\beta} \right) \] (2)

In this equation, \( p \) is the proton differential across the liposome membrane, recorded in mV, \( k \) is the instantaneous rate of pumping, \( \hat{y} \) is the average rate of proton backflow, and \( \beta \) describes the distribution of liposome sizes. The initial rate \( k \) is in mV/sec, the calibration was used to calculate the number of protons per mV in each experiment. This was transformed using the calibration values and the known concentration of protein to calculate the initial rate of H⁺ sec⁻¹.
molecule\(^{-1}\); the rate of proton pumping was averaged over a minimum of two experiments and compared to that of the wild-type. All rates reported are in per second, analogous to the turnover rate of an enzyme catalyst.

4.2. Results

4.2.1 Varied Conditions Affect Proton Pumping Ability

The soybean liposomes used are somewhat permeable to protons, leading to the backflow rate incorporated in the curve fit. During an illumination cycle, the pumping of protons from the exterior to the interior of the liposomes outpaces the rate of leakage. Eventually the experiment reaches a steady-state, where the rate of leak becomes equal to the rate of pumping. At this point the light is turned off and the leakage allows the mV to gradually return to baseline (Figure 4.1).

![Wild-Type Proton Pumping](image)

Figure 4.1: Wild-Type Proton Pumping: A representative trace of WT activity in soybean lipids. The period of illumination is shown in white.
Other lipids were investigated to determine if they are be less permeable to protons. Proton pumping activity was detected using neither *E. coli* polar lipids nor the synthetic lipids POPE/POPG. However, activity was observed using archaeal lipids extracted from *H. salinarum*. This was the only alternative to soybean lipids that resulted in proton pumping activity, and the initial rate was slightly reduced compared to the soybean lipids; in soybean lipids the average rate was 0.05 s\(^{-1}\) and in halophilic lipids it was 0.04 s\(^{-1}\) at pH 7.0. Unfortunately, these lipids are just as permeable to protons as the soybean lipids.

<table>
<thead>
<tr>
<th>pH</th>
<th>Average Rate s(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5</td>
<td>0.16</td>
</tr>
<tr>
<td>7.0</td>
<td>0.05</td>
</tr>
<tr>
<td>7.5</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 4.1: pH Dependence of Proton Pumping:
The rate of proton pumping is highest at pH 6.5.

Proton pumping activity in the wild-type is highest using soybean lipids with 40 ug protein per mg lipid at pH 6.5, so this condition was used for all mutants. The inclusion of 1 mM KP\(_i\) buffer in the external solution was also attempted, but even this amount of buffer prevented detection of the small changes in pH due to the proton pumping activity.

The increased stability of the wild-type and point mutants in alternate detergents discussed in Section 3.2, especially DM, indicates that perhaps purification in DM would result in higher activity in the reconstituted liposomes. Additionally, the smaller size and higher CMC of DM relative to DDM allows the incubation time of each diaysis step to be dramatically decreased, hopefully preventing loss of activity due to denaturation of the protein. With the exception of the swap mutants, protein purified with DM used in all buffers subsequent to the
membrane extraction step has reduced or no activity. In the wild-type, DM reduced activity to less than half of that observed when using DDM, and similar reductions were observed in helix swapping mutants. DDM was therefore used for all experiments reported here unless otherwise noted.

During the illumination cycle, the import of protons by bacteriorhodopsin causes a buildup of positive charge inside the liposomes. This charge gradient can be diffused by the addition of valinomycin, which facilitates the movement of potassium ions down the charge gradient. Experiments using the wild-type protein demonstrate that proton pumping activity is possible in the absence of valinomycin, but is reduced to ~80% of that observed when valinomycin was added. The ionophore was therefore added to all subsequent experiments.

4.2.2. Activity of Helix Swapping Mutants

All of the helix swapping mutants that caused a color change upon expression had proton pumping activity, except the AW mutant (Figure 4.2). The rate of pumping varied, from only about 40% of the wild-type rate to almost double (Figure 4.3). These results represent at least two different preparations of each mutant. In some trials the AW mutant had very low proton pumping activity, but this result was not reproducible. This mutant also had very low stability compared to the wild-type, which likely contributed to the variable results.
Figure 4.2: Light Induced Activity of Fold Mutations: Representative traces of each mutant, showing the initial rise in pH at the start of the illumination.

![Overlay of All Mutants](image)

**Proton Pumping Activity**

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Rate H⁺ s⁻¹ bR⁻¹</th>
<th>Relative Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBCDEFA</td>
<td>0.132</td>
<td>176</td>
</tr>
<tr>
<td>GFABCDE</td>
<td>0.080</td>
<td>106</td>
</tr>
<tr>
<td>Wild Type</td>
<td>0.076</td>
<td>100</td>
</tr>
<tr>
<td>CDEFGBA</td>
<td>0.050</td>
<td>66</td>
</tr>
<tr>
<td>DW</td>
<td>0.049</td>
<td>64</td>
</tr>
<tr>
<td>CDEFABG</td>
<td>0.047</td>
<td>62</td>
</tr>
<tr>
<td>BW</td>
<td>0.03</td>
<td>40</td>
</tr>
<tr>
<td>FW</td>
<td>0.03</td>
<td>40</td>
</tr>
</tbody>
</table>

Figure 4.3: Calculated Relative Activity: The activity of each helix swap mutant charted relative to wild-type, with the rate reported in the accompanying table.
4.2.3. Activity of Lysine Repositioning Mutants

Of the swap mutants explored, only the A53K/K216A resulted in reproducible proton pumping activity, at about 11% of the wild-type. The results with A53K/K216I and T89K/K216A are ambiguous; a very low level of proton pumping activity was observed in some experiments but was difficult to reproduce. This activity may be below the detection threshold of the instrument in the conditions used. All of the point mutants are much less stable than the wild-type, and this stability is probably also a factor in the difficulty of observing proton pumping activity. The extended dialysis required at room temperature may contribute to the lack of activity.

Preparation in DM may be a better option for the lysine point mutants, since it both improves their stability in SDS denaturation trials, and requires much shorter dialysis times. Initial results detected a low level of activity when DM was used as the detergent during the IMAC elution step and subsequent steps for the A53K/K216A and the A53K/K216I mutants. In fact, using DM increased the rate of the A53K/K216A mutant to 22.5% of wild-type from 11% in DDM.
Chapter Five

Discussion
5.1 Type I and Type II Rhodopsins are Homologous

The potential homology between type I and type II rhodopsins has been debated for decades and has not yet been satisfactorily resolved. The argument against homology hypothesizes that the proteins share an identical fold due to convergence on a unique architecture required for function. The work I present here is an experimental test of the key prediction of convergence. In addition, it is the first successful non-circular remodeling of a transmembrane protein.

I have identified seven alternative folds that maintain proton pumping activity. Additionally, one point mutant, which changes the location of the strictly conserved lysine residue, is also functional. These eight alternative architectures contradict the main prediction of convergence. I propose, therefore, that type I and type II rhodopsins evolved from an ancient common ancestor; even though the cellular role and other features have been altered, the shared fold is preserved.

5.2 Structural Constraints on the 7TM Fold

5.2.1 Mutant bR Targets to the E. coli Membrane

All of the novel fold mutants described produced pink E. coli cells when the protein was expressed, indicating that the mutants are targeted to the membrane. The set of permutation mutants characterized use six out of the seven native helices in the N-terminal position (Figure 5.1). Two mutants using the E helix at the N-terminus were attempted, one swap mutant with the sequence EFABCDG and the other using the WALP helix to make an EW mutant. Neither of these caused a color change when expressed in E. coli, and these mutants were not explored.
beyond expression trials. These results are inconclusive about the function of these proteins, but are an interesting indication that the E helix may be unable to facilitate membrane targeting.

![Figure 5.1: Various N-terminal Sequences Target the Membrane](image)

Figure 5.1: Various N-terminal Sequences Target the Membrane: Six out of the seven possible N-termini are capable of targeting to the *E. coli* membrane, the terminal helix is shown on the left and the mutant sequence on the right.

Previous work has established that bacteriorhodopsin from *H. salinarum* binds to the signal recognition particle (SRP) when the protein is expressed in *E. coli* (51). It is apparent from the color change in pelleted *E. coli* that bR from *H. turkmenica* targets and folds correctly in the cell membrane (74). Although SRP binding has not been experimentally confirmed for *H. turkmenica* bR, it is likely that this bR also requires SRP recognition and interaction with the translocation machinery to facilitate membrane insertion. Since these mutants are capable of insertion in *E. coli*, it is likely that they target to the membrane of the native bacteria, just as *Hs* bR uses the same machinery in *E. coli* and the native archeon (51, 147). It is apparent from these results that the SRP mechanism is not an evolutionary bottleneck; the requirements for SRP...
binding are not so stringent as to prevent the adoption of alternative folds. A stretch of hydrophobic residues alone may be sufficient to act as a signal; interestingly, the C helix has only marginal hydrophobicity, but can still initiate membrane targeting.

![Diagram showing folding pathway of GBCDEFA](image)

**Figure 5.2 Folding Pathway of GBCDEFA:** a) Helix A cannot insert without the concurrent insertion of helix F unless b) it goes in backward. This would prevent formation of functional protein, so the current model of bR folding cannot apply to these mutants (52).

To achieve the observed pink color, the mutants must be folded with retinal bound in a competent tertiary structure embedded in the cell membrane. Interestingly, the folding and insertion pathways of some mutants contradict the current folding hypothesis (52, 93). This model predicts that helices A-E form a stable folding core, while F and G form a second unit which inserts after the first five helices are correctly packed inside the membrane. Two of the mutants, GBCDEFA and CDEFABG, cannot fold according to this mechanism because the F and G helices are separated in primary sequence. The retention of these two helices outside the membrane would by necessity prevent the formation of the proposed core, as illustrated in Figure 5.2. The EFA sequence contained in both mutants requires that helix F inserts prior to or
simultaneously with the A helix, effectively preventing the formation of the A-E core. The proposed folding mechanism may be specific to the native helix connectivity, or may require revision. In either case, it is apparent from these results that neither the insertion nor the folding pathway presents a major barrier to the adoption of alternative folds during evolution.

5.2.2 Mutant bR Forms Helical Contacts and Binds to Retinal

When designing the mutants, the goal was to allow the helical contacts necessary for function despite dramatic rearrangements in primary structure. The ability of mutants to fold and bind retinal in the active conformation can be inferred from the absorbance spectrum; even small perturbations in the packing of retinal will shift the spectrum (148). The helix permutation mutants examined here have absorption maxima within 2 nm of the wild-type \( \lambda_{\text{max}} \). Although no crystal structures were obtained, the absorbance data is good evidence that the \( \alpha \)-helices in the reconfigured bR mutants make contacts similar to the wild-type and are binding retinal in the hydrophobic pocket.

The loops used to span the helices were designed to allow just enough flexibility for the helices to interact correctly. The first mutants explored had relatively few alterations; if the artificial loops facilitated correct retinal packing as indicated by formation of a colored pigment, these sequences were used in subsequent mutants. This allowed more complex mutants to build on successful precursors; for example the GB linkage originally designed for the GBCDEFA was also used for CDEFGBA.

Adding an artificial helix did not perturb the structure enough to change the absorbance; all of the mutants incorporating the WALP helix also had absorption maxima within 2 nm of the
wild-type. The WALP helix is positioned near the lysine required for retinal binding by virtue of its attachment to the G and A helices, but must not be interacting with either helix in such a manner as to affect the binding of retinal.

In contrast, the three lysine point mutants dramatically altered the absorbance spectrum. The change in $\lambda_{\text{max}}$ was anticipated, based on the distance between retinal and the new lysine; to form the covalent bond, retinal must be pulled into reach of the $\varepsilon$-amino group. The native lysine is positioned within 5 Å of carbon 15 in retinal, while A53 and T89 are further away at 7 Å. The spectral shifts observed in these mutants correspond to the protein conformation remodeling to accommodate the formation of the Schiff base, warping the hydrophobic pocket or local electrostatic environment and altering the binding of retinal.

Retinal can insert into the core of bacteriorhodopsin even in the absence of a covalent bond, but additional experiments are required to test for the formation of the bond. Acid trapping can detect this linkage. At low pH, the Schiff base cannot hydrolyze; the absence of nucleophilic species prevents attack on the electrophilic nitrogen. Retinal bound to a positively charged Schiff base absorbs at 440 nm (73), and this species is observed in bacteriorhodopsin unfolded using SDS after a brief incubation at low pH. The absorbance at 440 nm confirms that the protein has formed a covalent bond to retinal. The control experiment, where the protein is denatured at neutral pH, releases retinal by hydrolysis to yield the native absorbance of free retinal at 380 nm.

The positively charged Schiff base is observed in all mutants explored when they are acid trapped. A53K/K216A, A53K/K216I, and T89K/K216T all absorb at 440 nm when denatured at low pH, which confirms formation of the covalent bond in these spectrally shifted mutants. The acid trap experiment in combination with the spectral data is strong evidence that all of the
mutants are folding so that native helical contacts are formed, binding to retinal, and packing similarly to the wild-type.

A crystal structure would be beneficial to confirm the folding of these mutants. Although *H. salinarum* bR has been crystallized using many different conditions, bR from *H. turkmenica* proved to be less tractable. Several initial hits, consisting of small birefringent objects, were observed using the bicelle method, but did not yield any crystals that diffracted. The lipid cubic phase method has not yet been exhaustively explored and warrants further investigation.

### 5.2.3 Mutant Folds are Functional

The majority of bR mutants explored in this thesis, including seven novel folds, were able to pump protons in reconstituted lipid systems. The AW mutant, which had the added complication of being very unstable, did not demonstrate reproducible pumping activity. The other seven fold mutants, including the three other WALP mutants, had a significant level of activity which was reproduced across multiple purifications. Of the three lysine point mutants, only A53K/K216A has reproducible activity.

In the absence of conclusive results demonstrating the orientation of the mutants in liposomes, the observe rate of proton pumping constitutes a lower limit of activity. The canonical bacteriorhodopsin from *H. salinarum* orients 100% inside-out in reconstituted systems (66), and preliminary work indicates that the same is true for bR from *H. turkmenica*. Additional work is needed to confirm that the mutant folds possess the same ability to orient 100% in liposomes, but the rate of proton pumping can be considered a baseline level of activity.
The WALP helix mutants are especially interesting. The ability of the additional helix to integrate into functional folds adds an additional set of eight helix topologies to the 144 possible using the seven native helices. The AW mutant did not have reproducible proton pumping activity and was among the least stable of the mutants explored. Although the other three WALP mutants were less stable than the wild-type, they still maintain activity. It is likely that having two points of connection to the rest of the protein helps to stabilize the WALP helix, holding it in close proximity to the A and G helices. In contrast, the WALP helix in the AW mutant has only one point of connection, and in the absence of either polar or hydrophobic interactions to hold it in place near the core of the protein, this helix is free to explore alternative conformations. It is possible that this additional freedom of movement decreases the stability of this mutant.

In contrast to the helix mutants, where colored cell pellets indicated the presence of functional protein, two of the three lysine point mutants did not have observable proton pumping activity even though a color change occurred upon expression. These mutants were also very unstable, and it is unclear whether they lack activity, or if they are degrading so quickly that activity is undetectable after reconstitution. However, the A53K/K216A activity was reproducible, which is an encouraging sign that alternative lysine positions can form functional proteins. Compensatory mutations could help to stabilize the structure and improve pumping activity. This possibility could be explored further using directed evolution and is a candidate for future experiments.
5.3 Conclusions

5.3.1 Functional Landscapes

This work presented here tests for the presence of a functional constraint on the 7TM fold; convergence predicts that only one or a very small set of folds will support the native function. Additional folds that are capable of activity indicate that multiple independent lineages are not likely to converge on only one fold. Evolution has access to at least eight alternatives to the native topology, and based on this evidence a divergent relationship between type I and type II rhodopsins is a more parsimonious explanation. Although some mutants have reduced stability and activity, the ability of these permuted non-native folds to function is remarkable. This is the first empirical test of the predictions of convergent evolution, as well as the first successful dramatic structural rearrangement of an entire integral membrane protein.

In addition to the seven novel folds described, the conserved lysine position is not functionally necessary. This lysine, midway through the seventh helix, is an extraordinarily conserved feature of all known bR homologs (except a few fungal homologs of unknown function) and is one of the critical similarities between type I and type II rhodopsins. However, this location is not required in either family; other positions can form functional protein. A recent study on bovine rhodopsin relocated this critical lysine to three positions on different secondary structural elements of the protein, and preserved the ability to activate its G-protein transducin in a light-dependent manner (149). This unexpected result, along with the work presented here, suggests that rhodopsins are less structurally constrained than convergence predicts.

The reduced activity of even the least active mutants is nevertheless significant; in the entire history of life on earth only two systems have evolved the ability to convert solar energy
into chemical energy (150). Bacteriorhodopsin is finely tuned to couple proton transport to a series of conformational changes. This reaction has many moving parts and critically, has many more ways to fail than to succeed (151). The chromophore must absorb at the correct wavelength and transform the energy into physical changes, without decomposing and without dissipating as heat. The proton must be picked up and transferred by the correct residues to avoid being transported in the wrong direction. Bacteriorhodopsin must prevent the proton from flowing back down the gradient back through the channel it just traversed. Any detectable proton pumping activity is a remarkable accomplishment, and an alternative fold could easily be further mutated over time to fulfill the needs of a particular organism.

Most importantly, the proton pumping rate observed in the slowed mutants is within the range of naturally occurring bacteriorhodopsin. The green proteorhodopsins, found in aquatic and marine microbes, have a much slower photocycle than *H. salinarum* bR (152-155). However, this reduced rate of pumping is sufficient for the energy needs of the microbes. Interestingly, the absorbance maximum of the very slow A53K/K216A mutant is shifted into the same range where these proteorhodopsins absorb (154). The absorbance maximum is readily shifted by other mutations in the binding pocket; if the absorbance spectrum of a particular mutant is not optimal for the organism it could be tuned over time.

There are other biochemical constraints on bacteriorhodopsin that were not addressed in this work, including thermodynamic stability, folding kinetics, and protein degradation. However, membrane targeting, insertion, folding, retinal binding, and proton pumping are all achieved, making it clear these alternative topologies could be accessible to further selection,
rather than being eliminated from the population. Even a low rate of proton pumping, if sufficient to keep the organism alive, would allow the novel fold to be optimized.

Given the functional alternative folds, I propose that the 7TM topology observed in both type I and type II rhodopsins, as well as the lysine position, are artifacts preserved from an ancient common ancestor. Proteins already serving as a critical energy source in an organism have a large barrier to overcome to change folds, while proteins with independent origins could easily select an alternative topology. The native bacteriorhodopsin structure is apparently confined to a local fitness maximum. Another local maximum could be obtained with the lysine positioned in helix B. The intermediate steps with two lysines or no lysine are highly unstable, based on preliminary experiments performed by Adam Drake (personal communication, 2013), so it is highly unlikely that the double mutant would spontaneously occur, effectively confining type I rhodopsins to a structure with the lysine in the G helix.

5.3.2 Proposed Phylogeny of Rhodopsins

Given that rhodopsin function can be achieved using non-native folds, the native 7TM architecture appears to be phylogenetically constrained rather than physically constrained (156). It is much more difficult for an existing protein to change its structure than for a novel fold to arise. This shared topology is most likely inherited from an ancient common ancestor, and has been preserved even as accumulated mutations have obliterated any other indication of homology.

The phylogenies of type I and type II rhodopsins have been extensively studied (100, 102, 110, 157). Type I homologues are hypothesized to have arisen from a basal rhodopsin
possessing transport function, and have since diversified and spread throughout prokaryotes, often by lateral gene transfer (7). The protein phylogeny is dramatically different from the species tree, highlighting the interdomain transfers and the ability of this protein fold to rapidly change to fill the needs of the organism (7).

The evolution of type II rhodopsins is somewhat more complex. GPCRs served various receptor functions, and all share the rhodopsin topology. The current accepted phylogeny places cyclic AMP receptors at the base of the tree. Class A GPCRs, including rhodopsins, emerged directly from the cAMP family relatively late during the diversification of GPCR function (110). We propose that type I rhodopsins emerged directly from the type II rhodopsins in opisthokonts roughly 1-2 billion years ago; this phylogeny is illustrated in Figure 5.3 (158). The ancestral type I rhodopsin, freed from the G-protein interaction, underwent rapid sequence change and selection for proton-pumping activity. This hypothesis is consistent with the inferred rate of sequence evolution in type II rhodopsins and predicts the observed loss of identifiable sequence similarity.
Additional evidence in favor of this ancient common ancestor is the recently discovered phylogenetic overlap of these two rhodopsin families in fungi. Previously, the type II rhodopsins had only been identified in eumetazoa, but are now known to be present in microbial eukaryotes, specifically fungi (110, 111). Type I rhodopsins were thought to be exclusively prokaryotic but have recently been detected in Viridiplantae (algae), Alveolata, and Fungi. The two families
share a phylogenetic overlap in fungi, and more overlaps may yet be identified in other basal organisms.

5.4 Summary of Thesis

My experiments provide significant evidence that type I and type II rhodopsins are homologs that diverged from an ancient common ancestor so long ago that all sequence similarities have been wiped out by accumulated mutations. Convergent evolution predicts that only one fold will be functional, but I have shown that seven alternative topologies and one alternative lysine position all recapitulate the wild-type light-activated proton pumping function. The convergent hypothesis is contradicted by the work presented here, and therefore I propose that type I rhodopsins are descended from type II rhodopsins.
Appendix One

Mutant Bacteriorhodopsin Sequences
Table One: Protein Segments: The annotated sequences of primary structure segments used to create mutants. Alpha helical portions are highlighted and the uncolored loop sequences were associated with the indicated helices.

<table>
<thead>
<tr>
<th>Helix</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix A</td>
<td>AALAPPMAATVGPESIWLWIGTIGMTLGTLYFVGRGRGVRDR</td>
</tr>
<tr>
<td>Helix B</td>
<td>KMQEFYIITITITTIAAAAMYFAMATGFGVTEVM</td>
</tr>
<tr>
<td>Helix C</td>
<td>VGDEALTIY WARYADWLFITTPLLLDLSLAG</td>
</tr>
<tr>
<td>Helix D</td>
<td>ANRNTIATLIGLDFMIGTGAIAALSST</td>
</tr>
<tr>
<td>Helix E</td>
<td>PGTRIAWWAISTGALLALLYVVGTLSENARNR</td>
</tr>
<tr>
<td>Helix F</td>
<td>APEVASLFGRRLNALWFLYPVVWILTGTEGF</td>
</tr>
<tr>
<td>Helix G</td>
<td>GILPLYWETAAFMVLDLSAKVGFGVILLOSRSVLERVATPTAAPT</td>
</tr>
<tr>
<td>WALP23</td>
<td>GSGWWLALALALALALALALALALALALWWASG</td>
</tr>
</tbody>
</table>
Table Two: Mutant Sequences: The alpha helical portions are highlighted as in Table One. Glycine and serine residues added to span extramembrane loops are bolded.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild-Type bacterio-rhodopsin</strong></td>
<td></td>
</tr>
<tr>
<td>GCCDEFGA</td>
<td></td>
</tr>
<tr>
<td>GBCDEFA</td>
<td></td>
</tr>
<tr>
<td>CDEFGBA</td>
<td></td>
</tr>
<tr>
<td>GFABCDE</td>
<td></td>
</tr>
<tr>
<td>CDEFABG</td>
<td></td>
</tr>
<tr>
<td>FGWABCD</td>
<td></td>
</tr>
<tr>
<td>DEFGWABC</td>
<td></td>
</tr>
<tr>
<td>BCFEDGWA</td>
<td></td>
</tr>
</tbody>
</table>
Appendix Two

Derivation of Denaturation Equations
The equilibrium of bacteriorhodopsin unfolding is described in one of two way. The experiments described in section 3 have no added retinal, and the equilibrium equation in this case is:

\[ \text{bR}_f \Leftrightarrow \text{bR}_u \Leftrightarrow \text{bO}_u \] (1)

The unfolding curve is described by:

\[ \text{Abs}_{547} = \text{Abs}_{547}^0 \times F_f \] (2)

Where \( \text{Abs}_{547}^0 \) is the absorbance of the subdenaturant line and its extension over the experimental \( \chi_{SDS} \) range, which represents the theoretical absorbance if all the bR is folded, as in Cao et al. 2012 (84). In this case, we assume the absorbance is logarithmically dependent on \( \chi_{SDS} \):

\[ \text{Abs}_{547} = a \times \chi_{SDS} + b \] (3)

and \( F_f \) is the fraction of the folded state in each mutant. That is, \( F_f = [\text{bR}_f]/c \), where \( c \) is the total concentration of protein. Combining equations (2) and (3) yields:

\[ \text{Abs}_{547} = (a \times \chi_{SDS} + b) \times F_f \] (4)

The unfolding free energy is defined in terms of \( K_{eq} \), and using the linear assumption in the absence of added retinal the equation for \( \Delta G \) is:

\[ \Delta G = -RT \times \ln((c/F_f)(F_f)^2) \] (5)

At the midpoint of the unfolding transition, \( F_f \) is equal to 0.5, with gives the equation for \( \Delta G_u \):

\[ \Delta G_u = m \times \ln(\chi_{SDS}/C_m) - RT \times \ln(c/2) \] (6)

\( F_f \) can be written as a function of \( \Delta G_u \) and \( c \), and the results is solved using equation (6) to give the fraction folded in terms of \( \chi_{SDS} \) and \( C_m \):
\[
F_r = 1 + \left( \frac{1}{4} \right) \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{m/RT} - \sqrt{\left( \frac{1}{16} \right) \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{-2m/RT} + \left( \frac{1}{2} \right) \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{-m/RT}}
\] (7)

Plugging equation (7) into equation (4) gives the final equation (8) that can be fit using KaleidaGraph, as discussed in Section 3 of the main text.

Abs_{560} = (a \chi_{\text{SDS}} + b) \left( 1 + 0.25 \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{-m/RT} - \sqrt{0.0625 \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{-2m/RT} + 0.5 \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{-m/RT}} \right)
+ 0.5 \left( \frac{\chi_{\text{SDS}}}{C_m} \right)^{-m/RT})
\] (8)

Where \( \chi_{\text{SDS}} \) is the independent variable a, b, m, and C\( m \) are independent, and RT = 0.592 kcal mol\(^{-1} \).
Bibliography


discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria, *ISME J* 3, 726-737.


