Directed evolution of bacteriorhodopsin for applications in bioelectronics

Nicole L. Wagner1, Jordan A. Greco2, Matthew J. Ranaghan1 and Robert R. Birge1,2

1Department of Molecular & Cell Biology, University of Connecticut, Storrs, CT 06269-3125, USA
2Department of Chemistry, University of Connecticut, Storrs, CT 06269-3060, USA

In nature, biological systems gradually evolve through complex, algorithmic processes involving mutation and differential selection. Evolution has optimized biological macromolecules for a variety of functions to provide a comparative advantage. However, nature does not optimize molecules for use in human-made devices, as it would gain no survival advantage in such cooperation. Recent advancements in genetic engineering, most notably directed evolution, have allowed for the stepwise manipulation of the properties of living organisms, promoting the expansion of protein-based devices in nanotechnology. In this review, we highlight the use of directed evolution to optimize photoactive proteins, with an emphasis on bacteriorhodopsin (BR), for device applications. BR, a highly stable light-activated proton pump, has shown great promise in three-dimensional optical memories, real-time holographic processors and artificial retinas.

1. Introduction

Scientists and engineers are continuously seeking new ways to harness the resources of the natural world to improve and advance the fields of nanotechnology and nanobiotechnology. The desire of humans to seek help from nature follows naturally from a recognition that nature has been operating at the nanoscale for billions of years. By using multi-disciplinary approaches, researchers are starting to gain control over the intrinsic properties of naturally occurring materials, significantly changing the way that scientists approach and solve complex problems [1]. The integration and unique interplay between protein engineering and biology can transform a number of scientific disciplines, as well as biologically inspired technologies.

Developments in nanoscience are characterized by the manipulation of biological and non-biological matter at the nanoscale regime. Starting with the early proposals of Richard Feynman in 1979 to the more recent innovations in materials science, the goals of nanoscience remain the same: to develop useful architectures with at least one dimension less than 100 nm [2]. Nanoelectronics, originally at the forefront of nanotechnology, centred on electrical, mechanical and materials engineering, and began with the continual minimization of the transistor feature sizes on integrated circuits, a trend described by Moore’s Law [3]. Today, the field of nanoelectronics encompasses the processing, transmission, storage and retrieval of information at the molecular level. As a consequence of the increasing demand for miniaturization, scientists are forced to search for alternatives to the more conventional and fundamental methods of research to meet the demands of modern technology [4,5].

Nanotechnological paradigms can be divided into two contrasting efforts: a top-down approach and a bottom-up approach [6–8]. Although commonly applied to information processing and knowledge ordering in software and computer development, these approaches can also be used in any scientific field. In a top-down approach, nanoscale objects are made through the processing of larger objects, such as silicon crystals or graphite [6]. The bottom-up approach focuses on the measurement, prediction and construction of nanostructures from smaller building blocks, such as atoms and molecules.
A classic and sophisticated analogy of nanotechnology can be found in nature, where biological systems have carried out the direct assembly of the structures necessary to sustain life [9, 10]. Biologists and engineers are working to imitate this natural capability to produce small clusters of atoms that can then self-assemble into more complex structures. Using a bottom-up approach, scientists can generate smaller structures than is possible via a top-down approach and can do so with greater efficiency and less material waste [10–12].

As the emphasis of nanotechnology switches from mechanical and materials engineering to a more biologically influenced field, researchers have started to redefine the definition of nanotechnology to stress the development of novel chemicals, as well as new materials and devices to study biological systems. Using a scale set up by nature, one made up of cells, organelles, proteins and molecules, researchers have started to broaden the scope of traditional nanotechnology to include the natural sciences and biology, a term loosely defined as nanobiotechnology [13, 14]. Nanobiotechnology refers to the application of nanotechnology to target problems directly relevant to biology [15]. The closely related term, bionanotechnology, is often used interchangeably with nanobiotechnology; however, when distinction is needed, bionanotechnology is used to refer to the general overlap of biology and nanotechnology, and exploits natural biomimetic systems, such as DNA nanotechnology and self-assembly systems. Bionanotechnology also encompasses the study of how nanotechnology can be directed through a better understanding of biology and the integration of biological ideas and motifs into improving nanotechnologies [16]. Conversely, nanobiotechnology highlights the advances in and applications of nanotechnology to biological problems [15]. Some of the target applications of these technologies include drug delivery [17–19], gene therapy [20] and regenerative medicine [21, 22].

This review explores some of the strategies for tailoring the electronic and photochemical properties of proteins for performance in nanotechnological devices. In particular, we will focus on directed evolution, as a method to genetically engineer microbial rhodopsins and photovoltaic proteins for application in non-native environments. Bacteriorhodopsin (BR), a photoactive protein found in the salt marsh archaeon, _Halobacterium salinarum_, is a leading candidate for application in devices, including three-dimensional optical memories, real-time holographic media, photovoltaic cells and artificial retinas [23–28].

2. Directed evolution

Directed evolution has emerged as a powerful tool for optimizing biological macromolecules to function outside of the native organism. By using directed evolution, researchers have been able to create novel proteins and biomolecules with properties not normally found in nature [29–32]. Combined with high-throughput screening methods, directed evolution has been successful in generating proteins to function as protein-based tools and devices, including biosensors, therapeutics and biocatalysts [33–38].

Evolution yields proteins that have been optimized to function in a particular biological system. Unfortunately, nature finds no comparative advantage in optimizing proteins for devices, and thus applied technologies using proteins and other biological molecules have only found limited success. The most that one can hope for is that nature has optimized a protein for a biological function that corresponds in some fashion to the intended application of the researcher. In nature, the slow accumulation of genetic mutations and adaptations is manifested in changes in the structure and function of biological macromolecules. The accumulation of genetic mutations can be positive, negative or have no effect on the organism. Beneficial mutations often enable the host organism to overcome a selective environmental pressure, while lethal mutations lead to cell death. Most protein structures are so complex that predicting how to alter the structure and function of the protein for a specific function is impossible [39].

Directed evolution (otherwise referred to as laboratory evolution, molecular evolution, evolution _in vitro_ or evolution in a test tube) can be used to optimize the inherent properties of biological macromolecules via iterative rounds of diversification and differential selection [40]. Classically, investigators have used directed evolution to enhance and alter specific traits, such as thermal and chemical stability, which are not apparent in the parent molecule [41]. Directed evolution employs both chemical biology and combinatorial chemistry via multiple rounds of mutagenesis to improve or alter the efficiency and specificity of proteins, with the ultimate goal of generating new molecules with useful properties for applied technologies.

In nature, environmental pressures, such as temperature and availability of sunlight, often affect the overall fitness of the organism. Applying selective environmental pressures in the laboratory can alter the behaviour or fitness of an organism, thereby leading to atypical cell phenotypes or proteins, which thrive outside of the context of biology. However, in the research laboratory this type of directed evolution often requires the development of new technologies, which have the potential to improve the design process, such as novel screening systems and computer programs [33]. With rapid advancements in synthetic and computational tools, the functional space of protein engineering is only limited by the ingenuity of the protein scientist. Innovations in nanobiotechnology are providing new opportunities for the directed evolution of proteins.

Three main steps characterize directed evolution: diversification, selection and amplification. The first step involves the diversification of a mutant library and is accomplished using mutagenesis to create a large library of gene variants. High-throughput screening and various other selection methods are then used to test for the presence of mutants with a particular property, trait or characteristic. Finally, amplification of the mutant library is used to allow for further characterization of the mutant DNA. Using a combinatorial library, a diverse pool of mutants are generated randomly in a test tube, and then screened for a particular property or phenotype through expansive screening and selection methods [38].

With the emergence of biotechnology and nanotechnology, investigators are looking to optimize photoactive proteins to function in optoelectronic, photovoltaic and biomedical devices. This review will explore some of the advancements in genetic engineering and describe some of the advantages of combining biology and nanotechnology. More specifically, we will describe the application of directed evolution and other genetic engineering strategies on the optimization of photoactive proteins, using BR as the model system.
3. Bacteriorhodopsin

BR is one of the most studied proteins for use in bioelectronic and biomimetic devices [42,43]. Located in the outer membrane of the salt marsh archaeon, *H. salinarum*, this 27-kDa protein functions as a light-transducing proton pump. When environmental oxygen becomes scarce and drops below a level necessary to sustain oxidative phosphorylation, *H. salinarum* expresses BR to generate energy via photosynthesis [44]. The protein, comprised of seven transmembrane alpha helices, is arranged in a two-dimensional hexagonal lattice of trimers in the lipid bilayer, often referred to as the purple membrane [45–47]. The inherent stability of the semicrystalline matrix allows the protein to survive and remain functionally active in the most extreme environments [48–51].

The primary light-absorbing moiety of BR is an all-trans retinal chromophore that is bound to the protein via a protonated Schiff base linkage to lysine-216 [43]. When the retinal chromophore absorbs a photon, a photocycle is initiated (figure 1). Through a complex process involving a series of spectrally discrete intermediates, BR translocates a proton from the intracellular to extracellular milieu. This photocycle is used to generate a proton gradient that is used to drive membrane-bound ATPase to synthesize adenosine triphosphate (ATP) under anaerobic conditions [55–57].

After decades of research, the photocyclic mechanism of proton pumping of BR is well characterized and several models have been generated to help describe the complex nature of the BR photocycle [58–63].

The main photocycle of BR is made up of a series of transient photochemical and conformational intermediates labelled the bR, K, L, M, N and O photostates (figure 1) [64]. Each BR photocycle lasts approximately 10–15 ms and results in the net translocation of a single proton across the
membrane [65]. The primary photochemical event, which initiates the photocycle, operates with a quantum efficiency of 0.65 [66,67]. This high quantum efficiency is identical to that of the visual photopigment, rhodopsin, which is found in the photoreceptor cells of the retina [68,69].

One of the unique features of the BR photocycle is the branched photocycle [25,70–72]. During the classical photocycle of BR, the last and most red-shifted intermediate, designated as the 0 state, contains an all-trans configuration of the retinal chromophore (figure 1). Unless the 0 state is illuminated with a second red photon, it will thermally decay back to the resting state (BR). However, if the retinal chromophore absorbs a photon of red light during the 0 state, it will photomeromerize to form a 9-cis photoproduct that is unstable in the binding pocket [70]. This 9-cis chromophore is characteristic of the branched photocycle and is comprised of the P and Q states (figure 1). The P state contains a bound 9-cis chromophore, and was originally defined as having an absorption maximum (λ_{max}) of approximately 490 nm [73]; however, further research shows that the P state is actually a mixture of two photostates, described as P₁(λ_{max} = 525 nm) and P₂ (λ_{max} = 445 nm). These photostates undergo a dynamic equilibration after the absorption of a photon [54], and because 9-cis retinal is unstable in the binding pocket, the chromophore thermally decays to form the Q state (λ_{max} = 380 nm) [54,70,74,75]. The Q photoproduct contains a hydrolysed 9-cis retinal chromophore trapped inside the BR binding pocket (figure 1) [70]. With an activation barrier of approximately 190 kJ mol⁻¹ for the spontaneous formation of the Q state, the Q state is the only thermally stable photoproduct of the protein with a lifetime of approximately 7–12 years at ambient temperature [25]. Reversion of the Q state to BR is possible via absorption of a blue photon (360–400 nm) by the 9-cis chromophore [75].

The biological origin of the branched photocycle remains uncertain, although it is hypothesized that this photochemistry is an unwanted photochemical artefact minimized by years of evolution [54]. Additionally, the observation that a significant fraction of random mutants have a lower efficiency of formation suggests that nature has adjusted the efficiency of the formation of the branched photocycle rather than eliminate the formation of this state [41]. Possible explanations for this branching reaction are that the P and Q states serve as a way to minimize DNA photodamage [54]. Additionally, it is possible that at high light intensity, the protein converts part of the photoactive protein to the Q state to avoid pH damage owing to the generation of too large of a pH gradient [76]. Recall that formation of the branched photocycle and the Q photoproduct prevents the protein from carrying out the biological function of proton pumping. Although the origin of the Q state remains unclear, the presence of a branched photocycle provides an excellent template for biomimetic and biophotonic device applications [77].

The absorption of a photon by the retinal chromophore of BR is one of the most efficient and stable photochemical reactions found in nature [78,79]. The photochemical stability of BR, commonly referred to as the cyclicity, is a measure of the number of times that the protein can be converted between two states before 37 per cent (1/e) of the material denatures. At ambient temperatures, the cyclicity of BR has been found to exceed 10⁶ [25]. The thermal stability of BR, which is commonly cited to exceed 80°C [49–51], has also made BR a remarkable candidate for device applications.

With excellent photochemical efficiency and stability, coupled with the ability to withstand high levels of light flux and chemical stress from a self-induced pH gradient, BR has been a leading candidate for application in numerous biophotonic, biomimetic and biomedical applications.

4. Biophotonic device applications of bacteriorhodopsin

Shortly following the discovery of BR by Oesterhelt and Stoeckenius in 1971 [55], Soviet scientists first recognized the potential of using this photoactive material in protein-based computing and processing applications. At the time, interest in pursuing bioelectronic devices was tied to military development and international competition towards designing new nanobiotechnologies, which had first elicited a holographic processor based on BR thin films through the project, Biochrome [80,81]. In the subsequent decades of research, BR has emerged as one of the most rigorously investigated biomaterial for use in electronic applications, culminating in the development of numerous prototypes of devices that include Fourier-transform holographic associative processors [25,42,82–84], three-dimensional optical memories [25,82,85,86], biosensors [87,88], photovoltaic cells [89–91] and protein-based retinal prostheses [28,92]. Recent advances in genetic engineering and directed evolution have further invigorated the pursuit of optimizing these protein-based devices, thereby increasing the potential of introducing a commercial BR-based technology in the near future [38,41,76]. Here, we briefly highlight select device applications of BR, with a focus on those that incorporate the Q state photoproduct into the device architectures.

4.1. Holographic associative processors

Associative memories and processors differ significantly from the serial memories that dominate modern computing technologies. These memories mimic the function of the human brain, in which associative recall is used to search through the entire memory bank simultaneously to match an input data block [93,94]. In application, the memory would return the corresponding data block that matches the input criteria, or would return the closest match if no exact copy is identified. Because of this functional paradigm, associative processors are often considered a target design for the development of true artificial intelligence [95]. Fourier-transform holographic optical loops, which were first proposed by Paek and Psaltis, have long been considered an optimal architecture necessary to realize the implementation of this memory class [84,96–98]. Burkin et al. [80] had first proposed thin films of BR as a viable medium for the technology owing to the inherent holographic and photochromic properties of the photoactive protein. During the photocycle of BR, the biomaterial undergoes a series of spectrally distinct photointermediates, which are coupled with a significant transient modulation of the refractive index [64,99]. There are two options for the design of BR-based holographic memories: (i) a real-time processor based on the transient M state intermediate and (ii) a long-term memory based on the Q state photoproduct. We describe both schemes below.

Of particular interest for real-time holography is the formation and reversion of the M state, which is the photosensitive intermediate that displays the largest hypsochromic shift in
\( \lambda_{\text{max}} \) from the \( \text{bR} \) state (570 nm \( \rightarrow \) 410 nm) and contains a reasonable time constant (approx. 10 ms) for real-time holographic processing [100]. Recall that the primary photochemical event of BR is followed by a series of secondary dark reactions, corresponding to the formation and decay of the \( \text{K}, \text{L}, \text{M}, \text{N} \) and \( \text{O} \) states (figure 1) [64]. In addition to this thermal decay pathway, the \( \text{bR} \) state is capable of being regenerated directly from the \( \text{M} \) state by the absorption of light with a wavelength of approximately 410 nm [101]. This ability describes the behaviour of a photochromic pair, in which photochromism is defined as the reversible conversion between two chemical forms with different absorption maxima via photonic stimuli. Both photochemical reactions (i.e., \( \text{bR} \rightarrow \text{M} \) and \( \text{M} \rightarrow \text{bR} \)) exhibit unprecedented quantum efficiencies (approx. 0.65) among alternative media options for holographic associative processors [66,67]. In application, the exploitation of these photochromic properties circumvents the inherent limitation of depending on a finite photochemical intermediate. The lifetime of the \( \text{M} \) state, however, is prohibitively short in the native protein. Through the use of genetic engineering and chemical modification, Hampp [26] and Hampp et al. [102,103] have enhanced the diffraction efficiency of protein thin films and increased the \( \text{M} \) state lifetime up to the order of seconds, thereby enabling success in designing a real-time holographic interferometer. Target residues for site-directed mutagenesis involve those that are directly involved in proton translocation during the BR photocycle [26,102]. Using the directed evolution methodology outlined below, researchers have been able to identify BR mutants that further enhance \( \text{M} \) state characteristics, thus improving the prospect of optimizing the protein for this application [104].

Despite past successes in developing a Fourier-transform holographic associative processor based on BR, the transient nature of the \( \text{M} \) state precludes the design of a long-term data storage device based on this scheme. The branched photocycle and the \( \text{Q} \) state have been considered for use in holographic processing architectures because of a similarly blue-shifted \( \lambda_{\text{max}} \) (approx. 380 nm) and a temporal stability that nullifies the transient photochemistry of BR [25,75]. A recent review describes the incorporation of the \( \text{Q} \) state into holographic memory architectures by using chemical and mutational modification to generate the blue membrane form of the protein [105]. The photochromic pair that involves the \( \text{Q} \) state differs from the \( \text{bR/M} \) state system in two ways. First, the \( \text{Q} \) state is stable for months or years in contrast to the transient BR photocycle [25]. Secondly, the quantum efficiencies of the forward and reverse reactions to the branched photocycle are extremely small [71,106]. This feature, however, is easily circumvented by modern high-energy diode lasers and can even be advantageous for long-term storage applications. In the following section, we describe an alternate device architecture, which implements a volumetric approach to harnessing the sequential two-photon stimulation requirement of \( \text{Q} \) state formation in the purple membrane.

### 4.2. Three-dimensional optical memories

Within an increasingly technology-driven society, consumer demand for faster computing speeds and smaller device interfaces have led to a reevaluation of the efficacy of modern computers. One promising architectural motif, among a diverse collection of forthcoming molecular and biomolecular device structures, involves a three-dimensional (or volumetric) optical memory design [25,85]. BR is capable of serving as an efficient data storage medium when the purple membrane is fixed in a polymer-based suspension [25]. In combination with the \( \text{bR} \) resting state, the \( \text{Q} \) state photoproduce allows for the assignment of binary bit 0 and binary bit 1, respectively, within the volumetric matrix [25,85,86]. Because the branched photocycle is accessed using a sequential two-photon pulse sequence, the \( \text{BR} \) medium is placed within the beam path of orthogonally positioned diode lasers to selectively drive data blocks within the volume into the \( \text{Q} \) state. Unique volumetric patterns within the medium can then be used to store and retrieve information in a computing system based on differential absorptivity. The photochemistry of BR, the architectural scheme, and the nanoscale feature size facilitate a unique potential to write, read and erase data with high speed and a high storage capacity. In the following paragraphs, we briefly outline the specific routines to write, read and erase data; however, a more thorough review of this technology can be found in [25,82,85,86,107].

Both the writing and the reading processes begin by using a green paging laser (approx. 570 nm), which initiates the photocycle within a thin region of the BR medium. To write data within this page, a second laser pulse (approx. 640 nm) must be directed orthogonally to the page within the window of time determined by the BR photocycle lifetime. Ideally, the conversion to the branched photocycle via the second pulse should be accomplished once the \( \text{Q} \) state reaches a maximum concentration. Information is spatially encoded in the incident red writing laser using a spatial light modulator (SLM), which converts the protein into the \( \text{Q} \) state in the desired volumetric pattern. The writing function is an example of an optical AND gate, in which data are written if and only if the two required photochemical conditions are satisfied.

The reading operation begins with an identical first condition, in which the green paging laser activates the BR photocycle (figure 1) within a select region of the medium. In this case, the page being probed contains some volume that is already converted into the \( \text{Q} \) state and transparent to the incident laser pulses. The page is then exposed to red light during the \( \text{O} \) state, however, the intensity of the red writing laser is much lower and is not modulated using a SLM to encode information. A charge-coupled detector (CCD) array is then capable of reading the presence and absence of the \( \text{Q} \) state. In other words, the CCDs measure the pattern of binary bit 0 and binary bit 1 within the page.

To erase data written within the BR medium in the described architecture, the \( \text{Q} \) state must be reverted back to the \( \text{bR} \) resting state by using blue light (approx. 380 nm). The \( \text{Q} \rightarrow \text{bR} \) reversion can be achieved using two methods. First, a blue laser of coherent light can be used to selectively convert individual volumes within the BR medium. Second, the entire BR medium can be erased simultaneously by globally exposing the medium to blue incoherent light. Both options are available in the design of these volumetric optical memories. In general, the hardware and optics necessary to realize this architecture are widely available, and a number of prototypes have been developed that successfully implement the design [25,82]. Methods in directed evolution have now allowed \( \text{Q} \) state formation and reversion efficiencies to catch up to levels imposed by the hardware capabilities [38,41,76,104].
4.3. Protein-based retinal implants

BR has remained at the forefront of protein-based device development for the past several decades, however, this trend has only recently permeated into the field of biomedical devices and prosthetics [28,92]. More specifically, BR is currently being investigated as the photoactive element in a protein-based retinal prosthesis for patients suffering from retinal degenerative diseases such as age-related macular degeneration and retinitis pigmentosa. The proposed retinal implant mimics the light-absorbing capabilities of the homologous native visual pigments and generates a unidirectional proton gradient that is sufficient to stimulate the remaining neural network of the damaged retina [92]. The implant consists of a multilayered BR thin film that is anchored between two ion permeable and biologically inert membrane surfaces. The multiple layers of BR are adsorbed onto the membrane surface using an alternating layer-by-layer approach, with a polycation used to electrostatically interact with each applied layer of BR [108]. The oriented, multilayer thin film is necessary to absorb sufficient incident light and drive the proton-pumping mechanism that is required to generate a neural response [92]. In application, the Q state photoproduct can be used to further manipulate the active surface of the thin film. Retinal degenerative diseases are inherently difficult to treat with prosthetics because of extreme variability of pathologies, which are unique to the individual patient. Driving select areas of the film to the Q state can turn down or turn off extraneous pixels, which allows for the selective tuning of the active surface area of the implant that is required to generate meaningful vision.

5. Optimization strategies

The optimization of BR for a synthetic environment requires enhancing properties of the protein, which are either irrelevant or non-essential for biological function. Although researchers are actively studying the structural motifs and intramolecular interactions responsible for the photophysical properties of BR, many of these interactions remain unknown. Despite access to rigorous theoretical models and fast computers, neither software nor hardware can adequately predict which mutations will provide the desired properties. More expansive experimental methods and high-throughput screening techniques must be developed in order to make protein-based devices commercially viable. Consequently, devices, which are based on the Q state of BR, have been met with limited success because the native protein does not efficiently access the branched photocycle.

Advancements in genetic engineering techniques have changed the way scientists approach the field of nanotechnology. The redesign of proteins for use in synthetic environments has gained momentum owing to improvements in site-directed mutagenesis, site-specific saturation mutagenesis, semi-random mutagenesis, random mutagenesis and now directed evolution. Perhaps the most versatile method for improving protein performance is through the use of site-directed mutagenesis [109,110]. Site-directed mutagenesis is a targeted technique, which has been used to characterize structure–function relationships proteins from a variety of organisms. This technique is used to introduce a specific mutation at a point of interest on the primary sequence of the protein. Mutations are introduced using oligonucleotides that are complementary to a part of single-stranded DNA, but contain an internal mismatch, which directs the mutation. Site-directed mutagenesis is a powerful method when one knows what sites on the primary protein sequence to target. However, this technique is often inefficient when large protein sequences must be probed. Saturation mutagenesis is similar to site-directed mutagenesis, but has a higher efficiency of generating more mutations. With saturation mutagenesis, many mutations are generated via the saturation of a key residue or residues at a target locus on the gene of interest. In the case of site-specific saturation mutagenesis, doped oligonucleotides are used to introduce all 20 amino acids at the site of interest [111]. As a result of the limited efficiency of site-directed and saturation mutagenesis, many researchers have turned to more global techniques, implementing high-throughput screening and selection methods to optimize the genetic landscape of biological molecules. Semi-random mutagenesis and random mutagenesis have the capability to produce a large number of indiscriminate variants via chemical manipulation, ultraviolet light, error prone PCR, or the use of doped oligonucleotides [111]. These methods offer a selective advantage over classical mutagenesis strategies because they allow for a greater mutational landscape of the protein to be explored. In the case of mutagenesis for device applications, semi-random and random mutagenesis permit greater genetic diversification of the protein, allowing the researcher to narrow in on the diverse interactions, which contribute to a desired phenotype. An efficient and effective screening method must be available in order to make the generation and screening of such a large library of mutants time and cost effective.

For decades, researchers have been using semi-random or saturation mutagenesis to identify particular residues that play an integral role in protein function. Mutagenesis of this type can best be envisioned in terms of a mutational landscape, where the optimization of a single characteristic, such as the photophysical properties of the protein, is represented by fairly localized, irregular changes (figure 2b) [38]. More fluid properties, such as thermal stability, involve larger portions of the protein and a variety of complementary mutations, thus leading to a mutational landscape that is continuous and slowly varying (figure 2a). Unfortunately, for scientists trying to optimize proteins for function outside of the biological context of the organism, the challenge lies in predicting a priori which amino acid will generate the desired properties for the intended application.

Characteristic targets for the mutagenesis of phototransductive proteins are typically a group or series of residues that contribute to the photochemistry of the molecule. In the case of BR, a majority of mutagenesis has focused on optimizing the residues that contribute to the formation of the M and O photo-intermediates and the Q photoproduct because these states offer the greatest potential for use in device applications. In addition to enhancing the lifetime of different photocycle intermediates of BR, mutagenesis has been employed to enhance the innate dipole moment, gold-binding capabilities and specificity of ion pumping of BR. For a number of device platforms and architectures, the ability of BR to bind to gold is critical because gold is a chemically inert and electrically conductive [112–114]. The native protein contains no cysteine residues and thus the strategic addition of cysteine residues in the loop regions of BR via site-directed mutagenesis allows BR to
because the organism will either buffer the pH or die because accomplished through traditional cell-based techniques of the directed evolution process and cannot be adequately describe the importance of pH screening in the optimization of how to optimize a complex photochemical reaction, optimization of photoactive proteins, particularly the issue of pH fluctuations. In particular, we will focus on the optimization of the pH photoproduct of BR, a photochemical state that is rarely found in nature.

5.1. Optimization of the branched photocycle of bacteriorhodopsin by using directed evolution
In order to optimize proteins for applied technologies, a number of criteria must be satisfied. First, mutational vectors and methods must exist in order to introduce genetic variation into the coding sequence of the protein of interest. Next, an expression method must be implemented in order to obtain sufficient quantities of the macromolecule under investigation for testing. Finally, an assessment paradigm must be available to identify, screen and characterize the mutations generated. The ability to generate successful mutants relies heavily on the ability to identify the mutations of interest quickly, efficiently and inexpensively. High-throughput screening methods are necessary to detect improvements in the phenotype of interest. The methods and procedures presented below provide only the fundamental features of BR optimization, but provide a template for type I directed evolution of any protein for which expression methods, mutational vectors and screening methods are available.

The optimization of BR for devices that implement the branched photocycle requires the simultaneous modification of five variables: minimization of the formation lifetime of the O state, optimization of the decay of the O state, enhancement of the quantum efficiency of the O to P photochemical reaction, enhancement of the efficiency of the P to Q hydrolysis and an increase in the lifetime of the O state [38]. Prediction of all of the impactful mutations and interactions that would simultaneously lead to the optimization of each of these photochemical properties is nearly impossible. Although experimental data exist that would allow scientists to begin targeting residues of BR for specific photochemical enhancement, the time and cost associated with mutagenesis is prohibitive. Thus, directed evolution is an excellent method for the optimization of BR for a variety of applications.

Type I directed evolution, which implements automated screening methods and microgram protein characterization, is used to generate new BR mutants through a combination of region-specific semi-random mutagenesis, site-directed mutagenesis and saturation mutagenesis. These mutants are screened and evaluated with respect to the formation and reversion efficiencies of the branched photocycle, particularly the Q state. After selection and identification of the best Q state mutants, the most efficient Q state mutants serve as the parents to the next generation of progeny. This process is reiterative and gradually improves the efficiency of the Q state photochemistry at each stage of directed evolution. In order to maximize efficiency, only the best mutants from each round of automated testing need to be sequenced, and thus the procedures are cost and time efficient.

In type I directed evolution of BR, a diverse molecular library of BR mutants is generated using region-specific semi-random mutagenesis (figure 3). First, the bacterio-opsin (bop) sequence is divided into 17 regions, approximately 15–20 amino acids in length (figure 4, inset). Next, the sequence overlap extension method [116] is employed and doped oligonucleotides are designed to overlap each of the 17 regions of interest in the bop coding sequence. The resulting diversified mutant library is then transformed into Escherichia coli for covalently bind to gold. Enhancing the dipole moment of BR is critical for the application of the protein in photovoltaic devices [23,89,115]. Altering the intrinsic dipole moment through charge substitution in the helical loop regions of the protein allows the protein to pack more densely, while enhancing the photovoltaic output or signal from BR.

Genetic optimization of BR has led to the commercial development of a variety of applied technologies. In the following sections, we describe the use of type I directed evolution for the systematic optimization of BR for applied technologies. We explore the unique challenges faced in the optimization of photoactive proteins, particularly the issue of how to optimize a complex photochemical reaction, which gains stability from a two-dimensional lattice during protein expression in the native organism. Furthermore, we describe the importance of pH screening in the optimization of photoactive proteins. Enhancement of pH is a critical part of the directed evolution process and cannot be adequately accomplished through traditional cell-based techniques because the organism will either buffer the pH or die because

Figure 2. (a) Hypothetical mutational landscapes for fluid traits and (b) photochemical optimizations. The native protein is located at the squares located under A, the mutational coordinates \( \xi \) and \( \zeta \) are arbitrary, and the vertical axis measures an arbitrary Q (quality) factor. Beginning at A, the goal of directed evolution is to discover the optimal protein, represented by peak E. The \( A' \) and \( A' \) regions signify mutations that have little impact on the Q factor, the B regions are local maxima in Q and the optimal mutation is at the peak labelled E. Cyclical optimization will typically find E in the fluid case, but is unlikely to find E in the photochemical optimization without good modelling or operator intuition. For fluid characteristics such as thermal stability, less variation is expected (a). The landscape shown in (b) depicts more complex characteristics such as photochemical properties, in which most of the performance is localized in specific regions of the protein; after fig. 3 in [38].
amplification using the pBA1 expression vector. Following amplification, the resulting genetically distinct colonies are pooled and allowed to grow overnight. Next, mutant DNA is extracted and transformed into the native organism, *Halobacterium salinarum*, via the MPK 409 cell line. Transformants were selected using mevinolin resistance and recombinants were selected using 5-fluoroorotic acid (5-FOA). Individual colonies were picked from the 5-FOA plates and replicated in 96-well plates. An automated irradiator was used to provide data for computer analysis of $Q_{\text{formation}}$ and $Q_{\text{reversion}}$ and the best mutants were selected based on a computer algorithm. The result is a visual colorimetric display of the best $Q$ state formers with lighter colours representing high $Q$ state formation and darker colours signifying low or no $Q$ state formation. Mutants that have a high $Q_{\text{max}}$ are selected to serve as the parents to the next generation of mutants.

Screening of BR involves the characterization of select photophysical properties of the protein. An in-house automated irradiator is used to provide data for computer analysis of $Q$ formation ($bR \rightarrow Q$) and $Q$ reversion ($Q \rightarrow bR$) (figure 3). The automated irradiator system uses a set of 12 640 nm Luxeon III Lambertian LEDs, each driven at 850 mA to uniformly irradiate the 96-well plates containing purified BR mutants. The 640 nm light is important for coupling the absorption bands of the $bR$ resting state and the $O$ state. Each plate is then irradiated for 3 h under constant temperature (35 ± 0.1°C), and then scanned using a microplate reader to measure the absorption spectrum of the protein in each of the wells. The best mutants are selected based on a computer algorithm, which analyses the amount of $Q$ state formed for each mutant. Following $Q$ state formation, the plate is returned to the irradiator for further illumination via a set of 28 510 mcd LEDs at 395 nm for 2 h. The plate is then scanned using the microplate reader and the absorption spectrum is recorded for evaluation of the efficiency of reversion from ($Q \rightarrow bR$). Computer analysis of the resulting absorption spectra is used to calculate the formation and reversion efficiencies of the $Q$ state. The output of the computer analysis is a visual colorimetric display of the best $Q$ state mutants, where lighter colours represent a greater formation of the $Q$ state and darker colours signify low or no formation of the $Q$ state.

The computer algorithm functions by assigning a quality to each variable, $Q$, which should not be confused with the $Q$ state. The greater the $Q_{\text{total}}$, the more efficient the formation and reversion of the $Q$ state. $Q_{\text{formation}}$ represents the
multistep process of forming the $Q$ state ($bR \rightarrow Q$), while $Q_{\text{reversion}}$ represents the process of reverting the protein back to the resting state ($Q \rightarrow bR$):

$$Q_{\text{formation}} = \xi_{\text{experimental}} \times Q_{bR} \oplus Q_{O} \oplus Q_{P} \oplus Q_{Q},$$

$$Q_{\text{reversion}} = 1.0 + \xi_{\text{experimental}} \times Q_{Q} \oplus bR,$$

$$Q_{\text{total}} = Q_{\text{formation}} \times Q_{\text{reversion}}.$$

The $\xi$ multipliers are arbitrary multipliers, which define the scaling. The formation and reversion efficiencies are critical to the application of the $Q$ state of BR in read–write devices. The key is to use fully characterized proteins as internal standards to assign both $Q_{\text{formation}}$ and $Q_{\text{reversion}}$.

At the conclusion of each stage of mutagenesis, the top 10–20 mutants with the highest $Q_{\text{total}}$ values are sequenced. New mutants are then constructed using the best $Q$ state mutants from the previous round to serve as the parents to the next generation of progeny. The goal of this iterative process is to only pass on useful mutations and to avoid carrying silent mutations forward.

An important characteristic to consider when optimizing photoactive proteins is the inclusion of pH as a variable.

**Figure 4.** Type I directed evolution of BR to enhance formation of the $Q$ state. Progression of $Q_{\text{total}}$ efficiency through six stages of Type I directed evolution is shown. The shading corresponds to the parent mutation, as shown in the figure insert at upper left. Only those mutants with high $Q_{\text{total}}$ were sequenced so that these mutations could serve as a parent protein in the subsequent stage of evolution.

**Figure 5.** Number of mutations at each residue following six stages of directed evolution on BR using the methods and procedures outlined in this paper. Note that a majority of the mutations had no significant impact, and only a handful yielded significant improvement.
The pH of a solution has a significant role in the protonation states of the binding site residues. In stages 4–6 of the directed evolution of BR, pH was included as a screening parameter (figure 4). In order to implement this screening method, the best mutants at pH 7 were grown up at larger volumes and scanned from pH 6 to 10.5 in increments of 0.5, and the $Q_{\text{total}}$ value for each mutant was registered at the pH value that yielded the best result. Buffers below a pH of 6 were excluded to avoid formation of the blue membrane [71,118,119].

After six generations of optimization via directed evolution, involving over 10,000 mutants, a number of mutants have been discovered with excellent $Q$ state formation and reversion efficiency. The results of each of the six stages of mutagenesis to optimize $Q_{\text{total}}$ can be seen in figure 4. Important to note is the fact that many of the mutations to BR had little or no impact on protein function (figure 5). Additionally, the mutations that yielded a high $Q_{\text{total}}$ were mutations outside of the binding site (figure 6), which was an unexpected result. Moreover, without the inclusion of

---

**Figure 6.** The improvement in the variable $Q_{\text{total}}$ as a function of mutational position following six stages of directed evolution is shown in (a). The number of mutational hits ($N$) at each site during these stages is shown in (b). Note that only mutations that result in the formation of photochemically active (purple) membrane are included, and that hits at a given site may involve identical site mutations or are combined with other sites to yield a unique mutant. The residues in white were either not included in the search (238–248 on C-terminus end) or, when mutated, yielded inactive proteins or blue membranes.
pH as a variable, the best mutants would have not been discovered. The best mutant, V49A/I119T/T121S/A126T, was discovered after six stages of directed evolution and has a $Q_{\text{total}}$ of 977, which is approximately 70 times greater than wild-type BR (figure 7). The best single mutant, V49A, has a $Q_{\text{total}}$ of 924, nearly 62 times greater than wild-type BR. Both of these mutants offer significant advantages in the formation and reversion of the $Q$ state, and have allowed for improved commercialization of BR in devices.

Understanding the key mechanisms of $Q$ state enhancement in V49A and V49A/I119T/T121S/A126T is important to further enhance BR through directed evolution. We predict that the mechanism of $Q$ state enhancement in V49A is associated with stabilizing the hydrolysed 9-cis chromophore in the binding site. By providing a binding site that preferentially accommodates a 9-cis chromophore, V49A appears to not only enhance the stability of the $P$ state, but also appears to lower the activation barrier to the formation of the $P$ state. The V49A/I119T/T121S/A126T mutant adds hydroxyl groups, which alter the electrostatics of the binding site (figure 7) and lower the pH for optimal performance. The latter is important for making polymeric matrices, which are more stable at near-neutral pH values.

5.2. Optimization of protein stability

Any protein that is to be used as the photoactive element of a device must be resistant to both thermal and photochemical stress [25,101,120]. Type I directed evolution of BR includes cycllicity as a variable, and mutants with poor photochemical reversion will therefore have a low $Q_{\text{total}}$ and will not be selected for further investigation (see discussion above). However, the above methods are not designed to optimize the thermal stability of the protein, and thus a mutant with outstanding $Q_{\text{total}}$ may also diminish the structural integrity of BR. Moreover, the reversion efficiency optimized here does not translate directly into high photochemical cycllicity. Photochemical cycllicity measures how many times the protein can undergo a photocycle before $1/e$ (approx. 37%) of the protein ensemble has denatured [120]. A proper study of photochemical cycllicity is an arduous process, particularly when the protein has a high cycllicity [120]. We have found that the vast majority of BR mutants with a $\lambda_{\text{max}}$ between 540–570 nm have excellent photochemical cycllicity, comparable to the native protein [120]. The V49A and V49A/I119T/T121S/A126T mutants are both being studied in more detail to establish a formal cycllicity, but preliminary results suggest these mutants are within approximately 20 per cent of the native protein. The photochemical cycllicity required for a given device depends on the application and varies significantly. Volumetric memories can operate with a cycllicity as low as $10^3$, whereas most Fourier-transform optical associative processors require a cycllicity above $10^5$ [25,101,105,120]. This observation parallels the range observed in living organisms. *Halobacterium salinarum* requires a native protein with a high photochemical cyclicity (approx. $10^6$), whereas the deep sea

![Figure 7](http://rsif.royalsocietypublishing.org/) Location of the residues modified to form the high $Q$ mutant V49A/I119T/T121S/A126T, with the relevant residues marked with red lettering. Other residues that are important to the photochemical properties or proton translocation process are also shown. The proton channel is shaded. The contours of the electrostatic difference map associated with the quadruple mutation are shown in the background. The key electrostatic impact of the mutation is to increase negative charge in the Asp-85 (D85) and Asp-212 (D212) region, which stabilizes the protonation of D85 and prolongs the yield and lifetime of the $Q$ state. The V49A component of this mutation also preferentially stabilizes the 9-cis conformation. Contour lines are drawn at the following energies (J/mol): +1, +6, +21, +49, +96, +170, +260, +390, +560, +760, +1000, +1300, +1700, +2100, +2600, +3100, +3800, +4500, +5200, +6100.
bacteria that uses blue proteorhodopsin experiences a low enough light flux to require a protein with a much lower photochemical cyclicity (approx. 10^6) [120]. Organisms have a significant advantage in that they can express additional protein if the existing population is compromised. As a result, a device will place a higher burden on photochemical cyclicity than an organism. Hence, the remarkable photochemical cyclicity of BR is an important attribute that provides a comparative advantage over the majority of photochemical proteins and organic molecules.

Structural stability is the second important attribute of a photochromic material, particularly for devices that operate above ambient temperatures. Stability is often evaluated by measuring thermally induced conformational changes in the protein structure via differential scanning calorimetry. These transitions are often reported in terms of a melting temperature \(T_M\), or the temperature at which the transition exhibits the greatest change in heat. Two such transitions are observed for BR in an unbuffered solution (figure 8): a reversible relaxation of the rigid protein structure at approximately 80°C and an irreversible denaturation at approximately 98°C [49]. This stability is sensitive to several factors that include pH [121,122], manipulation of the lipid environment [123,124] and chemical effects [125–127]. Unsurprisingly, BR exhibits the greatest stability in the native membrane. Hence, the \(T_M\) of BR mutants is usually measured in the purified native membrane and suspended in deionized water. In general, most BR mutants yield data similar to the native protein, with \(T_M\) values at approximately 80°C and 90–100°C. Some mutants (e.g. V49F, R82G) exhibit complex profiles with a skewed baseline, which indicate that the protein is unstable (figure 8). These mutants are removed from the pool of candidate proteins.

BR is a kinetically stabilized protein, meaning that the irreversible \(T_M\) of BR is sensitive to the rate of heating. Measuring the \(T_M\) of mutants is therefore only the first step of screening the stability of candidates. Mutants with native-like \(T_M\) values are subjected to kinetic denaturation experiments that estimate the amount of total thermal energy required to denature the protein. Data from three to five traces, collected at various scanning rates, are fit to an Arrhenius model:

\[
\ln \left( \frac{v}{T_M^2} \right) = \text{const.} \frac{E_{\text{APP}}}{RT_M},
\]

where \(v\) is the scan rate (K min\(^{-1}\)), \(T_M\) is the melting temperature from differential scanning calorimetry data (K), \(E_{\text{APP}}\) is the apparent energy of thermal denaturation (kJ mol\(^{-1}\)) and \(R\) is the gas constant (8.314 J mol\(^{-1}\) K\(^{-1}\)) [128]. In water, native BR exhibits a stability of 900 kJ mol\(^{-1}\) and the V49A mutant, for example, demonstrates a similar stability of 850 kJ mol\(^{-1}\). Suspension in buffered solutions at slightly alkaline pH (e.g. pH 8.5), where \(Q_{\text{total}}\) is more efficient, roughly doubles this stability. This method consumes a considerable amount of protein and is only done for the top performing mutants.

6. Conclusions

Advances in genetic engineering and, more significantly, directed evolution have enhanced the potential of developing commercially viable protein-based devices. The approach presented in this review serves as a general template for researchers interested in creating a mutant library that would ultimately lead to a desired phenotype for application outside of the biological context of the organism. While the properties relevant to \(Q_{\text{total}}\) are capable of being measured by using absorption spectroscopy on BR mutants, we suggest that other spectroscopic or analytical methods could be exploited to monitor a similar series of stages of directed evolution when in pursuit of a non-native physical property. Using a combination of mutagenesis methods in conjunction with high-throughput screening techniques, the overall yield, lifetime and formation and reversion efficiencies of the \(Q\) state of BR have been enhanced for applications in bioelectronics. After six stages of directed evolution, nearly 10 000 BR mutants were generated, with a majority of those mutants offering improvements in \(Q_{\text{total}}\). Two of the best mutants, V49A and

![Figure 8. Differential scanning calorimetry traces of mutant BR relative to native wild-type (WT) BR. The relative heat capacity of each mutant is used as a comparative measure to highlight observed differences in melting temperature (\(T_M\)) and the general nature of the reversible and irreversible thermal transitions of the protein. Some mutants have stabilities roughly equivalent to WT (V49A, R82K and E204Q), while others can be considered less stable (I119T/T121S/A126T and T90A) or prohibitively unstable (V49F, R82G and L206P) for use in device applications. Note that V49A exhibits a \(T_M\) equivalent, if not superior, to WT, whereas I119T/T121S/A126T is slightly less stable. The stability of the V49A/I119T/T121S/A126T mutant is currently uncharacterized, however, the data suggest that the combination of V49A and the triple mutation will lead to a protein with sufficient stability for use in biophotonic devices.](http://rsif.royalsocietypublishing.org/Downloaded from)
V49A/II19T/T121S/A126T, have made it possible to use low intensity write lasers in three-dimensional optical memories. Efforts are being carried out to study these mutants in more detail and to further investigate the potential of these mutants as the photoactive element in photonic devices.

This research was supported in parts by grants from the National Science Foundation (EMT-0829916, EIA-0129731), the National Institutes of Health (GM-34548), DARPA (HR00011-05-1-0027), the Army Research Office (MURI, DAAD 19-99-1-0198) and the Harold S. Schwenk Sr. Distinguished Chair in Chemistry.

References


