Optimization of bacteriorhodopsin for bioelectronic devices

Kevin J. Wise, Nathan B. Gillespie, Jeffrey A. Stuart, Mark P. Krebs and Robert R. Birge

Bacteriorhodopsin (BR) is the photoactive proton pump found in the purple membrane of the salt marsh archaeon *Halobacterium salinarum*. Evolution has optimized this protein for high photochemical efficiency, thermal stability and cyclicity, as the organism must be able to function in a hot, stagnant and resource-limited environment. Photonic materials generated via organic chemistry have yet to surpass the native protein in terms of quantum efficiency or cyclicity. However, the native protein still lacks the overall efficiency necessary for commercial viability and virtually all successful photonic devices using bacteriorhodopsin are based on chemical or genetic variants of the native protein. We show that genetic engineering can provide significant improvement in the device capabilities of proteins and, in the case of bacteriorhodopsin, a 700-fold improvement has been realized in volumetric data storage. We conclude that semi-random mutagenesis and directed evolution will play a prominent role in future efforts in bioelectronic optimization.

DOI: 10.1016/S0167-7799(02)02023-1

Kevin J. Wise
Robert R. Birge
Dept of Chemistry and of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA.
e-mail: rbirge@uconn.edu

Nathan B. Gillespie
Jeffrey A. Stuart
W.M. Keck Center for Molecular Electronics, Syracuse University, Syracuse, NY 13244, USA.

Mark P. Krebs
Dept of Biological Sciences, Illinois State University, Normal, IL 61790, USA.

Bioelectronics is a field of molecular electronics that investigates the use of native, as well as modified, biological molecules (chromophores, proteins etc.) in electronic or photonic devices. Bioelectronics has shown considerable promise largely because evolution has often solved problems of a similar nature to those that must be solved in creating electronic devices from organic compounds and because self-assembly and genetic engineering provide sophisticated control and manipulation of large molecules or ensembles. Much of the current research effort in bioelectronics is directed towards self-assembled monolayers and thin films, biosensors and protein-based photonic devices.

Although several proteins have been explored for device applications [1–3], bacteriorhodopsin (BR) has received the most attention (reviewed in [1,4]). Historically, the late Yuri Ovchinikov deserves much of the credit for uncovering the potential of this material. As director of the Shemyakin Institute of Bioorganic Chemistry in Russia in the early 1970s, he proposed that Soviet science could leapfrog the West in computer technology by exploring bioelectronics and garnered significant funding for this proposal from the Soviet military under what was known as 'Project Rhodopsin'. One of the best-known accomplishments of this project was 'Biochrome' – a real-time photochromic and holographic film based on chemically modified polymer films containing BR [5]. More recent work has concentrated on genetic engineering to enhance holographic efficiency [6,7]. Genetic modification of BR has led to the successful commercial development of a real-time holographic image analysis system [8,9]. This review updates the use of genetic engineering to enhance the properties of BR for use in optical memories, and in particular, branched-photocycle 3D memories [4].

Bacteriorhodopsin-based protein optical memories offer great potential [1]. These memories can be configured for thin-film photochromic [10,11], 3D branched-photocycle [4,12], holographic-binary [7,13,14] or Fourier-transform holographic associative storage [4,12]. The storage medium is lightweight, radiation hardened and relatively inexpensive, and combined with the inherent quantum efficiency and cyclicity of the protein, provides comparative advantages over present organic and inorganic media. Nevertheless, the native protein is not optimal for any of the above-mentioned optical memories. In all studies to date, the protein has been modified chemically or genetically to enhance performance. In some cases, the viability of the memory system remains problematic pending significant improvement in the memory medium. Such is the case for the branched-photocycle 3D-memory system discussed. And although significant improvements in BR performance have been made using site-directed mutagenesis [15], it has become clear that further improvements will be much harder to achieve and will require more expansive methods. For this reason, random mutagenesis and directed evolution are currently under investigation and will be discussed here. To appreciate the goals of the genetic engineering effort requires some background on the photophysical properties of native BR and the architecture of the 3D memory.

Bacteriorhodopsin as a photonic material

Bacteriorhodopsin is a membrane-bound light-transducing protein that functions as a proton pump in the archaeon *Halobacterium salinarum* (previously known as *Halobacterium halobium* or *Halobacterium salinarium*). *H. salinarum* has adapted to high salt environments (5 M or 25% NaCl) and can use BR to absorb light energy and convert it to chemical energy. In response to low oxygen availability, *H. salinarum* produces the purple membrane, a natural crystal in which thousands of BR trimers are assembled in a 2D hexagonal lattice. Although BR is crucial for long-term cellular survival when oxygen is limited, the protein is not essential under aerobic conditions, because *H. salinarum* can also obtain energy from respiration.

  PII: S0167-7799(02)02023-1

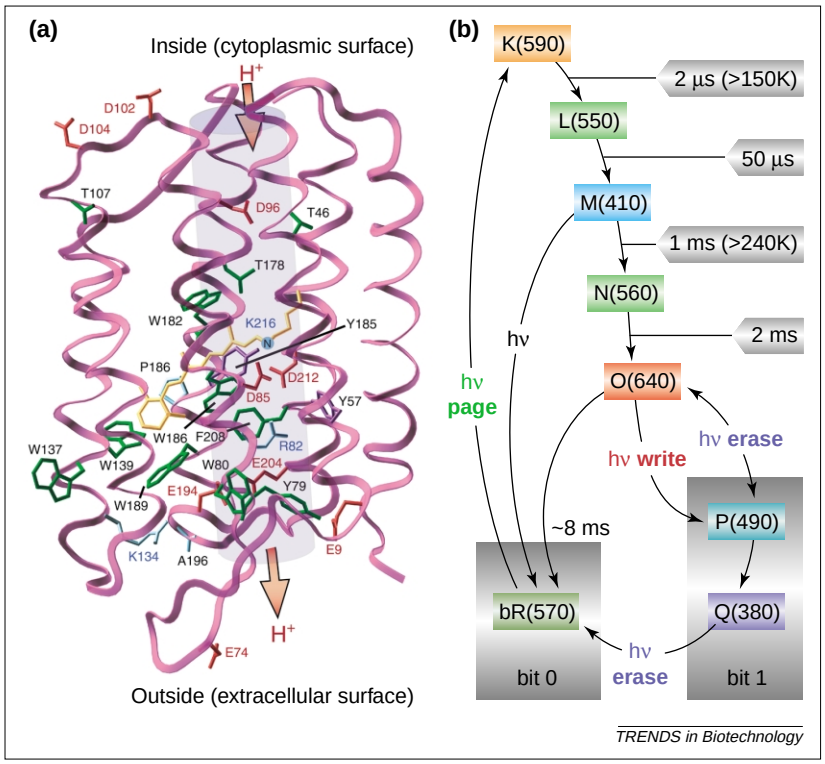


Fig. 1. A schematic representation of the bacteriorhodopsin tertiary structure is shown in (a) and the main photocycle and the branching reactions used to store data are shown in (b). The structure shown in (a) is based on the crystal coordinates [38], the approximate region of the proton pumping channel is shown in blue, and selected residues are shown.

Although the primary function of BR in *H. salinarum* is to pump protons, it is the ability of the molecule to absorb and convert light energy that is of particular importance for use in photonic devices. Each BR molecule consists of two components: a 248 amino acid polypeptide (bacterio-opsin or BO) and a chromophore (retinal) embedded in the middle of the protein (Fig. 1a). The proton pumping process of BR is initiated when the chromophore absorbs light and undergoes the photocycle schematically shown in Figure 1. Although many aspects of the proton-pumping mechanism remain to be discovered, much has been learned about the molecular details during the past few years [16]. The primary photochemical event involves isomerization of the chromophore from all-*trans* to 13-*cis*, producing the first trappable intermediate labeled K, shown in Figure 1. The proton-pumping mechanism is then performed via a series of dark reactions that form, in succession, the L, M, N and O intermediates (Fig. 1). The net result is transfer of a proton to the extracellular surface from a series of residues including Arg82, Glu194, Glu204 and Asp212 (all in three-letter amino acid code) and a hydrogen-bond network originating from water molecules (collectively referred to as the 'leaving group' or 'XH') [17,18]. Replacement of Asp96 with Asn (D96N, in single-letter amino acid code) lowers the efficiency of the reprotonation step and increases the lifetime of the M intermediate from a few milliseconds to 100 ms at ambient temperature. The D96N mutant is optimal for real-time holographic memories [6,8,9].

The branching reaction that is crucial to operation of the 3D memory is shown in Figure 1b. The branch involves photoactivation of the O state by red light, which induces all-*trans* to 9-*cis* photochemistry [19]. The electrostatic environment associated with protonation of Asp85 is responsible for selecting all-*trans* to 9-*cis* photochemistry (see Fig. 4 in [4] and also see [11]). The 9-*cis* chromophore is not stable in the binding site, however, and hydrolysis of the Schiff base takes place to produce 9-*cis* retinal trapped in the binding site [19]. The result is formation of the Q state ($\lambda_{max} \approx 380$ nm). In the absence of blue light or high temperatures, data stored using the Q state is stable for decades. However, under blue light irradiation, isomerization of the protein-constrained 9-*cis* retinal chromophore takes place to generate all-*trans* retinal, which spontaneously recombines with Lys216 to regenerate the light-adapted resting state, bR. Ongoing research efforts aim to enhance the efficiency of the branching reaction for use of the protein in device applications. Before examining the methods and procedures to do so, a discussion of the branched-photocycle optical architecture is necessary.

Branched-photocycle optical data storage

3D memories store information in a volumetric memory medium and offer a 300-fold improvement in data storage capacity for a given enclosure size. The branched-photocycle architecture explored here offers some important advantages. First, it requires two temporally separate beams to store information. This sequential excitation can rigorously exclude any photochemistry outside the irradiated volume, simplifying the optical design and improving reliability. Second, the use of linear, rather than nonlinear, excitation allows the use of inexpensive continuous wave (CW) lasers, which increases flexibility and decreases the cost of the read–write architecture. The total memory system gains additional comparative advantage from the inherently low cost of the storage medium. Large quantities of the protein can be produced through simple fermentation and isolation procedures.

The branching reaction of interest was discussed above and is shown in Figure 1. The memory functions by assigning the resting state (bR) to bit 0 and both P and Q to bit 1 (Fig. 1b). The P 'state' is actually a pair of states with absorption maxima at 445 and 525 nm that interconvert rapidly owing to protein relaxation processes [20]. For data storage, these species can be treated as a single intermediate. The above read, write and erase scheme can be realized using the optical layout shown in Figure 2. The device provides a method of storing and retrieving data within a volumetric memory medium consisting of BR in a polymer matrix sealed in a plastic cuvette (for a detailed discussion of the memory architecture see [4]). The key issue for this discussion is how genetic engineering can be used to optimize the protein with respect to maximizing the efficiency of the branching reaction shown in Figure 1.

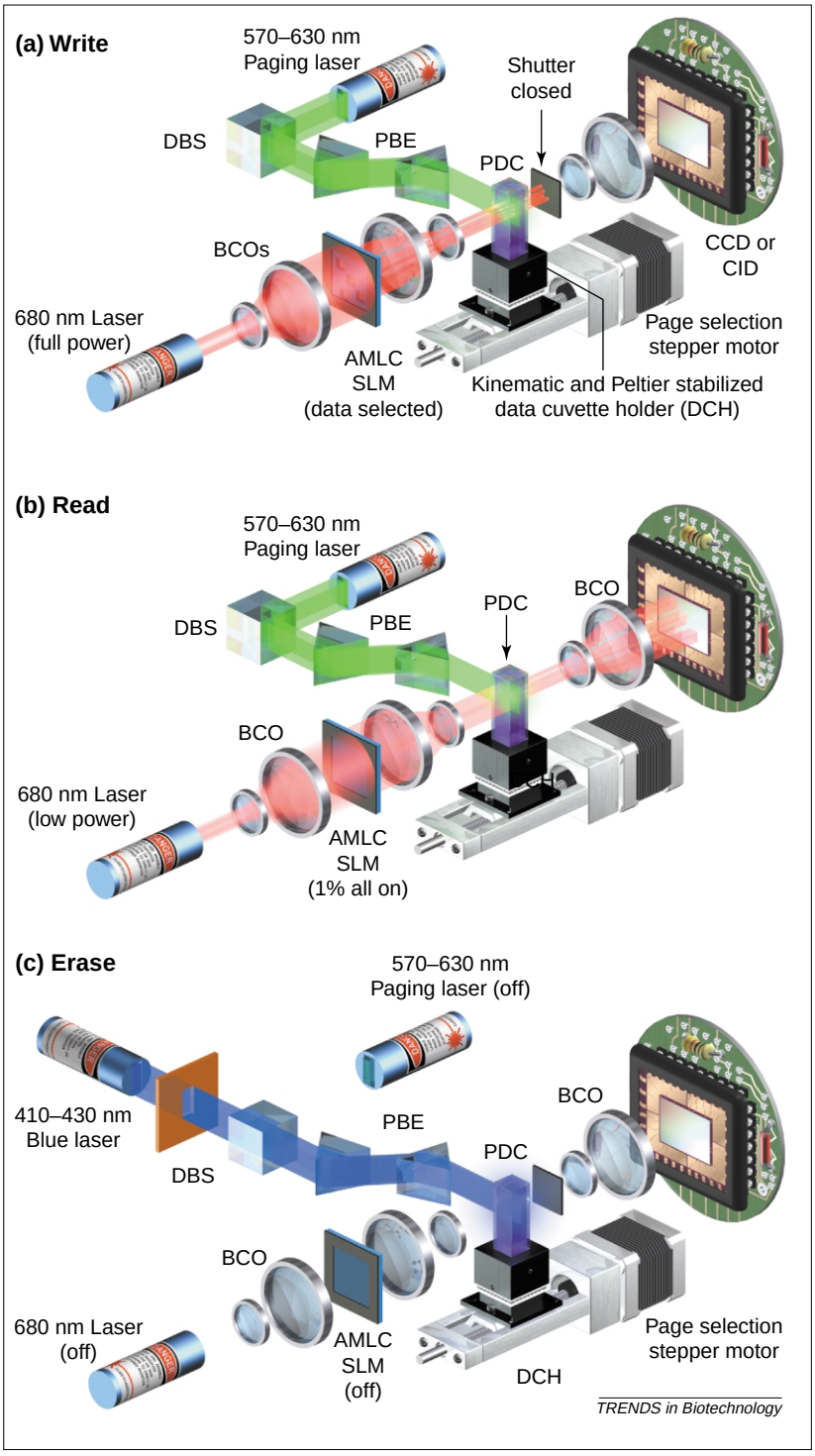


Fig. 2. The write, read and erase operations of the bacteriorhodopsin-based branched-photocycle memory. The write and read operations are both initiated by using a paging beam to activate the photocycle in a thin region within the memory medium (green light). On formation of the O state in this page, a write beam is activated and the data imposed on the beam using a spatial light modulator (SLM). The read process is similar, but in this case the modulator is turned off so that just enough light gets through to image the page onto the charge couple device (CCD) [or charge injection device (CID)] array. Because the O state is the only species that absorbs the write beam, the read and write processes only involve interaction of the light beam with the O state in the paged regions. A blue laser erases an entire page. These processes are shown schematically in Figure 1. Abbreviations: beam coupling optics (BCOs); DBS, dichroic bean splitter; PBE, prism beam expander; PDC, protein data cuvette.

Optimizing the protein for 3D data storage

This section outlines some of the current methods that are being used to enhance the yield of the O state and the quantum efficiency of the O to P photoreaction. Note that the O state and bit 0 are different entities (see Fig. 1b). Bacteriorhodopsin can be optimized in one of three ways: chemical modification of the chromophore and genetic and/or chemical modification of the protein. To assist these studies, strains of *H. salinarum* are used that are deficient in either the chromophore or the BO polypeptide, but not both. Retinal-deficient strains produce BO without incorporating a chromophore and provide an efficient method of generating analogue proteins with synthetically modified chromophores [13,21]. We also use the BO-deficient cell line L33 [22], which contains a DNA insertion within the gene encoding BO, *bop*. Production of native BR is abolished in L33, allowing expression of mutant proteins. Strains in which the *bop* gene is deleted or replaced with a selectable marker have also been used for genetic modification of BR [23].

The principal approaches to genetic engineering of proteins for device applications are compared in Table 1. We envision that the mutagenic optimization process surveys a mutational 'landscape', as shown in Fig. 3, which depicts the potential outcomes of optimizing a protein for device applications. Depending on the characteristics being measured several results are possible. Assuming that a triple mutant confers the optimal protein, a single mutation (peak B) might provide significant improvement. However, the optimal protein (peak E) might be surrounded by double-mutant troughs because of interference effects and these are very common when key mutations involve charged residues.

Site-directed mutagenesis

The ability to express BR and its site-directed mutants within the native organism, *H. salinarum*, is crucial to materials optimization [22]. The native organism provides the cellular machinery to express the protein within the purple membrane and it is the crystalline lattice of the purple membrane that confers the high photochemical and thermal stability that characterizes this system [24].

Although site-directed mutagenesis has been the primary technique for optimizing BR for device applications, it has a fundamental flaw. This technique is rarely useful without access to a good model for the molecular process that is to be optimized. The usefulness of site-directed mutagenesis is proportional to knowledge of protein function obtained from crystal structures and modeling programs. Recent reviews offer a perspective on the current status of BR modeling [25,26]. It is important to note that modeling can be based on theory or experiment and is at its best when operating in synergy. A good example was the development of the D96N mutant, which was optimized for holographic applications [8,9].

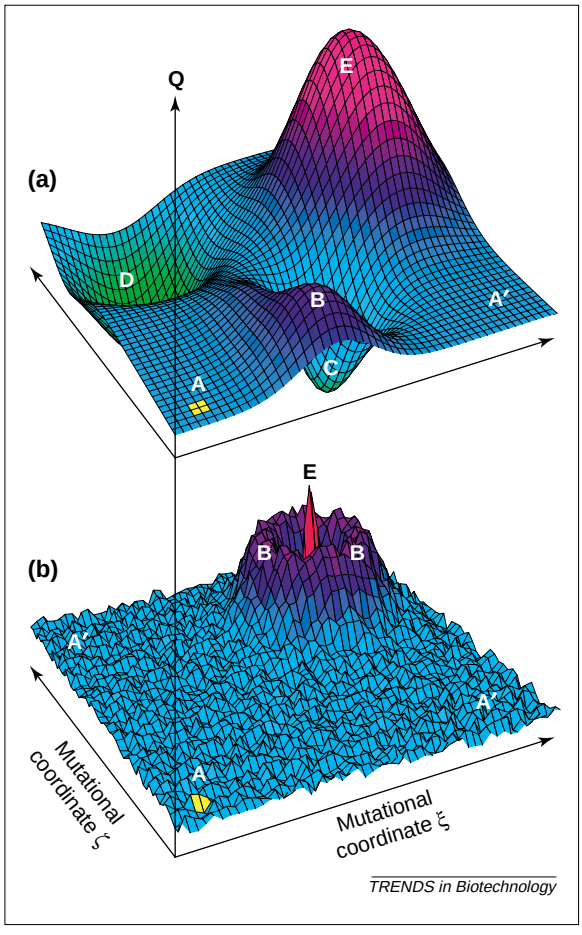
The application explored here requires mutations that optimize five variables simultaneously: the formation time of O state (minimize), the decay of

Table 1. Comparison of the five methods of optimizing proteins for device applications using genetic engineering

Technique methodology	Site-directed mutagenesis	Semi-random mutagenesis	Random mutagenesis	Directed evolution Type I	Directed evolution Type II
Efficiency	∝ modeling	Medium	Medium–High	High	Highest
Specificity	Single residue	Local region	Global *bop* gene	Colony or protein	*Halobacterium salinarum* cell
Variability generated	Low	Medium	High	N/A	N/A
Primary use	Fine-tuning	Narrowing in on optimal mutant	New search	Protein/colony screening	Organism selection
Requires	Modeling, or *a priori* knowledge	Some modeling	No *a priori* knowledge	High throughput and sensitive screening	Truncation selection

The efficiency of a method is a qualitative measure of how much operator effort and intervention is required to generate a successful mutation. A highly efficient method is one that requires little, if any, operator intervention during the optimization process. Type II directed evolution is the best choice in this regard. The specificity refers to the selectivity available with respect to modification of individual residues at specific locations within the primary structure. Highest specificity (single residue, specific replacement) is provided by site-directed mutagenesis. The variability generated parameter is a qualitative measure of the control one has (or perhaps loses) in going from a site-specific to a more randomized modification of the protein. Semi-random and random mutagenesis produce the highest degree of variability, but in some cases might provide too many degrees of freedom. Directed evolution methods provide no control over the amount of variability generated. The primary use parameter is self-explanatory but only one example for each method is provided and entries in this row should not be viewed as mutually exclusive.

O state (optimize), the quantum efficiency of the O to P photochemical transformation (maximize), the efficiency of the P to Q hydrolysis (maximize) and the lifetime of the Q state (maximize). It is virtually impossible to predict which single or double mutations will accomplish these tasks simultaneously. Complex systems such as the memory proposed here have too many variables to limit optimization to site-directed mutations. Nevertheless, one variable of importance to the memory was optimized based on modeling. A series of mutations were constructed in-house involving Glu194 and Glu204, both of which have been shown to be important participants in the photocycle. O state lifetimes and *Q* values for a sample of mutants are shown in Figure 4.



Screening versus selection

All mutational strategies used to create protein variants rely on a screening or a selection process to achieve the goals. Screening involves inspection of a population for a given characteristic (usually a phenotype of interest) but places no limits on the viability of the organisms that do (or more importantly do not) possess that given characteristic. A simple example of screening would be picking out only the gold coins from a mixed assortment of coins. The population is being screened for gold coins but all other phenotypes still exist (i.e. copper, nickel, silver). In selection, there is a predetermined mechanism that allows only a certain population to survive. An example of selection

Fig. 3. Hypothetical mutational landscapes for fluid traits (a) and spectrokinetic optimizations (b). The native protein is located at the central locus of the yellow squares, the mutational coordinates ξ and ζ are arbitrary, and the vertical axis measures an arbitrary *Q* factor. Starting at 'A', the ultimate goal of our mutational strategies is to discover the optimal protein, represented by peak 'E'. The 'A' and 'A'' regions represent 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 labeled 'E'. Cyclical optimization will typically find 'E' in the fluid case, but is unlikely to find 'E' in the spectrokinetic optimization without good modeling or operator intuition. For fluid characters such as temperature, less variation is expected (Fig. 3a). The overall temperature stability of the protein might decrease (troughs 'C' or 'D'), but has a comparable chance of increasing (peaks 'B' and 'E'). Figure 3b shows the mutational landscape for more complex characteristics such as spectrokinetic properties, in which most of the performance is localized in specific regions of the protein. The mutational surface is probably a rugged landscape with numerous local minima.

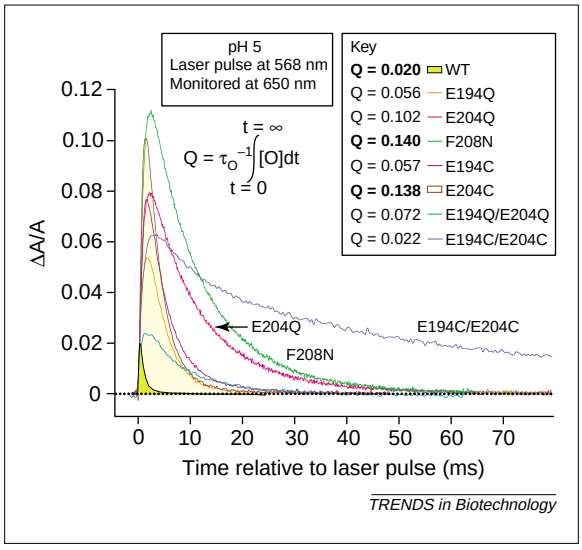


Fig. 4. The effect of mutations involving Glu194 and Glu204 on O state kinetics, measured at pH 5.0. Wild-type (WT) protein is shaded in green and has the smallest *Q* value of all proteins studied. The O state lifetime in WT is 6–8 ms. The O state lifetimes of E194Q and E204Q are 70 ms and 125 ms, respectively. The double mutant E194C/E204C has the longest O state lifetime reported yet, with a lifetime of ~1 s. This represents more than a 100-fold increase in the O state decay time. However, E204Q has the largest *Q* value.

can be found in construction of site-directed mutants. Only those organisms that possess the plasmid that confers novobiocin resistance will live and persist in the population. To follow the same example used earlier, if a coin sifter is used to allow all coins except the desired coin type to fall through, the desired coins have been selected for. It should be evident from this discussion that selection is the more efficient (and technically more difficult) of the two processes. The main advantage of selection is that smaller sample sizes need to be examined to detect an optimization.

Random mutagenesis

The next step in the genetic optimization process is to create random mutations and observe the impact of these mutations on the desired properties. This process requires screening of mutations and determining which should be used as templates for further exploration. If one is very lucky, a single random-mutation cycle can yield a significantly improved protein. More commonly, four or more cycles are required to achieve a significant improvement. Random mutations are more likely to be destructive or neutral than constructive. The total number of unique mutations for a protein the size of BR (248 residues) is 4712 single mutations and 11 056 708 double mutations. It is clear that a single mutagenesis cycle might well produce nothing of value. In the next section methods of restricting the location of random mutations to improve the probability of success will be examined.

There are many available methods to randomly mutate BR (for examples see [27–29]). The net result is a population of mutant proteins with no *a priori* bias of location in the sequence. A combination of methods can be used to generate controlled numbers of randomly distributed mutations.

Semi-random mutagenesis

Semi-random mutagenesis essentially combines site-directed and random mutagenesis to achieve high

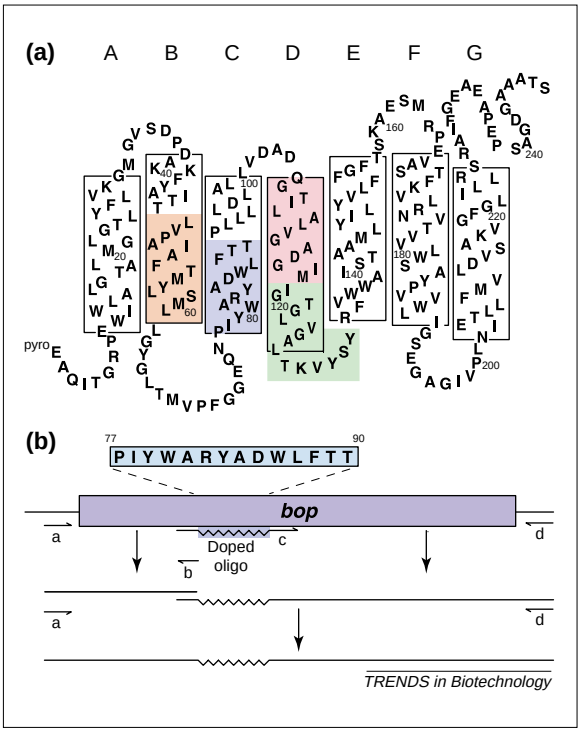


Fig. 5. Semi-random mutagenesis. (a) Secondary structural model of bacteriorhodopsin (BR) indicating four regions targeted for semi-random mutagenesis (*orange, blue, pink and green boxes)*. Amino acids 77 to 90 in helix C (*blue shading*) were targeted in initial studies. (b) Construction of mutant libraries by two-step PCR with a doped oligonucleotide. Mutagenesis of the region of the *bop* gene (*purple*) encoding amino acids 77–90 (*blue*) is shown. In the first step, two PCR products are generated from the *bop* template using primers 'a' and 'b' or 'c' and 'd'. Oligonucleotide primer 'c' is doped at a level of 20% within the region encoding amino acids 77–90 (*jagged line*). In the second step, the products of the first step, which overlap, are combined and amplified with primers 'a' and 'd'. The resulting PCR product is cut with restriction enzymes and ligated with vector fragments to generate an intact *bop* gene bearing the mutated region.

mutagenesis efficiencies over a limited region. The approach was pioneered by Sauer and co-workers in structure–function studies of soluble proteins [30], and has been used by others to examine membrane protein folding and function [31]. Krebs and co-workers use semi-random mutagenesis to determine the structural features that guide BR biogenesis in *H. salinarum* [32,33]. But this method also has significant advantages for bioelectronic optimizations, as described. In this technique, the complete *bop* sequence is divided into 17 segments of ~15 targeted amino acids that are mutated at a high rate without disturbing the surrounding sequence. Mutations are created by PCR using a 'doped' primer that is synthesized with a mixture containing 80% of the wild-type nucleotide and 6.6% of each non-wild-type nucleotide at each position within the targeted region. At this doping level, approximately five amino acid substitutions are predicted per mutant, based on Monte Carlo calculations [31]. As shown in Fig. 5, the doped oligonucleotide is combined with additional primers in a two-step PCR reaction. The resulting mutant *bop* fragment is digested with restriction enzymes and ligated with plasmid fragments to generate a plasmid

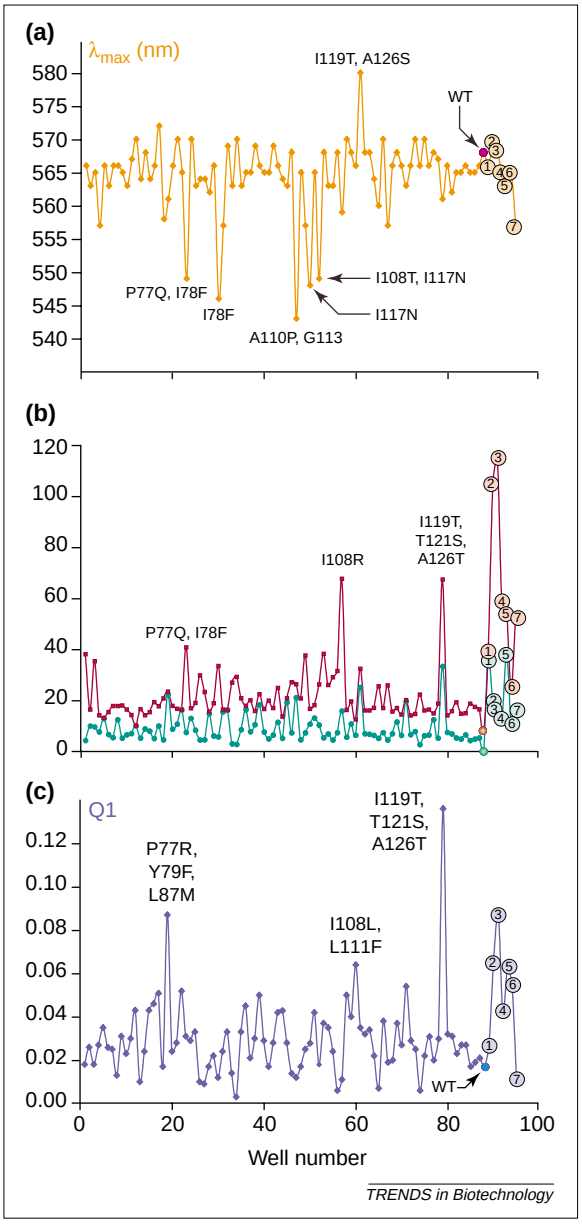


Fig. 6. Spectrokinetic screening of semi-random mutants for branched-photocycle memory applications. Semi-random mutant proteins were constructed and analysed in 96-well plate format. Wild-type protein and seven site-directed mutants (labeled 1–7 and positioned to the right of wild-type) are shown for comparison. 1 = E194Q; 2 = E204Q; 3 = F208N, 4 = E194C, 5 = E204C, 6 = E194C/E204C; 7 = E194Q/E204Q. Some of the semi-random mutant extrema are labeled. (a) The λ_{max} of mutant proteins and WT. (b) The O state intermediate (as measured at 650 nm) decay and formation times for the mutant proteins and WT. Formation times are shown in green, decay times in red. (c) The *Q* value for each mutant (see text for discussion). Note that some of the semi-random mutants show higher values in some categories than do the site-directed mutants designed from *a priori* knowledge of the protein.

library of at least 3000 mutant *bop* genes in *E. coli*. The pooled library is transformed into an *H. salinarum* strain developed specifically for high-efficiency *bop* mutagenesis [23], and recombinant colonies containing a single copy of the mutant *bop* genes are isolated by homologous gene replacement [23,33].

Because the mutations within the targeted region are random, the recombinant colonies exhibit a range of colours, reflecting a variation in BR spectral properties and expression levels among the mutants. The current analysis is limited to purple colonies. More than one hundred purple mutants in four regions of BR have been isolated (Fig. 4) and sequenced at the nucleotide level. An average of 1.2 amino acid substitutions is observed among purple mutants. This value is lower than the expected value of five substitutions per mutant, presumably because a higher number of substitutions interferes with formation of native BR.

To identify variants with potentially improved photonic properties, the purple membrane from semi-random mutants was partially purified in 96-well format by repeated ultracentrifugation. The BR in these samples was then screened for variation in spectral characteristics and photocycle intermediate lifetimes, particularly that of the O intermediate (Fig. 6). UV-visible spectra, O state formation and O decay times were collected for each mutant and the *Q* value was calculated (see Fig. 4). Despite the fact that these mutants were not developed *a priori*, some of them show *Q* values higher than those for site-directed mutants designed using modeling methods and structural information. The triple mutant I119T/T121S/A126T shows a *Q* value that is ~50% higher than the largest *Q* value for any site-directed mutant. Site-directed mutagenesis is currently being used to dissect the contribution of particular residues within this multiple mutant.

Directed evolution

One of the most interesting developments in materials research is the recent use of directed evolution and combinatorial methods to optimize materials [34–37]. One starts with the premise that nature has already produced some sophisticated materials through evolution. The challenge in using directed evolution for materials optimization is establishing a selection method that focuses on the desired properties of the material generated by the host.

The use of directed evolution cannot guarantee the creation of the ultimate material because the number of possible mutations and the time it would take to explore all the possibilities is well beyond that available in any grant cycle or even scientific career. But directed evolution does provide a method of exploring a large number of possible mutations in a systematic way which yields the highest probability of improving the properties of a biological material. We believe that this technique is potentially the single most important development in materials research in the past century.

Directed evolution can be divided into Types I and II, neither of which are discrete but rather represent a continuum of procedures that start with screening and end with selection. The major difference between Type I and Type II directed evolution is the level at which the screening or selection is implemented. Type I includes protein screening at the colony or protein level, whereas Type II includes screening or selection of BR in individual cells. Both methods entail identifying proteins with the highest *Q* values and using these mutants in the next round of screening.

Type I directed evolution involves screening of large numbers of colonies using a multi-well plate, or other high-throughput screening method. Because *H. salinarum* cells burst when placed in water, colonies can be placed in small aliquots of water and screened in the same 96-well plate format as would be done with isolated protein. Direct screening on colonies of *H. salinarum* avoids the need for protein

isolation but at the cost of introducing more scattering to the sample.

Type II directed evolution deals with detection of optimized forms at the microscopic organismal level and represents the most powerful technique. Screening and selection are used in tandem. A population of *H. salinarum* cells passes through an automated screening device consisting of laser diodes and a charge couple device (CCD) detector array. This device is essentially identical to the write process in the memory. The cells that show the most conversion to P and Q states (as monitored by a concomitant loss of O state) are separated from other cells and shunted to a collection vessel. The *bop* gene sequence from these cells is obtained using *in vivo* PCR, and favourable mutations are used for the next round of optimization.

Type II directed evolution is more efficient than Type I but is more difficult to implement. Because individual organisms are being monitored, cell to cell variation is problematic. In Type I directed evolution this variation is accounted for by looking at thousands of organisms or protein patches simultaneously. However, the rewards for Type II directed evolution are superior; this method provides the greatest possibility for high-throughput detection of the mutant that is best optimized for device applications.

Mutational strategies

The extent to which the variation in a population increases is dictated by the choice of mutagenesis method. Site-directed mutagenesis is used to explore small changes by changing one residue, and is used to fine-tune a particular mutant. For example, existing site-directed mutants can be combined with mutants found in Type I and Type II screening and selection. Current efforts include incorporation of the Glu194 and Glu204 mutants with the triple mutant I119T/T121S/A126T in an attempt to combine favourable properties from different sources.

Some techniques are designed to explore a greater area of the mutational landscape for a given protein than others. For semi-random mutagenesis, the mutational space explored is greater compared to site-directed mutagenesis. A new optimum might be reached instead of continuing to improve an original optimization. Random mutagenesis presents simultaneous advantages and disadvantages: it can be used to find new regions in the protein for optimization, but at the cost of neglecting the original optimization unless randomized libraries incorporating the desired mutation are used. Strategies for optimizing photochemical properties must take into account the localized nature of the mutational landscape (e.g. Fig. 3b). Once the key regions are discovered, semi-random mutagenesis provides the most productive approach.

Summary

Although evolution has optimized proteins for many characteristics relevant to device application, further optimization is usually desirable. In this article, five different genetic engineering approaches have been examined, ranging from single residue replacement through randomized replacement to directed evolution. Our discussion centred on the current problem of optimizing BR for a 3D memory system. This problem is challenging because the performance of the protein in this environment requires the simultaneous optimization of three to five variables. Modeling does not describe photochemical processes at a high enough level to permit the use of site-directed methods as the sole approach. Site-directed and semi-random mutagenesis have been used to achieve a 700-fold improvement in the performance of the protein in the 3D memory. Directed evolution offers the best potential for materials improvement, although the difficulty of *in vivo* measurements of photochemical properties remains problematic and leads to inefficient screening. A major emphasis of our current research is to optimize the speed and accuracy of *in vivo* screening. Although the discussion has concentrated on BR, the methods described herein are relevant to the optimization of any protein for device application. A key requirement is to determine the nature of the mutational landscape and choose a technique that is optimal for exploring a global or localized function. We anticipate that rapid improvements in the methods, procedures and equipment available for genetic manipulation and screening will lead to significant advances in bioelectronic materials during the next decade.

Acknowledgements
We thank Jason R. Hillebracht, Lavoisier Ramos, Duane Marcy and Richard Needleman for significant contributions to the research, and the NSF (EIA-0129731) and NIH (GM-34548) for support.

References

1 Vsevolodov, N.N. (1998) *Biomolecular Electronics. An Introduction via Photosensitive Proteins*, Birkhauser
2 Birge, R.R. (1990) Photophysics and molecular electronic applications of the rhodopsins. *Annu. Rev. Phys. Chem.* 41, 683–733
3 Lee, I. *et al.* (1997) Biomolecular electronics: vectorial arrays of photosynthetic reaction centers. *Phys. Rev. Lett.* 79, 3294–3297
4 Birge, R.R. *et al.* (1999) Biomolecular electronics: protein-based associative processors and volumetric memories. *J. Phys. Chem. B.* 103, 10746–10766
5 Vsevolodov, N.N. and Poltoratskii, V.A. (1985) Holograms in biochrome, a biological photochromic material. *Sov. Phys. Tech. Phys.* 30, 1235
6 Hampp, N. (2000) Bacteriorhodopsin: mutating a biomaterial into an optoelectronic material. *Appl. Microbiol. Biotechnol.* 53, 633–639
7 Millerd, J.E. *et al.* (1999) Improved sensitivity in blue-membrane bacteriorhodopsin films. *Opt. Lett.* 24, 1355–1357
8 Juchem, T. and Hampp, N. (2000) Interferometric system for non-destructive testing based on large diameter bacteriorhodopsin films. *Optics and Lasers in Engineering* 34, 87–100
9 Hampp, N. and Juchem, T. (2000) Fringemaker – the first technical system based on bacteriorhodopsin. In *Bioelectronic Applications of Photochromic Pigments* (Vol. 335) (Dér, A. and Keszthelyi, L., eds), pp. 44–53, IOS Press
10 Tallent, J.R. *et al.* (1996) Effective photochromic nonlinearity of dried blue-membrane bacteriorhodopsin films. *Opt. Lett.* 21, 1339–1341
11 Tallent, J.R. *et al.* (1998) Photochemistry in dried polymer films incorporating the deionized blue membrane form of bacteriorhodopsin. *Biophys. J.* 75, 1619–1634
12 Birge, R.R. *et al.* (1997) Protein-based three-dimensional memories and associative processors. In *Molecular Electronics* (Ratner, M.A. and Jortner, J., eds), pp. 439–471, Blackwell Science Ltd
13 Druzhko, A. *et al.* (1995) 4-keto bacteriorhodopsin films as a promising photochromic and electrochromic biological material. *BioSystems* 35, 129–132

14 Hampp, N. *et al.* (1992) Diffraction efficiency of bacteriorhodopsin films for holography containing bacteriorhodopsin wildtype BRwt and its variants BR_{D85E} and $BR_{D96N.}$ *J. Phys. Chem.* 96, 4679–4685

15 Lanyi, J. and Luecke, H. (2001) Bacteriorhodopsin. *Curr. Opin. Struct. Biol.* 11, 415–419

16 Lanyi, J. and Pohorille, A. (2001) Proton pumps: mechanism of action and applications. *Trends Biotechnol.* 19, 140–144

17 Rammelsberg, R. *et al.* (1998) Bacteriorhodopsin's intramolecular proton-release pathway consists of a hydrogen-bonded network. *Biochemistry* 37, 5001–5009

18 Zscherp, C. *et al.* (2001) Time-resolved FT-IR spectroscopic investigation of the pH-dependent proton transfer reactions in the E194Q mutant of bacteriorhodopsin. *Biochem. Biophys. Res. Commun.* 283, 57–63

19 Popp, A. *et al.* (1993) Photochemical conversion of the O-intermediate to 9-*cis*-retinal-containing products in bacteriorhodopsin films. *Biophys. J.* 65, 1449–1459

20 Gillespie, N.B. *et al.* Characterization of the branched-photocycle intermediates P and Q of bacteriorhodopsin. *J. Phys. Chem.* (in press)

21 Vought, B.W. and Birge, R.R., eds (1999) *Molecular Electronics and Hybrid Computers*, Wiley-Interscience

22 Ni, B. *et al.* (1990) An efficient system for the synthesis of bacteriorhodopsin in *Halobacterium halobium*. *Gene* 90, 169–172

23 Peck, R.F. *et al.* (2000) Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Mol. Microbiol.* 35, 667–676

24 Lukashev, E.P. and Robertson, B. (1995) Bacteriorhodopsin retains its light-induced proton-pumping function after being heated to 140°C. *Bioelectrochem. Bioenerg.* 37, 157–160

25 Tajkhorshid, E. *et al.* (2000) Molecular dynamics study of the nature and origin of retinal's twisted structure in bacteriorhodopsin. *Biophys. J.* 78, 683–693

26 Spassov, V. *et al.* (2001) pK(a) calculations suggest storage of an excess proton in a hydrogen-bonded water network in bacteriorhodopsin. *J. Mol. Biol.* 312, 203–219

27 You, L. and Arnold, F. (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* 9, 719

28 Wan, L. *et al.* (1998) *In vitro* evolution of horse heart myoglobin to increase peroxidase activity. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12825–12831

29 Callahan, M. and Jerpseth, B. (1996) An efficient random mutagenesis technique using an *E. coli* mutator strain. *Methods Mol. Biol.* 57, 375–385

30 Bowie, J.U. *et al.* (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* 247, 1306–1310

31 Wen, J. *et al.* (1996) Exploring the allowed sequence space of a membrane protein. *Nat. Struct. Biol.* 3, 141–148

32 Krebs, M.P. *et al.* (1997) Intramembrane substitutions in helix D of bacteriorhodopsin disrupt the purple membrane. *J. Mol. Biol.* 267, 172–183

33 Krebs, M.P. *et al.* (1993) Gene replacement in *Halobacterium halobium* and expression of bacteriorhodopsin mutants. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1987–1991

34 Whaley, S.R. *et al.* (2000) Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature* 405, 665–668

35 Anni, H. *et al.* (2001) Selection of phage-display library peptides recognizing ethanol targets on proteins. *Alcohol* 25, 201–209

36 Brust, M. and Kiley, C.J. (2002) Some recent advances in nanostructure preparation from gold and silver particles: a short topical review. *Colloids Surf. A Physicochem. Eng. Aspects* 202, 175–186

37 Mulholland, S.E. *et al.* (1999) Determination of nonligand amino acids critical to [4Fe-4S]2+/+ assembly in ferredoxin maquettes. *Biochemistry* 38, 10442–10448

38 Luecke, H. *et al.* (1999) Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* 291, 899–911