# Exploring the Effects of Proteorhodopsin on the Physiology of Native and Heterologous Hosts

by

Michael Valliere

Bachelor of Science, Biochemistry and Molecular Biology
Bachelor of Business Administration, Business Management
University of Massachusetts Amherst, 2009

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Environmental Biology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016





Signature of Author: Signature redacted

Michael Valliere
Department of Civil and Environmental Engineering

Certified by: Signature redacted

Edward F. DeLong
Department of Civil and Environmental Engineering and Biological Engineering

Accepted by: Signature redacted

Jesse Kroll
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Exploring the Effects of Proteorhodopsin on the Physiology of Native and Heterologous Hosts

by

Michael Valliere

Submitted to the Department of Civil and Environmental Engineering on August 18, 2016 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Environmental Biology

## ABSTRACT

Photoheterotrophic microbes use proteorhodopsin (PR) and other types of microbial rhodopsin photosystems to harness energy directly from sunlight. This thesis explores the photophysiology of PR in the context of a natural marine isolate, *Dokdonia* sp. strain MED134, and by establishing *Escherichia coli* as a heterologous host to experimentally determine conditions under which the PR photosystem is capable of enhancing growth rate and yield. In *Dokdonia* sp. MED134, PR and 11 other genes were discovered to be significantly induced by light but the induction of these genes increased growth rate and yield in an extremely carbon-limited environment; in richer media the induction of the genes by light had an inhibitory effect on growth. In *E. coli*, various genetic backgrounds were tested with PR expression, and one was discovered to exhibit slightly higher cell yields presumably as a result of light-driven proton pumping by PR. This work illustrates that PR is part of orchestrated response to light in a PR-containing isolate, but can also influence the growth of a heterologous host on its own. Further refinement of the genetic background of *E. coli* should unlock the full potential of PR as a cellular energy source for biotechnological applications—at least within this organism.

Thesis Supervisor: Edward F. DeLong
Title: Professor

Table of Contents

## I. The Ecological and Physiological Significance of Proteorhodopsin

### Biological light energy capture

Sunlight is—and always has been—an essential resource for life on Earth. The constant bombardment of sunlight energized the abiotic Earth long before life existed, and along with geothermal processes, catalyzed the formation of the reduced chemicals that served as electron donors to the first life forms. Hundreds of millions of years later, evolved versions of these early life forms gained the ability to directly harness the energy within sunlight by converting it to useful cellular energy, and this capability set Earth and its inhabitants on the trajectory that has given rise to the planet we experience today.

#### *Photosynthesis*

In its present form, life on Earth has invented and refined *only* two distinct systems for light mediated biochemical energy capture (Bryant and Frigaard, 2006). Chlorophyll-based photosynthesis is the better-known system, leveraged by organisms ranging from microscopic cyanobacteria to large multicellular plants. The many different forms of chlorophyll-based photosynthetic systems all work by exciting electrons to higher energy states within reaction centers using energy captured from light. Chlorophylls, bacteriochlorophylls, and other photosynthetic pigments (that can be structured in many different ways within cells) directly capture the energy from light and transfer the energy to the reaction centers. The electrons that become excited at the reaction centers come from water in the case of oxygenic photosynthesis, or may originate from a number of

different inorganic and organic compounds in anoxygenic photosynthesis. In either case, the electrons that become excited at the reaction centers ultimately provide reducing power and proton motive force (PMF) for ATP production.

The differences between photosynthetic systems are a result of evolution within discrete lineages of organisms that each adapted to a unique environment. Photosynthetic microbes, for example, are capable of using $H_2O$, $H_2S$, $H_2$, or reduced carbon compounds as electron donors for photosynthesis (Bryant and Frigaard, 2006); while water has the advantage of always being available, it's thermodynamically less favorable as an electron donor compared to more scarce compounds like hydrogen sulfide and biological hydrogen that can only be found in abundance in particular environments. The chlorophyll content within the reaction centers and antenna systems provides another example of divergent evolution with environmental selection among photosynthetic microbes. While photosynthetic microbes commonly employ bacteriochlorophylls a and b, there are over 100 known chlorophyll types (Scheer, 2006), with the most recent discovery of a chlorophyll (chlorophyll f) reported in 2010 (Chen et al., 2010). The different types are primarily distinguished by their primary structure, either a bacteriochlorin or phytochlorin ring, and the degree of saturation of the macrocycle, which gives each chlorophyll its characteristic spectral features (Scheer, 2006). In general, the absorption peaks of chlorophyll are between 400-500 nm and 600-700 nm. Chlorophyll f was found to be unique structurally, but also because its second adsorption peak extended beyond that of chlorophyll d into the infrared region (>700 nm) of the light spectrum (Chen and Blankenship, 2011).

*Rhodopsin photosystems*

Rhodopsin photosystems encompass the only other known biological system for light mediated biochemical energy capture. These photosystems are fundamentally different than chlorophyll-based photosystems because the electron excitation that occurs upon absorption of photons is confined to the retinal (or an accessory) chromophore embedded within the rhodopsin photosystem protein. The energy captured from light is directly absorbed by the chromophore, exciting electrons within the molecule in such a way that its conformation with respect to the rhodopsin photosystem protein changes and begins a photocycle that ultimately can result in the translocation of a proton through the rhodopsin photosystem protein or signal transduction.

The types of rhodopsin photosystems capable of generating a chemiosmotic electrochemical gradient across the cytoplasmic membrane of cells by light-mediated proton pumping is analogous to the potential generated across the thylakoid membrane during oxygenic photosynthesis, or the cytoplasmic and mitochondrial membranes that contain respiratory chains. Perhaps the most interesting facet of this form of energy conservation is that the process operates independent of complex electron transfer; thus, while light-driven proton pumping generates PMF, it does not generate reduction potential for cellular metabolism (Bryant and Frigaard, 2006).

The light-driven proton pumps are classified as microbial rhodopsin photosystems, or type I rhodopsin photosystems, and are described in detail in the following sections. Type II rhodopsins share a fundamentally conserved structure with the type I rhodopsins

(Kouyama and Murakami, 2010), but are differentiated by primary sequence alignment and by specific features of their secondary and tertiary structures (Spudich et al., 2000; Kouyama and Murakami, 2010). The type II rhodopsins are exclusively found in animals and other high level eukaryotes, where they serve as a photoreceptor in animal eyes (including the rod and cone pigments of humans), as well as the hypothalamus, pineal gland, and other tissues of non-mammalian vertebrates (Spudich et al., 2000; Kouyama and Murakami, 2010).

All types of rhodopsin photosystems use a photoactive pigment derived from carotenoids called retinal. Although retinal is bound in different confirmations within type I and type II rhodopsin photosystems (Spudich et al., 2000; Kouyama and Murakami, 2010), all of the natural rhodopsins that have been described absorb blue, green, and yellow light with an absorption peak ranging from 437 nm to 590 nm (Govorunova et al., 2013; Klapoetke et al., 2014; Engqvist et al., 2015). Peak absorbance for type I proton pumping microbial rhodopsin photosystems vary from 490 nm to 568 nm (Engqvist et al., 2015). Recently, the absorption range of a type I proton pumping microbial rhodopsin photosystem with natural peak absorption at 538 nm was expanded by the directed evolution of residues within the protein to shift the peak absorption hyposochromically to 458 nm and bathochromically to 619 nm; however, only variants with peak absorbance within the range of 490 nm to 560 nm retained the ability to pump protons (Engqvist et al., 2015). Another group of researchers recently established that retinal analogues can be used to shift the absorbance of proton-pumping rhodopsin photosystems as well, and their effect adds to a mutation induced spectral shift (Ganapathy et al., 2015).

*Rhodopsin photosystems fill the gap in PAR*

The microbes performing photosynthesis and leveraging rhodopsin photosystems are found in the same environments, and sometimes even within the same cells (Tsunoda et al., 2006; Slamovits et al., 2011). Considering a functional absorption range of 490 nm to 560 nm, the proton pumping rhodopsins capture light that is typically out of reach of chlorophyll-based photosynthetic systems that efficiently harvest visible light of wavelengths less than 500 nm and above 600 nm. As shown in Figure 1, the gap in photosynthetically active radiation (PAR) ensures that an abundance of light is available to the organisms that leverage proton pumping rhodopsin photosystems.

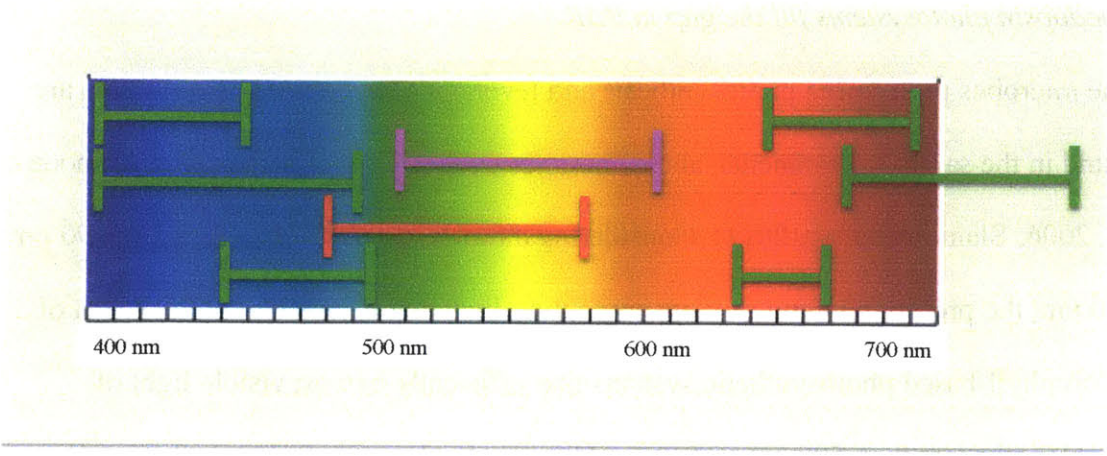



**Figure 1.** Rhodopsin photosystems fill the gap in PAR.

Rhodopsin photosystems absorb different wavelengths of light than photosynthetic systems, thus filling the gap in photosynthetic active radiation (PAR). The red and purple lines represent the absorbance range of the microbial rhodopsin photosystems PR and BR, respectfully. The green lines represent the absorption range of chlorophylls a, b, and d (top to bottom), which represent the strongest areas of PAR on the visible spectrum. Limitations associated with the different chromophores (chlorophyll and retinal) in the context of the light harvesting system may explain why each system is adapted to capture different wavelengths of light.

When considering the evolution of the two types of phototrophy, it is certainly more plausible that the much simpler rhodopsin photosystems emerged first. Since the retinal-based rhodopsin photosystems are limited in terms of their light harvesting capability, the abundance of visible light of more extreme wavelengths (than can be captured with retinal) may have favored the selection of chlorophyll pigments in the reaction centers and antenna systems.

It's also possible that ancestors of the first photosynthetic organisms contained chlorophyll for the purpose of reflecting green light. Assuming that photoheterotrophs emerged before oxygenic photosynthetic autotrophs, perhaps autotrophic organisms gained the ability to attract photoheterotrophs by reflecting green light because it somehow increased their relative fitness by symbiotic or commensalistic cross-feeding (Sarmento and Gasol, 2012; Sharma et al., 2014; Aylward et al., 2015). If proven to be true, such a relationship would suggest cross-feeding between photosynthetic autotrophs and photoheterotrophs is a long-term strategy for success in the marine environment, and may help to explain why the two types of phototrophy rely on different wavelengths of light.

**Microbial rhodopsin photosystems**

Microbial rhodopsin photosystems are light absorbing integral membrane proteins found within all three domains of life (Béjà and Lanyi, 2014). They are comprised of the

chromophore retinal covalently bound by the formation of a Schiff base to the lysine residue of a retinylidene protein, which is also referred to as a rhodopsin or an opsin (Oesterhelt, 1976; Beja et al., 2000; Spudich et al., 2000). Absorbed light triggers a conformational change in the retinal chromophore, inducing a cycle of conformational changes that ultimately regenerate the original chromophore within milliseconds. Depending on the type of microbial rhodopsin, conformational changes in the opsin protein resulting from retinal isomeration events are leveraged either to pump ions across a membrane (ion pumping rhodopsin), or to induce a linked protein to respond to the conformational change within the opsin by initiating a cellular signal (sensory rhodopsin).

Figure 2 shows a three-dimensional representation of the structure of proteorhodopsin (PR), a proton pumping microbial rhodopsin, with retinal (in green) bound by a Schiff base (in purple) at K231 along with the key proton donor E108 (in yellow) and proton acceptor D97 (in magenta) (Dioumaev et al., 2002; Reckel et al., 2011). The two forms of microbial (type I) rhodopsins have similar topological features within membranes, characterized by a seven transmembrane alpha helix that forms a binding pocket to covalently link retinal deep within the protein. In both cases, the structure of the opsin protein influences the wavelengths of light that can excite the retinal chromophore (Beja et al., 2001; Bielawski et al., 2004; Engqvist et al., 2015). Both forms of rhodopsin can also pump ions, but sensory rhodopsins are considerably slower. The best way to distinguish between the two types is by measuring photocycle kinetics; pumping rhodopsins are characterized by fast photocycle kinetics on the order of 10 to 20 ms, compared to >300 ms for sensory rhodopsins (Oesterhelt, 1976).

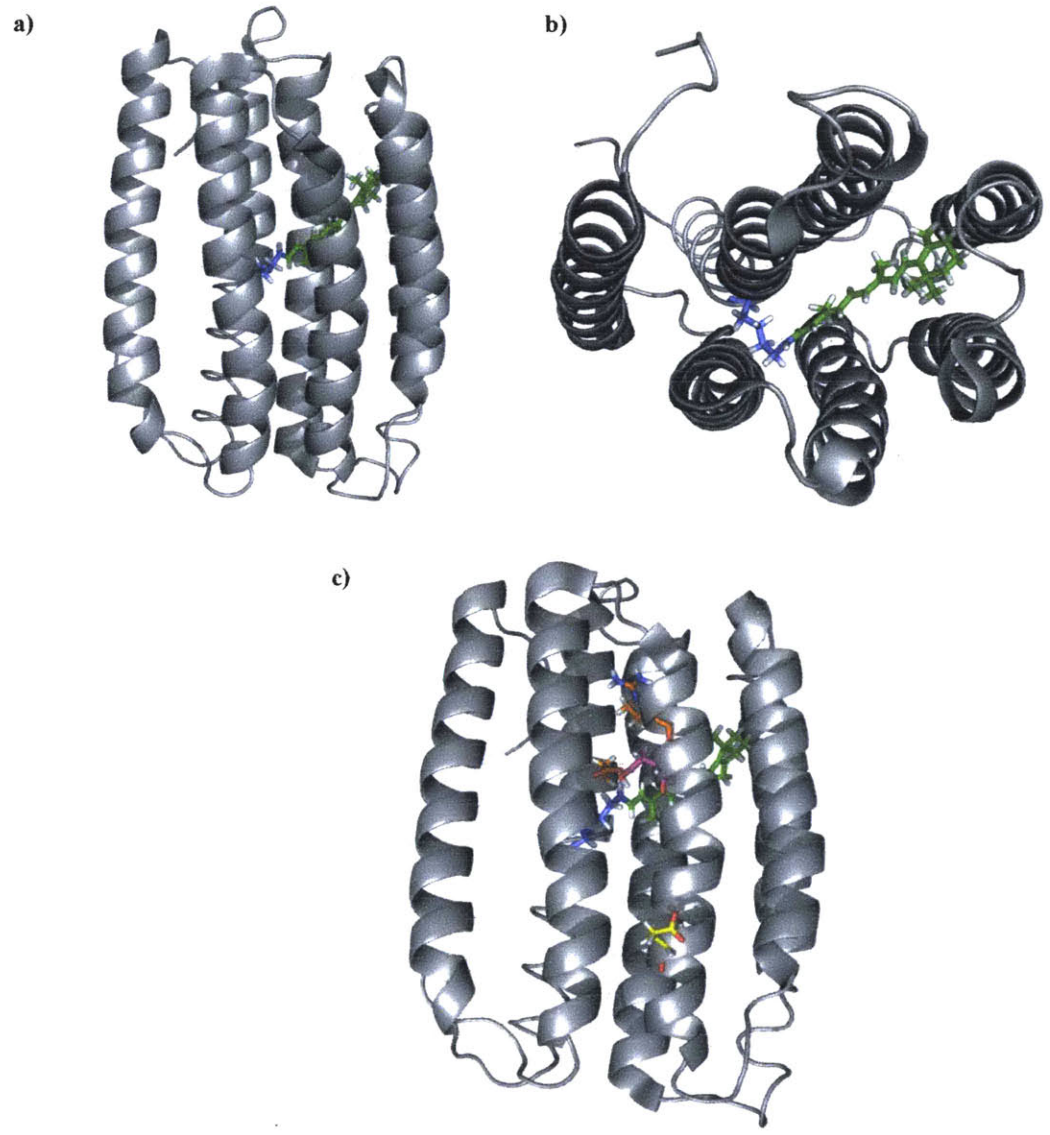



**Figure 2.** The structure of PR: a microbial proton pumping rhodopsin photosystem.

A three-dimensional representation of the structure of proteorhodopsin (PR) was created in PyMOL v.1.3 based on the solution NMR of EBAC31A08 available in the protein database under PDB 2L6X (Reckel et al., 2011). For structure a) the retinal in green bound by Schiff base in purple at K231. For b) the same features are observed from above the protein and looking down into the center. In c) the key proton donor E108 is represented in yellow and proton acceptor D97 in magenta. The counterions in orange are located in the retinal binding pocket and help to stabilize the Schiff base (Dioumaev et al., 2002; Reckel et al., 2011).

*Discovery of ion pumping rhodopsin photosystems*

The first ion-pumping rhodopsin was a proton pump discovered in the 1970s within *Halobacteria* that are commonly found in saline lakes and evaporating lagoons. The *Halobacteria* can reach high densities in these environments, and their membranes—composed of 75% bacteriorhodopsin (BR)—give the saltwater a reddish tint (Oesterhelt, 1976). Early work focused on defining the structure of the purple membranes and the functional properties of BR, with emphasis on probing photocycle kinetics, *in vitro* functional assays, and determining how BR function contributes to energy generation *in vivo*. To understand the effects of the photosystem *in vivo*, experiments were performed to measure pH of the media and ATP levels in the presence of light relative to darkness, with or without uncouplers and respiratory chain or ATPase inhibitors. In the presence of light, a change in pH of the media could be observed that was subsequently abolished in the absence of light (Oesterhelt, 1976). If ATP synthesis was uncoupled by addition of dicyclohexylcarbodiimide, the effect of proton pumping was more pronounced (Oesterhelt and Stoeckenius, 1971). When the uncoupler m-chlorophenylhydrazonomalononitrile was added, protons could diffuse back into the cells more quickly thus reducing the effect of proton pumping by BR (Blaurock and Stoeckenius, 1971). With respect to cellular ATP levels, light-dark experiments were conducted in the presence of cyanide to inhibit oxidative phosphorylation. Upon addition of KCN in a dark environment, ATP levels were reduced 80%; but when illuminated, ATP levels recovered to a level comparable to before the addition of KCN (Oesterhelt, 1976). Taken together, these experiments confirmed the proton pumping function predicted by photocycle kinetics and observed *in vitro*, and further illustrated the

functional significance of the BR photosystem in its ability to maintain ATP levels in the absence of oxidative phosphorylation.

Proteorhodopsin (PR), the first proton pumping rhodopsin originating from bacteria, was discovered by metagenomic analysis of a marine surface water sample at the turn of the millennium (Beja et al., 2000). The gene was found on a 130-kb bacterial artificial chromosome (BAC) that also encoded an rRNA operon from SAR86, an uncultivated gammaproteobacterium. The PR amino acid sequence was most similar to the BR gene from *Halobacteria*, but since no other genes from the BAC were similar to archaeal genes, it was assumed that the PR originated from SAR86. Once the PR gene was cloned into *E. coli* and over-expressed, functional assays confirmed that PR was in fact a light-driven proton pump. The results from Beja *et al.* clearly established a peak absorbance at 520 nm in PR-containing cells that was not present in control cells. This peak was also close to that of BR from purple membranes (560 nm), as was the rate of proton pumping. Finally, laser flash-induced absorbance changes in the cell suspensions of PR revealed an absorption difference spectrum characteristic of a type I ion pumping rhodopsin photosystem (Beja et al., 2000).

*Geography and Diversity of PR photosystems*

Since the discovery of the SAR86 PR gene, the evolution of next generation sequencing technology enabled the discovery of millions of novel PR-like microbial rhodopsin sequences. In fact, a combination of meta- genomic, transcriptomic, and proteomic surveys have established PR among the most abundant and highly expressed genes in

marine surface waters (Sabehi et al., 2005; Rusch et al., 2007; Campbell et al., 2008; Frias-Lopez et al., 2008; Poretsky et al., 2009; Morris et al., 2010; Wei, 2010). The PR gene can be found within most lineages of marine Proteobacteria (McCarren and DeLong, 2007), and PR-like genes are found in Actinobacteria (Sharma et al., 2008), Bacteroidetes (Gomez-Consarnau et al., 2007; Yoshizawa et al., 2012), Cyanobacteria (Sharma et al., 2006; McCarren and DeLong, 2007; Miranda et al., 2009), Plantomycetes (McCarren and DeLong, 2007), Firmicutes (Sharma et al., 2006), planktonic Euryarchaea (Frigaard et al., 2006) and eukaryotic dinoflagellates (Brown and Jung, 2006; Slamovits et al., 2011) as well. Other environments containing rhodopsin genes include freshwater (Atamna-Ismaeel et al., 2008; Sharma et al., 2009) and terrestrial (phyllosphere-associated) habitats (Atamna-Ismaeel et al., 2012).

Estimates of PR-containing genomes range anywhere from as low as 10% in Mediterranean Sea surface waters (Sabehi et al., 2005) to as high as 60% in the Sargasso Sea (Rusch et al., 2007; Campbell et al., 2008). The surveys demonstrated that abundance of most PR types (number of OTUs) was correlated with geographic location, specifically between costal and oligotrophic waters (Campbell et al., 2008; Morris et al., 2010; Wei, 2010). At any given site, a division was observed between widespread taxon-specific PR types (such as SAR11-PR) and unique, location-specific PR types originating from diverse lineages (Frigaard et al., 2006; McCarren and DeLong, 2007; Campbell et al., 2008; Cottrell and Kirchman, 2009; Morris et al., 2010; Wei, 2010). It is common to find PR genes "spectrally tuned" to match available light (Beja et al., 2001; Bielawski et al., 2004), but neither light intensity nor nutrient availability metadata explain why most

PR genes were more closely related to the environment they originated from than the organisms that possessed them (Campbell et al., 2008; Morris et al., 2010; Wei, 2010).

The existence of many unique and geographically distinct PR types originating from diverse taxa is cited as evidence of horizontal gene transfer (HGT) and strong positive selection for the phototrophy trait (Sabehi et al., 2005; Frigaard et al., 2006; McCarren and DeLong, 2007). As observed for horizontally transferred genes associated with the human microflora (Smillie et al., 2011), it is possible that local ecologies are shaping the PR phylogeny by selecting for the transfer of specific advantageous PR alleles between taxa. HGT is also expected to occur relatively frequently because the PR photosystem contains a minimum of only six genes (Martinez et al., 2007) that are commonly organized as an operon (McCarren and DeLong, 2007); or only a single gene in the case of an organism that already makes retinal or is capable of scavenging retinal from its environment. Regardless, it is apparent from the abundance of different organisms that incorporate PR into their metabolic strategy that PR is advantageous to genotypically and phenotypically different hosts from different ecologies, suggesting that energy from PR phototrophy can be leveraged for different physiological purposes (Fuhrman et al., 2008).

**Physiological effects associated with PR photosystems**

PR and other variations of proton pumping microbial rhodopsin photosystems directly increase the PMF and acidify the periplasm. The secondary effects of proton pumping come from the stimulation of cellular components that interact with the PMF. These systems include—but are not limited to—the F-type ATP synthase, rotary flagellar motor, electron transport chain (ETC), active transport systems, and voltage-gated and mechanosensitive ion channels. Stimulating any of these systems can be expected to culminate in tertiary physiological effects such as lag times, growth rates, cell yields, biomass yields, or cell survival in stationary phase.

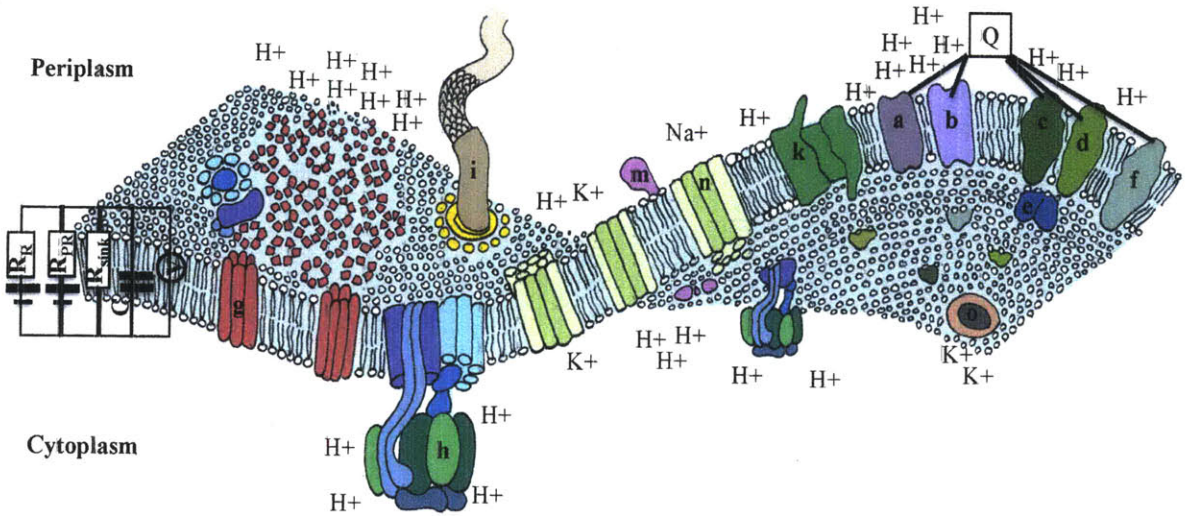



**Figure 3.** Representation of a PR-containing membrane with physiologically relevant PMF sources and sinks.

Various membrane components contribute to and draw from the PMF. The respiratory ETC (a-f) and proteorhodopsin (g) are the two major PMF sources, but the ATP synthase (h) can contribute by hydrolyzing ATP produced by substrate-level phosphorylation. The primary sinks include the ATP synthase (h), rotary flagella motor (i), voltage-gated and mechanosensitive channels (k), and active transporters like Ton-B (m), various symport and antiport proteins (n), and ATPases (o). The PMF can also be dissipated by chemical ionophores or by any mechanism that decreases the resistance of the membrane to protons. Protons are scattered across the membrane on both sides but will be more concentrated on the periplasmic side and even more so at the sources of proton pumping (Mulkidjanian et al., 2006). Some other cations are also included to illustrate transporter activities (divalent cations and other species are omitted for simplicity). The ETC is grouped in the membrane based on common supramolecular organization, with NDH-I (a) and NDH-II (b) forming a complex, and cytochrome $bo_3$ (c), cytochrome bd-I (d), and formate dehydrogenase (e) forming another complex (Sousa et al., 2011). Cytochrome *bd*-II (f) is left separate but performs a similar function as *bd*-I (Borisov et al., 2011). Q represents the quinone pool that is found within the membrane, but is depicted above to show the actors involved in its oxidation-reduction reactions. As indicated, the RC circuit proposed by Walter *et al* (2007) is represented on the diagram, with the respiration battery and its internal resistor $R_R$, the PR battery and its internal resister $R_{PR}$, the sink resistor in parallel with the batteries, C representing the membrane capacitance, and V representing the voltmeter.

*Natural isolates*

In accordance with the conclusions of survey experiments, all laboratory experiments conducted on marine isolates hosting PR photosystems suggest different uses of PR-derived energy (Fuhrman et al., 2008; Zubkov, 2009). Experimenting with marine isolates has provided evidence of photophosphorylation in Pelagibacter SAR11 (Giovannoni et al., 2005; Steindler et al., 2011), improved nutrient transport, growth rates, and cell yields in Flavobacteria (Gomez-Consarnau et al., 2007; Kimura et al., 2011; Gómez-Consarnau et al., 2016), and increased survival during—and recovery from—starvation conditions in Vibrio. In all cases, the effects linked to light utilization are only observed when the concentration of dissolved organic carbon (DOC) was extremely low and growth limiting.

While PR and other ion pumping rhodopsin photosystems are not known to be associated with carbon fixation, some researchers have suggested that PR stimulates the uptake of bicarbonate in Flavobacteria (Gonzalez et al., 2008; González et al., 2011; Palovaara et al., 2014). However, increased bicarbonate uptake in the presence of light could only be demonstrated for one strain, MED152, which did not experience light enhanced growth rates or cell yields (Gonzalez et al., 2008). Although light didn't seem to impact the growth phenotype in laboratory experiments, increased bicarbonate uptake in the presence of light does suggest that its PR photosystem impacts intracellular carbon fluxes. For a strain that demonstrates improved growth rates and yields in the presence of light, MED134, there is at least evidence that the *bicA* bicarbonate transporter gene is transiently induced by light (Palovaara et al., 2014). Since these photoheterotrophs have

the capacity to react bicarbonate and ATP with three carbon units (such as pyruvate and PEP) in order to supplement the pools of TCA cycle intermediates, it is possible that these Flavobacteria could fix a relatively small amount of inorganic carbon. And since these reactions require ATP and bicarbonate, light-driven proton pumping rhodopsins could accelerate the process of carbon fixation by raising the ratio of ATP to ADP and by increasing the active transport of bicarbonate.

*Heterologous hosts*

The optimal model system to study the physiological effects of PR would be a genetically tractable marine isolate that harbors a functional PR gene, expresses it constitutively or induces it in the presence of light, and shows evidence of improved growth when the cultures are exposed to light. As stated in the above section, increased growth rate and yields have been observed in the light, relative to dark, using the flavobacterium *Dokdonia sp.* MED134 (Gomez-Consarnau et al., 2007; Kimura et al., 2011); but attempts to establish a genetic system for this strain and other flavobacteria (Yoshizawa et al., 2012) have not yet been successful. Another option to consider was the Gammaproeteobacterium *Vibrio sp.* AND4 (Gomez-Consarnau et al., 2010); this strain has a functional PR and is genetically tractable, but (as stated in the section above) the PR seems to only benefit growth under extreme survival conditions or when recovering from survival conditions in media with an excessive amount of carbon and of higher osmolarity. Preliminary growth experiments with *Photobacterium angustum* S14, a phylogenetically similar strain (as AND4) that contains a functional PR, showed that the strain is highly susceptible to light upon entering stationary phase (results not shown).

The negative effect of the light on survival in stationary phase does not bode well for subsequent experimental work. Taken together, finding the appropriate marine isolate to serve as a model system has proven to be a substantial challenge, and there is no telling how quickly progress could be made on this front.

In lieu of a suitable native isolate, heterologous hosts represent an attractive alternative as a model system. Early experimental work has established that the PR photosystem is functional when heterologously expressed in *E. coli* (Beja et al., 2000; Martinez et al., 2007; Walter et al., 2007). Additionally, *E. coli* is well characterized with respect to its physiology and genetics, is relatively easy to experimentally manipulate, and grows extremely well in laboratory cultures. Finally, because PR is subject to extensive HGT in nature, using a heterologous system offers a unique opportunity to learn more about the immediate physiological effects of acquiring a new function, and subsequently, how evolutionary dynamics can be affected by acquisition of a new trait by HGT. For these reasons and in order to circumvent the challenges posed by studying marine isolates, *E. coli* should be further developed as a heterologous host.

Work in *E. coli* capable of heterologous PR expression has already confirmed some of the effects seen with PR-containing marine isolates, namely that expression of PRs and subsequent starvation of cells results in the ability to measure increases in proton pumping and ATP concentration of *E. coli* cell suspensions exposed to light (Beja et al., 2000; Wang et al., 2003; Martinez et al., 2007; Jung et al., 2008). Additional physiological effects associated with PRs expressed in *E. coli* include increased flagella

rotation in the presence of light (Walter et al., 2007) and increased hydrogen production in the presence of light by a strain co-expressing hydrogenase with PR (Kim et al., 2012).

Despite these experiments demonstrating PR functionality and its apparent contribution to cell bioenergetics in *E. coli*, increased growth rate or cell yield as a result of PR activity have never been reported. One might have expected to observe faster growth if the PR was capable of stimulating active nutrient transport, or perhaps higher cell or biomass yield if the PR activity was capable of redirecting some amount of carbon from catabolic to anabolic pathways. On the other hand, the experiments demonstrating PR functionality in *E. coli* (stated above) were not conducted under physiological conditions. For the proton pumping or ATP assay, cells had to be concentrated and placed at 4°C for at least three days prior to conducting either assay (Martinez et al., 2007). Even before the cold incubation, inducing the expression of PR during exponential phase abruptly halted cell growth. To demonstrate proton pumping from a blue light absorbing PR in *E. coli*, it was necessary to produce spheroplasts and treat them with potassium and valinomycin to dissipate the potassium gradient (Wang et al., 2003). The reason this treatment was required to observe proton pumping was likely due to the fact that the photocycle was much slower than typical proton pumping PRs.

Experiments that measured the swimming speed and angular rotation rate of cells were also performed after growth to exponential phase and induction of PR expression. Additionally, PR-containing cells had to be treated with sodium azide or transferred to an anaerobic environment to prevent simultaneous proton pumping from respiratory

complexes in order to observe increased swimming speeds or increased rates of angular rotation (Walter et al., 2007). Since flagella rotation is proportional to the PMF over a range from near zero to -150 mV (Gabel and Berg, 2003), these experiments established that PR contributes to PMF in the absence of respiration. The authors explained the requirement for sodium azide (or the anaerobic environment) by adopting an RC circuit-based model of PMF generation and consumption, stating that current is not produced from PR proton pumping unless the voltage from respiratory activity falls below a threshold level (Walter et al., 2007).

In this analogy, the PMF sources (respiration and PR) represent two batteries in parallel, each capable of producing current and each with its own internal resistance (in series with respect to each battery). The internal resistance is inversely related to the current (outward proton flux) through the respiratory complexes or the PRs. For respiration, the internal resistance depends on the number of active respiratory complexes. But for PR, the internal resistance depends on the voltage across the protein (higher voltage, higher resistance, smaller current, smaller outward flux of protons) and the light intensity (higher intensity, lower internal resistance, higher outward flux of protons, thus higher voltage), so the model incorporates a resistor for PR that varies depending on the light intensity (Friedrich et al., 2002; Walter et al., 2007). The model also incorporates a "sink" resistor (in parallel to the batteries and their resistors) to represent the cellular systems that consume PMF (resulting in inward proton flow) such as the flagella motor, active transporters, and ATP synthase. Finally, to account for the fact that membrane capacitance determines how fast changes to the batteries' internal resistors are reflected

as changes in PMF, a capacitor is placed in parallel with the batteries, the sink resistor, and the voltmeter (that measures the PMF). Figure 3 (above) provides a schematic of the model circuit with respect to a PR-containing membrane.

Applying the circuit model, Walter *et al* (2007) were able to use their data (measurements of PMF over time under different light conditions and azide concentrations) to predict that proton flux through PR under peak light conditions is at least an order of magnitude higher in the presence of azide; in the absence of azide under standard aerobic conditions the proton flux would to be too small (between $10^2$ and $10^3$ protons per second) to contribute to PMF (Walter et al., 2007). However, if the inward flux of protons significantly increases, or the outward flux by respiration decreases PMF below the threshold potential for PR pumping, then the model suggests that light-driven proton pumping by PR will begin to contribute to the PMF and further increase its contribution as the PMF declines.

The idea that PMF exerts a "back pressure" on charge transport reactions, such as proton pumping, originated from research that sought to determine whether the phenomenon of respiratory control applied to *E. coli* (Burstein et al., 1979). Burstein *et al.* confirmed previous reports that respiration of aerobically cultured *E. coli* in the presence of an exogenous carbon source was not affected by the addition of uncouplers such as CCCP, indicating that respiration was already operating at maximum capacity. However, cells that were starved for 10 minutes in substrate free media exhibited a (20 times) lower rate of respiration, and the addition of CCCP stimulated respiration of these cells by as much

as six-fold. After determining that the concentration of oxidizable substrate and electron acceptor did not limit the respiratory rate, it became apparent that respiratory capacity was being controlled by a mechanism that was sensitive to uncouplers (Burstein et al., 1979). Some 20 years later, liposomes reconstituted with ATP synthase were found to increase the rate of ATP hydrolysis (producing PMF) in the presence of uncouplers (valinomycin and nigericin) due to the release of back pressure from the PMF (Fischer et al., 2000). And in the following year, the notion that photosynthetic reaction centers behave like "a simple 0.2V battery" was put forward to explain an observed current-voltage relationship; regardless of light intensity, liposomes reconstituted with the reaction centers could no longer pump protons when the electrical potential reached 0.2V (van Rotterdam et al., 2002). The concept of back pressure was also illustrated by the blue absorbing PR, studied by Wang *et al*, which only pumped protons when the potassium gradient was abolished by valinomycin and excessive exogenous potassium (Wang et al., 2003).

Underlying the notion of back pressure and certainly pivotal to the RC circuit model by Walter *et al* is the fact the activity of PR depends on membrane potential, in addition to light intensity and pH (Friedrich et al., 2002; Pfleger et al., 2009). Friedrich *et al* reported that both acidic and alkaline forms of PR exist in a 2:1 ratio in liposomes at neutral pH, with the acidic form requiring only a single photon to generate an inward flux of protons while the basic form requires two photons to generate an outward flux of protons. Moreover, the authors reconstituted PR in *Xenopus* ooctyes and showed that PR acts as an outwardly directed proton pump at pH 7.5 at all potentials more positive than -260 mV

(analogous to BR); but at pH 5.5, PR functions at an inwardly directed proton pump when the membrane potential becomes more negative than -80 mV (Friedrich et al., 2002). In fact, PR activity is so dependent on the membrane potential that it was recently engineered to become a highly accurate fluorescent PMF sensor *in vivo* (Kralj et al., 2011).

Work within *E. coli* as a heterologous host for PR has also established that PR self assembles into oligomeric structures within cell membranes (Klyszejko et al., 2008). Moreover, recent evidence indicates that the assembly of PRs into oligomers alters their function by causing a shift in the $pK_a$ of the key proton acceptor residue D97 by as much as a full pH unit relative to the monomeric form (Hussain et al., 2015). Since the assembly of PR into membranes affects its functionality, it is possible that reconstitution of PR into liposomes impacts specific traits such as the reverse potentials at different pH values and the ratio of acidic to alkaline forms. Additionally, PR-containing marine bacteria have likely evolved mechanisms to optimize post-translational processing and membrane insertion of PR, as well as its membrane environment with respect to lipid length, head group/charge, and degree of saturation (Hussain et al., 2015; Lindholm et al., 2015). While the membrane environment of *E. coli* is generally well suited for PR (Lindholm et al., 2015), it is difficult to know *a priori* how deliberate modifications to the membrane will impact light-driven proton pumping. Experimenting with the membrane by adding or removing membrane complexes can directly affect its potential, as well as the behavior of protons in the membrane environment (Mulkidjanian et al., 2006).

*Engineered Systems*

Renewable fuels and electricity can be produced from engineered biological systems. To date, the most exciting progress toward the production of renewables has involved genetically engineering and adapting photosynthetic microbes to produce reduced chemicals for biodiesel, alcohols, and hydrogen (Vignais and Billoud, 2007; Johnson and Schmidt-Dannert, 2008; Lubitz et al., 2008; Lubner et al., 2010; Ducat et al., 2011). The metabolic versatility of *E. coli* and other common genetically tractable and well-understood heterologous hosts has also encouraged researchers to attempt to engineer photosynthetic machinery into these heterotrophs. While considerable progress has been made with regard to the functional assembly of reaction center subunits and antenna complexes in *E. coli* membranes (Johnson and Schmidt-Dannert, 2008), enabling functional photosynthetic light capture and energy conservation would require a minimum of 30 novel genes whose products must self-assemble into a foreign membrane environment (Bryant and Frigaard, 2006; Johnson and Schmidt-Dannert, 2008).

The capabilities of PR demonstrated thus far within natural isolates, heterologous hosts, and synthetic membranes such as liposomes have encouraged its use toward the engineering of heterotrophs into photoheterotrophs. While notably less efficient for photons than photosynthetic systems (Bryant and Frigaard, 2006), PR photosystems self-assemble into functional form in foreign membranes, and are better suited for schemes that necessitate an anaerobic environment for reduced chemical production (Ghirardi et al., 2005; Johnson and Schmidt-Dannert, 2008). As previously mentioned, PR or other light-driven proton pumps have not been proven to enhance growth rate or yield in *E.*

*coli*; nevertheless, successful applications of PR in engineered systems have demonstrated that PR can enhance the production of biohydrogen (Kim et al., 2012). But in another heterologous host, *Shewanella oneidensis* strain MR-1, PR was shown to enhance cell survival in culture, and the production of bioelectricity in an MFC setting (Johnson et al., 2010). And, for the first time, enhancement of growth rate and yield was demonstrated for an ATP synthase knockout mutant of MR-1 under anaerobic conditions (Hunt et al., 2010).

The enhancement of biohydrogen production in E. coli required the co-expression of PR and an oxygen-tolerant [NiFe]-hydrogenase from *Hydrogenovibrio marinus* (Kim et al., 2012). The native [NiFe]-hydrogenase of *E. coli* is also capable of evolving hydrogen under aerobic conditions but at rates inferior to that of *H. marinus* (Kim et al., 2010; Kim et al., 2012). After adequate amounts of PR and hydrogenase were produced in the membrane, cells were incubated with or without retinal, moved to production cultures, exposed to green light, and the evolution of biohydrogen was measured over time by sampling the headspace of sealed culture vials. The experiment showed that PR enhanced hydrogen production by 1.3-fold (Kim et al., 2012). While the authors did not discern how PR enhanced biohydrogen production, they speculated that an increased concentration of protons in the periplasmic space enhanced the activity of the hydrogenase. Certainly, it's also possible that consuming reduced electrons from the ETC reduced respiratory proton pumping and enabled PRs to pump protons in the first place. Follow up experiments to investigate the effects of hydrogenase activity on the PMF may

establish that its activity is limited by electron supply from the ETC rather than the proton supply from the periplasm.

The enhancement of cell survival and bioelectricity production was demonstrated in *S. oneidensis* strain MR-1, a heterotroph with amazing versatility with respect to respiratory substrates (Johnson et al., 2010). After heterologous expression of PR under anaerobic conditions with retinal, PR-containing cells were incubated in the dark without an electron donor to deplete the membrane potential, and subsequently incubated in the presence of green light or darkness, with or without lactate as the electron donor, and in the presence of a voltage sensitive dye. Since energized (highly fluorescent) cells occurred in the light without lactate but not in the dark cultures without lactate, the authors concluded that PR was contributing to PMF under these experimental conditions (Johnson et al., 2010). Additionally, PR was found to enhance the viability of starved cells (incubated with green light relative to darkness) in anaerobic stationary phase cultures containing lactate, fumarate, and retinal. After demonstrating physiological effects of PR in the nascent photoheterotroph, PR-containing cells were transplanted into a microbial fuel cell (MFC) setting, exposed to green light, and forced to respire by shuttling electrons through cytochromes and structural proteins to a graphite electrode (also known as the Mtr respiratory pathway). Current production was enhanced upon illumination to a level that depended on the intensity of green light (Johnson et al., 2010), indicating that membrane potential was sufficiently depleted to the point where lactate oxidation did not restrict light-driven proton pumping by PR. In a subsequent experiment, the presence of PR was shown to enhance growth rate and yield of an MR-1 F-type ATP

synthase knockout mutant that oxidized lactate and reduced fumarate (Hunt et al., 2010). Since PMF generation by ATP synthase is beneficial under these experimental conditions, it's likely that light-driven proton pumping partially compensated for the mutation. Taken together these results certainly establish that PR can improve current production of illuminated MFCs, but additionally the results prove PR can impact growth rates and yields of an energy-limited heterologous host.

**Other forms of ion pumping rhodopsin photosystems**

Many other novel light-driven proton pumps have been discovered in marine and freshwater environments since PR. The next variant discovered was another light-driven proton pump called xanthorhodopsin (XR) within the halophilic bacterium *Salinibacter ruber* (Balashov et al., 2005; Lanyi and Balashov, 2008). XR is differentiated from PR by subtle differences in the primary sequence and secondary and tertiary structures (Luecke et al., 2008), but most strikingly by the addition of a carotenoid antenna pigment called salinixanthin (Balashov et al., 2005), which extends the absorption of XR into higher energy wavelengths of visible light typically occupied by chlorophylls (Claassens et al., 2013). A single salinixanthin is found per XR, similar to retinal, and there is significant interaction between the salinixanthin and retinal molecules upon absorption of light due to their close proximity deep within the protein (Balashov et al., 2006). To date salinixanthin remains the only known accessory pigment associated with any microbial rhodopsin photosystem (Béjà and Lanyi, 2014), unless reconstituting the proton-pumping

rhodopsin photosystem from the cyanobacterium *Gloeobacter violaceus* with salinixanthin counts (Imasheva et al., 2009). While a native accessory pigment was not discovered for this rhodopsin, the authors did suggest that a novel pigment might exist in nature based on the structure of the retinal binding pocket and ease of reconstitution of salinixanthin.

Similar to the BR-containing *Halobacteria*, the response of XR-containing *S. ruber* to light includes an inhibition of the respiratory rate (Oesterhelt and Krippahl, 1973; Balashov et al., 2005). Thus, it appears that these salt-loving microbes have both realized through evolution that their light-driven proton pumping mechanisms are most effective in the absence of respiratory proton pumping. However, it remains unclear how fluxes of carbon change, specifically between catabolic and anabolic pathways, with respect to light, light-driven proton pumping, and photoinhibition of respiration.

Similar to the marine PRs, actinorhodopsins (AR) are light-driven proton pumps found predominately within freshwater and brackish environments (Atamna-Ismaeel et al., 2008; Sharma et al., 2008; Sharma et al., 2009). Since their function and expected physiological effects do not differ significantly from PR these variants will not be discussed in more detail. However, it is important to note that their discovery solidifies the importance of light-driven proton pumping across all sunlit aquatic environments.

To this point, only type I microbial rhodopsin photosystems that transduce a sensory signal or pump protons have been discussed. However, recent evidence indicates that

chloride pumping and sodium pumping rhodopsin photosystems are found within the toolkit of marine bacteria as well (Béjà and Lanyi, 2014). Like the progression of thought from the discovery of BR and then PR, the chloride-pumping halorhodopsin (HR) has been known for decades but was thought to be restricted to a specific (salty) niche (Lanyi, 1986); when in reality, inwardly directed light-driven chloride pumps—denoted ClR—are present in planktonic marine bacteria as well (Béjà and Lanyi, 2014; Yoshizawa et al., 2014).

For the case of sodium pumping rhodopsins, the first direct evidence of light-driven sodium pumping was reported in 2013 (Inoue et al., 2013). The rhodopsin was called "KR2", which can be denoted NaR as well, and seemed to possess a duel H+/Na+ pumping activity, switching to proton pumping in the absence sodium ions (Inoue et al., 2013; Inoue et al., 2015). This duel role has since been confirmed (Li et al., 2015), and a handful of similar pumps have been described in other marine bacteria (Inoue et al., 2015).

A common theme emerging from these recent discoveries of ion pumping rhodopsin photosystems is that the light-driven translocation of any charged ion across the membrane could theoretically be converted into useful cellular energy. Of course the most common ion pools manipulated by bacteria for energy conservation are $H^+$, $Na^+$, and $Cl^-$, but one cannot rule out the possibility of other monovalent cations such as $K^+$ or divalent cations or even other charged metabolites being leveraged for energy conservation. Ultimately, however, active transporters and specifically antiporters link

the ion gradients together so that energizing one gradient can increase the energy stored in the other gradients. Interestingly, a marine Flavobacteria has recently been isolated and proven to contain functional PR, NaR, and ClR –pumping rhodopsins (Yoshizawa et al., 2014). It is possible that this bacterium is capable of inducing the expression of the optimal light-driven ion pump based on the most energetically favorable ion gradient to exploit in a particular environment. And it is equally as likely that all three are constitutively expressed and their activities are coordinated at a higher level of organization.

**Photoheterotrophic metabolism and the carbon cycle**

It is now clear that PR photoheterotrophs use various forms of light-driven microbial ion pumps for energy generation, are abundant in nature, and are phylogenetically diverse which suggests many different lifestyles and metabolic schemes are impacted by the availability of light. This means that light can affect how a large fraction of heterotrophic bacteria behave in ecosystems, and collectively, how these bacteria impact the global carbon cycle.

The carbon cycle is driven by photo- or chemo-autotrophic organisms that convert inorganic carbon into reduced organic molecules in a process known as carbon fixation. The organic carbon created by carbon fixation can be stored in metabolically active forms, such as sugars, or converted into biomass. From there the organic carbon can

move through food chains in the form of biomass, but it will inevitably either be respired back to carbon dioxide by autotrophs and heterotrophs, or exported to soils and sediments. The inorganic carbon returned to the global carbon cycle in oxidized form ($CO_2$ and $HCO_3^-$) becomes available to enter ecosystems once again by way of autotrophic organisms, while the sequestered carbon becomes isolated from the cycle, potentially for hundreds of millions of years.

Organic carbon that is sequestered from the global carbon cycle is commonly referred to as fossil fuels. Global reserves of fossil fuels in the forms of coal, crude oil and natural gas are abundant, and advances in technology have improved the processes of harvesting and oxidizing the fossil fuels for energy in the forms of electricity and heat. These activities are increasing the concentration of carbon dioxide in the atmosphere, leading to serious climate concerns, but the total flux of carbon from anthropogenic sources is dwarfed by the natural fluxes of carbon.

Of course the carbon cycle does not operate independent from the global biogeochemical cycles of other elements. The global oxygen cycle is closely linked to the carbon cycle, as carbon fixation is typically associated with oxidative photosynthesis, and carbon oxidation yields the highest energetic returns in the presence of oxygen, which is subsequently reduced back to water during respiration. The surplus of oxygen in the atmosphere, in addition to the natural reserves of fossil fuels, indicates that photosynthesis has outpaced respiration in terrestrial and marine ecosystems over

geological time scales. If this were not the case, heterotrophs with oxidative metabolisms could not have emerged over a billion years ago.

The processes of photosynthesis (i) and respiration (ii) can be represented by the redox reactions:

$$\text{(i) } CO_2 + H_2O \rightarrow CH_2O + O_2$$

$$\text{(ii) } CH_2O + O_2 \rightarrow CO_2 + H_2O$$

This redox couple is a simplified but fairly accurate representation of the global carbon cycle, as well as how energy flows into and out of ecosystems. Photosynthetic autotrophs drive the capture and transformation of energy from sunlight into chemical energy and biomass (i), serving as primary producers for the majority of Earth's ecosystems. And these same organisms—as well as all of the heterotrophs that live within the ecosystem—are responsible for respiring the reduced carbon back to inorganic carbon while extracting energy for cellular processes.

Although the physiological data from natural isolates containing rhodopsin photosystems is limited, at the very least it is clear that most are photoheterotrophs, and that they are abundant on a global scale and so are predicted to significantly influence the carbon cycle. What remains unclear is the net effect that light has on catabolism under various conditions, since it can both accelerate catabolism by increasing nutrient accessibility and transport, or decelerate catabolism in favor increasing the flux of carbon toward biosynthetic pathways. A better understanding of how rhodopsin photosystems influence

the physiology of photoheterotrophic bacteria will lead to more accurate models of the carbon cycle that account for the influence of sunlight on the growth and respiration of heterotrophs, and ultimately on the net community production of ecosystems.

**Thesis Overview**

The objective of this thesis project was to explore how native PR-containing bacteria leverage PR to improve growth rate and yield, and to subsequently apply lessons from this research to successfully convert a heterotroph into a photoheterotroph capable of faster growth and/or improved cell and biomass yields. Ultimately, demonstrating the increased production of biomass (a carbon compound) from a heterologous host would prove that energy derived from PR integrates into its metabolic systems. Once integrated, renewable energy from PR could be applied in engineered systems to improve growth efficiency in terms of carbon, which would certainly reduce the costs associated with industrial-scale production of reduced chemicals.

Chapter 2 explores how light and carbon influences the growth phenotype of different native PR-containing Flavobacteria. One of these bacteria, *Dokdonia donghaensis* strain MED134, grows faster and to higher cell yields in the presence of light in media containing trace amounts of dissolved organic carbon. This result was confirmed, and an extensive RNA-seq experiment was performed to expose light induced genes in an attempt to better understand what changes in gene expression produce the observed light-enhanced growth phenotype.

Chapter 3 explores the applicability of *E. coli* as a model heterologous host for PR. Several knockout strains were tested as heterologous hosts in attempt to identify a more favorable genetic background for PR production and subsequent PR-derived energy integration than the wild-type. Additionally, plasmid constructs complementing key native genes or introducing novel genes were generated to further optimize the genetic background for growth with PR. This work lead to the creation of several unique strains with high potential to benefit energetically from PR, with one in particular evidenced by higher cell yields. Culture conditions certainly impacted how effective the PR was in terms of cell yield enhancement for this strain.

Chapter 4 summarizes the findings from Chapters 2 and 3, highlighting key data from the RNA-seq experiment with MED134 that motivated engineering efforts to improve the genetic background of E. coli for light-driven proton pumping. Future prospects for PR in engineered systems are also discussed with emphasis on applications that would create renewable sources of important reduced chemicals.

## II. The Physiological Effects of Proteorhodopsin in Flavobacteria

ABSTRACT

To better resolve how the marine proteorhodopsin (PR)-containing bacterium *Dokdonia donghaensis* strain MED134 responds to light, an extensive RNA-seq experiment was performed with triplicate light and dark cultures growing in low and high carbon complex media. Genes with significant differential expression between light and dark treatments in low and high carbon media were identified. These comparisons clearly showed that light induces the PR gene, phytoene dehydrogenase, and 11 other genes with unconfirmed or unknown functions. Furthermore, the data indicated that PR might enhance growth by specifically improving iron acquisition in low carbon media. Under high carbon conditions, the data suggested that fatty acid metabolism might influence the contribution of PR to cell energetics. Understanding the environmental conditions under which the light response translates into faster and more efficient growth for MED134—and other Flavobacteria (Kim, 2013)—will help progress efforts to determine the impact of light energy and PR activity on carbon and energy flows through marine ecosystems.

## Introduction

Marine environments host some of the largest and most diverse assemblages of life on Earth. Bacteria concentrated within the sunlit surface waters encompass most of the marine biomass and diversity, and in turn exert the greatest influence on Earth's biogeochemical cycles. The proteorhodopsin (PR) gene is among the most distributed, abundant and highly expressed genes found within surface-dwelling marine bacteria (Sabehi et al., 2005; Rusch et al., 2007; Campbell et al., 2008; Frias-Lopez et al., 2008; Poretsky et al., 2009; Morris et al., 2010; Wei, 2010), suggesting that PR and other rhodopsin-based light-harvesting proteins represent a major pathway for energy flow into ecosystems. Exactly how influential PR is to the metabolism of its large and diverse collection of host bacteria, and how these bacteria influence life on Earth en masse, are outstanding questions that must be addressed in order to fully understand marine ecosystem dynamics.

Since the discovery of PR photosystems (Beja et al., 2000), a small sample of marine bacteria with the PR gene have been isolated and studied. Of these, only three isolates of flavobacteria (Gomez-Consarnau et al., 2007; Feng et al., 2015; Gómez-Consarnau et al., 2016) have been reported to experience higher growth rates and cell yields as a result of light-driven proton pumping by PRs. Others have been reported to better sustain biomass and metabolic capacity during periods of starvation using energy derived from PR (Fuhrman et al., 2008; Gomez-Consarnau et al., 2010). Recent evidence from studying flavobacteria indicates that PR influences growth by improving nutrient acquisition

(Gómez-Consarnau et al., 2016), in addition to photophosphorylation (Martinez et al., 2007; Palovaara et al., 2014).

In effort to improve upon our understanding of the ecological impact of PR, I was interested in expanding the number of PR-containing isolates that demonstrated light-enhanced growth phenotypes. More recently, four strains isolated by Yoshizawa *et al* with enhanced growth rates and yields when grown in the presence of light (Yoshizawa et al., 2012; Kim, 2013). However, growth was only enhanced when a small amount of dissolved organic carbon (DOC) was used as growth substrates (Kim, 2013). At higher DOC concentrations, growth rates and yields of cultures maintained in the dark were not different or higher than cultures exposed to light (Kim, 2013).

In this work, I aimed to better resolve how a marine isolate that exhibits higher growth rate and yields—the PR-containing bacterium *Dokdonia donghaensis* strain MED134—responds to light by conducting an extensive RNA-seq experiment with triplicate light and dark cultures, in low and high DOC media, and with samples at two distinct points in the growth curve. A comprehensive statistical software package, DESeq, was applied to identify genes with significant differential expression between light and dark treatments in low and high DOC media (Anders and Huber, 2010). These comparisons clearly showed that the PR gene is induced by light in addition to the phytoene dehydrogenase (involved in the synthesis of the chromophore retinal) and other genes of unknown function. Furthermore, my results indicate that in low DOC media PR may enhance growth by specifically improving iron acquisition. At high DOC, the data suggest that

fatty acid metabolism might overshadow the contribution of PR to cell energetics. In this chapter, I discuss these results and compare with previously published transcriptome data from MED134 and other flavobacteria.

## Methods

*Growth experiments*

Experiments were carried out to determine the optimal dissolved organic carbon (DOC) concentration for light-enhanced growth of the PR-containing marine flavobacterium, *Dokdonia* sp. MED134. The growth media was composed of artificial seawater (ASW), 35 practical salinity units (prepared from Sea Salts; Sigma), which was filtered through a 0.2μ pore size filter (Millipore), subsequently autoclaved, and then supplemented with sterile solutions of $NH_4Cl$ and $Na_2HPO_4*12H_2O$ (Sigma) to final concentrations of 225 μM $NH_4Cl$ and 44.7 μM $Na_2HPO_4$, as previously reported (Kimura et al., 2011).

To generate the starter cultures, a colony originating from a frozen stock was picked from a Marine Broth Agar (BD) plate and inoculated into growth media that was further amended with 5.65 mL of full strength media (FSM) per liter. The FSM is five grams peptone and 1 gram yeast extract per liter of ASW, which reportedly contains about 242 mM of carbon (Gomez-Consarnau et al., 2007). After three days of growth to early stationary phase at 23°C without exposure to light or constant shaking, cells were diluted with ASW to a concentration of $10^6$ cells/mL and placed at 4°C for at least one week and

up to three months before use. Transfer of the stationary phase cells to ASW and subsequent incubation at 4°C was found to be an essential treatment in order to observe light-enhanced growth in subsequent experiments.

To test different DOC concentrations, the growth media was amended with either 0 µL, 700 µL, 2 mL, 5.65 mL or 250 mL of FSM per liter, inoculated with cells at a final concentration of 1000 cells/mL, split into triplicate 20mL cultures, and incubated with or without exposure to cool white fluorescent light (at about 150 µmol photons/$m^2$/s; refer to Appendix Figure 1a) for up to one week at 23°C without shaking. To determine the cell concentrations, daily samples were taken from each culture, diluted if necessary, filtered onto pre-blackened Isopore membrane filters (0.2µ pore size; Millipore), stained with SYBR Green I (1:100 dilution; Invitrogen) for at least 15 minutes, and then counted using an epifluorescence microscope (Axioskop Zeiss).

To avoid DOC contamination all glassware used for growing bacteria or storing media were new borosilicate bottles or tubes (VWR), and each was washed with 10% HCl prior to use.

*RNA-seq*

RNA-seq was used to identify significant differences between the transcriptomes of *Dokdonia* sp. MED134 cells incubated with and without light exposure in growth media containing low (700 µL of FSM per liter ASW) or high (250mL of FSM per liter ASW) DOC concentrations. Cells for this experiment were produced exactly as described for

growth experiments with MED134, and were subsequently inoculated into 4 L of growth media in 5 L bottles (VWR) amended with either low or high DOC at a final concentration of 1000 cells/mL. The 4 L cultures were incubated at 26°C in complete darkness for two days without shaking. After two days, each culture was well-mixed and then split into six 1 L bottles, each containing 500 mL of culture. Half of the bottles from each 4 L culture were then wrapped with aluminum foil, and all twelve bottles were placed in a light incubator (Beckman) modified to produce about 18 µmol photons/$m^2$/sec of green light (refer to Appendix Figure 1b). The 12 bottles were incubated without shaking for an additional three and a half days. Samples for RNA-seq analysis were taken at three points, denoted T0, T1, and T2 (indicated by the red arrows on Supplementary Figure S3). Each sample was removed from its respective culture through sterilized and acid-washed tubing (Masterflex) using a peristaltic pump. The tubing was connected to a sterile long pipette at the sampling end and a 0.22-µ pore size Sterivex filter (Millipore) at the other end. The volume of culture passed through the filter was 500 mL at T0 (sourced from the 4 L culture after 3 L was removed) and 200 mL at T1 and T2. After sampling, the Sterivex filters were bottom-sealed, filled to about 90% capacity with RNA*later* Solution (Ambion), capped, and then stored at -80°C until processing.

Total RNA was extracted from cells on the Sterivex filters using a previously described protocol (Kimura et al., 2011). The filters were thawed on ice, the RNA*later* was decanted, and Lysis/Binding Buffer from the *mir*Vana miRNA Isolation Kit (Ambion) was injected to refill the Sterivex filter. After a short incubation with mixing, the lysate was removed from the filters and subjected to Dnase treatment with the TURBO DNA-

free Kit (Ambion). Total RNA was then concentrated and purified with the RNeasy MinElute Cleanup Kit (Qiagen), and quantified with the RNA6000 Pico Total RNA Kit (Agilent). Total RNA was then treated to remove 16s and 23s rRNA with probes that were produced previously (Kimura et al., 2011). After hybridization and removal of the rRNA-probe complexes, the rRNA-subtracted RNA was concentrated, purified and quantified once again before proceeding to follow the manufacturer's instructions for the Scriptseq RNA-Seq Library Preparation Kit (Illumina) and subsequent analysis on Illumina's MiSeq platform.

The resulting libraries were subjected to Illumina's QC and demultiplexing software before the libraries were imported as paired end reads into the Genomics Workbench software package (CLCbio). The reads were trimmed and mapped to the MED134 annotated reference genome with the built in application and the read counts were exported as a .csv file. The read counts were normalized by importing the data into the R-package DESeq software (Anders and Huber, 2010). DESeq was used to normalize the read counts to account for library sizes and shot noise, and to subsequently identify differentially expressed genes between light and dark treatments. Genes identified as differentially expressed between light and dark treatments had an FDR-adjusted p-value $< 0.25$, which was chosen as an appropriate cutoff value considering the conservative nature of DESeq and that FDR-adjusted p-values increased significantly above this value for every comparison tested. Most genes considered to be differentially expressed had FDR adjusted p-values $< 0.05$. Phyre was used to analyze differentially expressed genes that encoded for uncharacterized proteins (Kelley et al., 2015).

## Results

*Growth of MED134 at low and high DOC concentrations*

Figure 1 shows how light influenced the growth of MED134 at low and high DOC concentrations. Increasing the concentration of DOC reduced the positive influence of light on growth to a point where the presence of light negatively influenced growth rate and yield. Because it has been previously shown that light enhanced growth of MED134 depends on the PR and the presence of green light (Gomez-Consarnau et al., 2007), I was interested in determining whether PR was responsible for the negative influence on growth in high DOC media. In addition to the PR, other factors such as growth rate and osmolarity of the media could influence the extent to which PR activity impacts growth.

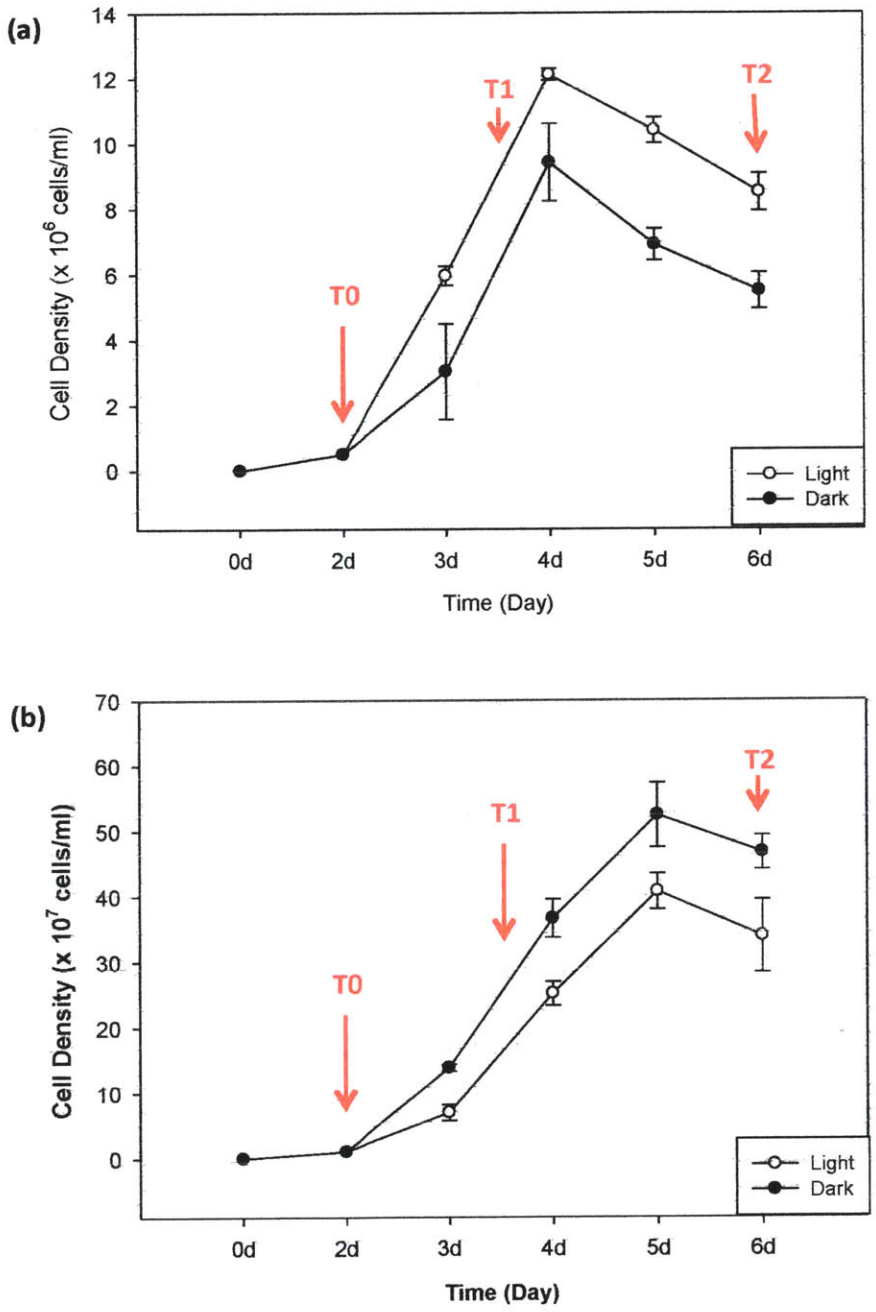



**Figure 1.** Growth of MED134 at low and high DOC concentrations.

MED134 exposed to light (open circles) or in darkness (filled circles) was grown in the presence of 0.7 mL (a) or 250 mL (b) of FSM per liter. Error bars denote the standard deviation for triplicate cultures. Sampling for the RNA-seq experiment took place at the sample time points indicated on the figure with red arrows.

*Light induced gene expression in MED134*

To better understand the influence of the DOC concentration on the growth physiology of PR-containing marine flavobateria, a replicated RNA-seq experiment was designed to identify differences in gene expression at low and high DOC concentrations in response to light. Despite the apparent effect of the DOC concentration on the growth physiology of MED134, I found that a similar set of genes was induced by light in both low and high DOC treatments.

Table 1 highlights transcripts that were found to be significantly more abundant among the light-exposed cultures relative to the dark-exposed cultures at the low DOC T2 and high DOC T1 sample time points. Remarkably, no transcripts were found to be significantly more abundant in the dark at low DOC T2 or high DOC T1.

**Table 1.** Average transcript abundances of the MED134 light induced genes

| Feature ID | Annotation | Low DOC T2 | | High DOC T1 | |
|---|---|---|---|---|---|
| | | Light | Dark | Light | Dark |
| MED134_07119 | Proteorhodopsin | 156 | 11 | 1197 | 84 |
| MED134_07089 | Putative uncharacterized protein | 25 | 2 | 73 | 9 |
| MED134_10201 | Probable bacterial cryptochrome | 41 | 4 | 337 | 45 |
| MED134_13081 | Transcriptional regulator, MerR family protein | 69 | 11 | 366 | 60 |
| MED134_14276 | Putative uncharacterized protein | 34 | 6 | 198 | 40 |
| MED134_04999 | Putative uncharacterized protein | 97 | 18 | 836 | 107 |
| MED134_01275 | Putative uncharacterized protein | 456 | 102 | 2537 | 317 |
| MED134_13076 | Phytoene dehydrogenase | 33 | 9 | 234 | 59 |
| MED134_03409 | Molecular chaperone (DnaK) | 161 | 66 | 467 | 91 |
| MED134_03969 | Putative adhesion lipoprotein | 590 | 268 | 1368 | 139 |
| MED134_03404 | ABC-1 | 206 | 99 | 1194 | 151 |
| MED134_11080 | Putative uncharacterized protein | 642 | 331 | 1587 | 472 |
| MED134_14271 | Short chain dehydrogenase | 15 | 4 | 149 | 25 |

Red numbers indicate that the difference in mean transcript abundance is not significant (FDR-adjusted p-value > 0.1).

The PR gene (MED134_07119) was confirmed to be more abundant among light exposed cultures, by a $\log_2$ fold change of 3.8, at both low DOC T1 and high DOC T2. This finding validates the DESeq analysis and suggests that induction of PR by light is independent of the DOC concentration. In addition, phytoene dehydrogenase (MED134_13076) was induced by light and to a similar extent ($\log_2$ fold change of 2.9 and 3.0) in both low and high DOC treatments. The phytoene dehydrogenase plays an essential role in the retinal synthesis pathway. The probable bacterial cyptochrome (MED134_10201) and the BLUF-containing putative uncharacterized protein (MED134_07089) were also induced by light regardless of the DOC concentration in my study.

As indicated by the annotations in Table 1, most of the remaining genes induced by light at low DOC T2 and high DOC T1 encode proteins of unknown function. However, many of these putative proteins were found to contain domains that suggest localization to the membrane and/or involvement in adhesion to particles. The proteins likely involved in adhesion include the putative adhesion lipoprotein (MED134_03969) and MED134_04999, which both contain fasciclin (FAS1) domains according to the Phyre analysis (see methods). The ABC-1 gene (MED134_03404) is probably localized to the membrane and shares 85% identity to a transferase whose function is linked to the cytochrome bc1 complex. And finally, the MED134_14276 gene encoded a protein similar (87%) to a nucleoside-diphosphate sugar epimerase.

Of the remaining light induced genes of unknown function, a putative uncharacterized protein (MED134_01275) containing the DUF4199 domain (of unknown function) was

identified as light induced at low DOC T1 and high DOC T2. Analysis of the corresponding protein sequence with the Phyre application revealed that the gene product is most similar (31%) to the transmembrane ionophore protein colicin Ia (low condfidence). The MerR-like transcriptional regulator, DnaK-like chaperone, and the putative uncharacterized protein (MED134_11080)—identified by Phyre as a small DNA/RNA binding protein—were all identified as significantly induced by light exclusively for low DOC T2, while the short chain dehydrogenase (MED134_14271), which is 70% identical to a 3-oxoacyl-ACP reductase, was only significantly induced by light at high DOC T1 (as indicated with red type in Table 1). The $\log_2$ fold change and p-values associated with the insignificant differences are 2.6/0.75, 2.4/1, 1.7/0.26, and 2.1/0.26, respectively (see Supplementary Table S2).

*Differential gene expression during late exponential phase in low DOC media*

We attempted to identify differences in transcript abundance between exponential phase cultures of MED134 exposed to light or darkness in low DOC media, but a library representing one of the three samples from the light-exposed cultures failed to generate an adequate amount of reads and was discarded (refer to Supplementary Table S1 for the total count of mapped reads per library). As a result of the reduced statistical power, only three genes were identified as differentially expressed at low DOC T1 (refer to red arrows on Figure 1). Transcripts of two of the three differentially expressed genes (MED134_06839 and MED134_06834) were found to be more abundant in light-exposed cultures, and encoded proteins involved in iron assimilation. The first of these proteins (MED134_06839) had an average normalized transcript abundance of 107 among the

light-exposed cultures compared to 27 among the dark cultures ($log_2$ fold change of 2 and FDR adjusted p-value of .017) and likely encodes a HmuY heme transport protein; while the second (MED134_06834) had 1798 transcripts among light-exposed cultures relative to 463 among the dark cultures ($log_2$ fold change of 2 and FDR adjusted p-value of $1.95x10^{-5}$) and encodes a TonB-dependent outer membrane receptor that identified as a HmuR heme receptor. The transcript of the third differently expressed gene at low DOC T1 (MED134_01225; hlyD) was more abundant in the dark-exposed cultures at 360 reads relative to 109 in light-exposed cultures ($log_2$ fold change of 1.7 and FDR-adjusted p-value of $5x10^{-6}$). In *Escherichia coli*, HlyD has been reported to aid in the secretion of the hemolysin protein HlyA (Pimenta et al., 2005).

*Differential gene expression during stationary phase in high DOC media*

Light or dark-exposed cultures grown to stationary phase in high DOC media resulted in the identification differentially expressed genes that were not found to be significantly different in low DOC media (T1 or T2) or during exponential/early stationary phase (T1) in high DOC media. In fact, the only gene from high DOC T2 that was differentially expressed at other time points was PR (MED134_07719), which had an average transcript abundance of 27 among light-exposed cultures compared to 8 among dark-exposed cultures ($log_2$ fold change of 1.7 and FDR adjusted p-value of 0.051). The only other gene whose expression was induced by light in high DOC cultures at T2 was the purine nucleoside phosphorylase (EC 2.4.2.1; MED134_00750), which had an average of 85 transcripts among light-exposed cultures compared to 46 among the dark-exposed cultures ($log_2$ fold change of 0.9 with FDR adjusted p-value of 0.075).

In contrast to growth in the low DOC media or at T1 in the high DOC media, the majority of differently expressed genes at high DOC T2 were more abundant among cultures that were maintained in darkness, and were translated into proteins involved in β-oxidation or the heat shock stress response (refer to Supplementary Table S3). The transcripts associated with β-oxidation originated from etfA (MED_00125) and etfB (MED134_00130), encoding an electron transport flavoprotein, as well as genes that encode Enoyl-CoA hydratase (MED134_06229), and a thiolase family protein (MED134_06234), with light to dark average transcript abundance of 407 to 694 ($log_2$ fold change of 0.8 and FDR adjusted p-value of $5x10^{-3}$), 311 to 574 ($log_2$ fold change of 0.9 and FDR adjusted p-value of $3.3x10^{-4}$), 409 to 779 ($log_2$ fold change of 0.9 and FDR adjusted p-value of 0.049), and 136 to 257 ($log_2$ fold change of 0.9 and FDR adjusted p-value of 0.09), respectively. The transcripts associated with heat shock originate from clpB (MED134_05174), grpE (MED134_13401), and a putative heat-shock related protein (MED134_04044), with light to dark average transcript abundance of 2604 to 4669 ($log_2$ fold change of 0.8 and FDR adjusted p-value of $1.1x10^{-4}$), 737 to 1344 ($log_2$ fold change of 0.9 and FDR adjusted p-value of $1.1x10^{-4}$), and 1158 to 2381 ($log_2$ fold change of 1 and FDR adjusted p-value of $2x10^{-10}$), respectively. In addition to these two general classes, two additional genes were up-regulated among dark cultures that encode a putative recognition particle-docking protein (MED134_12996) with light to dark transcript abundance of 333 to 545 ($log_2$ fold change of 0.7 and FDR adjusted p-value of $4.8x10^{-5}$), and a putative uncharacterized protein (MED134_03499) with an average of

189 transcripts among the illuminated cultures to 364 among cultures incubated in darkness ($log_2$ fold change of 0.9 and FDR adjusted p-value of 0.063).

**Discussion**

I confirmed previous reports that growth rates and yields of MED134 are enhanced in the presence of green light when the DOC concentration of the culture media is extremely low (Gomez-Consarnau et al., 2007; Kimura et al., 2011). As the DOC concentration was increased, I observed less of a difference in the growth phenotypes between cultures maintained in the presence of light, relative to darkness; in fact, at relatively high DOC concentrations cultures maintained in the dark experienced higher growth rates and yields than the light-exposed cultures (Figure 1). This phenomenon was not restricted to MED134, as other flavobacteria have shown a similar light effect at DOC concentrations (Kim, 2013). An RNA-seq experiment was performed to better understand how the DOC concentration impacts the growth phenotypes.

*Light induced gene expression in MED134*

I identified a set of genes that are strongly induced when MED134 was grown in the presence of green light (Table 1). Most of the genes that were significantly up-regulated among the light exposed cultures at low DOC were found to be up-regulated under high DOC conditions as well. The PR gene is among this set of light induced genes, and its level of induction ($log_2$ fold change of 3.8) is identical for the low and high DOC

conditions, indicating that induction of PR by light is independent of the DOC concentration. If induction of the PR gene is costly in the presence of high DOC, as the growth phenotype in Figure 1 suggests, one may expect there to be a feedback mechanism to reduce the level of induction under these conditions. However, it is also possible that the high concentration of DOC used in my experiment is not encountered frequently enough in nature to elicit the evolution of a more complex regulatory mechanism.

In addition to the PR, MED134 up-regulates the phytoene dehydrogenase among light exposed cultures independent of the DOC concentration. The fact that this enzyme is the only member of the retinal synthesis pathway identified as significantly induced by light suggests that it could be responsible for the rate-limiting step in retinal production. In an earlier transcriptome study of MED134 by Kimura *et. al.* (2011), the authors also found that PR and phytoene dehydrogenase were significantly induced by light, along with the geranylgeranyl pyrophosphate synthase gene. The use of replicates in the study design, coupled with the conservative nature of DESeq (Anders and Huber, 2010), supported light induction of PR and phytoene dehydrogenase transcripts, but up-regulation of geranylgeranyl pyrophosphate synthase was not statistically significant (Supplementary Table S4). Light induction of the phytoene dehydrogenase appeared to be linked to a MerR family transcriptional regulator protein that was significantly induced by light in low DOC (Table 1), and that lies immediately upstream of the phytoene dehydrogenase in the MED134 genome.

A set of light sensing proteins were also described by Kimura *et al.* as being potentially involved in the light response of MED134. We confirmed that the probable bacterial cryptochrome, DASH family (MED134_10201) was significantly induced by light. We also observed significant light induction from an uncharacterized protein containing a BLUF domain (MED134_07089). In contrast, I did not find significant light induction of the putative DNA photolyase /cryptochrome (6-4) photolyase family genes (MED134_10206, MED134_10211), but I did observe greater average transcript abundances among cultures exposed to light (of at least $\log_2$ fold change of 1.5) relative to dark for the high DOC T1 comparison (Supplementary Table S4). Additionally, Kimura *et al.* and Gomez-Consarnau *et al.* (2016) reported significant light induction of a histidine kinase gene (MED134_10396). Although not statistically significant in my experimental design, I also observed a higher abundance of this transcript among light exposed cultures ($\log_2$ fold change of 1.6) in low DOC media at late exponential phase, but the average number of reads was low as compared with the number of reads reported from these earlier transcriptome studies (Supplementary Table S4). Because the second library from low DOC cultures during stationary phase contained even less reads originating from the histidine kinase, it is possible that significant light induction occurred during early or mid-exponential phase.

Most of the remaining light induced genes that I identified encode proteins with unknown functions. But as previously noted (see Results), at least two of these proteins are likely exported from the cytoplasm and involved in particle adhesion. A proteomics study on a PR-containing flavobacteria isolated from sea ice also found higher abundances of

proteins containing fasciclin repeat domains upon illumination, leading the authors to suggest that these proteins are involved with adhesion to algal surfaces (Feng et al., 2015). I commonly observed the formation of clumps of MED134 cells in cultures, and these clumps tended to be larger among cultures exposed to light; this could be a result of higher concentrations of cells in light-exposed cultures, or clumping may be accelerated by increased production of adhesive proteins and exopolysaccharides. Future experiments designed to observe how MED134 behaves in the presence of light with algal or other types of particles may demonstrate a link between the presence of light and particle adhesion.

Unlike previous RNA-seq experiments with MED134 and other flavobacteria, I did not observe significant induction of any known metabolic genes in response to light. Many of the genes identified as significantly differentially expressed between light and dark treatments in prior transcriptome experiments are included in Supplementary Table S4 with gene abundance data from the experiment. Genes reported to be light induced in experiments with MED134 and other flavobacteria grown under similar conditions are involved with glycolysis and gluconeogenesis, the pentose phosphate pathway, the TCA cycle and the glyoxylate shunt, and the respiratory chain (Kimura et al., 2011; Palovaara et al., 2014; Feng et al., 2015; Gómez-Consarnau et al., 2016).

The glyoxylate shunt in particular has been suggested to be important for light-enhanced growth of MED134 (Palovaara et al., 2014). I observed a $\log_2$ fold change of 1.2 at low DOC T1 for the isocitrate lyase gene (MED134_01780), and found significantly less

reads, and less of a difference between light and dark abundance, associated with the gene at T2. This result contradicts RT-PCR data from Palovaara *et al.* (2014) that demonstrates higher expression levels and significant light induction of isocitrate lyase during stationary phase in low DOC media. Otherwise, the RNA-seq data is similar to the RT-PCR results from several other metabolic genes.

Samples for RNA-seq were not taken during the first 36 hours of growth, which is unfortunate because the data from Palovaara *et al.* suggests that some light induced gene expression only occurs within early to mid-exponential phase (prior to 36 hours of growth). These authors identified significant light induction of the bicarbonate transporter BicA (MED134_10061) and the carbonic anhydrase (MED134_10056) during the first 36 hours of growth when these genes were experiencing relatively high levels of induction, but total expression and light induced expression significantly decreased between 36 and 45 hours of growth. In the low carbon condition I found that the average transcript abundance of BicA and the carbonic anhydrase decreased by more than half between T1 and T2; however, in the high carbon condition the levels of these enzymes remained constant between T1 and T2 suggesting a higher requirement (relative to the low carbon condition) for one-carbon units during stationary phase. The pyruvate carboxylase (*pyc*) and PEP carboxylase (*ppc*) both affect the intracellular bicarbonate concentration, and expression of these genes followed a similar pattern as BicA and the carbonic anhydrase; expression of *pyc* and *ppc* during stationary phase was less than exponential phase, with insignificant (this study) or significant (Palovaara et al., 2014) light induced expression of both genes occurring during stationary phase in the low carbon condition. Interestingly,

my data from the high carbon condition shows that the expression of *pyc* and *ppc* increases by about 25% throughout stationary phase, in accordance with maintaining higher expression of the genes responsible for importing or synthesizing bicarbonate. In contrast, Kimura *et al.* reported significant down-regulation of *pyc* and *ppc* in the light treatment when MED134 was grown in low carbon conditions. The available data suggests both the level of DOC and the time of sampling likely explain the observed differences in *ppc* and *pyc* expression between these experiments.

*Transporter expression in low DOC cultures*

Membrane-embedded transporter proteins that actively import or export compounds from the cell require energy stored within ATP and the transmembrane ion gradients. Because PR photosystems generate proton motive force (PMF), PR can directly stimulate transporters energized by the proton gradient. Ton-B dependent transporters (TBDTs) are energized by the PMF and have been implicated in the light response of MED134 (Gómez-Consarnau et al., 2016). While light induction or high expression of the vitamin-B1 TBDT gene was not observed, I confirmed that at least one *susC*-like TBDT (MED134_05219) was among the most highly expressed genes in MED134 during exponential growth in low DOC media (Supplementary Table S4 and S5). Additionally, I found that a putative heme TBDT was significantly induced among light exposed cultures (MED134_06834), along with an excreted HmuY-like protein (MED134_06839) that is likely involved in iron acquisition. To my knowledge this is the first report of a PR-stimulated increase in the expression of iron acquisition genes in MED134. This finding suggests that iron, rather than vitamin B1, may be limiting growth when the low DOC

cultures were sampled during exponential growth, thus expanding the general role of TBDTs in the light-response of MED134.

Previous transcriptome experiments with MED134 also suggested that sodium transporters are integral to the light response. Kimura *et al.* reported significant light induction of the $Na^+$-translocating NQR respiratory complex, along with two $Na^+$/alanine symporters and a $Na^+$/phosphate symporter. Significant differences in the expression of these genes were not observed, but the average number of transcripts for two of the $Na^+$ symporters was higher among the light cultures maintained in low DOC at T1 (Supplementary Figure S4). It is possible that genes involved in sodium translocation were only significantly induced in the light treatment while alanine was being used preferentially as the carbon source during growth on the DOC. Previous growth experiments using DL-alanine as the sole carbon source required sodium translocation via NQR to observe a light-enhanced growth effect (Kimura et al., 2011). Several metabolic genes such as the isocitrate lyase and malate synthase are also involved in the light response of MED134 growing on alanine (Palovaara et al., 2014).

I did not observe significant differential expression of any genes encoding known $Na^+$-ATPases or $Na^+/H^+$ antiporters, but a $K^+/H^+$ KefC-like antiporter (MED134_06844) was among the statistically insignificant light induced genes with a $\log_2$ fold change >1 at low DOC T1 (statistical comparison based on only two light exposed replicates). The *kefC*-like transcripts were more abundant among light cultures relative to dark at T1 and T2 in low DOC media, and were equal or more abundant among the dark cultures in high DOC

media at T1 and T2 (Supplementary Table S2). KefC is gated by glutathione in *E. coli*, and its purpose is to reduce the intracellular pH to protect the cell from toxic electrophiles such as methylglyoxyl (Miller et al., 1997). Involvement of a KefC-like antiporter in the MED134 light response in low DOC media would suggest that $K^+$ efflux may be used to prevent excessive alkalization of the cytoplasm during PR-driven proton pumping, which would also effectively spill excess energy stored within the $K^+$ and $H^+$ gradients in order to prevent hyperpolarization of the membrane. Such a system would presumably be more effective in the low DOC media (relative to the high DOC media) where there is less cytoplasmic water loss and thus a lower requirement for intracellular $K^+$ (Record et al., 1998).

Indeed, experiments with the PR-containing flavobacteria isolated from sea ice showed that the media osmolarity (adjusted by the NaCl concentration) exerted a strong influence on the light response (Feng et al., 2015). The optimal osmolarity or salinity for PR expression and light-enhanced growth has not been determined for MED134. Future experiments comparing growth phenotypes at different DOC concentrations should be designed to control for differences in media osmolarity, and should also be conducted at the optimal experimentally determined osmolarity for light-enhanced growth. The effect of osmoprotectants ($K^+$ substitutes) such as glycine betaine should be explored as well.

*The influence of beta-oxidation in high DOC media*

Transcripts of three genes involved in fatty acid degradation (MED134_00130, MED134_00125, and MED134_06229) were found to be significantly more abundant

among high DOC cultures maintained in the dark during stationary phase (see Results). I found that the average transcript abundance of each of the genes increased from T1 to T2 among dark cultures in high DOC media, while the average transcript abundance among light cultures decreased over the same period. This pattern of gene expression was not observed for cultures growing in low DOC media; in fact, the average transcript abundances were higher in the light condition at both T1 and T2 for all three of the genes. Additionally, average transcript abundances of the three beta-oxidation genes decreased from T1 to T2 for light and dark cultures grown in the low DOC media (Supplementary Table S2). Taken together, MED134 grown in high DOC media and maintained in the dark exhibited a higher capacity for beta-oxidation in stationary phase than cells maintained in the light, or cells that were grown in lower DOC media.

A higher capacity for beta-oxidation at stationary phase may indicate that fatty acid pool sizes were also relatively larger, likely resulting from increased fatty acid synthesis during exponential growth. In addition to identifying and quantifying the fatty acids that MED134 synthesizes, stores, and subsequently metabolizes, future experiments should seek to determine if less fatty acids are synthesized and/or metabolized from the media among the cultures maintained in the light at high DOC, and if so, why. One possibility is that fatty acid metabolism maybe down regulated in the light to enable PRs to contribute to the proton motive force. MED134 may regulate the beta-oxidation genes differently in the dark condition since light was not influencing the PMF. If similar levels of beta-oxidation occurred in the light as observed in the dark it is possible growth would become less efficient in terms of carbon and energy, since excessive membrane

hyperpolarization can lead to the activation of energy spilling mechanisms like voltage-gated and mechanosensitive channels or the induction of genes that encode ionophores (as potentially observed—see MED134_01275). This hypothesis can also be addressed by future experiments designed quantify the membrane potential and the associated effects of hyperpolarization such as the realignment of membrane embedded proteins and the activation of mechanosensitive channels (Kralj et al., 2011).

*The ecopysiological impact of PR photosystems*

Meta-genomic, -transcriptomic, and -proteomic surveys of marine bacteria inhabiting the surface ocean environment establish PR as one of the most abundant genetic features, highly transcribed genes, and most abundant proteins. However, very little is known about how PR is integrated into the metabolism of marine bacteria and subsequently leveraged to improve their fitness.

I aimed to discover how light influences the growth phenotype under low and high DOC conditions using an RNA-seq approach. The data showed that PR was part of an orchestrated response to light that was not influenced by the DOC concentration of the media, yet the level of DOC dictated whether the light response enhanced or reduced growth rates and yields.

While there were some obvious differences between cultures maintained in the light or darkness in high DOC media, such as enhanced beta-oxidation among the dark cultures, future experiments are required to explain how the light response negatively impacts

growth in high DOC media while clearly enhancing growth in low DOC media. Understanding the environmental conditions under which the light response translates into faster and more efficient growth for MED134—and other Flavobacteria (Kim, 2013)—will help progress efforts to determine the impact of light energy and PR activity on carbon and energy flows through marine ecosystems.

## III. The Physiological Effects of Proteorhodopsin in *Escherichia coli*

ABSTRACT

The membrane potential produced by the process of respiration is suspected to limit the impact of light-driven proton pumps. Engineering applications of proteorhodopsin (PR) and other light-driven transmembrane ion pumps could be limited by the backpressure from respiration. New genetic backgrounds of *Escherichia coli* were created as part of this work to modify pools of NADH, ATP, and tricarboxylic acid cycle intermediates in order to reduce the steady-state proton motive force (PMF) of the cells by lowering respiration. Growth experiments to determine whether heterologous expression of PR improved specific growth rate or cell yield showed that one of the engineered backgrounds—a *fadR* knockout mutant of *E. coli* strain BW25113 with pRHA4 and pMCL8—exhibited higher cell yields in rich complex media as a result of light-driven proton pumping by PR. Based on this evidence, bioengineering applications that burden cells by depleting PMF and reducing equivalents may actually be required to reveal the full potential contribution of PR to cellular energetics in heterologous hosts.

## Introduction

Rhodopsin photosystems are a common trait among marine bacteria. As the simplest known light-harvesting protein complexes in nature, these photosystems can power cells by pumping a specific ion across the cytoplasmic membrane in a way that builds its electrochemical gradient. The proteorhodopsin photosystem (PR) was the first light-driven proton pump discovered within marine bacteria (Beja et al., 2000); and since then, an astonishing abundance and diversity of proton pumping rhodopsin photosystems have been found among heterotrophic bacteria that inhabit the photic zone of marine, freshwater, and terrestrial ecosystems (Bryant and Frigaard, 2006; Zubkov, 2009). In fact, there are now many examples of autotrophic and heterotrophic Bacteria, Archaea, and Eukaryotes that leverage PR to directly harness energy from sunlight (Béjà and Lanyi, 2014).

But despite their apparent utility, how light-driven proton pumps influence the lifestyle of the photoheterotrophs that deploy them is not well understood. They could be leveraged to provide maintenance energy during periods of starvation and otherwise do not influence cell energetics (Fuhrman et al., 2008; Gomez-Consarnau et al., 2010). It is also possible that photoheterotrophs utilize light-driven proton pumping to grow faster and more efficiently (Gomez-Consarnau et al., 2007; González et al., 2011), or to improve their fitness by increasing motility and transporter activities (Walter et al., 2007; Gómez-Consarnau et al., 2016). Ultimately the physiological effects of PR seem to depend on the metabolic capacity and lifestyle of its host.

*Escherichia coli as a heterologous host for PR gene expression*

So far PR-containing marine isolates that exhibit higher growth rates and cell yields in light are not genetically tractable. While there is still much to learn by studying how they operate (see Chapter 2), heterologous hosts offer well-known metabolic systems as a testing ground for observing the physiological effects of light-driven proton pumping by PR. *Escherichia coli* is a good host because its membrane environment is known to be well suited for PR production and function (Klyszejko et al., 2008; Lindholm et al., 2015), and because learning how PR can be leveraged in *E. coli* should aid engineering efforts to exploit PR for bioenergy production in other heterotrophs (Johnson and Schmidt-Dannert, 2008; Walter et al., 2010).

Expression of PR in *E. coli* has resulted in measurable increases in proton pumping and the concentration of ATP when cell suspensions were exposed to light (Beja et al., 2000; Wang et al., 2003; Martinez et al., 2007; Jung et al., 2008). Additional physiological effects associated with PR in *E. coli* include increased flagella rotation in the presence of light (Walter et al., 2007), and increased hydrogen production in the presence of light by a strain co-expressing hydrogenase with PR (Kim et al., 2012; Kuniyoshi et al., 2015). But despite these experiments demonstrating PR functionality and its apparent contribution to cell bioenergetics in *E. coli*, an increased growth rate or cell yield as a result of PR activity has never been reported.

The inability of PR to affect the growth phenotype of *E. coli* under standard aerobic conditions in minimal or complex media could be caused by respiratory inhibition of

light-driven proton pumping *in vivo* (Walter et al., 2007). This theory predicts that proton translocation by the respiratory complexes of the electron transport chain (ETC) establishes a membrane potential high enough to inhibit light-driven proton pumping by PR, which reportedly can only occur below a threshold potential that is never realized while cells grow aerobically on reduced carbon substrates (Friedrich et al., 2002; Walter et al., 2007). The back pressure from the membrane potential is also suspected to limit the current conducted through the ETC enzymes, as well as the reaction centers of photosynthetic cells (Burstein et al., 1979; van Rotterdam et al., 2002). Thus, it's likely that the respiratory rate needs to be reduced under standard aerobic growth conditions in order to observe the physiological effects of light-driven proton pumping by PR in *E. coli*. The following introductory sections will describe a number of practical approaches to modify the respiratory rate and other key variables that may influence PR production and function in *E. coli*. In this work, I aimed to identify a genetic background that allows light-driven proton pumping to positively influence the growth phenotype.

*Influence of the genetic background on respiration*

The respiratory rate can be assumed to be at its maximum capacity during aerobic growth under standard conditions (Burstein et al., 1979). The genetic background can be altered to reduce it, but any changes to existing genes or the addition of novel genes that have an effect on respiratory rate likely also impact fluxes of carbon, ATP and cellular redox ratios, and subsequently the activity of many other processes. Recombinant protein production is certainly one of the processes affected by a lower respiratory rate, so most

changes to the genetic background that reduce respiratory rate will also negatively impact PR production.

One of my objectives was to determine whether knockouts of catabolic genes that reduce $NAD^+$ or respiratory genes that oxidize NADH are suitable backgrounds for PR production, and to subsequently determine if PR enhances growth rate and yield in the mutants. The Δ*pykF*, Δ*icd*, Δ*fadR* and Δ*lpd* strains are single gene knockout mutants of the pyruvate kinase, isocitrate dehydrogenase, fatty acid degradation repressor, and the lpd subunit of several enzymes, respectively. The Δ*pykF* strain has been shown to reduce carbon flux toward pyruvate (Siddiquee et al., 2004), while the Δ*icd* and Δ*fadR* strains reduce flux toward the TCA cycle enzymes that evolve $CO_2$ and reduce NAD (Kabir and Shimizu, 2004; Peng and Shimizu, 2006). In the case of Δ*lpd*, flux toward pyruvate increases as a result of effectively knocking out three complexes that reduce $NAD^+$: pyruvate dehydrogenase, alpha-ketogluterate dehydrogenase, and the glycine cleavage system (Li et al., 2006).

Another objective of this work was to construct plasmids to complement native genes and/or heterologously express foreign genes that could potentially lower the $NADH/NAD^+$ ratio and subsequently reduce respiration. Two constructs expressing the foreign *nox* (NADH oxidase) and *pyc* (pyruvate carboxylase) genes were used to lower the NADH concentration, affecting respiration by competing with the ETC for NADH or by restricting the production of reducing equivalents.

*Influence of the genetic background on membrane properties*

Various properties of *E. coli* membranes can be modified at the genomic level to potentially create a membrane environment that enables PR to contribute to cellular energetics in ways that increase the growth rate or yield. The overexpression of membrane protein creates a bottleneck at Sec translocons (Wagner et al., 2007), which may be problematic in expressing the PR transmembrane protein in heterologous hosts like *E. coli*. All nascent membrane proteins and their associated ribosomes are directed to the translocons for insertion into the membrane, where they are forced to compete for available translocons. Flooding these systems with recombinant protein affects the insertion of native proteins, and increases the amount of native and recombinant proteins that become misfolded and subsequently degraded (Wagner et al., 2007).

A major downstream effect of the bottleneck is that respiratory enzymes become diluted in the membrane causing the respiratory rate of cells to decrease. While this could be viewed as a positive effect with respect to PR function, decreasing the respiratory rate in this way results in a more reduced quinone pool that subsequently induces the expression of *arcA*. And since ArcA represses the expression of most TCA cycle enzymes, the ultimate effect of overexpressing recombinant membrane protein such as PR is the shut down of the TCA cycle and the up-regulation of respirofermentive pathways that generate ATP by substrate-level phosphorylation and produce acetate (Wagner et al., 2007). If growth is limited by the highly reductive state of the cell and subsequent TCA cycle shutdown, PR production will be negatively effected and growth may stop before

cells become nutrient-limited or reach a growth state that is primed to benefit from light-driven proton pumping.

Increasing the rate of membrane protein insertion via the Sec translocons would likely delay the onset of the bottleneck upon induction of the recombinant protein and the subsequent buildup of reducing equivalents (Wagner et al., 2007; Nannenga and Baneyx, 2011). It has previously been shown that complementing the native *yidC* gene using a multicopy plasmid increased the production of recombinant membrane proteins, one of which being the proton-pumping deltarhodopsin HtdR from the cyanobacterium *Gloeobacter violaceus* (Nannenga and Baneyx, 2011). The increased production of HtdR was accompanied by faster growth to higher yields with *yidC* complementation than without it, indicating that YidC improved the insertion rate of HtdR and subsequently relieved (or delayed the formation of) the bottleneck at the translocons. An additive beneficial effect on HtdR production was observed when *yidC* was expressed within Δ*tig* backgrounds, which seemed to delay the degradation rate of nascent HtdRs bound to ribosomes essentially waiting for an available translocon (Nannenga and Baneyx, 2011).

Increasing the abundance of PR in the membranes should result in increased light-driven proton pumping when the PMF falls below the inhibiting threshold potential. And since proton pumping by PR is likely regulated by back pressure from the membrane potential (Friedrich et al., 2002; Walter et al., 2007), the goal in terms of PR production should be to maximize it so long as expression levels do not significantly interfere with the insertion of native proteins.

*Influence of the genetic background on acetate production*

Acetate production generates ATP by substrate-level phosphorylation and wastes carbon that could otherwise be channeled into biosynthesis or engineered pathways. The Δ*icd* and Δ*fadR* backgrounds reduce the flux of carbon directed towards respiration by inducing the expression of the glyoxylate shunt enzymes, which effectively redirect carbon from the $CO_2$ and NADH producing enzymes of central metabolism. Acetate production is reportedly reduced in these backgrounds because acetyl-coA is consumed as substrate in the glyoxylate shunt (Feng et al., 2015). The flux of carbon toward the glyoxylate shunt can also be modified by creating and *aceK* mutant allele (*aceK**) that only functions as an isocitrate dehydrogenase kinase, knocking down the activity of isocitrate dehydrogenases (Ikeda et al., 1992). Expressing the *aceK** allele should not have an effect in Δ*icd,* but could certainly increase flux through the glyoxylate shunt in the Δ*fadR* background.

Acetate production should also be reduced in a Δ*pta* background. In the case of Δ*pta*, the ATP yielding step of acetate synthesis is prevented so most of the acetate produced is a result of PoxB activity, which does not yield ATP but does produce $CO_2$ (Chang et al., 1999). Finally, another strategy that has worked successfully in *E. coli* to prevent acetate accumulation is the overexpression of *acs* (Lin et al., 2006; Valgepea et al., 2010). The acetyl-CoA synthetase encoded by *acs* consumes ATP and regenerates acetyl-coA from acetate. Thus, its introduction into backgrounds with an active glyoxylate shunt may further enhance the flux of carbon toward biosynthesis.

## Methods

*Bacterial strains and plasmids*

Table 1 lists the bacterial strains and plasmids that have been acquired or created for this project. Single gene knockout strains descended from BW25113 were acquired from the Keio Collection and the Km marker was removed with pCp20 (Baba et al., 2006). The ZK126 strains were a kind gift from R. Kolter (Zambrano et al., 1993), as was MB44 from K. Hellingwerf (Bekker et al., 2009), and ECOM4 from B. Palsson (Portnoy et al., 2010).

The pRHA801 Rhamnex vector (provided by DNA 2.0) and pMCL200 cloning vector (originally received from JGI) are compatible vectors that were used to host the various constructs created to manipulate the genetic background of BW25113 and related mutant strains. Table 2 lists the full collection of oligos (Invitrogen) designed for the construction of plasmids listed in Table 1, and includes notes regarding how each oligo was used. The synthetic promoter (SP1 and 2) was identical to what was used by Koebmann *et al.* to induce *atpAGD* for ATP hydrolysis from a multicopy plasmid (Koebmann et al., 2002). The synthetic terminator (ST1 and 2) B1002 has been previously characterized (Huang, 2008).

The Phusion PCR Kit (NEB) was used for all PCR reactions and the In Fusion Cloning Kit (Clontech) was used for all cloning reactions—except for the insertion of the constitute promoters to pMCL200, which was done by annealing two complementary

oligos with overhangs designed for efficient ligation (NEB). All PCR products and vectors linearized by restriction enzymes were gel-purified prior to cloning (Qiagen). When a ribosome binding site was required–for *kefFC** in pRHA4 and for all genes added to pMCL200 except for *yidC*—the De Novo DNA RBS calculator was used to design a unique RBS sequence specifically for the downstream gene to maximize translation initiation; this sequence could then be inserted by integration into the forward primer upstream of the start codon (Salis et al., 2009; Espah Borujeni et al., 2014).

Every construct was confirmed to be correct at each stage of cloning by restriction enzyme digests (NEB), with the exception of *kefFC** and *aceK** which were verified to be correct by Sanger sequencing (Quintara Biosciences). Plasmid maps and sequences for all constructs listed in Table 1 can be found in the Supplementary Materials.

*Growth experiments*

The cells for every growth experiment originated from a colony on a fresh agar plate containing the appropriate antibiotic(s) and 0.2% glucose. I found that PR expression was generally more reproducible using freshly transformed cells rather than cells that had been stored in glycerol and revived from -80°C. Conditions such as media composition and growth temperature of the starter culture were typically identical to media that would be used for the subsequent growth experiment. After growth to stationary phase, starter cultures were be diluted 1:50 into fresh media and grown to mid-exponential phase, at which time 500 μM Rhamnose was added for induction of PR-containing RHA plasmids.

Growth was allowed to continue to stationary phase before cells were used for subsequent growth experiments.

Growth experiments were performed in a light incubator refrigerated shaker (Beckman) modified to produce about 18 μmol photons/$m^2$/sec of green light (refer to Supplementary Material Figure S1). Cultures were maintained in 50 mL long screw top culture tubes (Pyrex) and were always angled at 45° to the light source. Rhamnose (typically 100-500 μM) was mixed into the fresh media with the antibiotics before 10 mL of media was transferred into each of the culture tubes. All-trans retinal (5 μm) was always added at t=0 just before cells from the induced pre-culture were inoculated into the media, and then again after 24-48 hours of growth to prevent any effects of photodegradation on the pool of all-trans retinal. An equal volume of ethanol (used to solubilize retinal) was added to cultures without retinal. Cells from the induced pre-culture were also incubated 45 minutes with or without retinal, and then washed once with cold culture media before resuspending and allocating equal volumes of each to the culture tubes. The exact culture conditions for each reported experiment can be found in the figure legends or within the text of the Results and Discussion section. Typically when 2xYT was used at pH 7.4, the media was buffered with 50mM MOPS (Sigma) and 50 mM MOBS (Sigma), supplemented with 100 ng/mL biotin for experiments with *pyc* (Gokarn et al., 2001), and the pH was adjusted with NaOH. When 2xYT was used at pH 7, 200mM of MOPS was necessary to restrict the pH of the medium from increasing above the intracellular pH (of about 7.6).

Samples taken during growth were immediately fixed with glutaraldehyde (5 μL/mL of 25% Grade I; Sigma), and subsequently stored up to one week before analysis at 4°C. To obtain cell counts, the fixed samples were diluted to a concentration within the range of accurate detection by the flow cytometer (roughly $5x10^4$ to $10^6$ cells/mL) and stained with Sybr Green (Invitrogen). Counting was performed using a Guava EasyCyte (Millipore) flow cytometer with the Incyte software package, and the data compared favorably with cfu counts plated at the time of sampling (see Supplementary Figure 2 and Figure 3).

**Results and Discussion**

*PR expression in wild-type backgrounds*

PR expression from pRHA1 and pRHA2 was tested in the *Escherichia coli* str. K-12 derivatives MG1666, ZK1142 and ZK1143, and BW25113 (Table 1). For the RHA vectors, the concentration of inducer (rhamnose) is correlated with the level of expression at the level of individual cells (Giacalone et al., 2006), enabling strict control over PR gene expression. Judging by pellet colors, PR expressed well in all wild-type genetic backgrounds; and for at least BW25113 and ZK1143, the flow cytometry data compared as expected to cfu counts (Supplementary Figure S1 and Figure S2). I found that expression of pRHA1 imposed less of a burden on growth relative to RHA2 in ZK1143 (Supplementary Figure S3), so RHA1 was used for all future growth experiments.

I measured the growth phenotype of BW25113 and ZK1143 in complex media including LB (Difco), 2xYT (Teknova), and Azure (Teknova), and minimal medias such as M9 (Sigma) and Medium KA (Ahmed and Booth, 1983a). LB and 2xYT were tested at full strength and at various dilutions with water, and all medias were tested with a range of carbon substrates at different concentrations, including various sugars and single amino acids. I also tested different growth temperatures, pH, shaking speeds, and green light intensities. Despite my efforts, I could not identify conditions that led to higher growth rate or yield for cultures that were capable of light-driven proton pumping (Beja et al., 2000; Martinez et al., 2007).

In case variable culture conditions between replicates were preventing the observation of a physiological PR effect, I used ZK1143 and the nearly isogenic strain ZK1142 to perform competition assays. The strains only differ by a neutral antibiotic marker (Zambrano et al., 1993; Finkel and Kolter, 1999), so any difference in growth rate or yield between the strains in co-culture should be a result of either PR expression or function. Since PR expression impacts growth, ZK1143+pRHA1 with or without retinal was compared to ZK1142 carrying empty pRHA. The relative abundance of each strain in co-culture was measured by cfu counts on LB plates with no antibiotic, with naladixic acid or with streptomycin. Despite these efforts, I was unable to identify a condition where PR clearly benefited the fitness of ZK1143+pRHA1 with retinal.

I did, however, identify one condition where the addition of retinal reduced growth rate compared with cultures that did not contain retinal. Table 3 provides growth rates and

generation times for BW25113+pRHA1 growing in Medium KA at pH 6, with or without the addition of retinal and lactose under constant illumination by green light. The data showed that light-driven proton pumping by PR decreased growth rate, but not by as much as the addition of lactose to cultures containing cells incapable of light-driven proton pumping. When lactose was added to cells capable of light-driven proton pumping by PR, its effect further amplified the negative influence that PR had on growth. These experiments were originally performed to determine if light-driven proton pumping could counteract the effect of lactose addition, which is known to decrease the rate of respiration by as much as 64% by diffusing across the membrane and dissociating into anions and protons in the cytoplasm (Ahmed and Booth, 1983a). Not only did PR fail to partially counteract the negative effect of lactose, its function seemed to have enhanced it at pH 6. The effect of light-driven proton pumping by PR on growth rate in KA medium was not observed at pH 7 (data not shown). These data may indicate for the first time that inward proton pumping by PR occurs at acidic pH *in vivo*. To date, inward proton pumping by PR at acidic pH has only been observed in liposomes reconstituted with PR (Friedrich et al., 2002; Pfleger et al., 2009).

**Table 1.** Bacterial strains and plasmids

| Strain or plasmid | Genotype or description | Hosted Genes | Reference |
|---|---|---|---|
| Strains | | | |
| BW25113 | *Escherchia coli* str. K-12 | | Baba *et al.* |
| MG1655 | Escherchia coli str. K-12 | | Portnoy *et al.* |
| ZK1142 | K-12 str. ZK126 nalR | | Zambrano *et al.* |
| ZK1143 | K-12 str. ZK126 strR | | Zambrano *et al.* |
| *ΔfadR* | JW1176 (BW25113 Δ*fadR*, Km marker removed) | | Baba *et al.* |
| *Δicd* | JW1122 (BW25113 Δ*icd*, Km marker removed) | | Baba *et al.* |
| *ΔpykF* | JW1666 (BW25113 Δ*pykF*, Km marker removed) | | Baba *et al.* |
| *Δlpd* | JW0112 (BW25113 Δ*lpd*, Km marker removed) | | Baba *et al.* |
| *ΔarcA* | JW4364 (BW25113 Δ*arcA*, Km marker removed) | | Baba *et al.* |
| *Δtig* | JW0426 (BW25113 Δ*tig*, Km marker removed) | | Baba *et al.* |
| *Δpta* | JW2294 (BW25113 Δ*pta*, Km marker removed) | | Baba *et al.* |
| *ΔnuoB ΔcyoB ΔappB ΔcydB* | MB44 (BW25113 Δ*nuoB* Δ*cyoB* Δ*appB* Δ*cydB*, Km marker removed) | | Bekker *et al.* |
| *ΔygiN ΔcyoABCD ΔcbdAB ΔcydAB* | ECOM4 (MG1655 Δ*ygiN* Δ*cyoABCD* Δ*cbdAB* Δ*cydAB*) | | Portnoy *et al.* |
| Plasmids | | | |
| **RHA801** | Medium copy number Rhamnex vector; KmR | | Giacalone *et al.* |
| **RHA1** | pRHA801+31a08 PR (green type) | PR | This study |
| **RHA2** | pRHA801+19p19 PR (blue type) | | This study |
| **RHA3** | RHA1+*kefFC* | PR, *kefFC* | This study |
| **RHA4** | RHA1+*kefFC** | PR, *kefFC** | This study |
| **MCL200** | Medium copy number general cloning vector; CmR | | Nakano *et al.* |
| MCL1 | pMCL200+SP1 | | This study |
| **MCL2** | pMCL200+yidCorf | *yidC* | This study |
| MCL3 | pMCL200+yidCorf+SP1 | | This study |
| MCL4 | pMCL200+yidCorf+SP2 | | This study |
| MCL5 | pMCL4+*nox* | | This study |
| **MCL6** | pMCL5+ST2 | *yidC*, *nox* | This study |
| **MCL7** | pMCL4+*atpAGD*+ST2 | *yidC*, *atpATG* | This study |
| **MCL8** | pMCL4+*pyc*+ST2 | *yidC*, *pyc* | This study |
| MCL9 | pMCL3+*acs* | | This study |
| **MCL10** | pMCL9+ST1 | *yidC*, *acs* | This study |
| **MCL11** | pMCL10+SP2+*pyc*+ST2 | *yidC*, *acs*, *pyc* | This study |
| **MCL12** | pMCL3+*aceK**+ST1+SP2+*pyc*+ST2 | *yidC*, *aceK**, *pyc* | This study |

* denotes allele with functionally significant point mutation
SP denotes synthetic promoter
ST denotes synthetic terminator
Bold type denotes expression plasmids (not cloning intermediates)
Parentheses denote plasmid alias

## **Table 2.** Oligonucleotides for plasmid engineering

| Oligo # | Oligo name | Sequence (5'->3') | Primer (Y/N) | Purpose |
|---|---|---|---|---|
| 1 | 31a08PR_f | **TCAGGAGGTGGTCGAC**ATGAAATTATTACTGATATTAGGTAGTGTTATTGCA | Y | In fusion cloning of green 31a08 PR w/ SalI & BamH1 linearized pRHA801 |
| 2 | 31a08PR_r | **GAGGGCGCGGGGATCC**TTAAGCATTAGAAGATTCTTTAACAGCAACATTCCA | Y | |
| 3 | 19p19PR_f | **TCAGGAGGTGGTCGAC**ATGTCAATAATTAAGTCTTTAAAGTCAGTCGCTAT | Y | In fusion cloning of blue 19p19 PR w/ SalI & BamH1 linearized pRHA801 |
| 4 | 19p19PR_r | **GAGGGCGCGGGGATCC**GCTGAATAAAGCGCAAGCACCCATT | Y | |
| 5 | kefFC_gDNA_f | **CCCCCTCGAGGTCGAC**ACAAAAGAACACAAAATAGAGGAGGTACCCAATGATTCTTATAATTTATGCGCATC | Y | In fusion cloning of the BW25113 kefFC genes w/ SalI & HindIII linearized pMCL3 |
| 6 | kefFC_gDNA_r | **ATTCGATATCAAGCTT**TTAGGATGAGGGTTTCGTTTC | Y | |
| 7 | kefFC_mut_f | TACCGGAAGTTTACTGCTCTCCAG | Y | Introduce R416S mutation (C->A) at position 1246 of kefC & In fusion clone |
| 8 | kefFC_mut_r | AGTAAACTTCCGGTAATCTGCCCAA | Y | |
| 9 | kefFC*_f | **TAATGCTTAAGGATCC**ACACAAAAGAACACAAAATAGAGGAGGT | Y | In fusion cloning of kefFC* w/ BamH1 & NdeI linearized pRHA1 |
| 10 | kefFC*_r | **ATTTCACACCGCATATG**TTAGGATGAGGGTTTCGTTTC | Y | |
| 11 | SP1 | **C**CATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTT**GGGCC** | N | Ligation of the Synthetic Promoter (SP) at position 1 w/ KpnI & ApaI linearized pMCL200 |
| 12 | SP1rc | **C**AACAGTACTGAACTATTATATCAGGCTCTACTTAATTGTCAAGAATAAACTACCCGATG**GGTAC** | N | |
| 13 | yidCorf_f | **TATACACTCCGCTAGC**GTTCAAGCTACGGAATTGAG | Y | In fusion cloning of the BW25113 yidC orf w/ NheI & KpnI linearized pMCL200 & pMCL1 |
| 14 | yidCorf_r | **TCGAGGGGGGGCCCGGTACC**TCGCTCATGATGTTCCTGT | Y | |
| 15 | SP2 | **AGCTT**CATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTT**A** | N | Ligation of the Synthetic Promoter (SP) at position 2 w/ HindIII & SpeI linearized pMCL2 |
| 16 | SP2rc | **CTAGT**AACAGTACTGAACTATTATATCAGGCTCTACTTAATTGTCAAGAATAAACTACCCGATG**A** | N | |
| 17 | nox_sp2_f | **AGTTCTAGAGCGGCCGC**TCCACATACGGAGCGGGGTCGAGGAGGTCAAATATGAGTAAAATCGTTGTAGTCGGT | Y | In fusion cloning of the S. pneumoniae nox gene w/ NotI & SacI linearized pMCL4 |
| 18 | nox_sp2_r | **TATAGGGCGAATTGGAGCTC[CCTGCAGG]**CTATTTTTCAGCCGTAAGGGCA | Y | |
| 19 | ST2 | **GGCTGAAAAATAGCCTGCAGG[CTTAAG]**CCCCGCTTCGGCGGGGTTTTTT**GAGCTCCAATTCGCCCTATA** | N | In fusion cloning of the Synthetic Terminator (ST) at position 2 w/ SbfI & SacI |
| 20 | ST2rc | **TATAGGGCGAATTGGAGCTCAAAAAACCCCGCCGAAGCGGGG[CTTAAG]CCTGCAGGCTATTTTTCAGCC** | N | |
| 21 | atpAGD_sp2_f | **AGTTCTAGAGCGGCCGC**GGGCCGCGACGACTAAGGAGGTACTAAATGCAACTGAATTCCACCGAAATCAG | Y | In fusion cloning of the BW25113 atpAGD genes w/ NotI & SbfI linearized pMCL6 |
| 22 | atpAGD_sp2_r | **AGCGGGGCTTAAGCCTGCAGG**TTAAAGTTTTTTGGCTTTTTCCACAGCT | Y | |
| 23 | pyc_sp2_f | **AGTTCTAGAGCGGCCGC**TACAAGCCAATCAAAATCCCCTAAGGAGGTAATACATGAAAATTAAAAAAGTCCTCG | Y | In fusion cloning of the MED134 pyc gene w/ NotI & SbfI linearized pMCL6 |
| 24 | pyc_sp2_r | **AGCGGGGCTTAAGCCTGCAGG**TTATTCAAGAGTTACAACGAGATC | Y | |
| 25 | acs_sp1_f | **TGGGCCCCCCCTCGAG**TAATCAACTTCCAAACGTACTTAAGGGGGTACTAC**ATGAGCCAAATTCACAAACACA** | Y | In fusion cloning of the BW25113 acs gene w/ XhoI & SalI linearized pMCL3 |
| 26 | acs_sp1_r | **TATCGATACCGTCGAC[ATGCAT]**TTACGATGGCATCGCGATA | Y | |
| 27 | ST1 | **GTAAATGCATGTCGAC[CTTAAG]**CCCCGCTTCGGCGGGGTTTTTT**ATCGATAAGCTTGATA** | N | In fusion cloning of the Synthetic Terminator (ST) at position 1 w/ SalI & BspDI linearized pMCL9 |
| 28 | ST1rc | **TATCAAGCTTATCGAT**AAAAAACCCCGCCGAAGCGGGG[CTTAAG]**GTCGACATGCATTTAC** | N | |
| 29 | aceBAK_gDNA_f | **CCCCCTCGAGGTCGAC**AAACGGATAAGACACGAAGATAAGGAGGTACCAATGACTGAACAGGCAACAAC | Y | In fusion cloning of the BW25113 aceBAK genes w/ SalI & HIndIII linearized pMCL3 |
| 30 | aceBAK_gDNA_r | **ATTCGATATCAAGCTT**TCAAAAAAGCATCTCCCCAT | Y | |
| 31 | aceK_mut_f | TGATTACAATGAAATTTGCTACATGACG | Y | Introduce D477N mutation (G->A) at position 1429 of aceK & In fusion clone |
| 32 | aceK_mut_r | ATTTCATTGTAATCATAAAAAACCACACGC | Y | |
| 33 | aceK*_sp1_f | **TGGGCCCCCCCTCGAG**GATACTCATATCCCTAAATAAGGAGGTTTACATATGCCGCGTGGCCTGGAA | Y | In fusion cloning of the aceK* allele w/ XhoI & SalI linearized pMCL11 |
| 34 | aceK*_sp1_r | **GGCTTAAGGTCGACATGCAT**TCAAAAAAGCATCTCCCCATACCGTACGCTA | Y | |

Bold type denotes regions of homology w/plasmid (not included in Tm)
Underlined type denotes synthetic rbs designed by DeNovo DNA
Brackets denote a novel restriction site added to the oligo

**Table 3. Growth rates of BW25113+pRHA1 in Medium KA at pH 6 exposed to light**

| Condition | Specific Growth Rate | Generation Time |
|---|---|---|
| -PR -lactose | 0.111 | 6.2 |
| +PR -lactose | 0.085 | 8.2 |
| -PR +lactose | 0.062 | 11.2 |
| +PR +lactose | 0.047 | 14.7 |

Addition of lactose to cultures reduces the specific growth rate of BW25113+pRHA1 induced with 1 mM Rhamnose. Reduction in growth rate by lactose is enhanced when retinal was added exogenously (+PR condition).
The growth rate is reported as $hr^{-1}$ and the generation times are in hours.

*PR expression in gene knockout backgrounds*

I tested several single gene knockout mutants derived from the Keio collection (Baba et al., 2006). The mutants I tested are listed in Table 1, and rationale for choosing these backgrounds has already been described (see Introduction). Unfortunately, I was not able to identify any single gene knockout mutants that experienced higher growth rate or yield as a result of PR activity.

I also investigated the effect of two mutant backgrounds with multiple gene deletions. Rather than conduct anaerobic growth experiments, I acquired two strains—derived from MG1665 (ECOM4) or BW25113 (MB44)—lacking all respiratory cytochromes that behave like obligate anaerobes (Bekker et al., 2009; Portnoy et al., 2010). These strains grew slowly, but I was able to transform pRHA1 into both and perform growth experiments with the extransformants. PR expression was not obvious from observing pellet color suggesting that lacking the capacity for aerobic respiration significantly reduced PR production. Thus, it was not surprising that I was unable to identify a condition where PR benefited the growth rate or yield of either of these strains. It is possible that the buildup of reducing equivalents restricted the growth (and PR production) of these strains, especially since fumarate or other anaerobic electron acceptors were not added to the growth media.

*The effects of yidC, atpAGD, and nox on BW25113 and ΔpykF strains expressing PR*

Since efforts to demonstrate a physiological effect of PR in wild-type and mutant genetic backgrounds were unsuccessful, I created plasmids designed to modify the background in

ways that would either reduce the ATP/ADP ratio by uncoupled ATP hydrolysis (*atpADG)*, or reduce the NADH/NAD+ ratio by uncoupled NADH oxidation (*nox)*.

ATP hydrolysis by the membrane-embedded $F_1F_o$-ATP synthase typically does not occur during aerobic respiration under standard conditions in rich media. Under these conditions the PMF is usually near the maximum (~0.2V) that can be supported by the membrane (Burstein et al., 1979; Walter et al., 2007). However, ATP hydrolysis can be engineered to occur under these conditions by complementing and overexpressing *atpAGD*—the $F_1$ portion of the ATPase—using a multicopy plasmid (Koebmann et al., 2002). This scheme results in the production of soluble $F_1$ ATPase in the cytoplasm, which decouples ATP hydrolysis from the generation of PMF.

Overexpression of *atpAGD* in wild-type *E. coli* growing on glucose minimal media increased glycolytic flux by as much as 70%, while reducing the growth rate by 24% and biomass yield by 55% (Koebmann et al., 2002). Since respiration cannot be increased to compensate for the reduction of ATP relative to ADP, the observed increase in glycolytic flux likely occurred to regenerate ATP by substrate-level phosphorylation reactions. And while the reduction in growth rate confirmed a reduced steady-state ATP/ADP ratio, the reduction in biomass yield provided the most direct evidence for ATP hydrolysis *in vivo* since biomass yield directly correlates to the amount of ATP produced (Russell and Cook, 1995; Koebmann et al., 2002).

While the effects of *atpAGD* overexpression in rich complex media lacking glucose are unknown, I reasoned that reducing the ATP/ADP ratio under standard aerobic conditions might induce the expression or stimulate the activity (turnover) of the native $F_1F_o$-ATP synthase (von Meyenburg et al., 1984; Jensen et al., 1993). In this case, the PMF could be decreased below the threshold potential necessary to observe light-driven proton pumping by PR despite respiration operating near full capacity.

Introducing *nox* from *Streptococcus pneumoniae* has been previously shown to directly lower the ratio of NADH to NAD by uncoupled oxidation (Vemuri et al., 2006b). While it may seem illogical to deplete cellular energy in this way, expression of *nox* in a Δ*arcA* mutant has actually been shown to increase recombinant protein production by delaying the ArcA-mediated induction of acetate production (Vemuri et al., 2006a).

Figure 1 shows how addition of these plasmids to the BW25113+pRHA1 and Δ*pykF*+pRHA1 backgrounds influenced their growth phenotypes in the presence of green light with or without exogenous retinal addition. The *yidC* plasmid (MCL2) was included as a control since *yidC* is also present on the plasmids carrying *nox* (MCL6) and *atpAGD* (MCL7). While neither the wild-type nor Δ*pykF* backgrounds containing pRHA1 with pMCL2, 6 or 7 resulted in significant differences with respect to the presence of retinal, the data suggested that reducing the flux of carbon from PEP to pyruvate by deletion of *pykF* enhanced the effect of the MCL6 plasmid (evidenced by lower growth rates). For Δ*pykF* backgrounds, the enhanced effect of pMCL6 as well as pMCL7 results in larger

but not statistically significant differences between cultures with and without retinal relative to the BW25113 background.

Perhaps more surprising was the observation that addition of pMCL6 actually increased cell yields relative to the control (pMCL2) in both BW25113 and Δ*pykF* backgrounds containing pRHA1. We did not expect that depleting cellular energy by uncoupled NADH oxidation would increase cell yields, nor has this been observed in previous experiments that involved the expression of *nox* in wild-type or mutant *E. coli* backgrounds. In fact, previous reports of heterologous *nox* expression in *Escherichia coli* and *Pseudomonas putida* have resulted in significantly lower growth rates and cell yields (Vemuri et al., 2006b; Vemuri et al., 2006a; Ebert et al., 2011). We interpreted this as evidence that the buildup of NADH upon induction of PR from pRHA1 limited growth yields in both backgrounds. Unfortunately, a constitutive promoter regulated *nox* expression in my experiments, so I could not increase its effect experimentally by increasing the concentration of inducer. I also could not generate extransformants of Δ*arcA* or Δ*fadR* containing pMCL6, likely due to the immediate effects of *nox* expression on colony formation (on selective agar plates). Thus, future growth experiments that test a modified expression vector containing a tightly regulated inducible promoter in the place of the constitutive promoter of pMCL6 should enable a larger host range for the *nox* gene with the potential to optimize expression to balance the NADH buildup originating from PR expression.

Like the Δ*arcA* background, the use of *nox* to reduce the NADH/$NAD^+$ ratio during PR expression likely only delays the eventual cession of growth. From a biotechnological perspective, excessive NADH is not a bad problem to have; the energy within NADH can be transferred uphill by reverse electron transfer to ferredoxin (Biegel et al., 2011), and subsequently used by hydrogenases and nitrogenases (Jeong and Jouanneau, 2000), or it could be directed to biosynthesis pathways by elevating conversion to NADPH via the addition of genes such as *nadK*, *pncB*, *gapA* and *pntAB* (Berríos-Rivera et al., 2002; San et al., 2002; Sauer and Eikmanns, 2005; Heuser et al., 2007; Ebert et al., 2011; Liang et al., 2013; Wang et al., 2013; Cui et al., 2014).

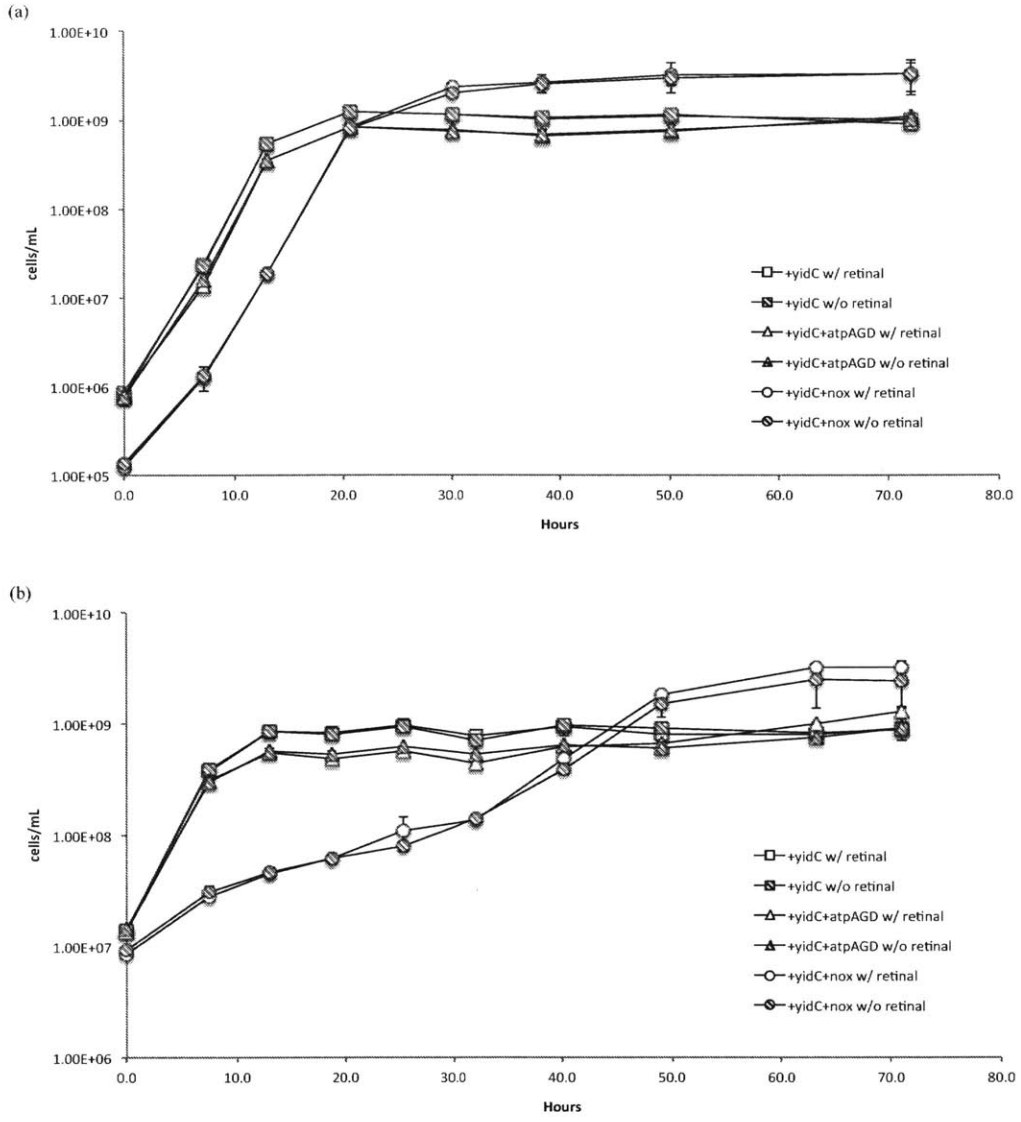



**Figure 1.** The effect of co-expressing pRHA1 with MCL2, MCL6, or MCL7 on the growth of BW25113 and Δ*pykF* (JW1666) with and without retinal

Constitutive expression of the *nox* gene (pMCL6) with PR (pRHA1) enhanced growth yields of BW25113 (a) and Δ*pykF* (b). Constitutive expression of *atpAGD* (pMCL7) reduced growth yields as expected during late exponential and early stationary phase. Growth was conducted in 2xYT media, pH 7.4, at 23°C. Error bars denote the standard deviation between duplicate cultures grown with or without retinal.

*The effects of yidC, pyc, and kefFC* on ΔfadR strains expressing PR*

The Δ*fadR* genetic background has modified TCA cycle activity due to the induction of the glyoxylate shunt, as well as changes to fatty acid metabolism and ultimately the fatty acid profile of the cytoplasmic membrane (Maloy et al., 1980; Campbell and Cronan, 2001; Peng and Shimizu, 2006; Chu et al., 2015).

Introducing a pyruvate carboxylase gene to the Δ*fadR* background should increase the flux of carbon to oxaloacetate and further reduce ATP/ADP and NADH/NAD+ ratios. The pyruvate carboxylase will compete with the pyruvate dehydrogenase for pyruvate, effectively redirecting carbon destined to become acetate—while producing NADH and ATP—to oxaloacetate instead (Gokarn et al., 2000; Gokarn et al., 2001). Modifying carbon flux with the pyruvate carboxylase should increase other anabolic reactions that support biosynthesis reactions as well (Sauer and Eikmanns, 2005). For example, expressing *pyc* in a Δ*fadR* background could result in even more carbon directed to malate and oxaloacetate for biosynthesis with a commensurate reduction in acetate (and ATP) production, similar to the introduction of *ppc* on a multicopy plasmid (Farmer and Liao, 1997). Since *Dokdonia donghaensis* strain MED134 was on-hand, its *pyc* gene was used for heterologous expression experiments in *E. coli*.

Figure 2 shows the results from introducing *pyc* (pMCL8) and PR (pRHA1) to Δ*fadR* strains and measuring growth with and without retinal. Although the average cell yields of stationary phase cultures containing retinal were higher than without retinal, the

differences were not highly significant; but samples taken after 16-hours of growth in the light exhibited student's t-test p-values < 0.15.

The yidC and *pyc*-modified Δ*fadR* genetic background was modified again by introducing *kefFC** to the pRHA1 plasmid, subsequently creating pRHA4. With the ultimate objective of reducing the back pressure from respiration in mind, the most obvious approach was to use an antiporter to spill energy stored within a dominant ion gradient. The native KefC $K^+/H^+$ antiporter is an attractive choice in this regard since its function simultaneously reduces the gradients of protons and potassium, but its natural gating prevents it from functioning under standard growth conditions (Ferguson et al., 1995; Ferguson and Booth, 1998). To abolish gating and constitutively activate KefC, a mutant allele (*kefFC**) was introduced to the pRHA1 vector (Miller et al., 1997). If *kefFC** expression ultimately results in a reduction in PMF by depleting the proton gradient, co-expression of PR should be able to partially compensate for this negative effect of *kefFC** by light-driven proton extrusion.

In addition to its affect on energy stored in the ion gradients, the expression of *kefFC** likely impacts turgor pressure and osmoregulation in general (Epstein, 1986; Epstein, 2003). If turgor is reduced by KefC* activity, *E. coli* could stimulate activity of inwardly directed $K^+$ pumps, such as the Kdp ATPase, which would essentially introduce a futile cycle of potassium ions when their concentration in the external media is low enough (Malli and Epstein, 1998; Roe et al., 2000). And due to the similarity between the

charges, sizes and behavior of $K^+$ and $NH4^+$, a futile cycle of ammonium ions can also be induced by supplementation with 100mM $NH_4Cl$ (Buurman et al., 1991).

I also anticipated that spilling energy stored in the potassium gradient by way of KefC would be more straightforward to interpret compared to the use of decoupling chemicals that require pretreatment of cells or potentially have off target effects (Ahmed and Booth, 1983b).

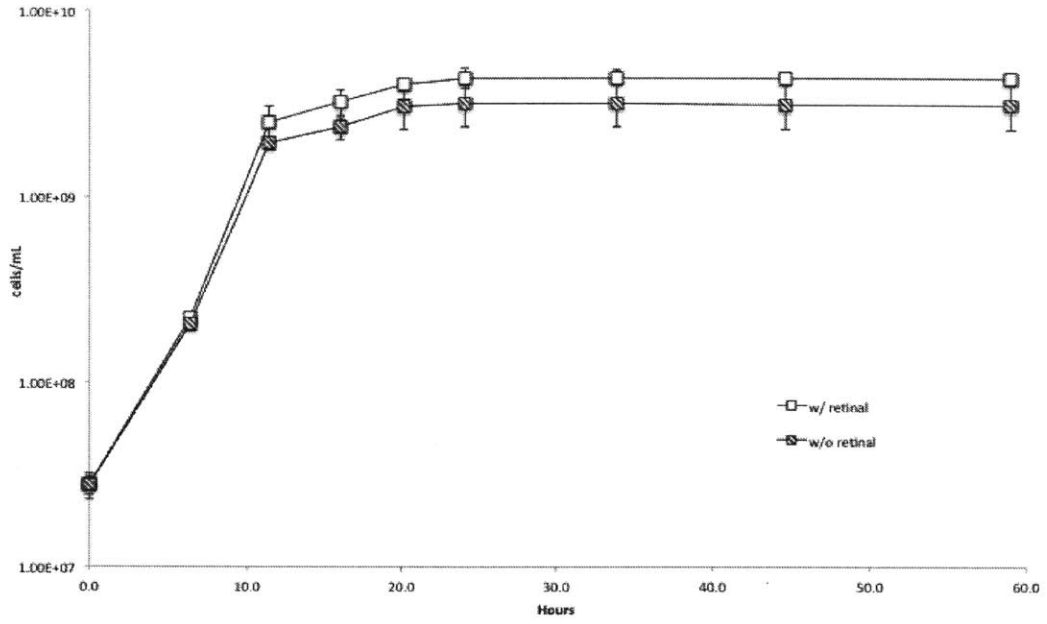



**Figure 2.** The effect of co-expressing pRHA1 and pMCL8 on the growth of Δ*fadR* (JW1176) with and without retinal

Constitutive expression of *pyc* (pMCL8) with PR (pRHA1) increased cell yields on average in the presence of retinal; but the difference was not highly significant ($p < 0.15$ after 16 hours of growth). Growth was conducted in 2xYT media, pH 7.4, at 33°C. Error bars denote the standard deviation between triplicate cultures grown with or without retinal.

To confirm that KefC* was acting constitutively as a KefC $K^+/H^+$ antiporter, I tested growth over a range of KCl additions to the 2xYT media. Figure 3 shows that without supplementation of 2xYT with KCl, the Δ*fadR* genetic background with pRHA4 and pMCL8 behaved similar as when pRHA1 was used (Figure 2), except that the difference between growth with and without retinal was determined to be significant for one sample (that at peak growth around 10 hours had a p-value of 0.085). However, when the 2xYT media was supplemented with 2 mM KCl, the cultures without retinal exhibited significantly higher yields ($p < 0.1$) throughout stationary phase relative to cultures with retinal. The same was true for cultures that were supplemented with 20 mM KCl (data not shown).

Figure 3 also shows that addition of 2 mM KCl marginally increased the yields of cultures that did not contain retinal, while the growth rates and cell yields of cultures containing retinal were inhibited (compare Figure 3a to 3b). Both this observation and the observation of increased cell yield on average for cultures with retinal (relative to without retinal) in the absence of KCl supplementation together suggested that KefC* constitutive activity was impacting the growth phenotype in a manner that was dependent on the presence of functional PR.

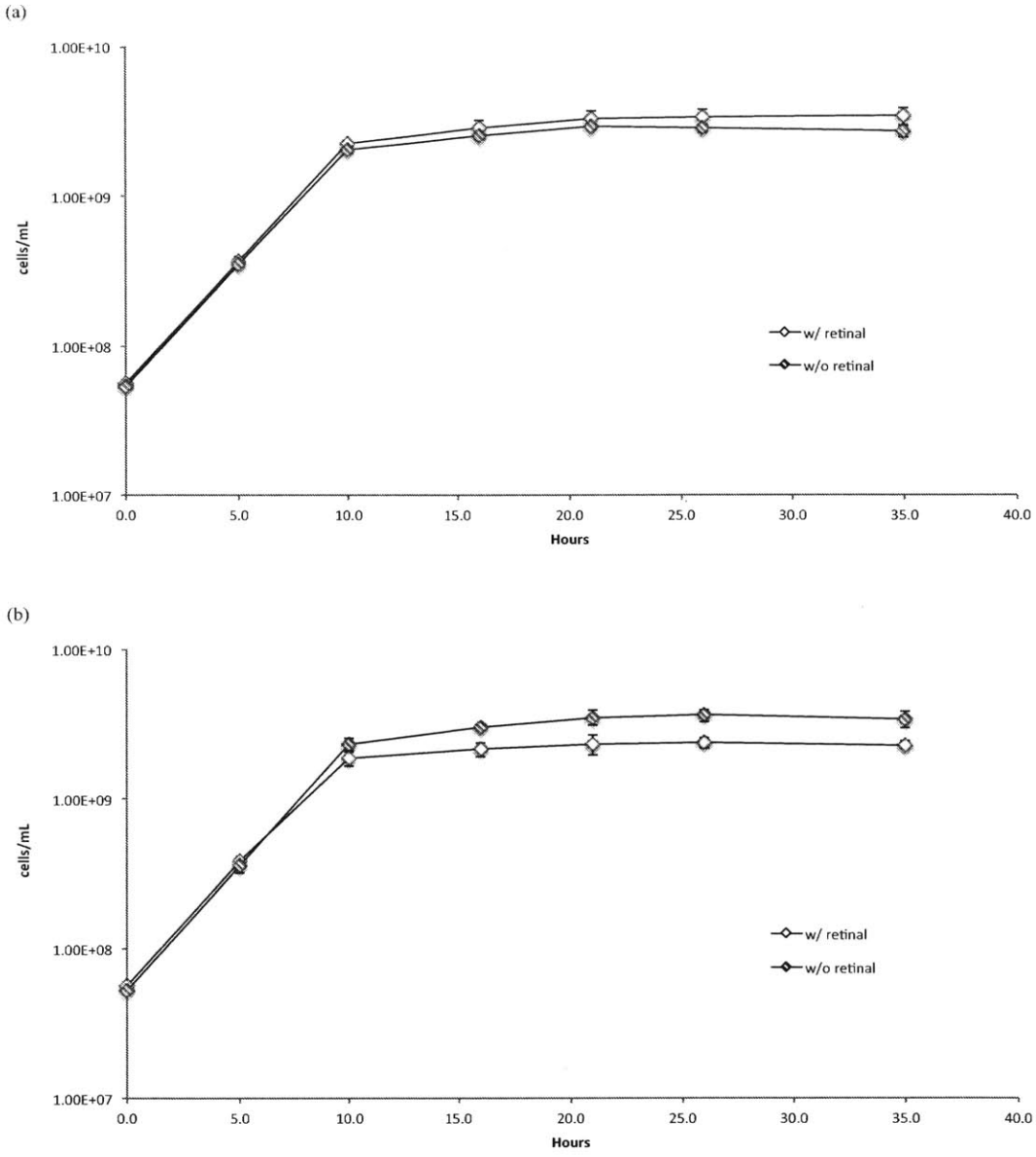



**Figure 3.** Effect of the media KCl concentration on growth of a *kefFC** expressing strain in the presence or absence of retinal with constant illumination

BW25113+pRHA4+MCL8 strain grown (a) without addition of KCl or (b) with addition of 2mM KCl. Growth was conducted in 2xYT media, pH 7.4, at 33°C. Error bars denote the standard deviation between duplicate cultures grown with or without retinal.

In line with my expectations, raising the extracellular potassium concentration likely resulted in less traffic through KefC* and thus a reduction in its effect on growth. Cytoplasmic acidification by KefC* may have reduced the contribution of the proton gradient to the PMF, lowering it, and activating light-driven proton pumping by PR. Alternatively, it is possible that KefC* reduced the concentration of intracellular potassium—a major intracellular osmolyte (Epstein, 1986)—enough to alter the properties of the cytoplasm and periplasm in a way that ultimately enhanced light-driven proton pumping by PR (Stock et al., 1977; Cayley et al., 2000; Akopyan and Trchounian, 2006).

Future experiments designed to assess *in vivo* activity of KefC* are required to prove its effect at different external KCl concentrations, optimally by incorporating a fluorescent sensor such as pHluorin or modified TorA to measure cytoplasmic pH, as well as measurements of cytoplasmic volumes (Cayley et al., 1991; Cayley et al., 2000; Wilks and Slonczewski, 2007; Kralj et al., 2011). Future experiments could also be designed to determine if common osmoprotectants such as glycine betaine and proline influence the growth phenotype of strains carrying pRHA4 over the same range of KCl concentrations (Record et al., 1998).

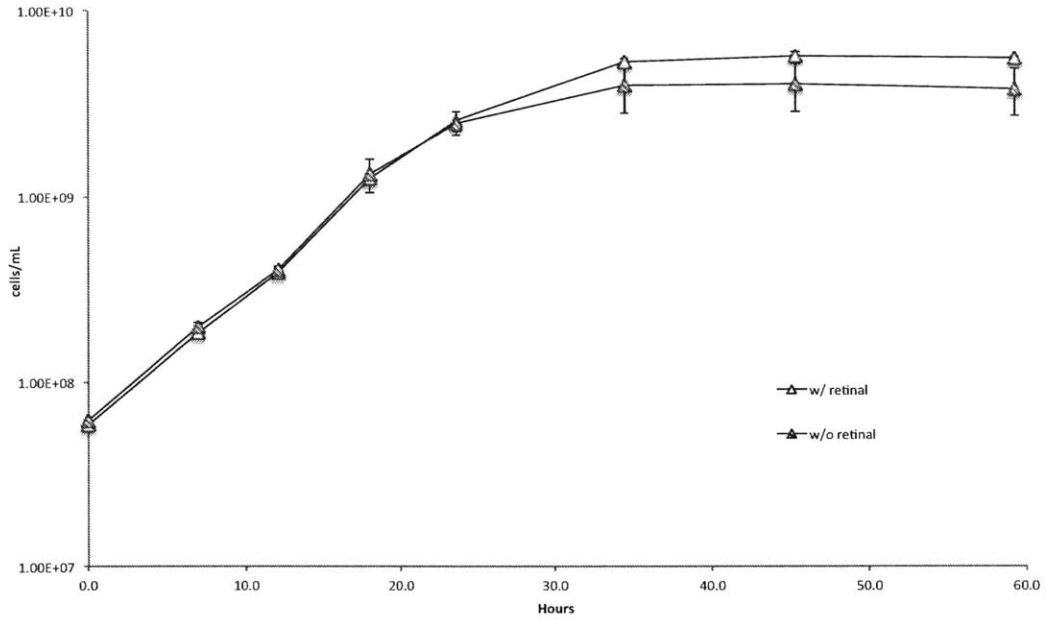



**Figure 4.** The effect of co-expressing pRHA4 and pMCL8 on the growth of Δ*fadR* (JW1176) with and without retinal in the presence of exogenous ammonium

The addition of 100 mM $NH_4Cl$ to the growth media resulted in significant differences between the stationary phase cell yields ($p<0.1$ for samples after 34 hours) of BW25113+pRHA4+pMCL8 cultures grown with and without retinal. Growth was conducted in 2xYT media, pH 7.0, at 28°C. Error bars denote the standard deviation between triplicate cultures grown with or without retinal.

Figure 4 shows that moderately significant differences ($p < 0.1$) were observed between the stationary phase cell yields of BW25113+pRHA4+pMCL8 cultures grown with and without retinal when 100 mM of ammonium chloride was added to the 2xTY media. Compared to the experiment that generated the results shown in Figure 3, the pH of the growth media used for this experiment was reduced from 7.4 to 7, and the growth temperature was reduced from 33°C to 28°C (in addition to the exogenous $NH_4Cl$). The excess of $NH_4Cl$ was expected to introduce a futile cycle to reduce the PMF by the active inward pumping of ammonium (via Kdp) and subsequent dissociation in the cytoplasm to freely diffusible ammonia and protons (Buurman et al., 1991). Published data indicates the influx of ammonium could take place under the growth conditions I used to serve as an osmoprotectant (Buurman et al., 1989).

Although introduction of *kefFC** to the genetic background together with the ammonium addition resulted in lower specific growth rates, as expected, these modifications actually improved cell yields (compare Figure 4 to Figure 2). Furthermore, the cell yields in the presence of ammonium were higher than without ammonium exclusively for cultures that contained retinal. Similar to the results shown for *nox* in Figure 1, the results shown for *kefFC** with ammonium in Figure 4 suggest that decreasing the energy level of the cells exposed the influence of light-driven proton pumping on growth physiology. However, since retinal was not shown to be an adequate control under these growth conditions, the experiment was repeated with the inclusion of cultures maintained in the dark.

Figure 5 shows that final cell yields were slightly higher on average for cultures with retinal than without retinal or that were maintained in the dark. Despite growing faster, the cultures maintained in the dark exhibited significantly lower cell yields than those maintained in the light (both containing retinal) with t-test p-values $< 0.1$ for the final two stationary phase samples. The difference between cell yields of stationary phase cultures grown in the light with and without retinal was less significant in this experiment ($p<0.2$) compared with the results presented in Figure 4 ($p<0.1$). The less significant difference between cultures grown with and without retinal in the light may have been a result of lowering the pH of the media in this experiment from 7 to 6.8 since this was the only difference between the growth conditions that produced the results for Figure 4.

In standard LB media the wild-type cultures expressing PR maintained in the dark had lower growth rates and yields than light-exposed cultures growing at 37°C (see Supplementary Figure 1 and Figure 2). But with the strain and growth conditions used to produce the results presented in Figure 5, the specific growth rates of cultures maintained in the dark were higher than those in the light. Published data suggests that the high concentration of MOPS buffer used in this experiment could have resulted in the production of enough hydrogen peroxide upon exposure to light to cause the observed inhibition of growth among the light-exposed cultures (Morris and Zinser, 2013). Since the light exposed cultures without retinal exhibited the same yields as the dark maintained cultures with retinal, the influence of MOPS in the light condition did not ultimately affect final yields, or by relation, the ratio of ATP/ADP of cells exposed to light.

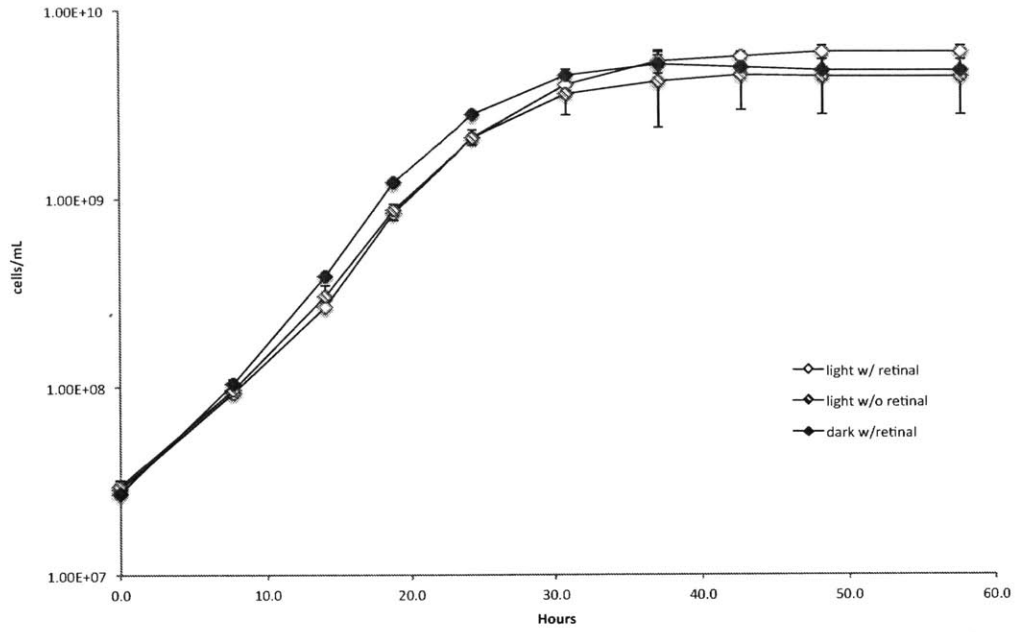



**Figure 5.** The effect of co-expressing pRHA4 and pMCL8 on the growth of Δ*fadR* (JW1176) in the presence of excess ammonium with and without retinal or light

Constitutive expression of *pyc* (pMCL8) with PR+*kefFC** (pRHA4) increased growth yields of Δ*fadR* in 2xYT in the presence of 100mM ammonium chloride with retinal relative to without retinal or without green light. Faster growth rates of the dark maintained cultures are likely the result of MOPS photoinhibition among the illuminated cultures. Growth was conducted in 2xYT media, pH 6.8, at 28°C. Error bars denote the standard deviation between triplicate cultures.

## Conclusion

Light-driven proton pumping contributes to cell energetics during aerobic growth in rich complex media if the genetic background is modified to restrict NADH production. Increased growth yield indicated that light-driven proton pumping elevated the ATP/ADP ratio in the modified strain (Russell and Cook, 1995; Koebmann et al., 2002), and data from *nox* gene expression in another modified strain suggested that cell yields (and thus ATP production) can be increased by the uncoupled oxidation of NADH as well. Based on this evidence, bioengineering applications—such as the introduction of hydrogenases and nitrogenases that burden cells by drawing PMF and reducing equivalents—may actually be required to reveal the full potential contribution of PR to cellular energetics.

Unfortunately, the collection of plasmids I built to manipulate the genetic background (MCL series) use constitutive promoters, with the exception of the native promoter used for *yidC*. And furthermore, *kefFC** cannot be induced separately from the PR because they share the same rhamnose inducible promoter. A slightly different configuration of these genetic elements is required to fully diagnose and improve upon the documented influence of PR in the genetic background of the fadR+pRHA4+pMCL8 strain. Ideally, PR would be moved to the chromosome by an established allele replacement method (Metcalf et al., 1996), and two expression vectors with different origins of replication and inducing agents would be used for *kefFC** and *pyc*, with the *yidC* native sequence following downstream of either gene. A light inducible promoter would be an intriguing choice for *kefFC** expression since its effects are only desired in the presence of light-

driven proton pumping, although this would prevent use of dark maintained control cultures (leaving only the without retinal control for relative functional comparisons). Further consideration for engineering *nox* into the genetic background with *kefFC** should also be considered.

*Future optimization experiments*

Various properties of *E. coli* membranes can be modified at the genomic level to potentially create a membrane environment that enables PR to contribute to cellular energetics in ways that increase the growth rate or yield. While my focus for optimization in this work was on the YidC chaperone protein, several other ways to modify the membrane properties should be explored to investigate their effects on PR functionality with the goal of optimizing the impact of PR on growth grates and yields.

Indole is a protonophore that directly reduces the resistance of the membrane to protons (Chimerel et al., 2012; Chimerel et al., 2013) effectively short circuiting the transmembrane proton gradient created by proton pumps like cytochromes and rhodopsins. The production of indole occurs naturally in *E. coli* during batch cultures, and its involved in the transition from exponential to stationary phase (Lelong et al., 2007). Indole synthesis can also be naturally induced by plasmid dimers through a regulatory RNA that upregulates tryptophanase activity, and in this case indole is produced to high enough levels to halt cell division (Chant and Summers, 2007). Thus complementing rcd, the regulatory RNA, on a multicopy plasmid offers a way to control indole synthesis. Alternatively, the addition of exogenous indole to a final concentration

of 3-5 mM can restrict cell division (Chimerel et al., 2012). In the case of *hns* mutants, 3-5 mM of exogenous indole can render cells into a quiescent but metabolically active "Q-Cell" state that is reportedly favorable for recombinant protein production and subsequent chemical synthesis because resources are not channeled into unwanted biomass (Chen et al., 2015).

The specific composition of fatty acids that form the cytoplasmic membrane can also be optimized for PR functionality. Membrane fatty acids directly impact PR assembly within the membrane and its subsequent function (Hussain et al., 2015; Lindholm et al., 2015). Although the membrane of *E. coli* is generally regarded as a favorable environment for PR, it seems that reducing lipid chain lengths and increasing the degree of saturation may improve membrane conditions for hosting PR (Lindholm et al., 2015). In addition to affecting carbon flux through the TCA cycle (Maloy et al., 1980; Peng and Shimizu, 2006), the Δ*fadR* mutant also increases the degree of saturation of its membrane fatty acids by inhibiting the expression of *fabB* (Campbell and Cronan, 2001; Chu et al., 2015)

A significant finding by Linholm and colleagues was that the charge of lipid head groups seems to be important for PR organization in the membrane (Lindholm et al., 2015). Additionally, the amount and types of anions bound to the cytoplasmic membrane, especially in the periplasmic space, have a significant role in creating the Donnan potential (negative inside) that impacts osmoregulation and the association of free monovalent and divalent cations with the membrane (Stock et al., 1977), which may affect the function of PR (Hussain et al., 2015; Lindholm et al., 2015). While efforts were

not focused particularly on modifying lipid charges or fixed periplasmic anions in this project, these could be targets for further enhancement of any observed physiological effects of PR in *E. coli*.

Introducing novel active transport systems or increasing the expression of native systems by complementation on a multicopy plasmid could reduce cellular PMF or the ATP/ADP ratio, creating an opportunity for light-driven proton pumping to influence cell growth rates and yields. The active transport of sodium, alanine, and vitamins, among other chemicals, seem to be involved in the natural growth response of Flavobacteria to light (Kimura et al., 2011; Gómez-Consarnau et al., 2016), suggesting that light-driven proton pumping can be leveraged to compensate for elevated transport functions.

The activation of passive channel proteins, mechanosensitive proteins, or voltage-gated channels that enable protons and other ions to flow down their concentration gradients in uncoupled fashion also represent potential strategies for depleting the membrane potential and relieving the back pressure associated with it (DeCoursey and Hosler, 2014), but the utility of these transporters was not explored as a part of this project.

In addition to optimizing membrane properties, future experiments that make use of a small RNA knockdown library of metabolic enzymes could speed efforts to find additional combinations of gene knockdowns that enhance the effect of light-driven proton pumping on the growth phenotype of *E. coli* (Nakashima et al., 2014).

*Future applications of PR and other ion pumping rhodopsin photosystems*

Bioenergy—renewable energy produced by organisms—is an insignificant piece of the industrial energy landscape in today's world (EIA, 2016). But the future potential of technologies capable of leveraging bioenergy to produce chemical fuels and electricity is extremely bright. A primary goal of current research efforts is to transition away from technologies that harvest bioenergy by brute force methods, such as burning, to processes that utilize design principles from biological systems to more efficiently extract bioenergy from biomass sources (Stephanopoulos, 2007; Mawhood and Gazis, 2016). While some sources of biomass come from waste products like wood chips and used cooking oils, large-scale biomass production from crops like switchgrass is typically required to obtain high yields of ethanol or other useful chemicals. In either case, collecting the biomass and transporting it to production facilities increases per unit costs and decreases economic feasibility relative to other energy sources.

Engineering organisms to directly produce renewable fuels and electricity from less starting materials would avoid some problems associated with traditional biomass sources (Lovley, 2006; Keasling, 2008; Mukhopadhyay et al., 2008; Carothers et al., 2009; Lovley, 2012). Naturally, algae and photosynthetic bacteria are a good starting point because they do not require expensive reduced carbon sources to grow (Ghirardi et al., 2005; Hannon et al., 2010; Rosenbaum et al., 2010; Paul Abishek et al., 2014). However, using algae and photosynthetic bacteria in engineered systems does limit the range of chemicals that can be sustainably produced.

The ability to convert heterotrophic bacteria to photoheterotrophs by the addition of PR or other light-driven ion pumps would enable synthetic biologists to leverage the extensive metabolic versatility of heterotrophic bacteria for the renewable production of reduced chemicals and electricity (Johnson and Schmidt-Dannert, 2008; Johnson et al., 2010; Kim et al., 2012). For example, oxygen-sensitive enzymes like [FeFe]-hydrogenases and nitrogenases could be more easily incorporated into production schemes since heterotrophs do not evolve oxygen and most are capable of growing anaerobically or microaerobically. Ultimately, the utility of light-driven ion pumps will depend on a better understanding of the conditions under which they are functional and significantly contribute to cell energetics.

## IV. Conclusions and Future Directions for Research

The main findings of this thesis are

1) Confirmation that proteorhodopsin (PR) was induced by light in the PR-containing marine isolate *Dokdonia donghaensis* strain MED134.
2) Induction of PR by light in MED134 occurred to a similar extent in complex media supplemented with low and high amounts of dissolved organic carbon (DOC).
3) Most of the other genes induced by light in low DOC media were induced in high DOC media as well, indicating that the genetic response of MED134 to light is largely independent of DOC concentration of its environment.
4) Most of the other genes involved in the MED134 light response translate into products with unconfirmed or unknown functions that might contribute to the physiological influence of PR in the marine environment.
5) Heterologous expression of PR in *Escherichia coli* wild-type strains did not measurably improve growth rates or yields of aerobically grown cells cultured under my experimental conditions in minimal or complex medias with low or high amounts of carbon and nutrients.
6) Heterologous PR expression caused reducing equivalents to accumulate to an extent that ultimately limited cell yields in batch culture.
7) A modified genetic background of *E. coli* exhibited slightly higher cell yields presumably as a result of light-driven proton pumping by PR.

**Light enhanced growth in natural PR-containing bacteria**

As a product of this thesis project, the effect of light on growth rates and yields of the PR-containing flavobacterium MED134 was confirmed in low and high DOC media. The presence of light increased growth rates and yields in the low DOC media, but reduced rates and yields in the high DOC media. I confirmed that expression of PR was induced by light in low DOC media, and reported for the first time that it is also induced to a similar extent in high DOC media. Along with the PR gene, light significantly induced the expression of 10 other genes in low and high DOC media. These findings indicated that the response of MED134 to light is independent of the DOC concentration (and osmolarity) of the growth media.

Given that the response to light on the level of gene expression was the same in low and high DOC media, it is interesting that growth was enhanced in the low DOC media but inhibited in the high DOC media in the presence of light. Clearly, the induction of these genes by light had different effects on the physiology of MED134 depending on the level of DOC in the culture media. New experiments are required to determine the function of most of these genes, and subsequently how each may impact the growth phenotype in a DOC-dependent fashion.

Some other genes were found to be differentially expressed either before or after the suite of light-induced genes were discovered. At an earlier time point in low DOC media none of the light-induced genes were differentially expressed; instead the light-exposed cultures up regulated iron acquisition genes whose products depend on TonB transport.

This finding expands the range of TonB-dependent transport processes that have been implicated in the MED134 light response (Gómez-Consarnau et al., 2016). After the observation of the light-induced genes in high DOC media, differential expression analysis of light and dark cultures at a later time in stationary phase showed that cultures maintained in the dark up regulated beta-oxidation and heat shock related genes, as opposed to light-exposed cultures which up regulated only PR and the purine nucleoside phosphorylase genes. While it is clear that light and dark cultures relied upon different metabolic strategies to persist in stationary phase, exactly how the MED134 light response resulted in the observation of divergent metabolic activities is unknown. Repeating the RNA-seq experiment with MED134 in high DOC media with more samples for RNA-seq analysis between T1 and T2 might reveal the progression of changes in gene expression that led to the differences I observed at T2. However, slower growth rates and lower growth yields among light-exposed cultures in high DOC media were observed even before the T1 sample was taken Chapter II, Figure 1), so its likely that the behavior of these cultures in stationary phase is simply a result of metabolic differences between light and dark cultures that culminated within the first 24 hours of growth.

It is also important to mention that the pre-culture for the RNA-seq experiment was maintained in the dark, and that cells used in this experiment grew together for two days (also in the dark) before being split into the light and dark treatments. Thus, it is possible that growth inhibition by light in high DOC was simply the result of strong induction of the light-induced genes; and it just so happens that many of these genes contain domains

that are contained within known membrane-embedded or secreted proteins that require translocons (similar to the observed effect on growth of PR overexpression *E. coli*; see Chapter III). Perhaps exposing the pre-cultures to light would equalize the burden of expression associated with the light-induced genes and result in less inhibition of growth between the light and dark treatments in high DOC media.

If strong induction of the light-induced genes influenced the growth phenotype negatively in high DOC, then in low DOC media the observed growth phenotype implies that any burden associated with the expression of the light-induced genes was compensated for by the effects deriving from those gene products, including the enhancement of light-driven proton pumping by PR. But the energy state of the cell growing in low DOC media is presumably less reduced than in the high DOC media, so the burden associated with the expression of the light-induced genes in low DOC media may simply be lessened (because reduced quinone and NADH never reach high enough concentrations to stimulate *arcA* and reduce carbon flux through the TCA cycle). If MED134 were genetically tractable, gene deletions or additions that reduce NADH production or increase its oxidation could be tested to determine if a smaller NADH pool subsequently lessened the effect of light (i.e. gene induction) on the growth phenotype in high DOC.

Along the same lines of reasoning, the importance of a cold incubation before starting a growth experiment with MED134—or other Flavobacteria (Kim, 2013) known to exhibit light-enhanced growth in low DOC media—cannot be overstated. If cells are transferred directly from the pre-culture in stationary phase to the experimental cultures (with or

without washing) the light-enhanced effect on growth may not be observed. In fact, I cultured several of the Flavobacteria from Yoshizawa *et al.* (2012) following the established protocols (Gomez-Consarnau et al., 2007; Kimura et al., 2011; Kim, 2013) and did not observe light-enhanced growth in low DOC media after a three-day cold incubation between harvesting the cells from the pre-culture and subsequently using the cells to begin a growth experiment. The growth curves obtained from these experiments (not reported in this thesis) looked different relative to published experiments; a long lag phase of at least 24 hours and subsequently slow growth was replaced by nearly instantaneous exponential growth to peak yields (which were similar to those published for a given carbon level) within the same 24 hour period.

The requirement for incubating the cells at 4°C in order to observe light-enhanced growth may be related to potential depletion of endogenous stores of carbon and energy stockpiled by cells during growth in the pre-culture. The influence of endogenous carbon stores may even explain the observation that MED134 and other flavobacteria grow in ASW without carbon supplementation (Gomez-Consarnau et al., 2007; Kim, 2013). If the depletion of endogenous carbon and energy is the underlying reason that a cold incubation is required to observe light-enhanced growth in low DOC media, then future experiments should consider replacing the cold incubation with several daily serial transfers at the standard growth temperature (where cells are transferred into fresh ASW supplemented with non-carbon nutrients at 24 hour increments) until the cells no longer grow without exogenous carbon supplementation. Treating the cells in this way prior to

their use in growth experiments could increase the consumption of endogenous energy stores and dilute unconsumed energy stores by cell division.

**Engineering *E. coli* for light-enhanced growth with PR**

Light-driven proton pumping only enhances growth of MED134 in low DOC media; and without genetic tools, it's difficult to determine *how* PR enhances growth *and why* it only enhances growth in low DOC media. Switching from the natural system to heterologous expression of PR in *E. coli* presents a unique opportunity to study PR in the context of a genetic background that is easily manipulated. Of all possible heterologous hosts, *E. coli* was the best choice since large amounts of functional PR can be produced, its metabolism has been well characterized, and the physiological effects of thousands of physical and genetic perturbations have been reported (see Chapter I).

From a reverse engineering perspective, *E. coli* offers a novice genetic background with respect to the PR gene—and by this I mean evolution hasn't acted to regulate PR expression or the membrane environment that its plugged into. Since PR is functional when expressed (Martinez et al., 2007), engineering the membrane should be reserved for optimizing a desired phenotype (once it's discovered). Focusing on PR expression led to the realization that overexpression of PR (on its own, not considering any physiological effects of the membrane-embedded protein) affects metabolism and can ultimately limit cell growth through the activation of *arcA* (Wagner et al., 2007). Several single-gene

knockout mutants were tested as genetic backgrounds for PR expression but many of these seemed to only delay the eventual cession of growth caused by the induction of PR. Obviously lower inducer concentrations were tested but since the effect of PR on growth presumably depends on its abundance in the membrane, a middle ground had to be discovered.

One of the most efficient approaches to reduce the apparent effect of PR overexpression was to introduce the *nox* gene from *Streptococcus pneumoniae* on a multicopy plasmid downstream of a constitutive promoter (pMCL6; see Chapter 3). This construct reduced the specific growth rate of wild-type *E. coli* and the mutants I tested it within (Chapter 3, Figure 1), but the yields of strains expressing PR with pMCL6 were significantly higher than those without it. Since introducing the *nox* gene decouples NADH oxidation from the ETC and reduces the NADH/$NAD^+$ ratio, it has only been reported to significantly reduce cell yields in *E. coli* (Vemuri et al., 2006b). Exactly how *nox* increases cell yield when co-expressed with PR is unknown, but this finding supports the notion of growth inhibition by *acrA* (Wagner et al., 2007).

While useful for modulating cellular levels of NADH, *nox* isn't likely to be useful in engineered applications with PR because its product, NADH oxidase, essentially wastes useful energy. On the other hand, if a strain was engineered to depend on PR and *nox* to grow in a particular environment, nox could conceivably then be swapped out of the background for another more useful NADH consuming enzyme. The RNF complex (Biegel et al., 2011) uses the PMF to convert NADH to the more reduced electron carrier

ferredoxin, which can serve as a source of electrons for a hydrogenase or nitrogenase (Jeong and Jouanneau, 2000). Since PR has already been shown to enhance hydrogenase activity in *E. coli* (Kim et al., 2012; Kuniyoshi et al., 2015), *rnf* could be added to a similar background and efficacy of the novel system could be easily determined by comparing hydrogenase activity in the presence and absence of green light and with or without functional RNF.

Another strategy worth investigating is to introduce *nox* and PR to an NADH dehydrogenase (NDH-I; *nuo*) mutant background. This approach could reduce proton translocation by respiratory complexes without significantly affecting the $NADH/NAD^+$ ratio. If PR influenced growth in this background, deleting one or two of the respiratory cytochromes might further reduce coupled oxidation of NADH, so long as this modification doesn't lead to *arcA* activation by the reduced quinone pool. Some or all of these modifications could reduce the PMF enough to create a genetic background that becomes dependent on light-driven proton pumping to maintain flux through the TCA cycle; because without active PR, cells would be restricted to generating most of their PMF from ATP produced (less efficiently) by substrate-level phosphorylation.

If the goal is to increase biomass yield with PR, other ways to modify the genome in order to manipulate NADH levels should also be explored. As mentioned in Chapter III, converting the NADH to NADPH could increase carbon flux to biosynthesis pathways relative to catabolism. The *aceK** allele could also be useful in this regard by redirecting carbon from isocitrate directly toward malate and oxaloacetate if the glyoxylate shunt

enzymes are also expressed. My attempts to test *aceK** seemed to increase the biomass content of Δ*fadR* cells (based on forward scatter measurements of single cells, data not reported in this thesis) but did not result in differences in cell yields between cultures with and without retinal. The same result was also found when *acs* was overexpressed in the Δ*fadR* background. However, co-expression of *pyc* (from MED134) and PR in Δ*fadR* very slightly increased cell yields of cultures containing retinal relative to those without retinal (Figure 2, Chapter III). Presumably the slight enhancement of cell yield by light-driven proton pumping under these conditions was a result of lowering NADH production by the pyruvate dehydrogenase by redirection of some portion of the pyruvate to oxaloacetate via Pyc. In addition to reducing NADH production and enhancing carbon flux to oxaloacetate, Pyc prevents subsequent substrate-level phosphorylation of acetyl-coA from the Ack-Pta pathway and increases ATP consumption (since ATP is a substrate for Pyc).

The addition of *kefFC** to the genetic background enabled control over the potassium content of cells, which showed that higher levels of intracellular potassium negatively influenced cell yields with retinal relative to without it (when the potassium gradient was abolished by *kefFC**; Figure 3, Chapter III). The reason(s) that retinal (i.e. light-driven proton pumping) increased cell yield relative to cultures without retinal when no potassium was added to 2xYT, but the opposite was observed with 2mM or 20mM KCl addition remains unclear. The result could be explained by the level of KefC* activity, which is assumed to be highest at the lowest extracellular potassium concentration; since $K^+$ is concentrated in cells and KefC* enables it to flow down the concentration gradient,

the flow will stop when the extracellular concentration of potassium is equal to the intracellular. And since the flow of potassium out is opposed electronically by the inward flow of protons, spilling energy concentrated in the potassium gradient with KefC* also spills energy from the proton gradient. If the intracellular potassium concentration drops low enough that the cells experience a reduction in turgor or general osmotic stress, a futile cycle with potassium ions could be established by the activation of the potassium ATPase (Kdp), which is inhibited by high ~1mM extracellular KCl (Malli and Epstein, 1998; Roe et al., 2000); the futile cycle would consume ATP to actively transport potassium inward just to have it subsequently flow out with inward proton pumping via KefC*. The potassium content of the 2xYT media was not determined, but if the media could be formulated with lower amounts of potassium (or free potassium can be bound by a media additive) than the data suggests that the effect of KefC* on the growth phenotype may be further enhanced. While I have not mentioned the role of PR in the system with KefC*, it should be clear that its activity could reduce the acidification of the cytoplasm and contribute to the membrane potential and proton gradient (i.e. PMF) by electrogenic $H^+$ transport across the membrane. Since respiration is likely operating at max capacity in aerobic *E. coli* (Burstein et al., 1979; Walter et al., 2007)—and thus unable to be increased in the absence of functional PR—the most likely way that cultures without retinal compensate for KefC* activity is by coupling hydrolysis of ATP to proton pumping.

Finally, the icing on the cake—so to speak—was the addition of ammonium chloride to the culture media, which was expected to introduce another futile cycle. The basic

concept is that $NH_4^+$ is imported in response to osmotic stress (through dedicated importers and Kdp) and subsequently dissociates to $NH_3$ and $H^+$ in the more basic cytoplasm followed by diffusion of the $NH_3$ out of the cell (Buurman et al., 1991). Together the additions of *kefFC** to the genetic background and 100 mM $NH_4Cl$ to the culture media resulted in moderately significant differences between cultures maintained in the light with retinal compared to cultures maintained in the light without retinal and maintained in the dark with retinal (Chapter III, Figure 4 and Figure 5). Ultimately, the results presented in Chapter III and summarized within the above section indicate that PR can positively affect growth yield in a strain that is depleted with respect to the pools of ATP, NADH, and some TCA cycle intermediates.

**Applying results from MED134 to engineer *E. coli***

MED134 exhibited higher growth yields as a result of light-driven proton pumping but only when the availability of carbon and energy was restricted. This discovery led to the common hypothesis that light-driven proton pumping can only affect physiology when the cell is energy-depleted (Fuhrman et al., 2008; Gomez-Consarnau et al., 2010; Gómez-Consarnau et al., 2016). Sure enough, growing MED134 in high DOC media abolishes any positive effect of proton pumping and even results in a negative affect on growth (Kim, 2013).

While attempts to demonstrate light-enhanced growth as a result of light-driven proton pumping by PR in *E. coli* failed in low DOC experiments, the results did not discourage

me from pushing forward with engineering *E. coli* to heterologously express PR and grow in rich media. My rationale for proceeding to use 2xYT media was simply that the genetic background could be modified to produce cells in a variety of different energetic states regardless of the carbon concentration of the media. This hypothesis led to the creation of a genetic background that presumably influenced the pools of ATP, NADH, and some TCA cycle intermediates enough (during cell growth or stationary phase survival) to create an environment that enabled light-driven proton pumping to energetically contribute to physiology, analogous to MED134 in low DOC with respect to these common metabolite pools.

But the goal of engineering *E. coli* to a reduced energy state was not the only applied lesson from work with MED134. Introduction of *pyc* to *E. coli* was also inspired by earlier reports about its potential utility in MED134 (González et al., 2011) and MED152 (Gonzalez et al., 2008) flavobacteria. In addition, when DL-alanine was used as the sole carbon substrate in place of DOC, *pyc* was induced by light (Palovaara et al., 2014). Of course, previous experiments with *pyc* and other anabolic genes such as *ppc* also established the functionality and utility of heterologous *pyc* expression in *E. coli* under various growth conditions (Farmer and Liao, 1997; Gokarn et al., 2001).

Perhaps the most interesting and important lesson from MED134 was to consider the implications of the potassium gradient and function of KefC with respect to light-driven proton pumping by PR. As started in Chapter II, the MED134 *kefC* transcript was more abundant in light cultures than dark cultures on average (nearly statistically significant) in

the first RNA-seq samples from low DOC media. And investigating the potential effects of inducing KefC in the light led to the realization that it could be used to regulate PMF production as light-driven proton pumping activities commenced. I was also intrigued by data from Feng *et al.* (Feng et al., 2015) that suggested osmolarity of the media impacted the light-enhanced growth phenotype of a different flavobacteria, which prompted consideration of how KefC activity also impacts turgor pressure and osmoregulation.

There is still much that can be learned by observing the behavior of natural flavobacteria grown under different conditions with and without light. But without genetic systems, the best way to fully realize the potential applications of PR in engineered systems is to progress the development of model heterologous hosts.

**Engineering the genetic background with serial transfer evolution**

Several deliberate genetic manipulations to *E. coli* have been performed or described that either have been observed or have the potential to increase the physiological effect of light-driven proton pumping on cellular energetics. But perhaps the best engineering approach—especially when it comes to improving an existing phenotype—is to use the process of evolution to optimize the growth of a novel (engineered) photoheterotroph.

The growth phenotype is an integration of many different physiological processes, each affected by different combinations of genetic elements that interact in complex ways.

Evolution provides order to these interactions by constantly selecting for the most optimally regulated system in a given environment. The process improves fitness—the reproductive rate of the organism—by fixing beneficial spontaneous mutations in the population (Elena and Lenski, 2003). Every mutation that establishes itself in the population as a result of evolution improves the system by directly optimizing regulation of at least one physiological process, which indirectly influences the effects of other physiological processes. From the perspective of a single gene, this implies that any change to the 'genetic background' has the potential to influence its effect on growth and fitness (Phillips, 2008; Chou et al., 2009).

Experimental evolution could be used to generate genetic backgrounds that leverage PR phototrophy to grow faster or more efficiently in terms of carbon. Of course novel selective pressures resulting from the presence of the PR gene during growth would be required to observe significant differences in the growth phenotype of lineages evolving with PR, relative to the ancestor or to the lineages evolved without PR. To experience positive selection, the PR gene would need to improve the fitness of its host; in the case of cells in batch cultures, this means that the PR protein must contribute to cell growth by shortening lag time, increasing the specific growth rate, or improving stationary phase survival. Otherwise, evolving strains with PR would probably elicit strong negative selection on the PR gene for reasons associated with the expression of a foreign genetic sequence (Welch et al., 2009); for example, expressing foreign DNA can cause depletion of amino acids, increased ribosome time, or protein folding problems. Additionally,

foreign gene products can interact with other genes and proteins, crowd the membrane at the sec translocons, or induce protein aggregation (Wagner et al., 2007).

Data from this thesis (Chapter 3) suggests that a fitness advantage is not conferred by PR expression in the ZK1143 background. Further work in BW25113 suggested that PR does not influence traits that could improve fitness in several mutant backgrounds either. PR was only found to enhance growth yield of *E. coli* when genetic tools were applied to manipulate its metabolism in ways that presumably reduced cellular energy levels. Thus, generating a strain suitable for serial transfer evolution with PR isn't amenable to a simple plug-in and go approach. Instead, lessons learned from this thesis should be applied to engineer the genome in a way that primes the strain to benefit from PR phototrophy.

In the case of *E. coli*, engineering attempts to improve yield should test Δ*fadR* strains modified by replacement of the native *kefC* locus with *kefC**, and genomic integration of *pyc* and *nox*. It's important that these changes are introduced to the genome for evolution experiments since plasmids can become unstable due to the burden they impose on replication. If a higher fitness peak can be reached with PR in this engineered background before the background evolves to reduce the effects of PR, *pyc* and *kefC** and *nox*, then these changes can become fixed as the lineages embark on a novel fitness landscape. In this case novel mutations could occur directly to the PR gene, or to any other genetic element underlying a physiological process that is influenced by the presence of the PR gene, that enhance the impact of phototrophy on metabolism. Since engineering a strain

with serial transfer evolution to improve phototrophic growth enables backtracking to discover and characterize the mutations that improved it (Elena and Lenski, 2003), new targets for engineering could be discovered from evolution experiments and subsequently leveraged to convert *E. coli* and other heterotrophs into useful photoheterotrophs. Or alternatively, the novel strains produced by the evolutionary process may be primed for direct use in certain applications as well. Finally, it's also possible to use serial transfer to effectively increase the metabolic dependency of a particular strain on a particular gene, like *nox*, and then replace the *nox* gene with a more favorable gene of similar function, like *rnf*, before using the evolved background in an application.

**Future applications of PR and other ion-pumping rhodopsin photosystems**

However the strain is created, engineered photoheterotrophs could be leveraged to produce more biomass from a given input of exogenous carbon or even engineered to fix carbon (Bryant and Frigaard, 2006), so that it can be recycled from the air or waste $CO_2$ stream to regenerate industrially useful carbon-based compounds. The term "productive biomass" usefully describes the notion that *E. coli* or other heterologous hosts could be cultured for energy applications that require the input of biomass, but they could also be engineered to produce useful commodities (including electricity) while they grow (and produce biomass—a carbon compound itself) to maximize outputs from engineered systems.

While using live cells certainly comes with its advantages (Bryant and Frigaard, 2006; Johnson and Schmidt-Dannert, 2008; Walter et al., 2010; Claassens et al., 2013), the disadvantages of using cells in applications can restrict the feasibility of engineered systems. To leverage PR in a production environment, cells would need to be grown with sunlight and their media would need to be sterile, set to an appropriate temperature, and contain all of the necessary trace nutrients. The standard difficulties of self-shading and light limitation would also need to be overcome. Obviously the economics of any production process ultimately determine its practical utility.

Perhaps the best future applications of PR and other light-driven ion pumps involve producing them with recombinant technology and reconstituting them into artificial membranes. This has already been done with respect to small scale *in vitro* production of ATP (Hara et al., 2011), and future applications can expand upon this principle to use synthetic membranes containing PR to potentially drive the production of biohydrogen and fixed nitrogen more efficiently than in a cell-based production strategy, and certainly in a fashion that is less costly to the environment than currently utilized technologies.

In conclusion, biological applications leveraging PR have the future potential to replace environmentally damaging and unsustainable processes that industries currently deploy for the production of fuels, fertilizers, and many other useful organic chemicals.

## References


Ahmed, S., and Booth, I.R. (1983a) The effect of beta-galactosides on the protonmotive force and growth of Escherichia coli. *J Gen Microbiol* 129: 2521-2529.

Ahmed, S., and Booth, I.R. (1983b) The use of valinomycin, nigericin and trichlorocarbanilide in control of the protonmotive force in Escherichia coli cells. *Biochem J* 212: 105-112.

Akopyan, K., and Trchounian, A. (2006) Escherichia coli membrane proton conductance and proton efflux depend on growth pH and are sensitive to osmotic stress. *Cell Biochem Biophys* 46: 201-208.

Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106.

Atamna-Ismaeel, N., Sabehi, G., Sharon, I., Witzel, K.P., Labrenz, M., Jurgens, K. et al. (2008) Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. *ISME J* 2: 656-662.

Atamna-Ismaeel, N., Finkel, O.M., Glaser, F., Sharon, I., Schneider, R., Post, A.F. et al. (2012) Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environ Microbiol* 14: 140-146.

Aylward, F.O., Eppley, J.M., Smith, J.M., Chavez, F.P., Scholin, C.A., and DeLong, E.F. (2015) Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc Natl Acad Sci U S A* 112: 5443-5448.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M. et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2: 2006.0008.

Balashov, S.P., Imasheva, E.S., and Lanyi, J.K. (2006) Induced chirality of the light-harvesting carotenoid salinixanthin and its interaction with the retinal of xanthorhodopsin. *Biochemistry* 45: 10998-11004.

Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Anton, J., Wang, J.M., and Lanyi, J.K. (2005) Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science* 309: 2061-2064.

Beja, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* 411: 786-789.

Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P. et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289: 1902-1906.

Bekker, M., de Vries, S., Ter Beek, A., Hellingwerf, K.J., and de Mattos, M.J. (2009) Respiration of Escherichia coli can be fully uncoupled via the nonelectrogenic terminal cytochrome bd-II oxidase. *J Bacteriol* 191: 5510-5517.

Berríos-Rivera, S.J., San, K.Y., and Bennett, G.N. (2002) The effect of NAPRTase overexpression on the total levels of NAD, the NADH/NAD+ ratio, and the distribution of metabolites in Escherichia coli. *Metab Eng* 4: 238-247.

Biegel, E., Schmidt, S., González, J.M., and Müller, V. (2011) Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell Mol Life Sci* 68: 613-634.

Bielawski, J.P., Dunn, K.A., Sabehi, G., and Beja, O. (2004) Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc Natl Acad Sci U S A* 101: 14824-14829.

Blaurock, A.E., and Stoeckenius, W. (1971) Structure of the purple membrane. *Nat New Biol* 233: 152-155.

Borisov, V.B., Murali, R., Verkhovskaya, M.L., Bloch, D.A., Han, H., Gennis, R.B., and Verkhovsky, M.I. (2011) Aerobic respiratory chain of Escherichia coli is not allowed to work in fully uncoupled mode. *Proc Natl Acad Sci U S A* 108: 17320-17324.

Brown, L.S., and Jung, K.H. (2006) Bacteriorhodopsin-like proteins of eubacteria and fungi: the extent of conservation of the haloarchaeal proton-pumping mechanism. *Photochem Photobiol Sci* 5: 538-546.

Bryant, D.A., and Frigaard, N.U. (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* 14: 488-496.

Burstein, C., Tiankova, L., and Kepes, A. (1979) Respiratory control in Escherichia coli K 12. *Eur J Biochem* 94: 387-392.

Buurman, E.T., Teixeira de Mattos, M.J., and Neijssel, O.M. (1991) Futile cycling of ammonium ions via the high affinity potassium uptake system (Kdp) of Escherichia coli. *Arch Microbiol* 155: 391-395.

Buurman, E.T., Pennock, J., Tempest, D.W., Teixeira de Mattos, M.J., and Neijssel, O.M. (1989) Replacement of potassium ions by ammonium ions in different micro-organisms grown in potassium-limited chemostat culture. *Arch Microbiol* 152: 58-63.

Béjà, O., and Lanyi, J.K. (2014) Nature's toolkit for microbial rhodopsin ion pumps. *Proc Natl Acad Sci U S A* 111: 6538-6539.

Campbell, B.J., Waidner, L.A., Cottrell, M.T., and Kirchman, D.L. (2008) Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environ Microbiol* 10: 99-109.

Campbell, J.W., and Cronan, J.E. (2001) Escherichia coli FadR positively regulates transcription of the fabB fatty acid biosynthetic gene. *J Bacteriol* 183: 5982-5990.

Carothers, J.M., Goler, J.A., and Keasling, J.D. (2009) Chemical synthesis using synthetic biology. *Curr Opin Biotechnol* 20: 498-503.

Cayley, D.S., Guttman, H.J., and Record, M.T. (2000) Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress. *Biophys J* 78: 1748-1764.

Cayley, S., Lewis, B.A., Guttman, H.J., and Record, M.T. (1991) Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *J Mol Biol* 222: 281-300.

Chang, D.E., Shin, S., Rhee, J.S., and Pan, J.G. (1999) Acetate metabolism in a pta mutant of Escherichia coli W3110: importance of maintaining acetyl coenzyme A flux for growth and survival. *J Bacteriol* 181: 6656-6663.

Chant, E.L., and Summers, D.K. (2007) Indole signalling contributes to the stable maintenance of Escherichia coli multicopy plasmids. *Mol Microbiol* 63: 35-43.

Chen, C.C., Walia, R., Mukherjee, K.J., Mahalik, S., and Summers, D.K. (2015) Indole generates quiescent and metabolically active Escherichia coli cultures. *Biotechnol J* 10: 636-646.

Chen, M., and Blankenship, R.E. (2011) Expanding the solar spectrum used by photosynthesis. *Trends Plant Sci* 16: 427-431.

Chen, M., Schliep, M., Willows, R.D., Cai, Z.L., Neilan, B.A., and Scheer, H. (2010) A red-shifted chlorophyll. *Science* 329: 1318-1319.

Chimerel, C., Field, C.M., Piñero-Fernandez, S., Keyser, U.F., and Summers, D.K. (2012) Indole prevents Escherichia coli cell division by modulating membrane potential. *Biochim Biophys Acta* 1818: 1590-1594.

Chimerel, C., Murray, A.J., Oldewurtel, E.R., Summers, D.K., and Keyser, U.F. (2013) The effect of bacterial signal indole on the electrical properties of lipid membranes. *Chemphyschem* 14: 417-423.

Chou, H.H., Berthet, J., and Marx, C.J. (2009) Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. *PLoS Genet* 5: e1000652.

Chu, K.H., Huang, G., An, T., Li, G., Yip, P.L., Ng, T.W. et al. (2015) Photocatalytic inactivation of Escherichia coli—The roles of genes in *Catalysis Today*.

Claassens, N.J., Volpers, M., dos Santos, V.A., van der Oost, J., and de Vos, W.M. (2013) Potential of proton-pumping rhodopsins: engineering photosystems into microorganisms. *Trends Biotechnol* 31: 633-642.

Cottrell, M.T., and Kirchman, D.L. (2009) Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Appl Environ Microbiol* 75: 4958-4966.

Cui, Y.Y., Ling, C., Zhang, Y.Y., Huang, J., and Liu, J.Z. (2014) Production of shikimic acid from Escherichia coli through chemically inducible chromosomal evolution and cofactor metabolic engineering. *Microb Cell Fact* 13: 21.

DeCoursey, T.E., and Hosler, J. (2014) Philosophy of voltage-gated proton channels. *J R Soc Interface* 11: 20130799.

Dioumaev, A.K., Brown, L.S., Shih, J., Spudich, E.N., Spudich, J.L., and Lanyi, J.K. (2002) Proton transfers in the photochemical reaction cycle of proteorhodopsin. *Biochemistry* 41: 5348-5358.

Ducat, D.C., Sachdeva, G., and Silver, P.A. (2011) Rewiring hydrogenase-dependent redox circuits in cyanobacteria. *Proc Natl Acad Sci U S A* 108: 3941-3946.

Ebert, B.E., Kurth, F., Grund, M., Blank, L.M., and Schmid, A. (2011) Response of Pseudomonas putida KT2440 to increased NADH and ATP demand. *Appl Environ Microbiol* 77: 6597-6605.

EIA (2016) *International Energy Outlook 2016*. In. DOE/EIA-0484 (ed).

Elena, S.F., and Lenski, R.E. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4: 457-469.

Engqvist, M.K., McIsaac, R.S., Dollinger, P., Flytzanis, N.C., Abrams, M., Schor, S., and Arnold, F.H. (2015) Directed evolution of Gloeobacter violaceus rhodopsin spectral properties. *J Mol Biol* 427: 205-220.

Epstein, W. (1986) Osmoregulation by potassium transport in Escherichia coli. *FEMS Microbiology Reviews* 39: 73-78.

Epstein, W. (2003) The roles and regulation of potassium in bacteria. *Prog Nucleic Acid Res Mol Biol* 75: 293-320.

Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* 42: 2646-2659.

Farmer, W.R., and Liao, J.C. (1997) Reduction of aerobic acetate production by Escherichia coli. *Appl Environ Microbiol* 63: 3205-3210.

Feng, S., Powell, S.M., Wilson, R., and Bowman, J.P. (2015) Proteomic Insight into Functional Changes of Proteorhodopsin-Containing Bacterial Species Psychroflexus torquis under Different Illumination and Salinity Levels. *J Proteome Res* 14: 3848-3858.

Ferguson, G.P., and Booth, I.R. (1998) Importance of glutathione for growth and survival of Escherichia coli cells: detoxification of methylglyoxal and maintenance of intracellular K+. *J Bacteriol* 180: 4314-4318.

Ferguson, G.P., McLaggan, D., and Booth, I.R. (1995) Potassium channel activation by glutathione-S-conjugates in Escherichia coli: protection against methylglyoxal is mediated by cytoplasmic acidification. *Mol Microbiol* 17: 1025-1033.

Finkel, S.E., and Kolter, R. (1999) Evolution of microbial diversity during prolonged starvation. *Proc Natl Acad Sci U S A* 96: 4023-4027.

Fischer, S., Graber, P., and Turina, P. (2000) The activity of the ATP synthase from Escherichia coli is regulated by the transmembrane proton motive force. *J Biol Chem* 275: 30157-30162.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* 105: 3805-3810.

Friedrich, T., Geibel, S., Kalmbach, R., Chizhov, I., Ataka, K., Heberle, J. et al. (2002) Proteorhodopsin is a light-driven proton pump with variable vectoriality. *J Mol Biol* 321: 821-838.

Frigaard, N.U., Martinez, A., Mincer, T.J., and DeLong, E.F. (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439: 847-850.

Fuhrman, J.A., Schwalbach, M.S., and Stingl, U. (2008) Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* 6: 488-494.

Gabel, C.V., and Berg, H.C. (2003) The speed of the flagellar rotary motor of Escherichia coli varies linearly with protonmotive force. *Proc Natl Acad Sci U S A* 100: 8748-8751.

Ganapathy, S., Bécheau, O., Venselaar, H., Frölich, S., van der Steen, J.B., Chen, Q. et al. (2015) Modulation of spectral properties and pump activity of proteorhodopsins by retinal analogues. *Biochem J* 467: 333-343.

Ghirardi, M.L., King, P.W., Posewitz, M.C., Maness, P.C., Fedorov, A., Kim, K. et al. (2005) Approaches to developing biological H(2)-photoproducing organisms and processes. *Biochem Soc Trans* 33: 70-72.

Giacalone, M.J., Gentile, A.M., Lovitt, B.T., Berkley, N.L., Gunderson, C.W., and Surber, M.W. (2006) Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* 40: 355-364.

Giovannoni, S.J., Bibbs, L., Cho, J.C., Stapels, M.D., Desiderio, R., Vergin, K.L. et al. (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* 438: 82-85.

Gokarn, R.R., Eiteman, M.A., and Altman, E. (2000) Metabolic analysis of Escherichia coli in the presence and absence of the carboxylating enzymes phosphoenolpyruvate carboxylase and pyruvate carboxylase. *Appl Environ Microbiol* 66: 1844-1850.

Gokarn, R.R., Evans, J.D., Walker, J.R., Martin, S.A., Eiteman, M.A., and Altman, E. (2001) The physiological effects and metabolic alterations caused by the expression of Rhizobium etli pyruvate carboxylase in Escherichia coli. *Appl Microbiol Biotechnol* 56: 188-195.

Gomez-Consarnau, L., Gonzalez, J.M., Coll-Llado, M., Gourdon, P., Pascher, T., Neutze, R. et al. (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* 445: 210-213.

Gomez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L. et al. (2010) Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol* 8: e1000358.

Gonzalez, J.M., Fernandez-Gomez, B., Fernandez-Guerra, A., Gomez-Consarnau, L., Sanchez, O., Coll-Llado, M. et al. (2008) Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proc Natl Acad Sci U S A* 105: 8724-8729.

González, J.M., Pinhassi, J., Fernández-Gómez, B., Coll-Lladó, M., González-Velázquez, M., Puigbò, P. et al. (2011) Genomics of the proteorhodopsin-containing marine flavobacterium Dokdonia sp. strain MED134. *Appl Environ Microbiol* 77: 8676-8686.

Govorunova, E.G., Sineshchekov, O.A., Li, H., Janz, R., and Spudich, J.L. (2013) Characterization of a highly efficient blue-shifted channelrhodopsin from the marine alga Platymonas subcordiformis. *J Biol Chem* 288: 29911-29922.

Gómez-Consarnau, L., González, J.M., Riedel, T., Jaenicke, S., Wagner-Döbler, I., Sañudo-Wilhelmy, S.A., and Fuhrman, J.A. (2016) Proteorhodopsin light-enhanced growth linked to vitamin-B1 acquisition in marine Flavobacteria. *ISME J* 10: 1102-1112.

Hannon, M., Gimpel, J., Tran, M., Rasala, B., and Mayfield, S. (2010) Biofuels from algae: challenges and potential. *Biofuels* 1: 763-784.

Hara, K.Y., Suzuki, R., Suzuki, T., Yoshida, M., and Kino, K. (2011) ATP photosynthetic vesicles for light-driven bioprocesses. *Biotechnol Lett* 33: 1133-1138.

Heuser, F., Schroer, K., Lutz, S., Bringer-Meyer, S., and Sahm, H. (2007) Enhancement of the NAD(P)(H) Pool in *Escherichia coli* for Biotransformation. *Engineering in Life Sciences* 7: 343-353.

Huang, H. (2008) Design and Characterization of Artificial Transcriptional Terminators. In *Department of Electrical Engineering and Computer Science*. Cambridge, MA: MIT, p. 104.

Hunt, K.A., Flynn, J.M., Naranjo, B., Shikhare, I.D., and Gralnick, J.A. (2010) Substrate-level phosphorylation is the primary source of energy conservation during anaerobic respiration of Shewanella oneidensis strain MR-1. *J Bacteriol* 192: 3345-3351.

Hussain, S., Kinnebrew, M., Schonenbach, N.S., Aye, E., and Han, S. (2015) Functional consequences of the oligomeric assembly of proteorhodopsin. *J Mol Biol* 427: 1278-1290.

Ikeda, T.P., Houtz, E., and LaPorte, D.C. (1992) Isocitrate dehydrogenase kinase/phosphatase: identification of mutations which selectively inhibit phosphatase activity. *J Bacteriol* 174: 1414-1416.

Imasheva, E.S., Balashov, S.P., Choi, A.R., Jung, K.H., and Lanyi, J.K. (2009) Reconstitution of Gloeobacter violaceus rhodopsin with a light-harvesting carotenoid antenna. *Biochemistry* 48: 10948-10955.

Inoue, K., Kato, Y., and Kandori, H. (2015) Light-driven ion-translocating rhodopsins in marine bacteria. *Trends Microbiol* 23: 91-98.

Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., and Kandori, H. (2013) A light-driven sodium ion pump in marine bacteria. *Nat Commun* 4: 1678.

Jensen, P.R., Westerhoff, H.V., and Michelsen, O. (1993) Excess capacity of H(+)-ATPase and inverse respiratory control in Escherichia coli. *EMBO J* 12: 1277-1282.

Jeong, H.S., and Jouanneau, Y. (2000) Enhanced nitrogenase activity in strains of Rhodobacter capsulatus that overexpress the rnf genes. *J Bacteriol* 182: 1208-1214.

Johnson, E.T., and Schmidt-Dannert, C. (2008) Light-energy conversion in engineered microorganisms. *Trends Biotechnol* 26: 682-689.

Johnson, E.T., Baron, D.B., Naranjo, B., Bond, D.R., Schmidt-Dannert, C., and Gralnick, J.A. (2010) Enhancement of survival and electricity production in an engineered bacterium by light-driven proton pumping. *Appl Environ Microbiol* 76: 4123-4129.

Jung, J.Y., Choi, A.R., Lee, Y.K., Lee, H.K., and Jung, K.H. (2008) Spectroscopic and photochemical analysis of proteorhodopsin variants from the surface of the Arctic Ocean. *FEBS Lett* 582: 1679-1684.

Kabir, M.M., and Shimizu, K. (2004) Metabolic regulation analysis of icd-gene knockout Escherichia coli based on 2D electrophoresis with MALDI-TOF mass spectrometry and enzyme activity measurements. *Appl Microbiol Biotechnol* 65: 84-96.

Keasling, J.D. (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol* 3: 64-76.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10: 845-858.

Kim, H. (2013) Metatranscriptomic and physiological analyses of proteorhodopsin-containing marine flavobacteria. In *Civil and Environmental Engineering*. Cambridge, MA: MIT, p. 81.

Kim, J.Y., Jo, B.H., and Cha, H.J. (2010) Production of biohydrogen by recombinant expression of [NiFe]-hydrogenase 1 in Escherichia coli. *Microb Cell Fact* 9: 54.

Kim, J.Y., Jo, B.H., Jo, Y., and Cha, H.J. (2012) Improved production of biohydrogen in light-powered Escherichia coli by co-expression of proteorhodopsin and heterologous hydrogenase. *Microb Cell Fact* 11: 2.

Kimura, H., Young, C.R., Martinez, A., and Delong, E.F. (2011) Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. *ISME J* 5: 1641-1651.

Klapoetke, N.C., Murata, Y., Kim, S.S., Pulver, S.R., Birdsey-Benson, A., Cho, Y.K. et al. (2014) Independent optical excitation of distinct neural populations. *Nat Methods* 11: 338-346.

Klyszejko, A.L., Shastri, S., Mari, S.A., Grubmuller, H., Muller, D.J., and Glaubitz, C. (2008) Folding and assembly of proteorhodopsin. *J Mol Biol* 376: 35-41.

Koebmann, B.J., Westerhoff, H.V., Snoep, J.L., Nilsson, D., and Jensen, P.R. (2002) The glycolytic flux in Escherichia coli is controlled by the demand for ATP. *J Bacteriol* 184: 3909-3916.

Kouyama, T., and Murakami, M. (2010) Structural divergence and functional versatility of the rhodopsin superfamily. *Photochem Photobiol Sci* 9: 1458-1465.

Kralj, J.M., Hochbaum, D.R., Douglass, A.D., and Cohen, A.E. (2011) Electrical spiking in Escherichia coli probed with a fluorescent voltage-indicating protein. *Science* 333: 345-348.

Kuniyoshi, T.M., Balan, A., Schenberg, A.C., Severino, D., and Hallenbeck, P.C. (2015) Heterologous expression of proteorhodopsin enhances H2 production in Escherichia coli when endogenous Hyd-4 is overexpressed. *J Biotechnol* 206: 52-57.

Lanyi, J.K. (1986) Halorhodopsin: a light-driven chloride ion pump. *Annu Rev Biophys Biophys Chem* 15: 11-28.

Lanyi, J.K., and Balashov, S.P. (2008) Xanthorhodopsin: a bacteriorhodopsin-like proton pump with a carotenoid antenna. *Biochim Biophys Acta* 1777: 684-688.

Lelong, C., Aguiluz, K., Luche, S., Kuhn, L., Garin, J., Rabilloud, T., and Geiselmann, J. (2007) The Crl-RpoS regulon of Escherichia coli. *Mol Cell Proteomics* 6: 648-659.

Li, H., Sineshchekov, O.A., da Silva, G.F., and Spudich, J.L. (2015) In Vitro Demonstration of Dual Light-Driven $Na^+/H^+$ Pumping by a Microbial Rhodopsin. *Biophys J* 109: 1446-1453.

Li, M., Ho, P.Y., Yao, S., and Shimizu, K. (2006) Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by 13C-labeling experiments. *J Biotechnol* 122: 254-266.

Liang, L., Liu, R., Chen, X., Ren, X., Ma, J., Kequan, C. et al. (2013) Effects of overexpression of NAPRTase, NAMNAT, and NAD synthetase in the NAD(H) biosynthetic pathways on the NAD(H) pool, NADH/$NAD^+$ ratio, and succinic acid production with different carbon sources by metabolically engineered*Escherichia coli*. *Biochemical Engineering Journal* 81: 90-96.

Lin, H., Castro, N.M., Bennett, G.N., and San, K.Y. (2006) Acetyl-CoA synthetase overexpression in Escherichia coli demonstrates more efficient acetate assimilation and lower acetate accumulation: a potential tool in metabolic engineering. *Appl Microbiol Biotechnol* 71: 870-874.

Lindholm, L., Ariöz, C., Jawurek, M., Liebau, J., Mäler, L., Wieslander, Å. et al. (2015) Effect of lipid bilayer properties on the photocycle of green proteorhodopsin. *Biochim Biophys Acta* 1847: 698-708.

Lovley, D.R. (2006) Bug juice: harvesting electricity with microorganisms. *Nat Rev Microbiol* 4: 497-508.

Lovley, D.R. (2012) Electromicrobiology. *Annu Rev Microbiol* 66: 391-409.

Lubitz, W., Reijerse, E.J., and Messinger, J. (2008) Solar water-splitting into H2 and O2: design principles of photosystem II and hydrogenases. *The Royal Society of Chemistry Energy & Environmental Science* 1: 15-31.

Lubner, C.E., Grimme, R., Bryant, D.A., and Golbeck, J.H. (2010) Wiring photosystem I for direct solar hydrogen production. *Biochemistry* 49: 404-414.

Luecke, H., Schobert, B., Stagno, J., Imasheva, E.S., Wang, J.M., Balashov, S.P., and Lanyi, J.K. (2008) Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc Natl Acad Sci U S A* 105: 16561-16565.

Malli, R., and Epstein, W. (1998) Expression of the Kdp ATPase is consistent with regulation by turgor pressure. *J Bacteriol* 180: 5102-5108.

Maloy, S.R., Bohlander, M., and Nunn, W.D. (1980) Elevated levels of glyoxylate shunt enzymes in Escherichia coli strains constitutive for fatty acid degradation. *J Bacteriol* 143: 720-725.

Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., and DeLong, E.F. (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc Natl Acad Sci U S A* 104: 5590-5595.

Mawhood, R., and Gazis, E.J., SierkHoefnagels, RicSlade, Raphael (2016) Production pathways for renewable jet fuel: a review of commercialization status and future prospects. *Biofuels, Bioprod Bioref* 10: 462-484.

McCarren, J., and DeLong, E.F. (2007) Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* 9: 846-858.

Metcalf, W.W., Jiang, W., Daniels, L.L., Kim, S.K., Haldimann, A., and Wanner, B.L. (1996) Conditionally replicative and conjugative plasmids carrying lacZ alpha for cloning, mutagenesis, and allele replacement in bacteria. *Plasmid* 35: 1-13.

Miller, S., Douglas, R.M., Carter, P., and Booth, I.R. (1997) Mutations in the glutathione-gated KefC K+ efflux system of Escherichia coli that cause constitutive activation. *J Biol Chem* 272: 24942-24947.

Miranda, M.R., Choi, A.R., Shi, L., Bezerra, A.G., Jr., Jung, K.H., and Brown, L.S. (2009) The photocycle and proton translocation pathway in a cyanobacterial ion-pumping rhodopsin. *Biophys J* 96: 1471-1481.

Morris, J.J., and Zinser, E.R. (2013) Continuous hydrogen peroxide production by organic buffers in phytoplankton culture media. *Journal of Phycology* 49: 1223-1228.

Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., and Rocap, G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* 4: 673-685.

Mukhopadhyay, A., Redding, A.M., Rutherford, B.J., and Keasling, J.D. (2008) Importance of systems biology in engineering microbes for biofuel production. *Curr Opin Biotechnol* 19: 228-234.

Mulkidjanian, A.Y., Heberle, J., and Cherepanov, D.A. (2006) Protons @ interfaces: implications for biological energy conversion. *Biochim Biophys Acta* 1757: 913-930.

Nakashima, N., Ohno, S., Yoshikawa, K., Shimizu, H., and Tamura, T. (2014) A vector library for silencing central carbon metabolism genes with antisense RNAs in Escherichia coli. *Appl Environ Microbiol* 80: 564-573.

Nannenga, B.L., and Baneyx, F. (2011) Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. *Protein Sci* 20: 1411-1420.

Oesterhelt, D. (1976) Bacteriorhodopsin as an example of a light-driven proton pump. *Angew Chem Int Ed Engl* 15: 17-24.

Oesterhelt, D., and Stoeckenius, W. (1971) Rhodopsin-like protein from the purple membrane of Halobacterium halobium. *Nat New Biol* 233: 149-152.

Oesterhelt, D., and Krippahl, G. (1973) Light inhibition of respiration in Halobacterium halobium. *FEBS Lett* 36: 72-76.

Palovaara, J., Akram, N., Baltar, F., Bunse, C., Forsberg, J., Pedrós-Alió, C. et al. (2014) Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. *Proc Natl Acad Sci U S A* 111: E3650-3658.

Paul Abishek, M., Patel, J., and Prem Rajan, A. (2014) Algae oil: a sustainable renewable fuel of future. *Biotechnol Res Int* 2014: 272814.

Peng, L., and Shimizu, K. (2006) Effect of *fadR* gene knockout on the metabolism of *Escherichia coli* based on analyses of protein expressions, enzyme activities and intracellular metabolite concentrations. *Enzyme and Microbial Technology* 38: 512-520.

Pfleger, N., Worner, A.C., Yang, J., Shastri, S., Hellmich, U.A., Aslimovska, L. et al. (2009) Solid-state NMR and functional studies on proteorhodopsin. *Biochim Biophys Acta* 1787: 697-705.

Phillips, P.C. (2008) Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9: 855-867.

Pimenta, A.L., Racher, K., Jamieson, L., Blight, M.A., and Holland, I.B. (2005) Mutations in HlyD, part of the type 1 translocator for hemolysin secretion, affect the folding of the secreted toxin. *J Bacteriol* 187: 7471-7480.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., and Moran, M.A. (2009) Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* 11: 1358-1375.

Portnoy, V.A., Scott, D.A., Lewis, N.E., Tarasova, Y., Osterman, A.L., and Palsson, B. (2010) Deletion of genes encoding cytochrome oxidases and quinol monooxygenase blocks the aerobic-anaerobic shift in Escherichia coli K-12 MG1655. *Appl Environ Microbiol* 76: 6529-6540.

Reckel, S., Gottstein, D., Stehle, J., Löhr, F., Verhoefen, M.K., Takeda, M. et al. (2011) Solution NMR structure of proteorhodopsin. *Angew Chem Int Ed Engl* 50: 11942-11946.

Record, M.T., Courtenay, E.S., Cayley, D.S., and Guttman, H.J. (1998) Responses of E. coli to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends Biochem Sci* 23: 143-148.

Roe, A.J., McLaggan, D., O'Byrne, C.P., and Booth, I.R. (2000) Rapid inactivation of the Escherichia coli Kdp K+ uptake system by high potassium concentrations. *Mol Microbiol* 35: 1235-1243.

Rosenbaum, M., He, Z., and Angenent, L.T. (2010) Light energy to bioelectricity: photosynthetic microbial fuel cells. *Curr Opin Biotechnol* 21: 259-264.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S. et al. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5: e77.

Russell, J.B., and Cook, G.M. (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* 59: 48-62.

Sabehi, G., Loy, A., Jung, K.H., Partha, R., Spudich, J.L., Isaacson, T. et al. (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* 3: e273.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27: 946-950.

San, K.Y., Bennett, G.N., Berríos-Rivera, S.J., Vadali, R.V., Yang, Y.T., Horton, E. et al. (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in Escherichia coli. *Metab Eng* 4: 182-192.

Sarmento, H., and Gasol, J.M. (2012) Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ Microbiol* 14: 2348-2360.

Sauer, U., and Eikmanns, B.J. (2005) The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev* 29: 765-794.

Scheer, H. (2006) Chlorophylls and Bacteriochlorophylls. In *An Overview of Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*. Grimm, B., Porra, R.J., Rudiger, W., and Scheer, H. (eds). Dordrecht, Netherlands: Springer, pp. 1-26.

Sharma, A.K., Spudich, J.L., and Doolittle, W.F. (2006) Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* 14: 463-469.

Sharma, A.K., Zhaxybayeva, O., Papke, R.T., and Doolittle, W.F. (2008) Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environ Microbiol* 10: 1039-1056.

Sharma, A.K., Sommerfeld, K., Bullerjahn, G.S., Matteson, A.R., Wilhelm, S.W., Jezbera, J. et al. (2009) Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. *ISME J* 3: 726-737.

Sharma, A.K., Becker, J.W., Ottesen, E.A., Bryant, J.A., Duhamel, S., Karl, D.M. et al. (2014) Distinct dissolved organic matter sources induce rapid transcriptional responses in coexisting populations of Prochlorococcus, Pelagibacter and the OM60 clade. *Environ Microbiol* 16: 2815-2830.

Siddiquee, K.A., Arauzo-Bravo, M.J., and Shimizu, K. (2004) Effect of a pyruvate kinase (pykF-gene) knockout mutation on the control of gene expression and metabolic fluxes in Escherichia coli. *FEMS Microbiol Lett* 235: 25-33.

Slamovits, C.H., Okamoto, N., Burri, L., James, E.R., and Keeling, P.J. (2011) A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat Commun* 2: 183.

Smillie, C.S., Smith, M.B., Friedman, J., Cordero, O.X., David, L.A., and Alm, E.J. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480: 241-244.

Sousa, P.M., Silva, S.T., Hood, B.L., Charro, N., Carita, J.N., Vaz, F. et al. (2011) Supramolecular organizations in the aerobic respiratory chain of Escherichia coli. *Biochimie* 93: 418-425.

Spudich, J.L., Yang, C.S., Jung, K.H., and Spudich, E.N. (2000) Retinylidene proteins: structures and functions from archaea to humans. *Annu Rev Cell Dev Biol* 16: 365-392.

Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., and Giovannoni, S.J. (2011) Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One* 6: e19725.

Stephanopoulos, G. (2007) Challenges in engineering microbes for biofuels production. *Science* 315: 801-804.

Stock, J.B., Rauch, B., and Roseman, S. (1977) Periplasmic space in Salmonella typhimurium and Escherichia coli. *J Biol Chem* 252: 7850-7861.

Tsunoda, S.P., Ewers, D., Gazzarrini, S., Moroni, A., Gradmann, D., and Hegemann, P. (2006) H+ -pumping rhodopsin from the marine alga Acetabularia. *Biophys J* 91: 1471-1479.

Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.J., Arike, L., and Vilu, R. (2010) Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst Biol* 4: 166.

van Rotterdam, B.J., Crielaard, W., van Stokkum, I.H., Hellingwerf, K.J., and Westerhoff, H.V. (2002) Simplicity in complexity: the photosynthetic reaction center performs as a simple 0.2 V battery. *FEBS Lett* 510: 105-107.

Vemuri, G.N., Eiteman, M.A., and Altman, E. (2006a) Increased recombinant protein production in Escherichia coli strains with overexpressed water-forming NADH oxidase and a deleted ArcA regulatory protein. *Biotechnol Bioeng* 94: 538-542.

Vemuri, G.N., Altman, E., Sangurdekar, D.P., Khodursky, A.B., and Eiteman, M.A. (2006b) Overflow metabolism in Escherichia coli during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol* 72: 3653-3661.

Vignais, P.M., and Billoud, B. (2007) Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* 107: 4206-4272.

von Meyenburg, K., Jørgensen, B.B., and van Deurs, B. (1984) Physiological and morphological effects of overproduction of membrane-bound ATP synthase in Escherichia coli K-12. *EMBO J* 3: 1791-1797.

Wagner, S., Baars, L., Ytterberg, A.J., Klussmeier, A., Wagner, C.S., Nord, O. et al. (2007) Consequences of membrane protein overexpression in Escherichia coli. *Mol Cell Proteomics* 6: 1527-1550.

Walter, J.M., Greenfield, D., and Liphardt, J. (2010) Potential of light-harvesting proton pumps for bioenergy applications. *Curr Opin Biotechnol* 21: 265-270.

Walter, J.M., Greenfield, D., Bustamante, C., and Liphardt, J. (2007) Light-powering Escherichia coli with proteorhodopsin. *Proc Natl Acad Sci U S A* 104: 2408-2412.

Wang, W.W., Sineshchekov, O.A., Spudich, E.N., and Spudich, J.L. (2003) Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J Biol Chem* 278: 33985-33991.

Wang, Y., San, K.Y., and Bennett, G.N. (2013) Improvement of NADPH bioavailability in Escherichia coli by replacing NAD(+)-dependent glyceraldehyde-3-phosphate dehydrogenase GapA with NADP (+)-dependent GapB from Bacillus subtilis and addition of NAD kinase. *J Ind Microbiol Biotechnol* 40: 1449-1460.

Wei, B. (2010) Diversity and distribution of proteorhodopsin-containing microorganisms in marine environments. *Frontiers of Environmental Science & Engineering* 6: 98-106.

Welch, M., Villalobos, A., Gustafsson, C., and Minshull, J. (2009) You're one in a googol: optimizing genes for protein expression. *J R Soc Interface* 6 Suppl 4: S467-476.

Wilks, J.C., and Slonczewski, J.L. (2007) pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol* 189: 5601-5607.

Yoshizawa, S., Kawanabe, A., Ito, H., Kandori, H., and Kogure, K. (2012) Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. *Environ Microbiol* 14: 1240-1248.

Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W. et al. (2014) Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc Natl Acad Sci U S A* 111: 6732-6737.

Zambrano, M.M., Siegele, D.A., Almirón, M., Tormo, A., and Kolter, R. (1993) Microbial competition: Escherichia coli mutants that take over stationary phase cultures. *Science* 259: 1757-1760.

Zubkov, M.V. (2009) Photoheterotrophy in marine prokaryotes. *Journal of Plankton Research* 31: 933-938.

## VI. Appendix

### Supplementary Materials for Chapter II

Table of Contents

**Supplementary Figure Legends**

**Figure S1.** Spectrum and intensity of light for growth experiments
For RNA-seq experiments, green filters were placed over four full-spectrum fluorescent lamps in a Beckman light incubator retrofitted with fans to improve air circulation.

**Supplementary Table Captions**

**Table S1.** Overview of Library Sequencing Results
This table contains the total number of reads from each library that mapped to the MED134 reference genome. The number in parentheses indicates the percent of total reads that mapped to MED134. The percent of total reads that could be mapped to the MED134 genome is indicated in parentheses. Reads that did not map to MED134 genes were found to map to rRNA genes, intergenic regions, or organisms that were confirmed to not be present in the cultures (data not shown). Every library was included with the relevant DESeq comparison with the exception of one library that contained an insufficient number of reads and was omitted from the analysis (Low DOC T1 L3).

**Table S2.** Annotated transcript abundance data
Raw read counts were normalized and tested for differential expression using DESeq (see Methods). The transcript abundance data from each comparison (low and high DOC, T1 and T2) is presented as a sortable Excel file. Genes identified as differentially expressed or potentially differentially expressed are highlighted on sheets 1 and 2, followed by the full gene annotations on sheet 3. Each of the remaining 4 sheets represents one of the 4 comparisons and is labeled accordingly. The table within each sheet can be sorted by the mean transcript abundance "baseMean", average transcript abundance of 3 light-exposed replicate cultures "light" or dark-exposed cultures "dark", the "fold change", the "log 2 fold change", the "p-value", and finally, by the "FDR adjusted p-value". The default sorting is by the FDR-adjusted p-values, and transcripts with significant differences are highlighted in green.

**Table S3.** Differentially expressed genes at High DOC T2
Most differentially expressed genes at High DOC T2 are more abundant in dark-exposed cultures and are annotated as enzymes involved in β-oxidation or heat shock. This table contains the differentially expressed genes and the average transcript abundance data associated with each, as well as the gene annotations. All genes listed in this table have FDR-adjusted p-values <0.25.

**Table S4.** Top 10 highly expressed genes
This table contains the top 10 most highly expressed genes for every comparison tested. Most of these genes encode for proteins of unknown function and proteins involved in heat shock, although some enzymes and membrane proteins of known function are highly expressed as well. The average transcript abundance data is included with their associated fold change values and p-values. Only 1 of the highly expressed genes for any comparison is significant (highlighted in green, High DOC T2, MED05174) with an FDR-adjusted p-value <0.1.

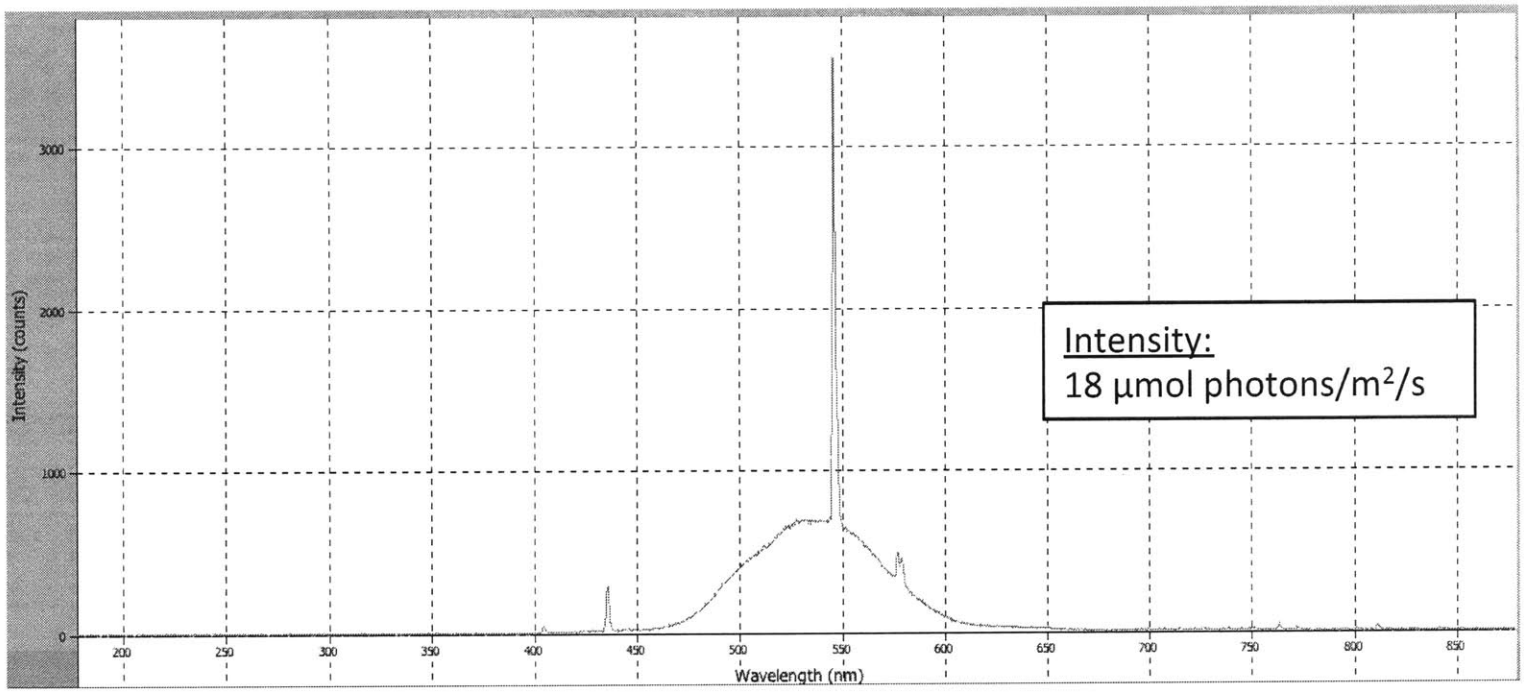



**Figure S1.** Spectrum and intensity of light for growth experiments

For RNA-seq experiments, green filters were placed over four full-spectrum fluorescent lamps in a Beckman refrigerated light incubator shaker retrofitted with accessory fans to improve air circulation.

**Table S1.** Overview of Library Sequencing Results

| | T0 | Low DOC T1 | | | Low DOC T2 | | | High DOC T1 | | | High DOC T2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Low nutrient** | 672551 (82) | 396472 (76) | 434562 (75) | 111 (76) | 454587 (77) | 316091 (75) | 391343 (69) | 408616 (63) | 81880 (66) | 152802 (62) | 68029 (53) | 431899 (70) | 254788 (59) |
| **High nutrient** | 381740 (69) | 388883 (69) | 400666 (75) | 248805 (64) | 364477 (65) | 246919 (60) | 102794 (42) | 241879 (67) | 363742 (74) | 355208 (74) | 394832 (71) | 356372 (68) | 437947 (73) |

This table contains the total number of reads from each library that mapped to genes within the MED134 reference genome. The number in parentheses indicates the percent of total reads that mapped to genes. Reads that did not map to MED134 genes were found to map to rRNA genes, intergenic regions, or organisms that were confirmed to not be present in the cultures (data not shown). Every library was included with the relevant DESeq comparison, with the exception of one library that contained an insufficient number of reads and was omitted from the analysis (Low DOC T1 L3).

**Table S3.** Differentially expressed genes at High DOC T2

| Feature ID | Annotation | High DOC T2 | | Notes |
|---|---|---|---|---|
| | | Light | Dark | |
| MED134_12996 | Putative recognition particle-docking protein | 333 | 545 | |
| MED134_00125 | Electron transfer flavoprotein alpha-subunit (etfA) | 407 | 694 | electron transfer protein in β-Oxidation. PHX gene |
| MED134_05174 | Putative heat shock ClpB protein | 2604 | 4669 | |
| MED134_13401 | GrpE protein (Hsp-70 cofactor) | 737 | 1344 | chaperonins and/or protein degradation |
| MED134_00130 | Electron transfer flavoprotein beta subunit (etfB) | 311 | 574 | electron transfer protein in β-Oxidation |
| MED134_06234 | Thiolase family protein | 136 | 257 | acetyl-CoA C-acyltransferase. Fatty acid and phospholipid metabolism. PHX gene |
| MED134_06229 | Enoyl-CoA hydratase | 409 | 779 | acyl-CoA dehydrogenase in β-Oxidation |
| MED134_03499 | Putative uncharacterized protein | 189 | 364 | putative cell adhesion protein, PHX gene, RpoN-linked |
| MED134_04044 | Putative heat-shock related protein | 1158 | 2381 | |
| MED134_07119 | Bacteriorhodopsin | 27 | 8 | |
| MED134_00750 | Purine nucleoside phosphorylase (EC 2.4.2.1) | 85 | 46 | |

Most differentially expressed genes at High DOC T2 are more abundant in dark-exposed cultures and are annotated as enzymes involved in β-oxidation or heat shock. This table contains the differentially expressed genes and the average transcript abundance data associated with each, as well as the gene annotations. All genes listed in this table have FDR-adjusted p-values <0.1.

**Table S4.** Light induced genes from published experiments with MED134 or related flavobacteria

| Locus ID | Annotation | Low DOC--T1 | | | Low DOC--T2 | | | High DOC--T1 | | | High DOC--T2 | | | Notes | Reference(s) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | light | dark | log2 FC | light | dark | log2 FC | light | dark | log2 FC | light | dark | log2 FC | | |
| MED134_07466 | Geranylgeranyl pyrophosphate synthase crtE | 153 | 162 | 0.1 | 6 | 7 | 0.1 | 62 | 62 | 0 | 144 | 141 | 0.1 | Involved in retinal synthesis, not siginifcantly induced in our experiments | 2 |
| MED134_10161 | Putatitve TBDT; vitamin B1 | 59 | 68 | 0.2 | 10 | 7 | 0.4 | 43 | 32 | 0.4 | 35 | 36 | 0 | Possible role in light-enhanced growth transporting vitamin B1 | 1 |
| MED134_04464 | ThiJ/Pfpl family; vitamin B1 | 206 | 162 | 0.4 | 25 | 27 | 0.1 | 71 | 67 | 0.1 | 544 | 399 | 0.4 | Possible role in light-enhanced growth processing vitamin B1 | 1 |
| MED134_05424 | ThiJ/Pfpl family; dimethylallytransferase | 252 | 266 | 0.1 | 66 | 27 | 1.3 | 255 | 260 | 0 | 465 | 393 | 0.2 | Possibly light-induced | 1 |
| MED134_12071 | pyruvate dehydrogenase | 238 | 290 | 0.3 | 12 | 10 | 0.3 | 137 | 154 | 0.2 | 244 | 219 | 0.2 | High Expression in MED134; vitamin B1 dependent | 1 |
| MED134_10396 | Sensory transduction histidine kinase | 9 | 3 | 1.6 | 1 | 0 | 2.6 | 11 | 9 | 0.3 | 2 | 4 | 0.7 | Differentally-expressed in MED134, DSW-1, and PRO95; PAS-Domain | 1&2 |
| MED134_04254 | Putative uncharacterized protein | 65 | 63 | 0 | 23 | 13 | 0.9 | 156 | 98 | 0.7 | 143 | 161 | 0.2 | Differentally-expressed in MED134, DSW-1, and PRO96; hypothetical protein | 1 |
| MED134_12381 | Putatitve TBDT; starch | 1516 | 1082 | 0.5 | 76 | 45 | 0.7 | 361 | 368 | 0 | 851 | 884 | 0.1 | High Expression in MED134 | 1 |
| MED134_05219 | Putatitve TBDT; starch | 2941 | 3841 | 0.4 | 10 | 13 | 0.4 | 44 | 48 | 0.1 | 2222 | 1806 | 0.3 | High Expression in MED134 | 1 |
| MED134_07711 | 2-oxoglutarate dehydrogenase sucB | 553 | 680 | 0.3 | 36 | 24 | 0.5 | 186 | 165 | 0.2 | 678 | 666 | 0 | High Expression in MED134; vitamin B1 dependent; central metabolism; induced by light in *P.torquis* | 1-4 |
| MED134_10061 | Bicarbonate transporter BicA | 48 | 50 | 0.1 | 18 | 16 | 0.1 | 39 | 44 | 0.1 | 33 | 39 | 0.2 | Possible role in light-enhanced growth transporting bicarbonate | 4 |
| MED134_10056 | Carbonic Anhydrase | 36 | 43 | 0.3 | 3 | 6 | 1 | 25 | 18 | 0.5 | 24 | 21 | 0.2 | Possible role in light-enhanced growth; catalyzes carbon dioxide to bicarbonate | 4 |
| MED134_06244 | Pyruvate carboxylase | 67 | 39 | 0.8 | 11 | 5 | 1.1 | 48 | 56 | 0.2 | 70 | 97 | 0.5 | Anaplerotic; induced under suboptimal salinity in *P.torquis* | 2,4 |
| MED134_10331 | PEP carboxykinase | 188 | 219 | 0.2 | 10 | 17 | 0.7 | 71 | 91 | 0.4 | 434 | 406 | 0.1 | Anaplerotic; induced by light in *P.torquis* | 2-4 |
| MED134_06089 | PEP carboxylase | 282 | 254 | 0.1 | 40 | 34 | 0.3 | 177 | 198 | 0.2 | 241 | 244 | 0 | Anaplerotic enzyme | 4 |
| MED134_11446 | Malic enzyme | 84 | 108 | 0.4 | 8 | 7 | 0.2 | 39 | 48 | 0.3 | 71 | 72 | 0 | Anaplerotic enzyme | 4 |
| MED134_14141 | Isocitrate dehydrogenase | 352 | 399 | 0.2 | 15 | 8 | 0.9 | 70 | 64 | 0.1 | 237 | 265 | 0.2 | Central metabolism | 4 |
| MED134_01780 | Isocitrate lyase | 128 | 57 | 1.2 | 11 | 9 | 0.3 | 51 | 43 | 0.2 | 62 | 73 | 0.2 | Glyoxylate Shunt | 4 |
| MED134_01770 | Malate synthase | 7 | 6 | 0.2 | 3 | 2 | 1 | 28 | 17 | 0.7 | 11 | 14 | 0.3 | Glyoxylate Shunt | 4 |
| MED134_12971 | Malate dehydrogenase | 224 | 236 | 0.1 | 6 | 7 | 0.4 | 32 | 40 | 0.3 | 290 | 262 | 0.1 | Anaplerotic; induced by light in *P.torquis* | 3 |
| MED134_04539 | Adenylosuccinate lyase | 27 | 46 | 0.8 | 12 | 14 | 0.2 | 22 | 17 | 0.3 | 37 | 35 | 0.1 | Anaplerotic; induced by light in *P.torquis* | 3 |
| MED134_06329 | Glyceraldehyde-3-phosphate dehydrogenase I | 73 | 88 | 0.3 | 26 | 30 | 0.2 | 198 | 163 | 0.3 | 81 | 61 | 0.4 | Central glycolytic pathway | 3 |
| MED134_01405 | Glyceraldehyde-3-phosphate dehydrogenase II | 42 | 48 | 0.2 | 26 | 19 | 0.5 | 19 | 15 | 0.4 | 45 | 61 | 0.4 | Central glycolytic pathway | 3 |
| MED134_03314 | Glyceraldehyde-3-phosphate dehydrogenase III | 210 | 269 | 0.4 | 40 | 32 | 0.3 | 42 | 38 | 0.2 | 374 | 396 | 0.1 | Central glycolytic pathway | 3 |
| MED134_13716 | Triosephosphate isomerase | 37 | 30 | 0.3 | 17 | 24 | 0.5 | 34 | 45 | 0.4 | 25 | 24 | 0.1 | Interconverts dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate | 3 |
| MED134_09321 | Phosphoglycerate kinase | 181 | 187 | 0.1 | 13 | 12 | 0 | 91 | 83 | 0.1 | 200 | 182 | 0.2 | Interconverts 1,3-BPG and ADP to 3-PG and ATP | 3 |
| MED134_08271 | Fructose-bisphosphatase fbp | 94 | 41 | 1 | 39 | 61 | 0.7 | 119 | 123 | 0.1 | 58 | 62 | 0.1 | Central glycolytic pathway | 2 |
| MEd134_14231 | Glucose-6-phosphate dehydrogenase zwf | 123 | 133 | 0.1 | 4 | 7 | 0.9 | 54 | 63 | 0.2 | 169 | 177 | 0.1 | Pentose phosphate cycle | 2 |
| MED134_14226 | Phosphogluconate dehydrogenase gnd | 152 | 147 | 0.1 | 15 | 18 | 0.2 | 116 | 144 | 0.3 | 166 | 169 | 0 | Pentose phosphate cycle | 2 |
| MED134_11155 | Succinate dehydrogenase sdhA | 501 | 649 | 0.4 | 23 | 16 | 0.5 | 300 | 270 | 0.2 | 325 | 325 | 0 | Central metabolism | 2 |
| MED134_11160 | Succinate dehydrogenase sdhB | 1116 | 1053 | 0.1 | 12 | 10 | 0.2 | 148 | 128 | 0.2 | 413 | 398 | 0.1 | Central metabolism | 2 |
| MED134_11150 | Succinate dehydrogenase sdhC | 48 | 59 | 0.3 | 6 | 5 | 0.2 | 31 | 30 | 0.1 | 42 | 42 | 0 | Central metabolism | 2 |
| MED134_07134 | Fumarate hydratase fumC | 364 | 429 | 0.2 | 10 | 7 | 0.5 | 97 | 80 | 0.3 | 313 | 377 | 0.3 | Central metabolism | 2&3 |
| MED134_02355 | Na+/alanine and glycine symporter | 268 | 280 | 0.1 | 20 | 22 | 0.2 | 72 | 87 | 0.3 | 313 | 377 | 0.3 | Transporter energized by the Na+ gradient | 2 |
| MED134_14567 | Na+/alanine and glycine symporter | 58 | 33 | 0.8 | 1 | 1 | 0 | 22 | 28 | 0.4 | 42 | 37 | 0.2 | Transporter energized by the Na+ gradient | 2 |
| MED134_11180 | Na+/phosphate symporter | 212 | 193 | 0.1 | 59 | 75 | 0.4 | 37 | 47 | 0.4 | 112 | 107 | 0.1 | Transporter energized by the Na+ gradient | 2 |
| MED134_00295 | Na+translocating NADH quinone oxidoreductase nqrA | 204 | 307 | 0.6 | 17 | 8 | 1 | 49 | 50 | 0 | 205 | 202 | 0 | Na+ pumping repiratory enzyme | 2 |
| MED134_00300 | Na+translocating NADH quinone oxidoreductase nqrB | 294 | 449 | 0.6 | 7 | 9 | 0.4 | 44 | 54 | 0.3 | 230 | 261 | 0.2 | Na+ pumping repiratory enzyme | 2 |
| MED134_00305 | Na+translocating NADH quinone oxidoreductase nqrC | 154 | 180 | 0.2 | 8 | 10 | 0.4 | 22 | 22 | 0 | 111 | 133 | 0.3 | Na+ pumping repiratory enzyme | 2 |
| MED134_00310 | Na+translocating NADH quinone oxidoreductase nqrD | 98 | 146 | 0.6 | 6 | 2 | 1.3 | 10 | 13 | 0.4 | 61 | 68 | 0.1 | Na+ pumping repiratory enzyme | 2 |
| MED134_00320 | Na+translocating NADH quinone oxidoreductase nqrF | 745 | 981 | 0.4 | 7 | 9 | 0.3 | 37 | 50 | 0.4 | 379 | 373 | 0 | Na+ pumping repiratory enzyme | 2 |
| MED134_10206 | DNA photolase/cryptochrome, (6-4) photolyase family | 14 | 14 | 0 | 20 | 10 | 1 | 149 | 39 | 1.9 | 13 | 8 | 0.7 | Putative light sensing protein potentially involved in gene regulation | 2 |
| MED134_10211 | DNA photolase/cryptochrome, (6-4) photolyase family | 14 | 15 | 0.1 | 7 | 5 | 0.3 | 66 | 23 | 1.5 | 7 | 9 | 0.4 | Putative light sensing protein potentially involved in gene regulation | 2 |
| MED134_01400 | Multi-sensor hybrid His kinase | 203 | 235 | 0.2 | 19 | 15 | 0.3 | 53 | 50 | 0.1 | 170 | 192 | 0.2 | Putative light sensing protein potentially involved in gene regulation | 2 |

Light and dark transcript abundance data is the average of three independent replicates, with the exception of low DOC T1-light for which 1 of 3 replicates was omitted

Reference 1: Gomez-Consarnau *et al.* (2007)

Reference 2: Kimura *et al.* (2011)

Reference 3: Feng *et al.* (2015)

Reference 4: Palovaara *et al.* (2014)

This table contains previously identified differentially expressed genes between light and dark conditions from earlier transcriptome studies of MED134 or related Flavobacteria, complete with gene abundance data from this study.

**Table S5.** Top 10 highly expressed genes

**Low DOC T1**

| Locus ID | Mean Read Abundance | Mean Read Abundance--Light | Mean Read Abuncance--Dark | Fold Change | Log2 Fold Change | p-value | FDR-adjusted p-value | Annotation |
|---|---|---|---|---|---|---|---|---|
| MED134_02735 | 4009.8 | 4270.6 | 3835.9 | 0.9 | -0.2 | 0.570 | 1.000 | Probable iron-sulfur protein |
| MED134_05014 | 3958.6 | 5116.3 | 3186.9 | 0.6 | -0.7 | 0.152 | 1.000 | Putative uncharacterized protein |
| MED134_05219 | 3481.1 | 2940.6 | 3841.5 | 1.3 | 0.4 | 0.407 | 1.000 | Putative outer membrane protein |
| MED134_04484 | 3433.9 | 2754.4 | 3887.0 | 1.4 | 0.5 | 0.320 | 1.000 | OmpA/MotB |
| MED134_09446 | 3233.8 | 3163.3 | 3280.9 | 1.0 | 0.1 | 0.911 | 1.000 | Putative uncharacterized protein |
| MED134_05084 | 3154.0 | 3678.8 | 2804.2 | 0.8 | -0.4 | 0.221 | 1.000 | Putative uncharacterized protein |
| MED134_14812 | 3126.0 | 2641.6 | 3448.9 | 1.3 | 0.4 | 0.258 | 1.000 | VCBS (Fragment) |
| MED134_05369 | 2705.3 | 2917.1 | 2564.1 | 0.9 | -0.2 | 0.567 | 1.000 | TPR repeat protein |
| MED134_13996 | 2639.3 | 2542.3 | 2703.9 | 1.1 | 0.1 | 0.754 | 1.000 | Putative uncharacterized protein |
| MED134_02925 | 2384.7 | 2135.9 | 2550.5 | 1.2 | 0.3 | 0.723 | 1.000 | Putative uncharacterized protein |

**Low DOC T2**

| Locus ID | Mean Read Abundance | Mean Read Abundance--Light | Mean Read Abuncance--Dark | Fold Change | Log2 Fold Change | p-value | FDR-adjusted p-value | Annotation |
|---|---|---|---|---|---|---|---|---|
| MED134_01970 | 6128.9 | 5220.7 | 7037.2 | 1.3 | 0.4 | 0.179 | 1.000 | Putative uncharacterized protein |
| MED134_10266 | 4958.3 | 4749.0 | 5167.5 | 1.1 | 0.1 | 0.792 | 1.000 | Thioredoxin |
| MED134_10910 | 4482.0 | 3897.0 | 5066.9 | 1.3 | 0.4 | 0.177 | 1.000 | Putative uncharacterized protein |
| MED134_03979 | 3600.9 | 2985.8 | 4216.0 | 1.4 | 0.5 | 0.113 | 1.000 | TRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase |
| MED134_05174 | 3310.8 | 3548.9 | 3072.7 | 0.9 | -0.2 | 0.916 | 1.000 | Putative heat shock ClpB protein |
| MED134_04044 | 3058.6 | 3401.3 | 2715.9 | 0.8 | -0.3 | 0.522 | 1.000 | Putative heat-shock related protein |
| MED134_11726 | 2984.2 | 2889.6 | 3078.8 | 1.1 | 0.1 | 0.695 | 1.000 | Putative uncharacterized protein |
| MED134_03319 | 2972.0 | 2338.0 | 3606.0 | 1.5 | 0.6 | 0.035 | 1.000 | Serine protease |
| MED134_14812 | 2874.1 | 2723.3 | 3024.9 | 1.1 | 0.2 | 0.728 | 1.000 | VCBS (Fragment) |
| MED134_00870 | 2742.7 | 2685.8 | 2799.7 | 1.0 | 0.1 | 0.804 | 1.000 | Putative uncharacterized protein |

**High DOC T1**

| Locus ID | Mean Read Abundance | Mean Read Abundance--Light | Mean Read Abuncance--Dark | Fold Change | Log2 Fold Change | p-value | FDR-adjusted p-value | Annotation |
|---|---|---|---|---|---|---|---|---|
| MED134_03679 | 4248.4 | 4377.4 | 4119.5 | 0.9 | -0.1 | 0.888 | 1.000 | Leucine-rich-repeat protein |
| MED134_05174 | 3421.2 | 5217.8 | 1624.7 | 0.3 | -1.7 | 0.009 | 1.000 | Putative heat shock ClpB protein |
| MED134_03414 | 3140.7 | 4625.3 | 1656.1 | 0.4 | -1.5 | 0.007 | 1.000 | DnaK protein |
| MED134_12871 | 2887.6 | 3223.2 | 2552.0 | 0.8 | -0.3 | 0.943 | 1.000 | Putative sigma factor A |
| MED134_10266 | 2881.9 | 3373.0 | 2390.9 | 0.7 | -0.5 | 0.778 | 1.000 | Thioredoxin |
| MED134_10910 | 2873.8 | 3228.9 | 2518.6 | 0.8 | -0.4 | 0.901 | 1.000 | Putative uncharacterized protein |
| MED134_09821 | 2850.5 | 2609.7 | 3091.4 | 1.2 | 0.2 | 0.605 | 1.000 | Putative uncharacterized protein |
| MED134_07421 | 2686.4 | 2180.5 | 3192.3 | 1.5 | 0.5 | 0.292 | 1.000 | Putative uncharacterized protein |
| MED134_04044 | 2616.3 | 3914.2 | 1318.4 | 0.3 | -1.6 | 0.013 | 1.000 | Putative heat-shock related protein |
| MED134_13401 | 2515.3 | 3366.5 | 1664.1 | 0.5 | -1.0 | 0.200 | 1.000 | GrpE protein (Hsp-70 cofactor) |

**High DOC T2**

| Locus ID | Mean Read Abundance | Mean Read Abundance--Light | Mean Read Abuncance--Dark | Fold Change | Log2 Fold Change | p-value | FDR-adjusted p-value | Annotation |
|---|---|---|---|---|---|---|---|---|
| MED134_02735 | 4152.2 | 3668.6 | 4635.8 | 1.3 | 0.3 | 0.278 | 1.000 | Probable iron-sulfur protein |
| MED134_03414 | 3771.2 | 3253.2 | 4289.3 | 1.3 | 0.4 | 0.009 | 0.594 | DnaK protein |
| MED134_09446 | 3682.0 | 3583.3 | 3780.6 | 1.1 | 0.1 | 0.746 | 1.000 | Putative uncharacterized protein |
| MED134_05174 | 3636.5 | 2603.7 | 4669.3 | 1.793 | 0.843 | 0.000 | 0.000 | Putative heat shock ClpB protein |
| MED134_01135 | 2895.6 | 2619.8 | 3171.5 | 1.211 | 0.276 | 0.066 | 1.000 | 60 kDa chaperonin |
| MED134_05704 | 2878.9 | 3051.1 | 2706.8 | 0.887 | -0.173 | 0.369 | 1.000 | Putative uncharacterized protein |
| MED134_02925 | 2541.2 | 2517.4 | 2565.0 | 1.019 | 0.027 | 0.791 | 1.000 | Putative uncharacterized protein |
| MED134_03344 | 2537.6 | 2066.0 | 3009.3 | 1.457 | 0.543 | 0.001 | 0.123 | Heat shock protein 90 |
| MED134_06084 | 2485.2 | 2696.0 | 2274.4 | 0.8 | -0.2 | 0.118 | 1.000 | Phosphoenolpyruvate carboxylase (EC 4.1.1.31) |
| MED134_04484 | 2444.3 | 2573.7 | 2314.9 | 0.9 | -0.2 | 0.418 | 1.000 | OmpA/MotB |

This table contains the top 10 most highly expressed genes for every comparison tested. Most of these genes encode for proteins of unknown function and proteins involved in heat shock, although some enzymes and membrane proteins of known function are highly expressed as well. The average transcript abundance data is included with their associated fold change values and p-values. Only 1 of the highly expressed genes for any comparison is statistically significant (highlighted in green, High DOC T2, MED05174).

**Supplementary Materials for Chapter III**

Table of Contents

**Supplementary Figure Legends**

**Figure S1.** Spectrum and intensity of light for growth experiments
For growth experiments, green filters were placed over four full-spectrum fluorescent lamps in a Beckman refrigerated light incubator shaker retrofitted with fans to improve air circulation.

**Figure S2.** Comparison of cell counts of BW25113+pRHA1 cultures grown in LB media at 37°C by cfu and flow cytometry methods
Growth curves based on colony-forming units (a) on LB agar plates containing Kanamycin (30 μg/mL) were compared to curves based on flow cytometry data (b) from fixed cell samples. Despite generally higher yields from flow cytometry counts, the data was otherwise identical. In both a and b, the data shows that growth of BW25113+pRHA1 in LB at 37°C results in higher cell yields in the light relative to the dark condition, but not compared to cultures maintained in the light without retinal. Thus, the PR was not responsible in the difference observed between the light and dark cultures with retinal. The same result was observed for ZK1143.

**Figure S3.** Comparison of cell counts of ZK1143+pRHA1 cultures grown in LB media at 37°C by cfu and flow cytometry methods
Growth curves based on colony-forming units (a) on LB agar plates containing Kanamycin (30 μg/mL) were compared to curves based on flow cytometry data (b) from fixed cell samples. Despite generally higher yields from flow cytometry counts, the data was otherwise identical. In both a and b, the data shows that growth of ZK1143+pRHA1 in LB at 37°C results in higher cell yields in the light relative to the dark condition, but not compared to cultures maintained in the light without retinal. Thus, the PR was not responsible in the difference observed between the light and dark cultures with retinal. The same result was observed for BW25113.

**Figure S4.** Effect of induction of PR expression (1mM Rhamnose) from pRHA1 and pRHA2 on the growth of ZK1143 growing in LB at 37°C
ZK1143 grown with 1% glucose resulted in similar cell yields regardless of whether it hosted pRHA1 or pRHA2. However, replacing the glucose with 1mM Rhamnose resulted in measurable differences between the effects of pRHA1 and pRHA2 on growth. Since pRHA2 imposed a more significant burden than pRHA1 when induced, pRHA1 was used for all of the reported growth experiments (Figures 1-4 and Table 1).

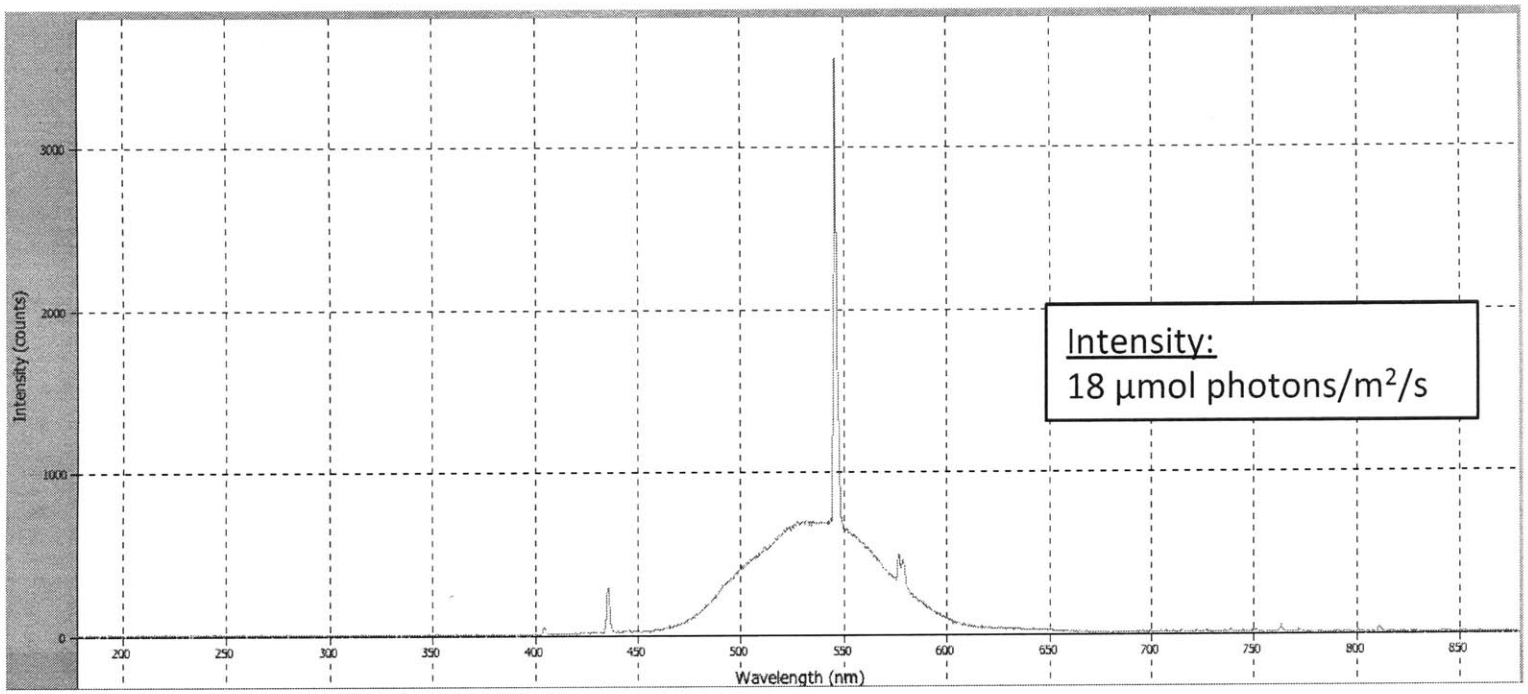



**Figure S1.** Spectrum and intensity of light for growth experiments

For growth experiments, green filters were placed over four full-spectrum fluorescent lamps in a Beckman refrigerated light incubator shaker retrofitted with fans to improve air circulation.

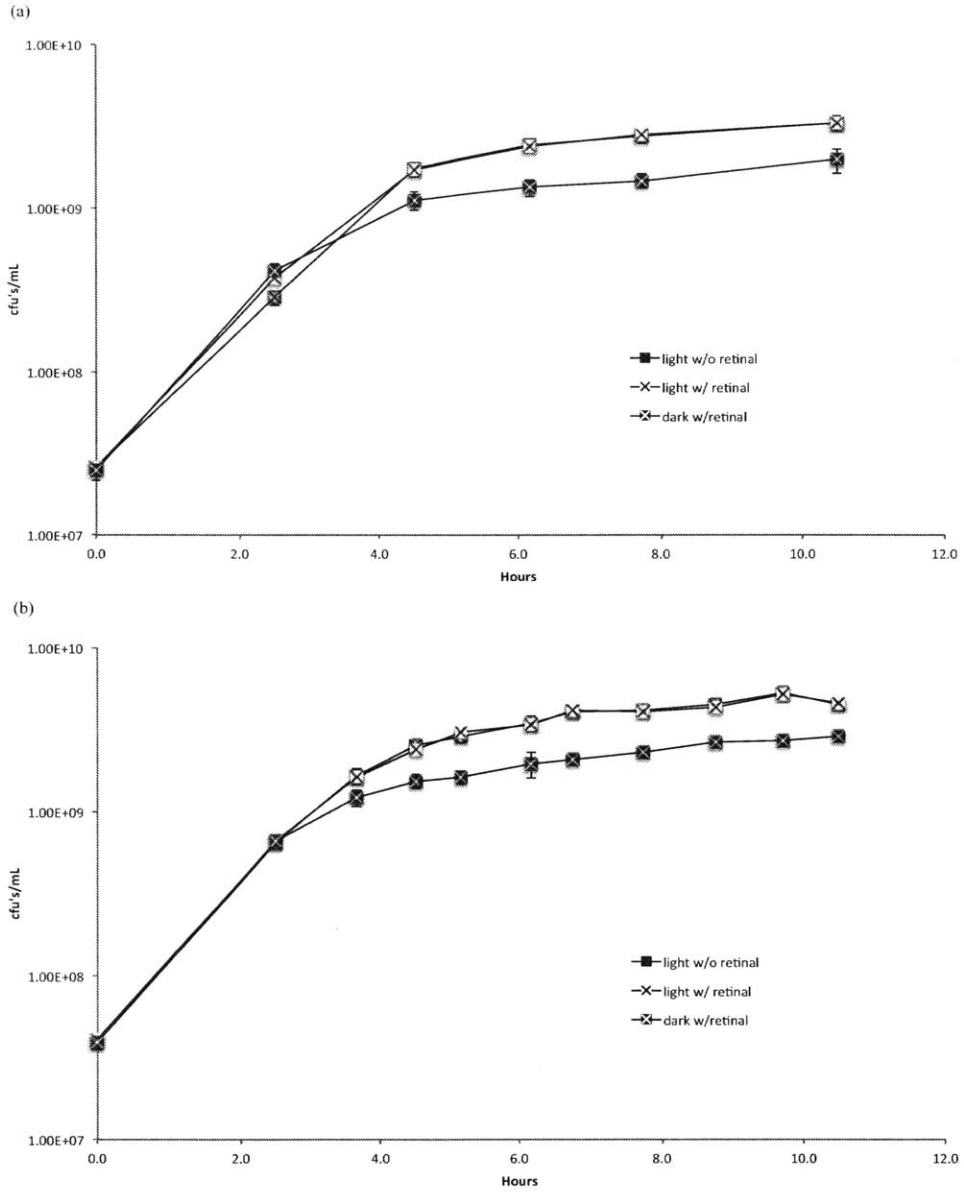



**Figure S2.** Comparison of cell counts of BW25113+pRHA1 cultures grown in LB media at 37°C by cfu and flow cytometry methods

Growth curves based on colony-forming units (a) on LB agar plates containing Kanamycin (30 μg/mL) were compared to curves based on flow cytometry data (b) from fixed cell samples. Despite generally higher yields from flow cytometry counts, the data was otherwise identical. In both a and b, the data shows that growth of BW25113+pRHA1 in LB at 37°C results in higher cell yields in the light relative to the dark condition, but not compared to cultures maintained in the light without retinal. Thus, the PR was not responsible in the difference observed between the light and dark cultures with retinal. The same result was observed for ZK1143.

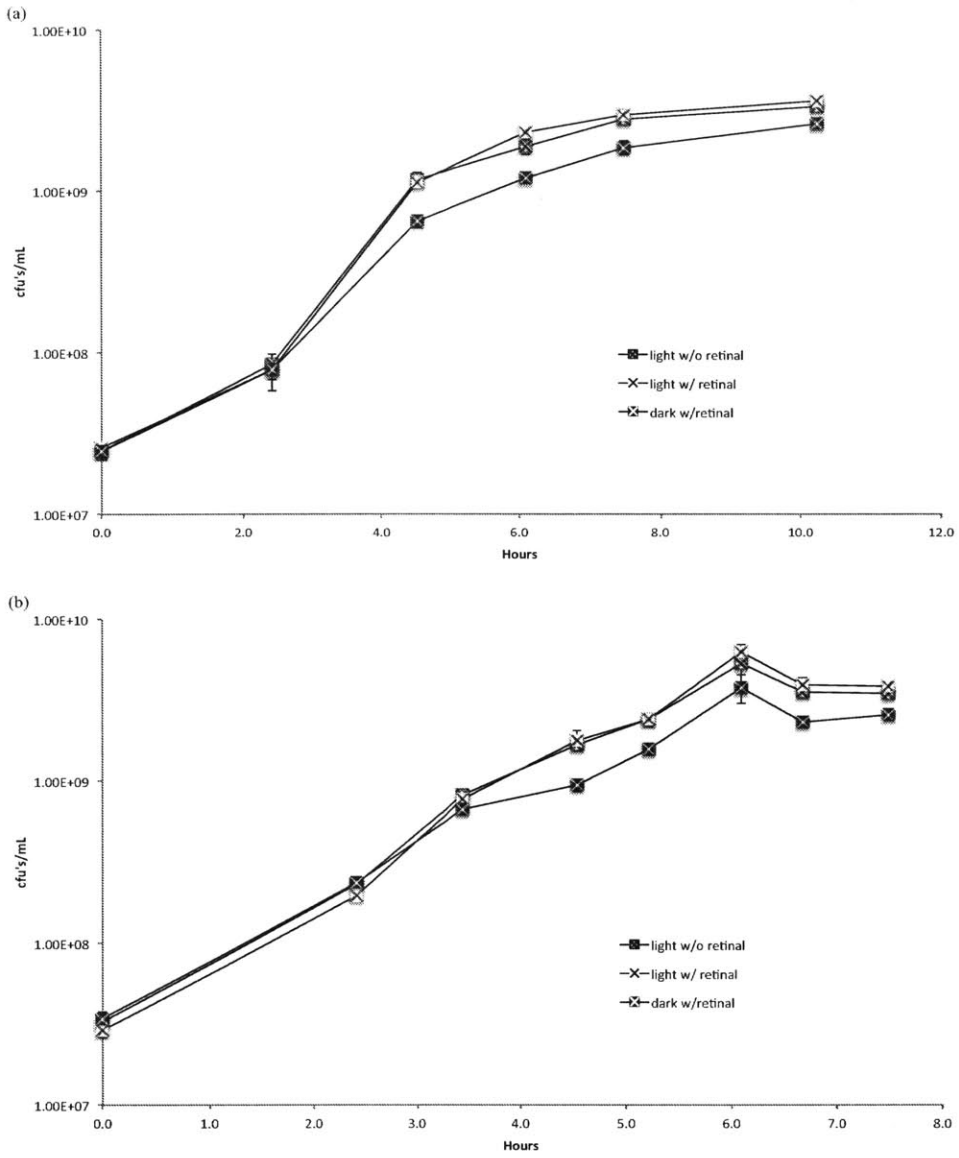



**Figure S3.** Comparison of cell counts of ZK1143+pRHA1 cultures grown in LB media at 37°C by cfu and flow cytometry methods

Growth curves based on colony-forming units (a) on LB agar plates containing Kanamycin (30 μg/mL) were compared to curves based on flow cytometry data (b) from fixed cell samples. Despite generally higher yields from flow cytometry counts, the data was otherwise identical. In both a and b, the data shows that growth of ZK1143+pRHA1 in LB at 37°C results in higher cell yields in the light relative to the dark condition, but not compared to cultures maintained in the light without retinal. Thus, the PR was not responsible in the difference observed between the light and dark cultures with retinal. The same result was observed for BW25113.

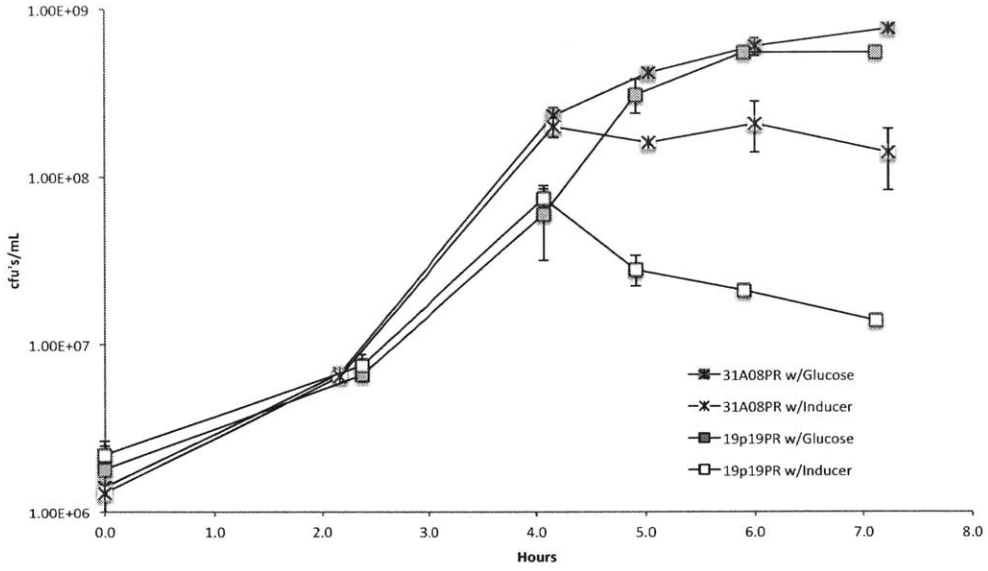



**Figure S4.** Effect of induction of PR expression (1mM Rhamnose) from pRHA1 and pRHA2 on the growth of ZK1143 growing in LB at 37°C

ZK1143 grown with 1% glucose resulted in similar cell yields regardless of whether it hosted pRHA1 (31a08PR) or pRHA2 (19p19PR). However, replacing the glucose with 1mM Rhamnose resulted in measurable differences between the effects of pRHA1 and pRHA2 on growth. Since pRHA2 imposed a more significant burden than pRHA1 when induced, pRHA1 was used for all of the reported growth experiments (Figures 1-4 and Table 1).

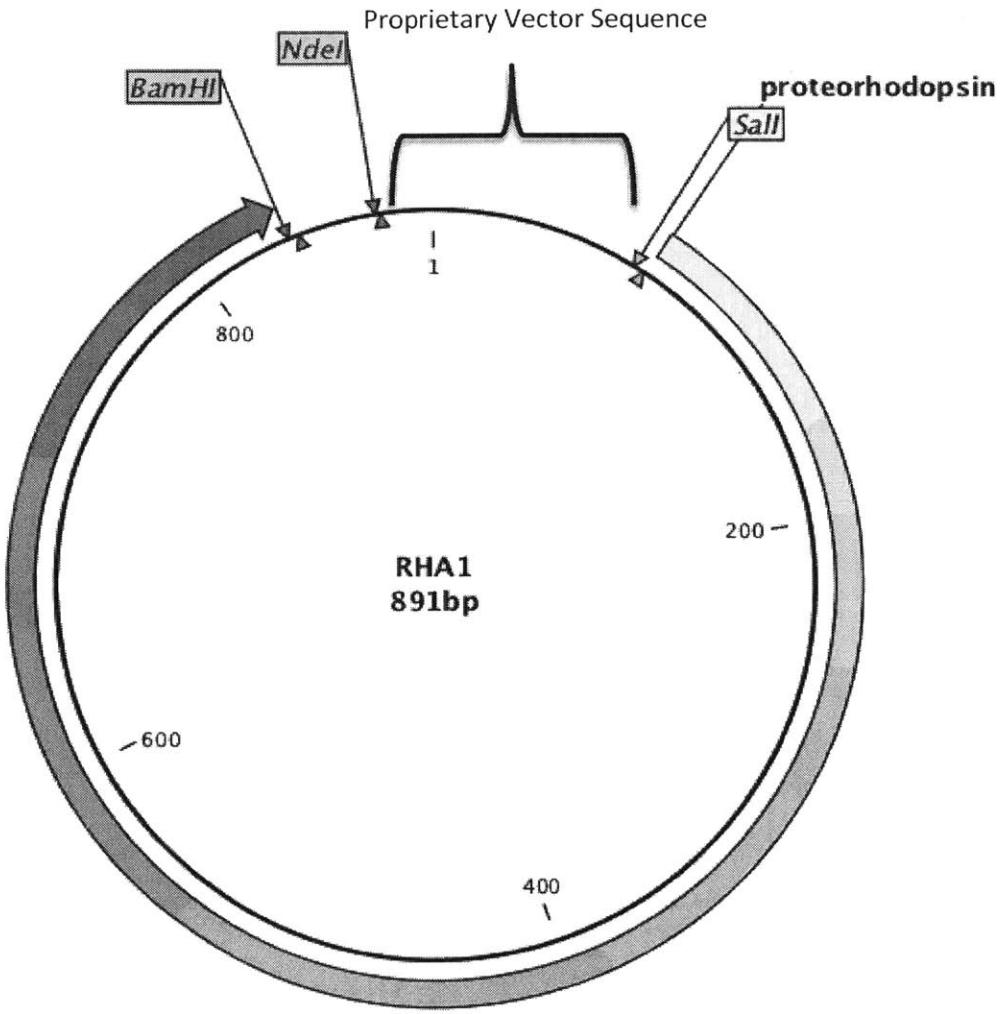



**Figure S5.** Map of pRHA1 and pRHA2

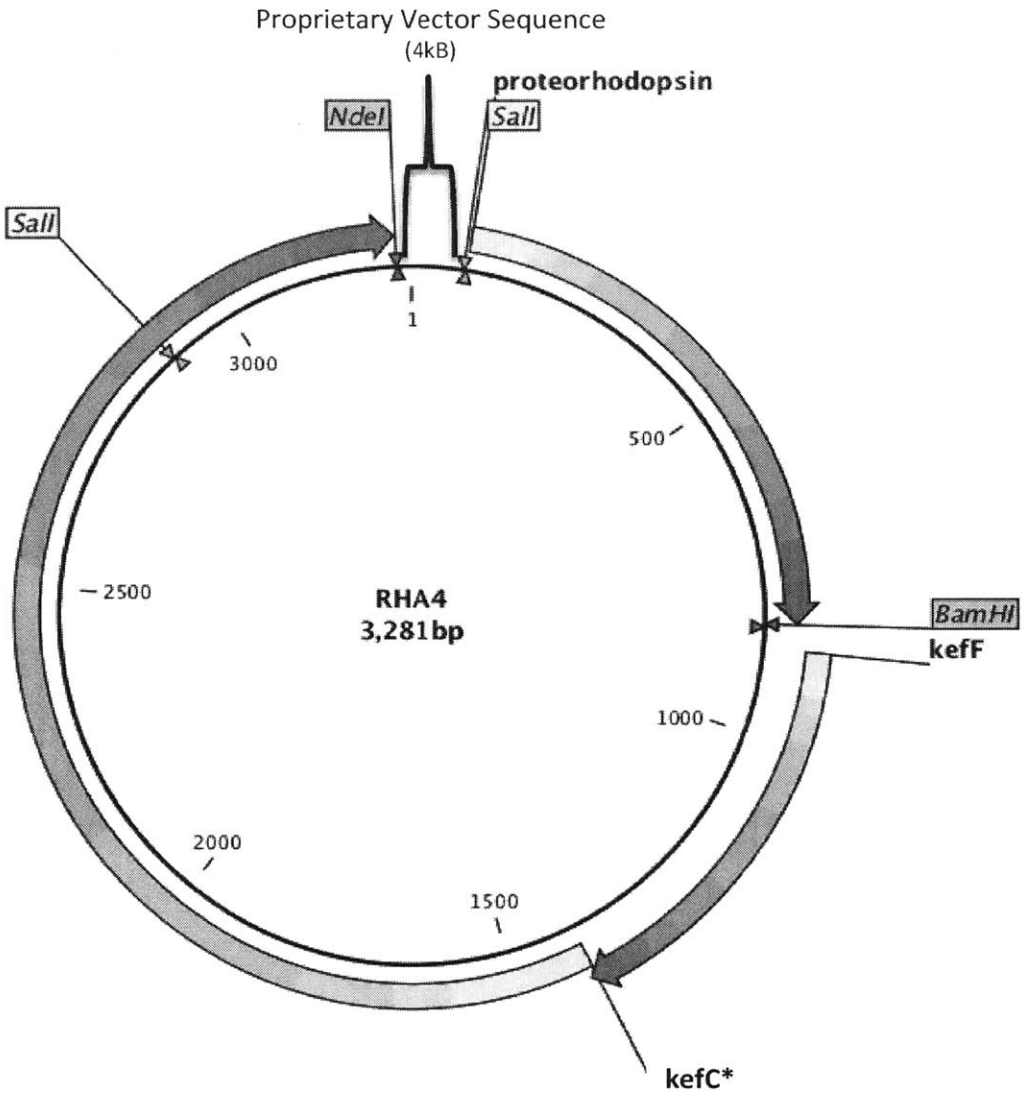



**Figure S6.** Map of pRHA4

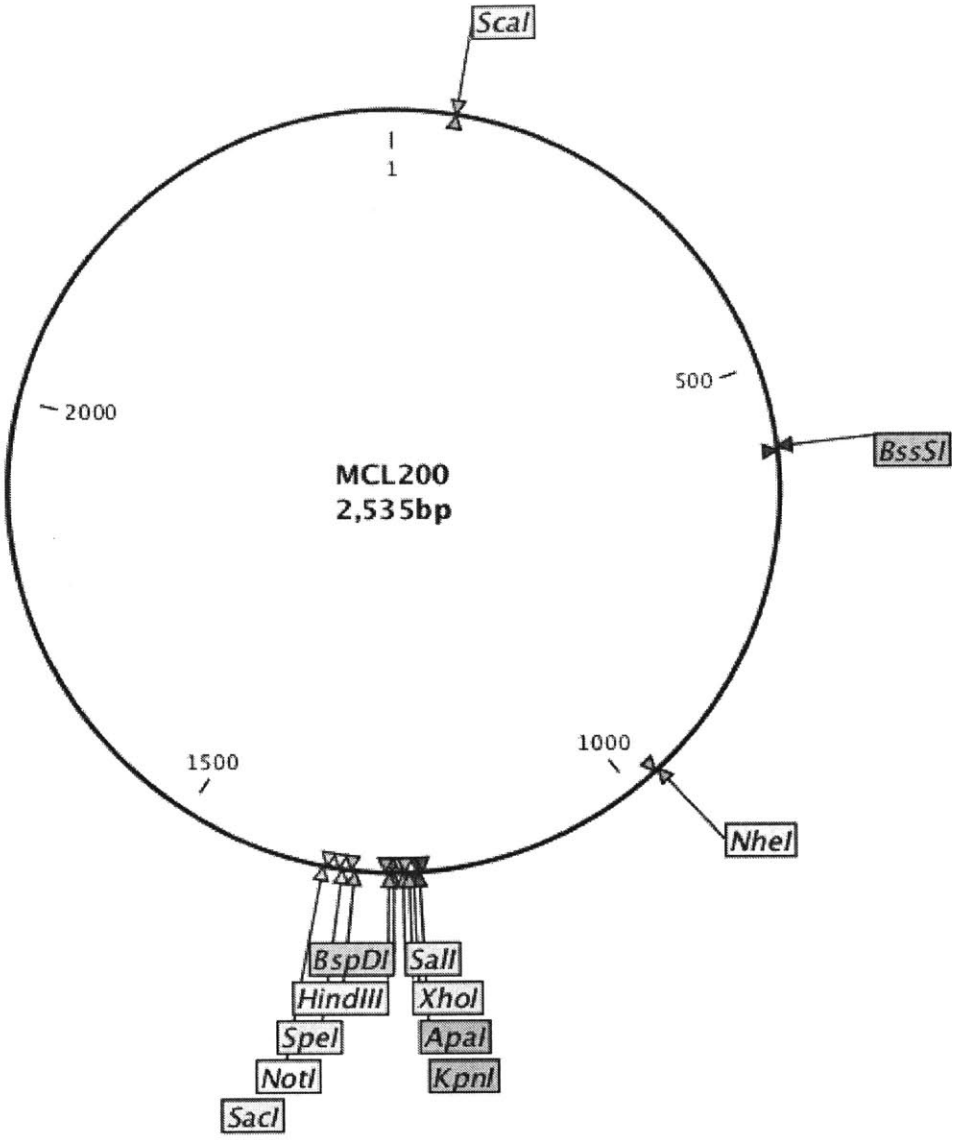



**Figure S7.** Map of pMCL200

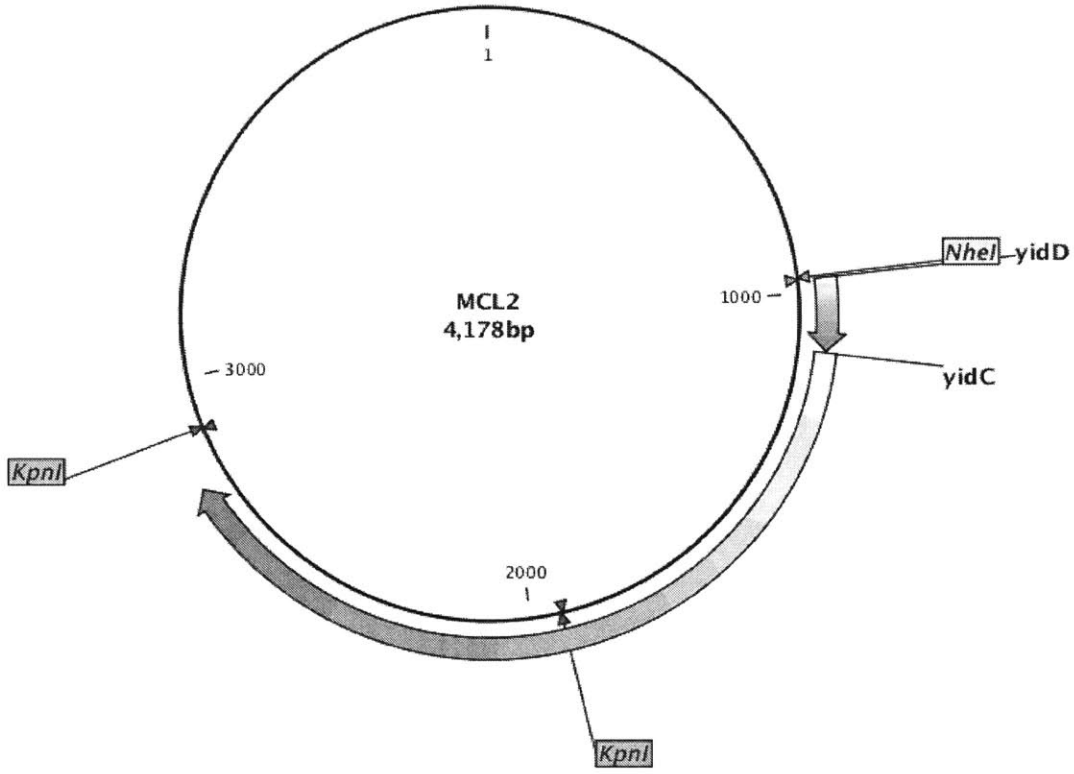



**Figure S8.** Map of pMCL2

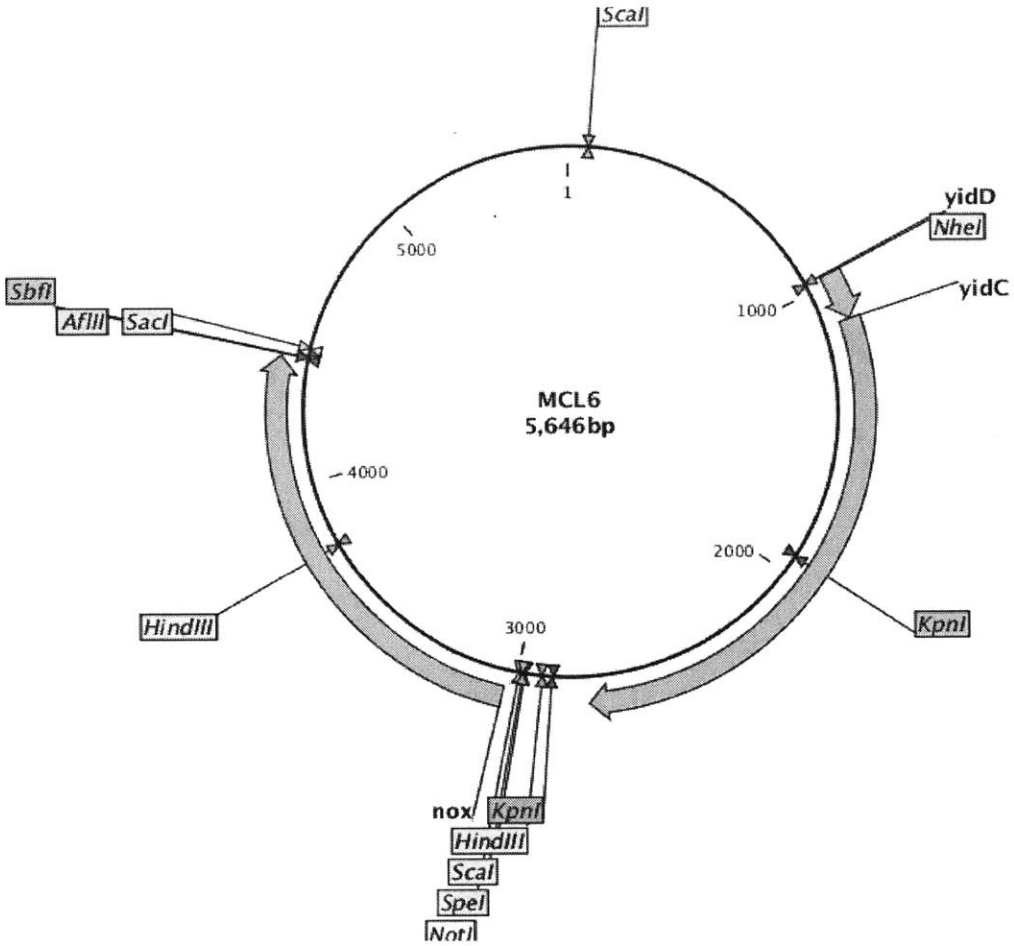



**Figure S9.** Map of pMCL6

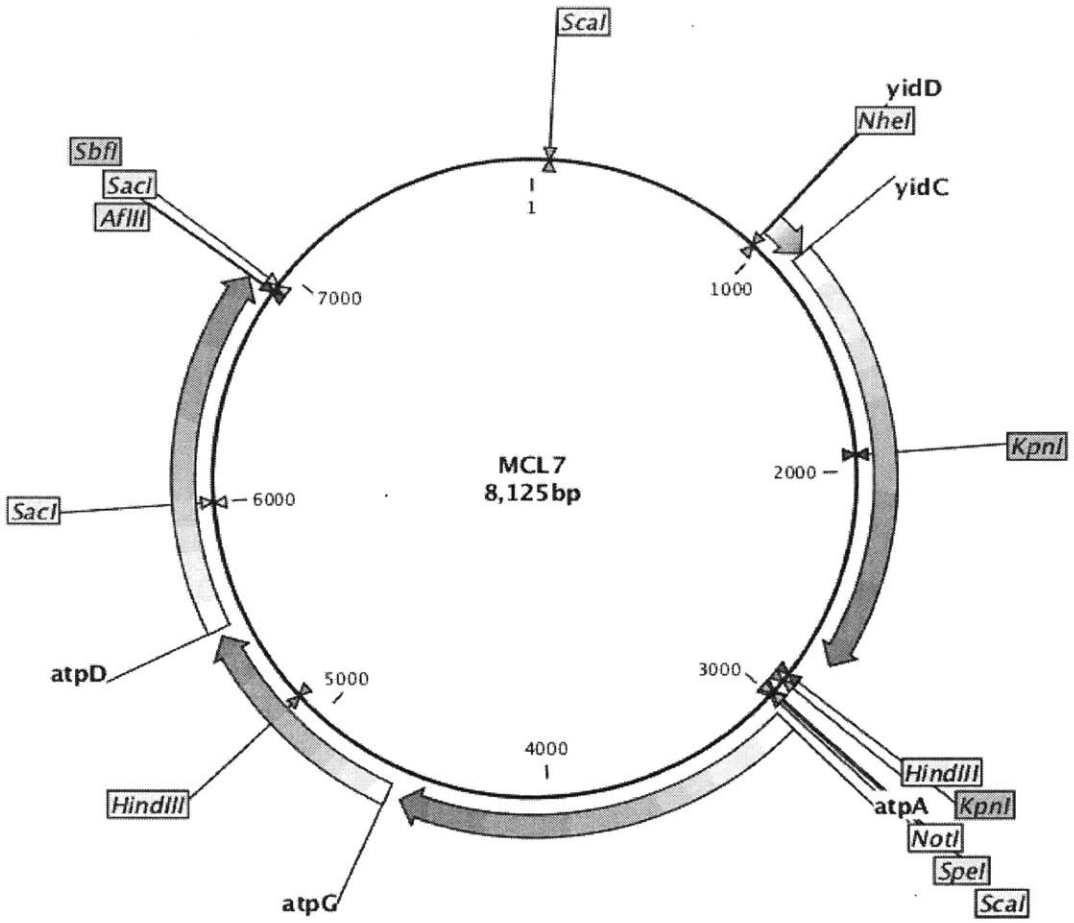



**Figure S10.** Map of pMCL7

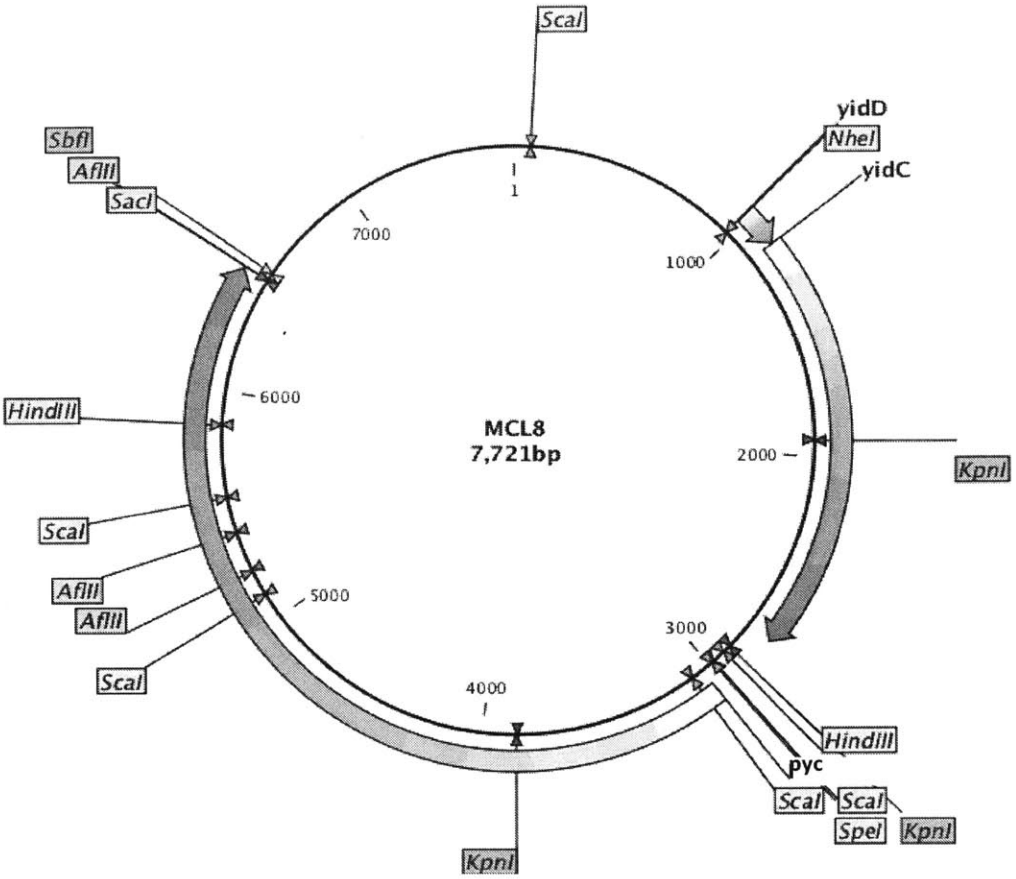



**Figure S11.** Map of pMCL8

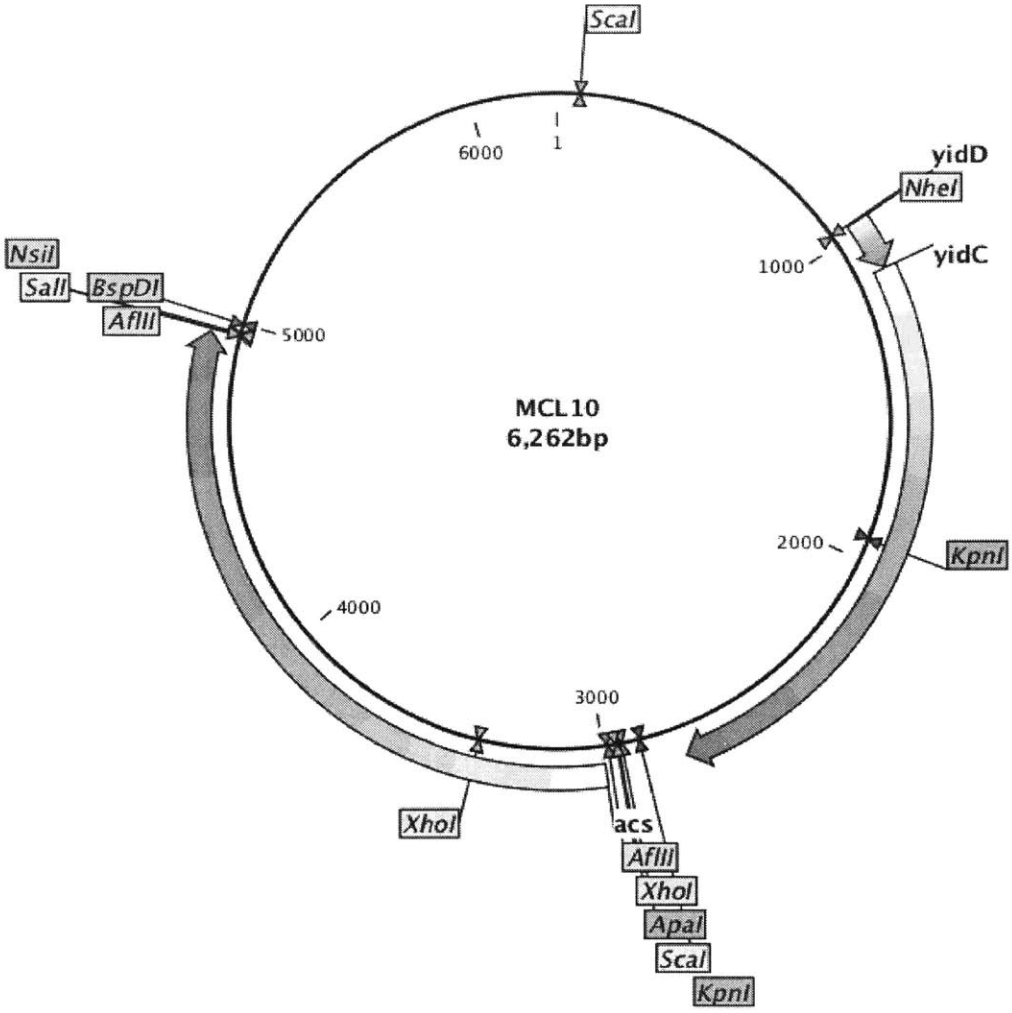



**Figure S12.** Map of pMCL10

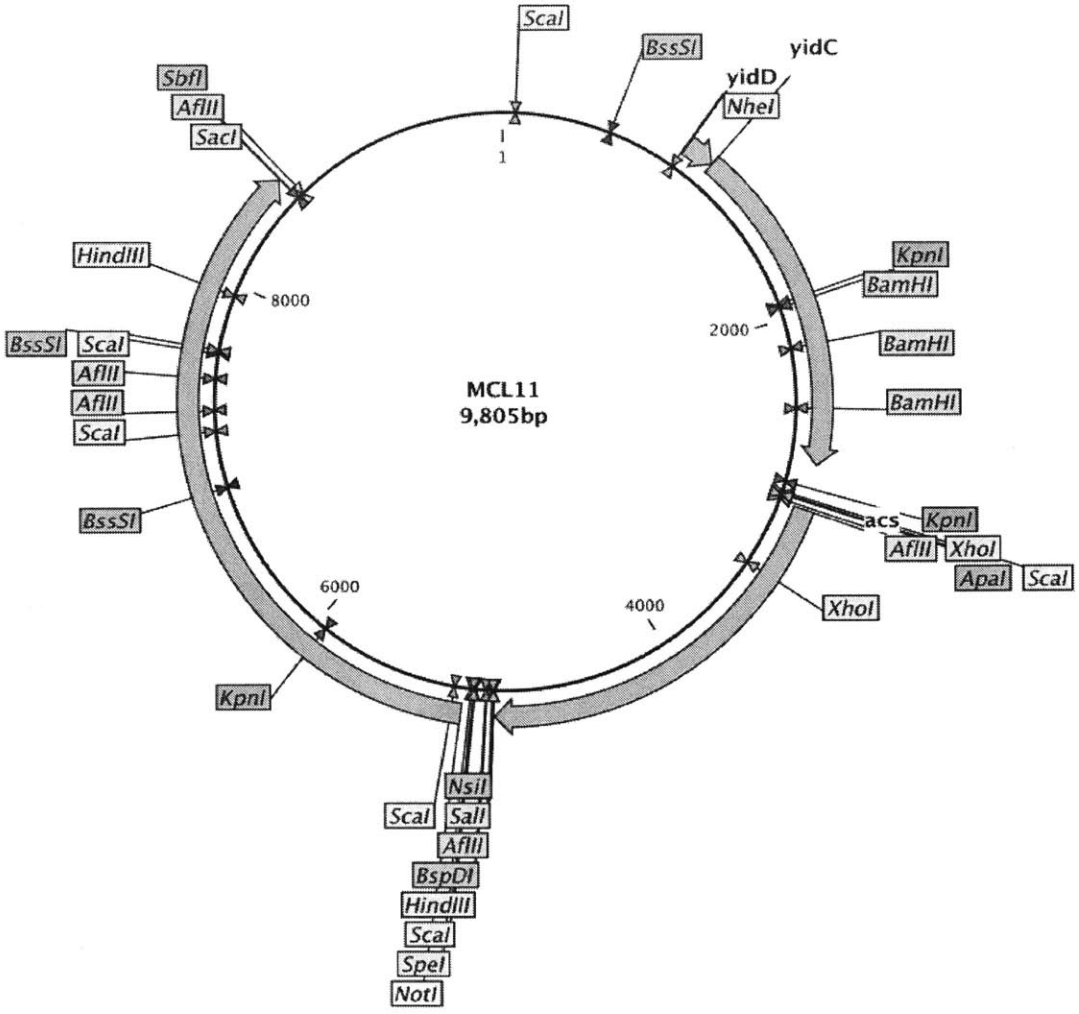



**Figure S13.** Map of pMCL11

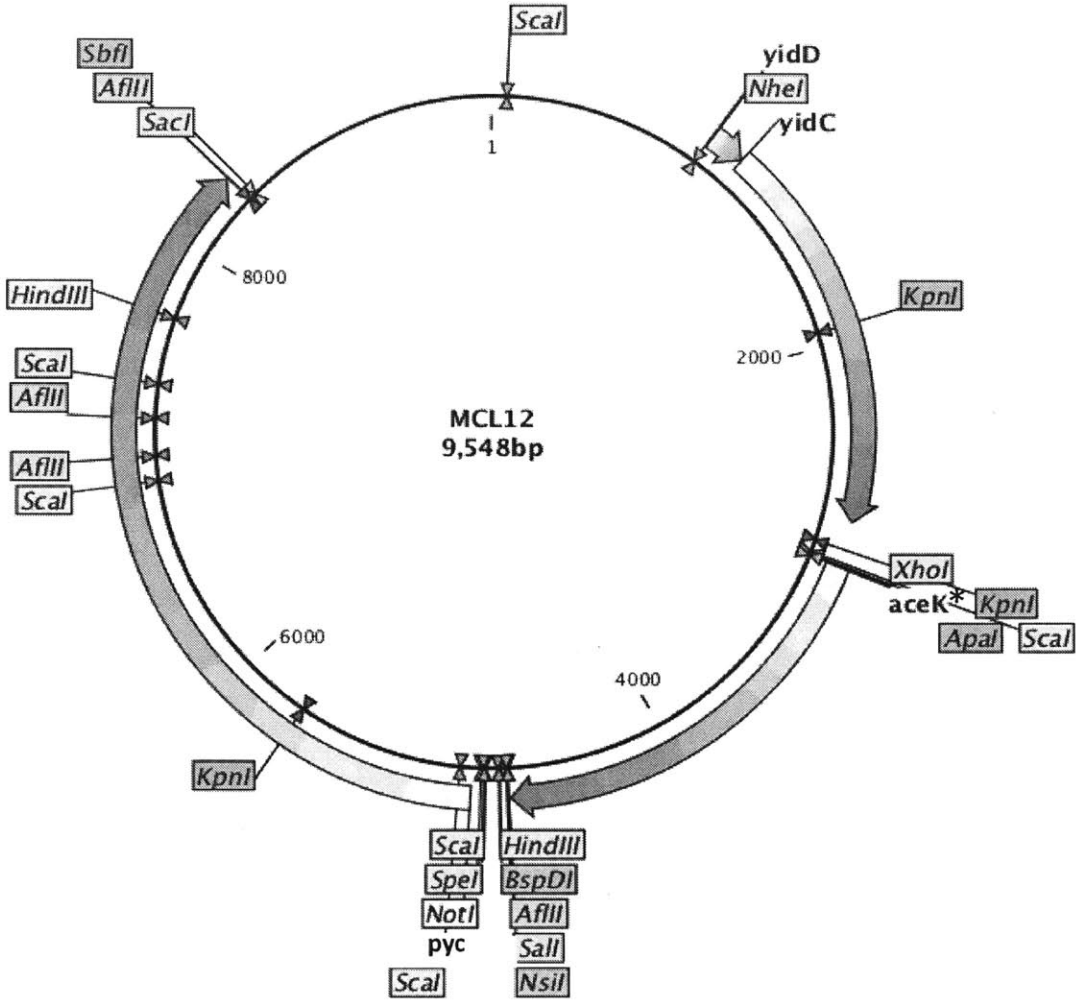
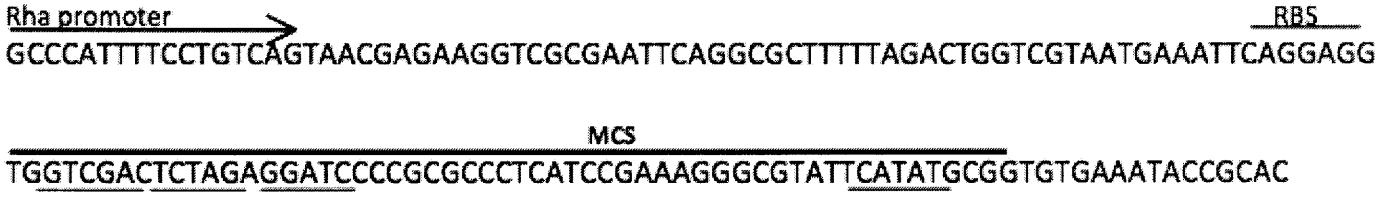



**Figure S14.** Map of pMCL12

**Table S1.** DNA sequence of pRHA1 and pRHA2




*Sequence provided by manufacturer

**RHA1:** 31a08 green PR in pRHA801

GCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTAATGAAATTCAGGAGGTGGTCGACATGAAATTA
TTACTGATATTAGGTAGTGTTATTGCACTTCCTACATTTGCTGCAGGTGGTGGTGACCTTGATGCTAGTGATTACACTGGTGTTTCTTTTTGGTT
AGTTACTGCTGCTTTATTAGCATCTACTGTATTTTTCTTTGTTGAAAGAGATAGAGTTTCTGCAAAATGGAAAACATCATTAACTGTATCTGGTC
TTGTTACTGGTATTGCTTTCTGGCATTACATGTACATGAGAGGGGTATGGATTGAAACTGGTGATTCGCCAACTGTATTTAGATACATTGATTG
GTTACTAACAGTTCCTCTATTAATATGTGAATTCTACTTAATTCTTGCTGCTGCAACTAATGTTGCTGGATCATTATTTAAGAAATTACTAGTTGG
TTCTCTTGTTATGCTTGTGTTTGGTTACATGGGTGAAGCAGGAATCATGGCTGCATGGCCTGCATTCATTATTGGGTGTTTAGCTTGGGTATAC
ATGATTTATGAATTATGGGCTGGAGAAGGAAAATCTGCATGTAATACTGCAAGTCCTGCTGTGCAATCAGCTTACAACACAATGATGTATATT
ATCATCTTTGGTTGGGCGATTTATCCTGTAGGTTATTTCACAGGTTACCTGATGGGTGACGGTGGATCAGCTCTTAACTTAAACCTTATCTATA
ACCTTGCTGACTTTGTTAACAAGATTCTATTTGGTTTAATTATATGGAATGTTGCTGTTAAAGAATCTTCTAATGCTTAAGGATCCCCGCGCCCT
CATCCGAAAGGGCGTATTCATATGCGGTGTGAAATACCGCAC

**RHA2:** 19p19 blue PR in pRHA801

GCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTAATGAAATTCAGGAGGTGGTCGACATGTCAATA
ATTAAGTCTTTAAAGTCAGTCGCTATTGCATCACTGGCTATTCTCATTCCATCAATTGCATTAGCAGCTGGTGGAAACTTGGAGCCGAACGATC
CTGTCGGCATTACCTTTTGGTTGATATCGATCGCAATGGTCGCCGCTACCGTATTTTTCTTGATGGAATCATTGAGAGTTGACGGCAAGTGGA
GAACATCAATGATTGTGGGCGGACTTGTTACACTAGTAGCTGCGGTCCACTACTTTTATATGCGTGATGTTTGGGTTGCCACAGGTGCATCGC
CCACGGTCTTTCGTTATGTAGATTGGCTAATCACAGTGCCTCTGCAGATGATTGAATTCTACCTTATTTTAGCTGCTTGCACTGCCATTGCTGTC
GGCGTCTTCTGGCGCCTCATGATCGGCACAATGGTAATGTTAATTGGCGGATATCTAGGTGAGGCGGGCTTCATCAACGCTACGGTTGGTTTC
GTGATCGGAATGGCCGGTTGGGGATATATATTGTATGAAATCTTTGCTGGTGAAGCTGGCAAGGTTGCTGCAGAAGGTGCGCCCCCATCTGT
ACAATCAGCTTTCAACACAATGCGTCTGATTGTGACTATTGGTTGGGCTATTTATCCACTTGGATATTTCTTTGGCTACATGACCGGAGGAGTT
GACGCAAATTCGCTGAACCTGATCTACAACGTTGCTGACGTTGTTAACAAGATTGGTTTCTGTTTAGCAATTTGGGCTGCTGCTACTTCACAAT
CTGAAGCGGCAAAGTAAGGATCCCCGCGCCCTCATCCGAAAGGGCGTATTCATATGCGGTGTGAAATACCGCAC

**Table S2.** DNA sequence of pRHA4

GCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTAATGAAATTCAGGAGGTGGTCGACATGAAATTATTACTGATATTAGGTAGTGTTATTGCACTTCCTA
CATTTGCTGCAGGTGGTGGTGACCTTGATGCTAGTGATTACACTGGTGTTTCTTTTTGGTTAGTTACTGCTGCTTTATTAGCATCTACTGTATTTTTCTTTGTTGAAAGAGATAGAGTTTCTGCAAAAT
GGAAAACATCATTAACTGTATCTGGTCTTGTTACTGGTATTGCTTTCTGGCATTACATGTACATGAGAGGGGTATGGATTGAAACTGGTGATTCGCCAACTGTATTTAGATACATTGATTGGTTACT
AACAGTTCCTCTATTAATATGTGAATTCTACTTAATTCTTGCTGCTGCAACTAATGTTGCTGGATCATTATTTAAGAAATTACTAGTTGGTTCTCTTGTTATGCTTGTGTTTGGTTACATGGGTGAAGC
AGGAATCATGGCTGCATGGCCTGCATTCATTATTGGGTGTTTAGCTTGGGTATACATGATTTATGAATTATGGGCTGGAGAAGGAAAATCTGCATGTAATACTGCAAGTCCTGCTGTGCAATCAGC
TTACAACACAATGATGTATATTATCATCTTTGGTTGGGCGATTTATCCTGTAGGTTATTTCACAGGTTACCTGATGGGTGACGGTGGATCAGCTCTTAACTTAAACCTTATCTATAACCTTGCTGACTT
TGTTAACAAGATTCTATTTGGTTTAATTATATGGAATGTTGCTGTTAAAGAATCTTCTAATGCTTAAGGATCCACAAAAGAACACAAAATAGAGGAGGTACCCAATGATTCTTATAATTTATGCGCAT
CCGTATCCGCATCATTCCCATGCGAATAAACGGATGCTTGAACAGGCAAGGACGCTGGAAGGCGTCGAAATTCGCTCTCTTTATCAACTCTATCCTGACTTCAATATCGATATTGCCGCCGAGCAGG
AGGCGCTGTCTCGCGCCGATCTGATCGTCTGGCAGCATCCGATGCAGTGGTACAGCATTCCTCCGCTCCTCAAACTTTGGATCGATAAAGTTTTCTCCCACGGCTGGGCTTACGGTCATGGCGGCAC
GGCGCTGCATGGCAAACATTTGCTGTGGGCGGTGACGACCGGCGGCGGGGAAAGCCATTTTGAAATTGGTGCGCATCCGGGCTTTGATGTGCTGTCGCAGCCGCTACAGGCGACGGCAATCTAC
TGCGGGCTGAACTGGCTGCCACCGTTTGCCATGCACTGCACCTTTATTTGTGACGACGAAACCCTCGAAGGGCAGGCGCGTCACTATAAGCAACGTCTGCTGGAATGGCAGGAGGCCCATCATGG
ATAGCCATACGCTGATTCAGGCGCTGATTTATCTCGGTTCGGCAGCGCTGATTGTACCCATTGCGGTACGTCTTGGTCTGGGATCGGTACTTGGCTACCTGATCGCCGGCTGCATTATTGGCCCGTG
GGGGCTGCGACTGGTGACCGATGCCGAATCTATTCTGCACTTTGCCGAGATTGGGGTGGTGCTGATGCTGTTTATTATCGGCCTCGAACTCGATCCACAAAGGCTGTGGAAGCTGCGTGCGGCAG
TGTTCGGCTGTGGCGCATTGCAGATGGTGATTTGCGGCGGCCTGCTGGGGCTGTTCTGCATGTTACTTGGGCTGCGCTGGCAGGTCGCGGAATTGATCGGCATGACGCTGGCGCTCTCCTCTACG
GCGATTGCCATGCAGGCGATGAATGAACGCAATCTGATGGTGACGCAAATGGGTCGCAGTGCCTTTGCGGTGCTGCTGTTCCAGGATATCGCGGCGATCCCGCTGGTGGCGATGATTCCGCTACT
GGCAACGAGCAGTGCCTCGACGACGATGGGCGCATTTGCTCTCTCGGCGTTAAAAGTGGCGGGTGCGCTGGTGCTGGTGGTATTGCTGGGGCGCTATGTCACGCGTCCGGCGCTGCGTTTTGTA
GCCCGCTCTGGCTTGCGGGAAGTGTTTAGTGCCGTGGCGTTATTCCTCGTGTTTGGCTTTGGTTTGCTGCTGGAAGAGGTCGGCTTGTCGATGGCGATGGGCGCGTTTCTGGCGGGCGTACTGCTG
GCAAGCTCGGAATACCGTCATGCGCTGGAGAGCGATATCGAACCATTTAAAGGTTTGCTGTTGGGGCTGTTTTTCATCGGTGTTGGCATGTCGATAGACTTTGGCACGCTGCTTGAAAACCCATTG
CGCATTGTCATTTTGCTGCTCGGTTTCCTCATCATCAAAATCGCCATGCTGTGGCTGATTGCCCGACCGTTGCAAGTGCCAAATAAACAGCGTCGTTGGTTTGCGGTGTTGTTAGGGCAGGGCAGTG
AGTTTGCCTTTGTGGTATTTGGCGCGGCGCAGATGGCGAATGTGCTGGAGCCGGAGTGGGCGAAATCGCTGACCCTGGCGGTGGCGCTGTCGATGGCAGCAACGCCGATTCTGCTGGTGATCCT
CAATCGCCTTGAGCAATCTTCTACTGAGGAAGCGCGTGAAGCCGATGAGATCGACGAAGAACAGCCGCGCGTGATTATCGCCGGATTCGGTCGTTTTGGGCAGATTACCGGAAGTTTACTGCTCT
CCAGCGGGGTGAAAATGGTGGTACTCGATCACGATCCGGACCATATCGAAACCTTGCGTAAATTTGGTATGAAAGTGTTTTATGGCGATGCCACGCGGATGGATTTACTGGAATCTGCCGGAGCG
GCGAAAGCGGAAGTGCTGATTAACGCCATCGACGATCCGCAAACCAACCTGCAACTGACAGAGATGGTGAAAGAACATTTCCCGCATTTGCAGATTATTGCCCGCGCCCGCGATGTCGACCACTA
CATTCGTTTGCGTCAGGCAGGCGTTGAAAAGCCGGAGCGTGAAACCTTCGAAGGTGCGCTGAAAACCGGGCGTCTGGCACTGGAAAGTTTAGGTCTGGGGCCGTATGAAGCGCGAGAACGTGCC
GATGTGTTCCGCCGCTTTAATATTCAGATGGTGGAAGAGATGGCAATGGTTGAGAACGACACCAAAGCCCGCGCGGCGGTCTATAAACGCACCAGCGCGATGTTAAGTGAGATCATTACCGAGG
ACCGCGAACATCTGTCATTAATTCAACGACATGGCTGGCAGGGAACCGAAGAAGGTAAACATACCGGCAACATGGCGGATGAACCGGAAACGAAACCCTCATCCTAACATATGCGGTGTGAAAT
ACCGCAC

**Table S3.** DNA sequence of pMCL200

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTT
ATTGGTGCCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCG
AAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTC
CTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTC
GTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGA
ACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGG
CGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTC
GAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTG
ACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTT
CCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGA
GCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGTACCGGGCCCCCCCTCGAGGTCGACGG
TATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTG
GCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATC
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGC
GCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACAC
TAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGC
CCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCA
GGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCT
ATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGG
AATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCAC
GACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCC
TGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTG

**Table S4.** DNA sequence of pMCL2

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTT
ATTGGTGCCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCG
AAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTC
CTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTC
GTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGA
ACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGG
CGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTC
GAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTG
ACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTACGGAATTGAGGCATTGCGCAGGTTTGGAGTGATAAAAGGCAGTTGGTTGACGGTGA
AACGCGTATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCCGCCCGGACCATTTGATACCAGAGAACACTAACGATGGATTCGCAACGCAATCTTTTAGTCATCG
CTTTGCTGTTCGTGTCTTTCATGATCTGGCAAGCCTGGGAGCAGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCGCCGCCG
ACCAGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACACCCGTGGTGGTGATGTTGAGCAAGCTCTGCTGCCTGCTT
ACCCGAAAGAGCTGAACTCTACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGCACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCGGCTAA
CGGCCCGCGTCCGCTGTATAACGTTGAAAAAGACGCTTATGTGCTGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAA
AACGTTTGTCCTGAAACGTGGTGATTACGCTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGAAACCGCTGGAAATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACT
CTGCCACCGCATCTCGATACCGGAAGCAGCAACTTCGCACTGCACACCTTCCGTGGCGCGGCGTACTCCACGCCTGACGAGAAGTATGAGAAATACAAGTTCGATACCATTGCCG
ATAACGAAAACCTGAACATCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCAACAACTTCTATACCGCTAA
TCTGGGTAACGGCATCGCCGCTATCGGCTATAAATCTCAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGCGATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAA
AATGGCAGCTGTTGCTCCGCACCTGGATCTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTGTGGGTAACTGGG
GCTTCTCCATTATCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCTGACCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATG
CGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCTGTACAAAGCTGAGAAGGTTAACCCGCTGGGCGGCTGCTTCCCGCTGCTGATCCAGATGCCA
ATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACATCCTGCCGATCCT
GATGGGCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGT
TCCCGTCAGGTCTGGTGCTGTACTATATCGTCAGCAACCTGGTAACCATTATTCAGCAGCAGCTGATTTACCGTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAA
ATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGGTCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCGGATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCATG
AGCGAGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCC
AATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCC
AGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGT
GGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAA
GCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAA
ATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCAC
TACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACA
TTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTAT
TCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAA
CTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGG
TTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATT
ATACGCAAGGCGACAAGGTGCTGATGCCGCTG

**Table S5.** DNA sequence of pMCL6

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTAGAAATATTTTATCTG
ATTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGG
AGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTT
ACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTA
AACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTA
TGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAG
GAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTAC
GGAATTGAGGCATTGCGCAGGTTTGGAGTGATAAAAGGCAGTTGGTTGACGGTGAAACGCGTATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCCGCCCGGACCATTTGATACCAGAGAACACTAAC
GATGGATTCGCAACGCAATCTTTTAGTCATCGCTTTGCTGTTCGTGTCTTTCATGATCTGGCAAGCCTGGGAGCAGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCG
CCGCCGACCAGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACACCCGTGGTGGTGATGTTGAGCAAGCTCTGCTGCCTGCTTACCCGAAAGAGCTGAACTC
TACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGCACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCGGCTAACGGCCCGCGTCCGCTGTATAACGTTGAAAAAGACGCTTATGTGC
TGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAAAACGTTTGTCCTGAAACGTGGTGATTACGCTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGA
AACCGCTGGAAATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACTCTGCCACCGCATCTCGATACCGGAAGCAGCAACTTCGCACTGCACACCTTCCGTGGCGCGGCGTACTCCACGCCTGACGAGAAGTATGAGAAA
TACAAGTTCGATACCATTGCCGATAACGAAAACCTGAACATCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCAACAACTTCTATACCGCTAATCT
GGGTAACGGCATCGCCGCTATCGGCTATAAATCTCAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGCGATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAAAATGGCAGCTGTTGCTCCGCACCTGGAT
CTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTGTGGGTAACTGGGGCTTCTCCATTATCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCTGA
CCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATGCGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCTGTACAAAGCTGAGAAGGTTAACC
CGCTGGGCGGCTGCTTCCCGCTGCTGATCCAGATGCCAATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACA
TCCTGCCGATCCTGATGGGCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGTTCCCGTCAGGTC
TGGTGCTGTACTATATCGTCAGCAACCTGGTAACCATTATTCAGCAGCAGCTGATTTACCGTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAAATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGG
TCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCGGATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCATGAGCGAGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTCATCGGGTAGTTTAT
TCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTACTAGTTCTAGAGCGGCCGCTCCACATACGGAGCGGGGTCGAGGAGGTCAAATATGAGTAAAATCGTTGTAGTCGGTGCTAACCACGCTGGTACA
GCATGTATCAATACCATGTTGGATAATTTTGGAAATGAGAACGAAATTGTTGTATTTGACCAAAACTCTAACATCTCTTTCCTAGGATGTGGAATGGCTCTTTGGATTGGTGAACAAATTGACGGTGCTGAAGGCTTGTTC
TATTCTGATAAAGAAAAATTGGAAGCTAAAGGTGCTAAAGTTTACATGAACTCACCTGTTCTTTCAATCGACTATGATAACAAAGTAGTTACAGCGGAAGTTGAAGGAAAAGAGCACAAAGAATCATACGAAAAATTGAT
TTTCGCTACAGGTTCTACACCAATCTTGCCACCAATCGAAGGTGTTGAAATTGTTAAAGGAAACCGCGAATTTAAAGCAACTCTTGAAAACGTACAATTCGTGAAATTGTACCAAAATGCTGAAGAAGTTATCAATAAACT
TTCTGACAAGAGCCAACACCTCGACCGTATCGCCGTTGTTGGTGGTGGTTACATCGGTGTTGAACTTGCTGAAGCCTTTGAACGTCTTGGAAAAGAAGTTGTCCTTGTTGATATCGTTGATACTGTCTTGAACGGTTACTA
TGACAAAGACTTCACACAAATGATGGCGAAGAACTTGGAAGATCACAACATCCGCTTGGCTCTAGGTCAAACTGTTAAAGCAATCGAAGGTGACGGTAAAGTTGAACGCTTGATTACTGACAAAGAAAGCTTTGACGTG
GATATGGTTATCCTTGCAGTTGGTTTCCGTCCAAACACAGCCCTTGCAGATGGTAAGATCGAACTCTTCCGCAACGGTGCCTTCCTTGTAGACAAGAAACAAGAAACATCTATCCCAGGCGTTTACGCTGTTGGTGACTGT
GCGACTGTTTATGACAATGCTCGTAAAGATACAAGCTATATCGCTCTTGCTTCAAATGCTGTTCGCACTGGTATCGTTGGTGCCTACAATGCTTGTGGACATGAATTGGAAGGAATCGGTGTTCAAGGTTCAAATGGTATC
TCAATCTACGGTCTTCACATGGTTTCAACTGGTTTGACTCTTGAAAAAGCGAAAGCTGCTGGTTACAACGCAACTGAAACAGGCTTTAACGATCTTCAAAAACCAGAATTCATGAAACATGACAACCATGAAGTAGCAATT
AAGATTGTCTTTGACAAAGATAGCCGTGAAATTCTTGGTGCCCAAATGGTTTCACATGATATTGCAATTAGCATGGGAATCCACATGTTCTCACTTGCTATCCAAGAGCATGTGACAATTGATAAATTGGCATTGACAGAC
CTCTTCTTCTTGCCACACTTCAACAAACCATACAACTACATCACAATGGCTGCCCTTACGGCTGAAAAATAGCCTGCAGGCTTAAGCCCCGCTTCGGCGGGGTTTTTTGAGCTCCAATTCGCCCTATAGTGAGTCGTATTAC
AATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTG
CGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGT
CAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATC
ACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTA
TTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTAT
AACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGT
GAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTAC
GGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCA
AATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTG

**Table S6.** DNA sequence of pMCL7

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTT
TGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGA
CTAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGT
GTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGA
TGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGC
GGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTACGGAATTGAGGCATTGCGCAGGTTTGGAGTGATAA
AAGGCAGTTGGTTGACGGTGAAACGCGTATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCGCCCGGACCATTTGATACCAGAGAACACTAACGATGGATTCGCAACGCAATCTTTTAGTCATCGCTTTGCTGTTCGTGTCTTTCATGATCTGGCAAG
CCTGGGAGCAGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCGCCGCCGACCAGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACACCCGTGGTGG
TGATGTTGAGCAAGCTCTGCTGCCTGCTTACCCGAAAGAGCTGAACTCTACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGCACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCGGCTAACGGCCCGCGTCCGCTGTATAACGTT
GAAAAAGACGCTTATGTGCTGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAAAACGTTTGTCCTGAAACGTGGTGATTACGCTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGAAACCGCTGGA
AATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACTCTGCCACCGCATCTCGATACCGGAAGCAGCAACTTCGCACTGCACACCTTCCGTGGCGCGGCGTACTCCACGCCTGACGAGAAGTATGAGAAATACAAGTTCGATACCATTGCCGATAACGAAAACCTGAACA
TCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCAACAACTTCTATACCGCTAATCTGGGTAACGGCATCGCCGCTATCGGCTATAAATCTCAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGC
GATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAAAATGGCAGCTGTTGCTCCGCACCTGGATCTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTGTGGGTAACTGGGGCTTCTCCATTA
TCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCTGACCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATGCGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCTGTACAAAGCTG
AGAAGGTTAACCCGCTGGGCGGCTGCTTCCCGCTGCTGATCCAGATGCCAATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACATCCTGCCGATCCTGATG
GGCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGTTCCCGTCAGGTCTGGTGCTGTACTATATCGTCAGCAACCTGGTAACCATTATTCAGCAG
CAGCTGATTTACCGTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAAATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGGTCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCGGATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCA
TGAGCGAGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTCATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTACTAGTTCTAGAGCGGCCGCGGGCCGCGACGACTAAGGAGGTACTAAATGCAACTGAATTC
CACCGAAATCAGCGAACTGATCAAGCAGCGCATTGCTCAGTTCAATGTTGTGAGTGAAGCTCACAACGAAGGTACTATTGTTTCTGTAAGTGACGGTGTTATCCGCATTCACGGCCTGGCCGATTGTATGCAGGGTGAAATGATCTCCCTGCCGGGTAACCGTTACGCT
ATCGCACTGAACCTCGAGCGCGACTCTGTAGGTGCGGTTGTTATGGGTCCGTACGCTGACCTTGCCGAAGGCATGAAAGTTAAGTGTACTGGCCGTATCCTGGAAGTTCCGGTTGGCCGTGGCCTGCTGGGCCGTGTGGTTAACACTCTGGGTGCACCAATCGACGGT
AAAGGTCCGCTGGATCACGACGGCTTCTCTGCTGTAGAAGCAATCGCTCCGGGCGTTATCGAACGTCAGTCCGTAGATCAGCCGGTACAGACCGGTTATAAAGCCGTTGACTCCATGATCCCAATCGGTCGTGGTCAGCGTGAATTGATCATCGGTGACCGTCAGACAG
GTAAAACCGCACTGGCTATCGATGCCATCATCAACCAGCGCGATTCCGGTATCAAATGTATCTATGTCGCTATCGGCCAGAAAGCGTCCACCATTTCTAACGTGGTACGTAAACTGGAAGAGCACGGCGCACTGGCTAACACCATCGTTGTGGTAGCAACCGCGTCTGA
ATCCGCTGCACTGCAATACCTGGCACCGTATGCCGGTTGCGCAATGGGCGAATACTTCCGTGACCGCGGTGAAGATGCGCTGATCATTTACGATGACCTGTCTAAACAGGCTGTTGCTTACCGTCAGATCTCCCTGCTGCTCCGTCGTCCGCCAGGACGTGAAGCATTCC
CGGGCGACGTTTTCTACCTCCACTCTCGTCTGCTGGAGCGTGCTGCACGTGTTAACGCCGAATACGTTGAAGCCTTCACCAAAGGTGAAGTGAAAGGGAAAACCGGTTCTCTGACCGCACTGCCGATTATCGAAACTCAGGCGGGTGACGTTTCTGCGTTCGTTCCGAC
CAACGTAATCTCCATTACCGATGGTCAGATCTTCCTGGAAACCAACCTGTTCAACGCCGGTATTCGTCCTGCGGTTAACCCGGGTATTTCCGTATCCCGTGTTGGTGGTGCAGCACAGACCAAGATCATGAAAAAACTGTCCGGTGGTATCCGTACCGCTCTGGCACAGT
ATCGTGAACTGGCAGCGTTCTCTCAGTTTGCATCCGACCTTGACGATGCAACACGTAAGCAGCTTGACCACGGTCAGAAAGTGACCGAACTGCTGAAACAGAAACAGTATGCGCCGATGTCCGTTGCGCAGCAGTCTCTGGTTCTGTTCGCAGCAGAACGTGGTTACCT
GGCGGATGTTGAACTGTCGAAAATTGGCAGCTTCGAAGCCGCTCTGCTGGCTTACGTCGACCGTGATCACGCTCCGTTGATGCAAGAGATCAACCAGACCGGTGGCTACAACGACGAAATCGAAGGCAAGCTGAAAGGCATCCTCGATTCCTTCAAAGCAACCCAATC
CTGGTAACGTCTGGCGGCTTGCCTTAGGGCAGGCCGCAAGGCATTGAGGAGAAGCTCATGGCCGGCGCAAAAGAGATACGTAGTAAGATCGCAAGCGTCCAGAACACGCAAAAGATCACTAAAGCGATGGAGATGGTCGCCGCTTCCAAAATGCGTAAATCGCAGG
ATCGCATGGCGGCCAGCCGTCCTTATGCAGAAACCATGCGCAAAGTGATTGGTCACCTTGCACACGGTAATCTGGAATATAAGCACCCTTACCTGGAAGACCGCGACGTTAAACGCGTGGGCTACCTGGTGGTGTCGACCGACCGTGGTTTGTGCGGTGGTTTGAACA
TTAACCTGTTCAAAAAACTGCTGGCGGAAATGAAGACCTGGACCGACAAAGGCGTTCAATGCGACCTCGCAATGATCGGCTCGAAAGGCGTGTCGTTCTTCAACTCCGTGGGCGGCAATGTTGTTGCCCAGGTCACCGGCATGGGGGATAACCCTTCCCTGTCCGAACT
GATCGGTCCGGTAAAAGTGATGTTGCAGGCCTACGACGAAGGCCGTCTGGACAAGCTTTACATTGTCAGCAACAAATTTATTAACACCATGTCTCAGGTTCCGACCATCAGCCAGCTGCTGCCGTTACCGGCATCAGATGATGATGATCTGAAACATAAATCCTGGGAT
TACCTGTACGAACCCGATCCGAAGGCGTTGCTGGATACCCTGCTGCGTCGTTATGTCGAATCTCAGGTTTATCAGGGCGTGGTTGAAAACCTGGCCAGCGAGCAGGCCGCCCGTATGGTGGCGATGAAAGCCGCGACCGACAATGGCGGCAGCCTGATTAAAGAGCT
GCAGTTGGTATACAACAAAGCTCGTCAGGCCAGCATTACTCAGGAACTCACCGAGATCGTCTCGGGGGCCGCCGCGGTTTAAACAGGTTATTTCGTAGAGGATTTAAGATGGCTACTGGAAAGATTGTCCAGGTAATCGGCGCCGTAGTTGACGTCGAATTCCCTCAG
GATGCCGTACCGCGCGTGTACGATGCTCTTGAGGTGCAAAATGGTAATGAGCGTCTGGTGCTGGAAGTTCAGCAGCAGCTCGGCGGCGGTATCGTACGTACCATCGCAATGGGTTCCTCCGACGGTCTGCGTCGCGGTCTGGATGTAAAAGACCTCGAACACCCGATT
GAAGTCCCGGTAGGTAAAGCGACTCTGGGCCGTATCATGAACGTACTGGGTCGAACCGGTCGACATGAAAGGCGAGATCGGTGAAGAAGAGCGTTGGGCGATTCACCGCGCAGCACCTTCCTACGAAGAGCTGTCAAACTCTCAGGAACTGCTGGAAACCGGTATCA
AAGTTATCGACCTGATGTGTCCGTTCGCTAAGGGCGGTAAAGTTGGTCTGTTCGGTGGTGCGGGTGTAGGTAAAACCGTAAACATGATGGAGCTCATTCGTAACATCGCGATCGAGCACTCCGGTTACTCTGTGTTTGCGGGCGTAGGTGAACGTACTCGTGAGGGTA
ACGACTTCTACCACGAAATGACCGACTCCAACGTTATCGACAAAGTATCCCTGGTGTATGGCCAGATGAACGAGCCGCCGGGAAACCGTCTGCGCGTTGCTCTGACCGGTCTGACCATGGCTGAGAAATTCCGTGACGAAGGTCGTGACGTTCTGCTGTTCGTTGACAA
CATCTATCGTTACACCCTGGCCGGTACGGAAGTATCCGCACTGCTGGGCCGTATGCCTTCAGCGGTAGGTTATCAGCCGACCCTGGCGGAAGAGATGGGCGTTCTGCAGGAACGTATCACCTCCACCAAAACTGGTTCTATCACCTCCGTACAGGCAGTATACGTACCT
GCGGATGACTTGACTGACCCGTCTCCGGCAACCACCTTTGCGCACCTTGACGCAACCGTGGTACTGAGCCGTCAGATCGCGTCTGGGACTCCACCAGCCGTCAGCTGGTGGTTGGTCAGGAACACTACGACA
CCGCGCGTGGCGTTCAGTCCATCCTGCAACGTTATCAGGAACTGAAAGACATCATCGCCATCCTGGGTATGGATGAACTGTCTGAAGAAGACAAACTGGTGGTAGCGCGTGCTCGTAAGATCCAGCGCTTCCTGTCCCAGCCGTTCTTCGTGGCAGAAGTATTCACCGG
TTCTCCGGGTAAATACGTCTCCCTGAAAGACACCATCCGTGGCTTTAAAGGCATCATGGAAGGCGAATACGATCACCTGCCGGAGCAGGCGTTCTACATGGTCGGTTCCATCGAAGAAGCTGTGGAAAAAGCCAAAAAACTTTAACCTGCAGGCTTAAGCCCCGCTTCG
GCGGGGTTTTTTGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATC
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACA
GGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCG
GGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCAC
CGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATG
AATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGT
GGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGA
CAAGGTGCTGATGCCGCTG

**Table S7.** DNA sequence of pMCL8

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTT
TGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGA
CTAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGT
GTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGA
TGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGC
GGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTACGGAATTGAGGCATTGCGCAGGTTTGGAGTGATAA
AAGGCAGTTGGTTGACGGTGAAACGCGTATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCCGCCCGGACCATTTGATACCAGAGAACACTAACGATGGATTCGCAACGCAATCTTTTAGTCATCGCTTTGCTGTTCGTGTCTTTCATGATCTGGCAAG
CCTGGGAGCAGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCGCCGCCGACCAGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACACCCGTGGTGG
TGATGTTGAGCAAGCTCTGCTGCCTGCTTACCCGAAAGAGCTGAACTCTACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGCACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCGGCTAACGGCCCGCGTCCGCTGTATAACGTT
GAAAAAGACGCTTATGTGCTGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAAAACGTTTGTCCTGAAACGTGGTGATTACGCTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGAAACCGCTGGA
AATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACTCTGCCACCGCATCTCGATACCGGAAGCAGCAACTTCGCACTGCACACCTTCCGTGGCGCGGCGTACTCCACGCCTGACGAGAAGTATGAGAAATACAAGTTCGATACCATTGCCGATAACGAAAACCTGAACA
TCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCAACAACTTCTATACCGCTAATCTGGGTAACGGCATCGCCGCTATCGGCTATAAATCTCAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGC
GATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAAAATGGCAGCTGTTGCTCCGCACCTGGATCTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTGTGGGTAACTGGGGCTTCTCCATTA
TCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCTGACCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATGCGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCTGTACAAAGCTG
AGAAGGTTAACCCGCTGGGCGGCTGCTTCCCGCTGCTGATCCAGATGCCAATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACATCCTGCCGATCCTGATG
GGCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGTTCCCGTCAGGTCTGGTGCTGTACTATATCGTCAGCAACCTGGTAACCATTATTCAGCAG
CAGCTGATTTACCGTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAAATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGGTCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCGGATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCA
TGAGCGAGGTACCGGGCCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTCATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTACTAGTTCTAGAGCGGCCGCTACAAGCCAATCAAAATCCCCTAAGGAGGTAATACATGAAA
ATTAAAAAAGTCCTCGTCGCAAATCGCGGGGAAATTGCTATTAGAGTACTACGTGCCTGTACAGAACTTAATATTGCAACCGTAGCTGTCTACACCTATGAAGATAGATACTCACAACATCGCAACAAAGCAGATGAGTCTTACCAGATAGGAGCAGATAATGAACCGC
TCAAACCCTATTTAGATATGGATCAAATCATTGATCTAGCAAAATCTAAAAATGTAGATGCCATTCACCCCGGCTACGGCTTCTTATCTGAAAACTCAGAATTTGCAAGAAGATGTGCAAAAAATGGCATTATATTTATTGGCCCTGAGCCAGAAGTAATGGATGCACTA
GGTGATAAAATCACCGCAAAAAAAGTAGCACAATCTTGTGGAGTACCTATAATAGAAAGTAATAAAAAAGCCTTAGATTCACTAGAAACGGCCCTAAGTGAAGCTGGAAAAATAGGATATCCAGTAATGCTCAAAGCTGCCTCTGGTGGAGGTGGTCGCGGTATGCGT
ATTATACGTAAGGAAGAAGATTTACAGCTTAGTTTTGACTCTGCACGTAATGAGGCGCTCAACTCTTTTGGAGATGATACAATGTTTTTAGAAAAATATGTAGAAGATCCAAAACATATAGAAGTACAAATAGTAGCAGACACACACGGTAATATAAGACACCTTTATGA
ACGTGATTGCTCTGTACAACGCAGGCACCAGAAGGTGGTAGAAATGGCTCCTTCATATAACCTATCACAAACGCTACGCGATAACCTTTATAAGTATGCTGTAGCTATTGCTACAGAGGTAAATTATAACAACATAGGTACCGTAGAGTTTCTTGTGGATGCAGATGATA
ATATTTACTTCATCGAGGTTAACCCACGTATACAAGTAGAGCACACAGTAACCGAAATGGTTACAGGTATTGACCTTGTAAAAACACAAATCTTTGTGGCTGGCGGTTATAAACTATCAGATAAGCAGATAAAGATTTATGGCCAAGAGTCTATAGGTACTTACGGCTTT
GCCATACAGTGTAGACTTACCACAGAAGACCCTACTAATAAATTTACACCAGATTACGGAAAGATAACAACCTACCGTAGCGCCTCTGGTATGGGTATACGTCTAGATGCCGGTAGCATTTATCAAGGGTATAAAGTGAGTCCGTTTTTTGATTCGATGCTTGTAAAAGT
ATCTGCTCACGGTAGAACCCTGGATGGTGCAATACGTAAGATGGTGCGCGCCTTAAAAGAATTTAGAATACGTGGTGTAGAGACCAACATTCATTTTCTTCAAAATGTCATACAGACAGACGTTTTTAAAAATGGAGAGACCACAGTAAATTTCATTCAGAATACTCCTT
CCCTATTTAATATTAAACTCTCTCAAGATCGTACCTCTAAGGTTGTTCAGTTTCTGGCAGATGTGAGTGTAAATGGCAATAGTGATGTAAAAACAAAAGATCCAAAGAAGGTATTTACAAACCCTAAGGTGCCACTTTACAGCCTGTCTGATCCGTTTCCAAAAGGGACAA
AAGATATGCTTACAGAGCTAGGTCCGGAGGCATTTTGTAAGTGGTTATTAGAGGAGAAAAAAATACACTTTACAGATACCACAATGCGTGACGCACACCAGTCTTTACTGGCAACGCGTATGCGCTCTTATGATATGCTCAAAGTAGCATCCAGTTTTGCAAAAAATCAC
CCTAACACCTTTAGTATGGAAGTATGGGGAGGTGCAACCTTTGATGTATGTATGCGATTCTTACACGAGAGCCCTTGGACACGTTTACGAGAGTTACGCAAGGCAGTACCTAACATCTTGTTCCAGATGCTCCTTCGTGGTTCTAATGGCGTGGGTTACAAAGCATATCC
AGATAACCTTATAGAAAAGTTTGTAGAGAAGTCTTGGGAGAATGGTGTAGACCTCTTCCGTATTTTTGACTCTCTTAACTGGGTAAAAGCAATGGAACCTAGTATACAATACGTACGTAATCTTACAGGAGGAATAGCCGAAGGTGCCATAAGTTATACGGGAGATATT
CTAGATACTAAGCAAACTAAGTACAACCTCAAGTACTACACACAACTTGCAAAAGATCTAGAAAATGCTGGGGCACATATGATTGCCATTAAGGATATGGCAGGATTATTAAAACCATATGCAGCGACAGAACTTGTTACAGCACTTAAGGACACTGTAAATGTACCTA
TACACTTACACACACACGACACGTCTTCTTTACAAACAGCTACTTACCTCAAGGCCATTGAGGCAAATGTAGATGTGGTAGATGTAGCTCTAGGAGGACTTTCTGGATTAACTTCTCAACCTAACTTTAATGCTGTAGTAGAGATGCTTAAGTATCAAGATAGAGAACACC
CATTTGAGATGGAAAAACTCAATCAGTTTTCTAATTATTGGGAAGATACGCGTGATCTCTACTACCCTTTTGAGTCTGGACTTAAAGCCGGAACGGCAGAAGTATTTCAGCACGAGATACCAGGTGGGCAGTACTCAAACTTGCGCCCTCAGGCAACTGCTCTAGGTCTT
GCAGATAGATTTACAGAGGTAAAAAAGATGTATGCCGCTGTAAATAAAATGTTTGGCAACCTAGTAAAGGTTACGCCTAGCTCTAAGGTGGTAGGTGATATGGCCATTTTTATGGTGACCAATAACCTCACACCAGAAGATGTGATGGAGCGTGGTGATACCATTTCCT
TCCCAGAGTCTGTTATAAACTTCTTTAAAGGAGATCTAGGTCAGCCAGTAGGTGGTTTCCCAAAAAAGCTACAAGCCATTATCTTAAAAAATAAAGAAGCTTATACAGATAGACCTAATGCACACCTAGCACCAGTAGATTTTGATAAGGAGTATAAAGTCTTTACAGAG
ACCTTCCAAAAAGGGTTTACCAGGCCACTAGAGATAGAAGATTTTCTATCATATATGCTCTACCCAAAGGTTTTTGAGCAAGCACACGAAAACTATAAGAAATACGGTAACCTAGCGCTCGTACCTACTAAGAATTTCTTCTTTGGTATGAAACTAGGTGAAGAAACACTT
ATTGAGCTAGAACCTGGAAAAACGGTGATCATCAAGTTACTCTCTATTGGTATTCCTAATGATGAGGGAATGCGCACGGTATTCTTTAAAGTAAATGGAGAAAATCGCTTTGTAGAAGTACACGACACTTCCCTTAATATCAAAAAGGTTGAAAACGTAAAAATTGATCC
AGATAATGAAAACCACATAGGTGCTCCATTACAAGGCTCTCTTTACAAAGTGCTGGTCAAAAAAGGGCAAGAAGTCAAAGAAAATGACCCACTTTTTATAATAGAAGCTATGAAAATGGAAACTACCGTTGTTGCCTTTAAAGATGGTACAGTACAATCTATAATCGTG
ACAGATGGAACAATGGTGTCTCAAGATGATCTCGTTGTAACTCTTGAATAACCTGCAGGCTTAAGCCCCGCTTCGGCGGGGTTTTTTGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCG
TTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACT
TTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGC
GTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTAC
GGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAAC
TGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTT
GATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTG

**Table S8.** DNA sequence of pMCL10

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTAGAAATATTTTATCTGATTAATAAG
ATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACT
TGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACT
GAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGG
AGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTA
TCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTT
TTCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTACGGAATTGAGGCATTGCGCAGGTTTGGAGTGATAAAAGGCAGTTGGTTGACGGTGAAACGCGT
ATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCCGCCCGGACCATTTGATACCAGAGAACACTAACGATGGATTCGCAACGCAATCTTTTAGTCATCGCTTTGCTGTTCGTGTCTTTCATGATCTGGCAAGCCTGGGAGC
AGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCGCCGCCGACCAGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACA
CCCGTGGTGGTGATGTTGAGCAAGCTCTGCTGCCTGCTTACCCGAAAGAGCTGAACTCTACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGCACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCG
GCTAACGGCCCGCGTCCGCTGTATAACGTTGAAAAAGACGCTTATGTGCTGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAAAACGTTTGTCCTGAAACGTGGTGATTACG
CTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGAAACCGCTGGAAATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACTCTGCCACCGCATCTCGATACCGGAAGCAGCAACTTCGCACTGCACACCTTCCGTGGCGCGGCG
TACTCCACGCCTGACGAGAAGTATGAGAAATACAAGTTCGATACCATTGCCGATAACGAAAACCTGAACATCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCA
ACAACTTCTATACCGCTAATCTGGGTAACGGCATCGCCGCTATCGGCTATAAATCTCAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGCGATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAAAATGGCAGCTGTTGCT
CCGCACCTGGATCTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTGTGGGTAACTGGGGCTTCTCCATTATCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCT
GACCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATGCGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCTGTACAAAGCTGAGAAGGTTAACCCGCTGG
GCGGCTGCTTCCCGCTGCTGATCCAGATGCCAATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACATCCTGCCGATCCTGA
TGGGCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGTTCCCGTCAGGTCTGGTGCTGTACTATATCGTCAGCA
ACCTGGTAACCATTATTCAGCAGCAGCTGATTTACCGTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAAATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGGTCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCG
GATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCATGAGCGAGGTACCCATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTGGGCCCCCCCTCGAGTAATCAACTTCCAAACGTAC
TTAAGGGGGTACTACATGAGCCAAATTCACAAACACACCATTCCTGCCAACATCGCAGACCGTTGCCTGATAAACCCTCAGCAGTACGAGGCGATGTATCAACAATCTATTAACGTACCTGATACCTTCTGGGGCGAACAGGGAAAAAT
TCTTGACTGGATCAAACCTTACCAGAAGGTGAAAAACACCTCCTTTGCCCCCGGTAATGTGTCCATTAAATGGTACGAGGACGGCACGCTGAATCTGGCGGCAAACTGCCTTGACCGCCATCTGCAAGAAAACGGCGATCGTACCGCCA
TCATCTGGGAAGGCGACGACGCCAGCCAGAGCAAACATATCAGCTATAAAGAGCTGCACCGCGACGTCTGCCGCTTCGCCAATACCCTGCTCGAGCTGGGCATTAAAAAAGGTGATGTGGTGGCGATTTATATGCCGATGGTGCCGGA
AGCCGCGGTTGCGATGCTGGCCTGCGCCCGCATTGGCGCGGTGCATTCGGTGATTTTCGGCGGCTTCTCGCCGGAAGCCGTTGCCGGGCGCATTATTGATTCCAACTCACGACTGGTGATCACTTCCGACGAAGGTGTGCGTGCCGGG
CGCAGTATTCCGCTGAAGAAAAACGTTGATGACGCGCTGAAAAACCCGAACGTCACCAGCGTAGAGCATGTGGTGGTACTGAAGCGTACTGGCGGGAAAATTGACTGGCAGGAAGGGCGCGACCTGTGGTGGCACGACCTGGTTGA
GCAAGCGAGCGATCAGCACCAGGCGGAAGAGATGAACGCCGAAGATCCGCTGTTTATTCTCTACACCTCCGGTTCTACCGGTAAGCCAAAAGGTGTGCTGCATACTACCGGCGGTTATCTGGTGTACGCGGCGCTGACCTTTAAATATG
TCTTTGATTATCATCCGGGTGATATCTACTGGTGCACCGCCGATGTGGGCTGGGTGACCGGACACAGTTACTTGCTGTACGGCCCGCTGGCCTGCGGTGCGACCACGCTGATGTTTGAAGGCGTACCCAACTGGCCGACGCCTGCCCGT
ATGGCGCAGGTGGTGGACAAGCATCAGGTCAATATTCTCTATACCGCACCCACGGCGATCCGCGCGCTGATGGCGGAAGGCGATAAAGCGATCGAAGGCACCGACCGTTCGTCGCTGCGCATTCTCGGTTCCGTGGGCGAGCCAATTA
ACCCGGAAGCGTGGGAGTGGTACTGGAAAAAAATCGGCAACGAGAAATGTCCGGTGGTCGATACCTGGTGGCAGACCGAAACCGGCGGTTTCATGATCACCCCGCTGCCTGGCGCTACCGAGCTGAAAGCCGGTTCGGCAACACGTC
CGTTCTTCGGCGTGCAACCGGCGCTGGTCGATAACGAAGGTAACCCGCTGGAGGGGGCCACCGAAGGTAGCCTGGTAATCACCGACTCCTGGCCGGGTCAGGCGCGTACGCTGTTTGGCGATCACGAACGTTTTGAACAGACCTACTT
CTCCACCTTCAAAAATATGTATTTCAGCGGCGACGGCGCGCGTCGCGATGAAGATGGCTATTACTGGATAACCGGGCGTGTGGACGACGTGCTGAACGTCTCCGGTCACCGTCTGGGGACGGCAGAGATTGAGTCGGCGCTGGTGGC
GCATCCGAAGATTGCCGAAGCCGCCGTAGTAGGTATTCCGCACAATATTAAAGGTCAGGCGATCTACGCCTACGTCACGCTTAATCACGGGGAGGAACCGTCACCAGAACTGTACGCAGAAGTCCGCAACTGGGTGCGTAAAGAGATT
GGCCCGCTGGCGACGCCAGACGTGCTGCACTGGACCGACTCCCTGCCTAAAACCCGCTCCGGCAAAATTATGCGCCGTATTCTGCGCAAAATTGCGGCGGGCGATACCAGCAACCTGGGCGATACCTCGACGCTTGCCGATCCTGGCG
TAGTCGAGAAGCTGCTTGAAGAGAAGCAGGCTATCGCGATGCCATCGTAAATGCATGTCGACCTTAAGCCCCGCTTCGGCGGGGTTTTTTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGC
CGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA
GCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCTGATGTCCGGC
GGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATA
AATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCAC
TACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCT
ATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGG
TGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTA
TTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTG
CTGATGCCGCTG

**Table S9.** DNA sequence of pMCL11

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGT
AATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAG
TGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAG
CGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCC
CTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACA
TGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTACGGAATTGAGGCATTGCGCAGGTTTGGAGTGATAAAAGGCAGTTGGTTGACGGTGAAACGCGTATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCC
GCCCGGACCATTTGATACCAGAGAACACTAACGATGGATTCGCAACGCAATCTTTTAGTCATCGCTTTGCTGTTCGTGTCTTTCATGATCTGGCAAGCCTGGGAGCAGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCGCCGCCGACC
AGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACACCCGTGGTGGTGATGTTGAGCAAGCTCTGCTGCCTGCTTACCCGAAAGAGCTGAACTCTACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGC
ACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCGGCTAACGGCCCGCGTCCGCTGTATAACGTTGAAAAAGACGCTTATGTGCTGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAAAACGTTTGTCCTGAAACGTG
GTGATTACGCTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGAAACCGCTGGAAATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACTCTGCCACCGCATCTCGACATCGCAGCCTTCGCACTGCACACCTTCCGTGGCGCGGCGTACTCCACGCCTGACGAGAAGTA
TGAGAAATACAAGTTCGATACCATTGCCGATAACGAAAACCTGAACATCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCAACAACTTCTATACCGCTAATCTGGGTAACGGCATCGCCGCTATCGGCTATAAATCT
CAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGCGATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAAAATGGCAGCTGTTGCTCCGCACCTGGATCTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTG
TGGGTAACTGGGGCTTCTCCATTATCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCTGACCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATGCGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCT
GTACAAAGCTGAGAAGGTTAACCCGCTGGGCGGCTGCTTCCCGCTGCTGATCCAGATGCCAATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACATCCTGCCGATCCTGATGG
GCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGTTCCCGTCAGGTCTGGTGCTGTACTATATCGTCAGCAACCTGGTAACCATTATTCAGCAGCAGCTGATTTACC
GTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAAATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGGTCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCGGATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCATGAGCGAGGTACCCATCGGGTAGT
TTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTGGGCCCCCCCTCGAGTAATCAACTTCCAAACGTACTTAAGGGGGTACTACATGAGCCAAATTCACAAACACACCATTCCTGCCAACATCGCAGACCGTTGCCTGATAAACCCTCAGCAGTACGAGGCGATGTATC
AACAATCTATTAACGTACCTGATACCTTCTGGGGCGAACAGGGAAAAATTCTTGACTGGATCAAACCTTACCAGAAGGTGAAAAACACCTCCTTTGCCCCCGGTAATGTGTCCATTAAATGGTACGAGGACGGCACGCTGAATCTGGCGGCAAACTGCCTTGACCGCCATCTGCAAGAAAA
CGGCGATCGTACCGCCATCATCTGGGAAGGCGACGACGCCAGCCAGAGCAAACATATCAGCTATAAAGAGCTGCACCGCGACGTCTGCCGCTTCGCCAATACCCTGCTCGAGCTGGGCATTAAAAAAGGTGATGTGGTGGCGATTTATATGCCGATGGTGCCGGAAGCCGCGGTTGCGA
TGCTGGCCTGCCGCCCGCATTGGCGCGGTGCATTCGGTGATTTTCGGCGGCTTCTCGCCGGAAGCCGTTGCCGGGCGCATTATTGATTCCAACTCACGACTGGTGATCACTTCCGACGAAGGTGTGCGTGCCGGGCGCAGTATTCCGCTGAAGAAAAACGTTGATGACGCGCTGAAAAACC
CGAACGTCACCAGCGTAGAGCATGTGGTGGTACTGAAGCGTACTGGCGGGAAAATTGACTGGCAGGAAGGGCGCGACCTGTGGTGGCACGACCTGGTTGAGCAAGCGAGCGATCAGCACCAGGCGGAAGAGATGAACGCCGAAGATCCGCTGTTTATTCTCTACACCTCCGGTTCTAC
CGGTAAGCCAAAAGGTGTGCTGCATACTACCGGCGGTTATCTGGTGTACGCGGCGCTGACCTTTAAATATGTCTTTGATTATCATCCGGGTGATATCTACTGGTGCACCGCCGATGTGGGCTGGGTGACCGGACACAGTTACTTGCTGTACGGCCCGCTGGCCTGCGGTGCGACCACGCTG
ATGTTTGAAGGCGTACCCAACTGGCCGACGCCTGCCCGTATGGCGCAGGTGGTGGACAAGCATCAGGTCAATATTCTCTATACCGCACCCACGGCGATCCGCGCGCTGATGGCGGAAGGCGATAAAGCGATCGAAGGCACCGACCGTTCGTCGCTGCGCATTCTCGGTTCCGTGGGCGA
GCCAATTAACCCGGAAGCGTGGGAGTGGTACTGGAAAAAAATCGGCAACGAGAAATGTCCGGTGGTCGATACCTGGTGGCAGACCGAAACCGGCGGTTTCATGATCACCCCGCTGCCTGGCGCTACCGAGCTGAAAGCCGGTTCGGCAACACGTCCGTTCTTCGGCGTGCAACCGGCGC
TGGTCGATAACGAAGGTAACCCGCTGGAGGGGGCCACCGAAGGTAGCCTGGTAATCACCGACTCCTGGCCGGGTCAGGCGCGTACGCTGTTTGGCGATCACGAACGTTTTGAACAGACCTACTTCTCCACCTTCAAAAATATGTATTTCAGCGGCGACGGCGCGCGTCGCGATGAAGATG
GCTATTACTGGATAACCGGGCGTGTGGACGACGTGCTGAACGTCTCCGGTCACCGTCTGGGGACGGCAGAGATTGAGTCGGCGCTGGTGGCGCATCCGAAGATTGCCGAAGCCGCCGTAGTAGGTATTCCGCACAATATTAAAGGTCAGGCGATCTACGCCTACGTCACGCTTAATCACG
GGGAGGAACCGTCACCAGAACTGTACGCAGAAGTCCGCAACTGGGTGCGTAAAGAGATTGGCCCGCTGGCGACGCCAGACGTGCTGCACTGGACCGACTCCCTGCCTAAAACCCGCTCCGGCAAAATTATGCGCCGTATTCTGCGCAAAATTGCGGCGGGCGATACCAGCAACCTGGGC
GATACCTCGACGCTTGCCGATCCTGGCGTAGTCGAGAAGCTGCTTGAAGAGAAGCAGGCTATCGCGATGCCATCGTAAATGCATGTCGACCTTAAGCCCCGCTTCGGCGGGGTTTTTTATCGATAAGCTTCATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAG
TACTGTTACTAGTTCTAGAGCGGCCGCTACAAGCCAATCAAAATCCCCTAAGGAGGTAATACATGAAAATTAAAAAAGTCCTCGTCGCAAATCGCGGGGAAATTGCTATTAGAGTACTACGTGCCTGTACAGAACTTAATATTGCAACCGTAGCTGTCTACACCTATGAAGATAGATACTCA
CAACATCGCAACAAAGCAGATGAGTCTTACCAGATAGGAGCAGATAATGAACCGCTCAAACCCTATTTAGATATGGATCAAATCATTGATCTAGCAAAATCTAAAAATGTAGATGCCATTCACCCCGGCTACGGCTTCTTATCTGAAAACTCAGAATTTGCAAGAAGATGTGCAAAAAATGG
CATTATATTTATTGGCCCTGAGCCAGAAGTAATGGATGCACTAGGTGATAAAATCACCGCAAAAAAAGTAGCACAATCTTGTGGAGTACCTATAATAGAAAGTAATAAAAAAGCCTTAGATTCACTAGAAACGGCCCTAAGTGAAGCTGGAAAAATAGGATATCCAGTAATGCTCAAAGCT
GCCTCTGGTGGAGGTGGTCGCGGTATGCGTATTATACGTAAGGAAGAAGATTTACAGCTTAGTTTTGACTCTGCACGTAATGAGGCGCTCAACTCTTTTGGAGATGATACAATGTTTTTAGAAAAATATGTAGAAGATCCAAAACATATAGAAGTACAAATAGTAGCAGACACACACGGTA
ATATAAGACACCTTTATGAACGTGATTGCTCTGTACAACGCAGGCACCAGAAGGTGGTAGAAATGGCTCCTTCATATAACCTATCACAAACGCTACGCGATAACCTTTATAAGTATGCTGTAGCTATTGCTACAGAGGTAAATTATAACAACATAGGTACCGTAGAGTTTCTTGTGGATGCA
GATGATAATATTTACTTCATCGAGGTTAACCCACGTATACAAGTAGAGCACACAGTAACCGAAATGGTTACAGGTATTGACCTTGTAAAAACACAAATCTTTGTGGCTGGCGGTTATAAACTATCAGATAAGCAGATAAAGATTTATGGCCAAGAGTCTATAGGTACTTACGGCTTTGCCAT
ACAGTGTAGACTTACCACAGAAGACCCTACTAATAAATTTACACCAGATTACGGAAAGATAACAACCTACCGTAGCGCCTCTGGTATGGGTATACGTCTAGATGCCGGTAGCATTTATCAAGGGTATAAAGTGAGTCCGTTTTTTGATTCGATGCTTGTAAAAGTATCTGCTCACGGTAGAA
CCCTGGATGGTGCAATACGTAAGATGGTGCGCGCCTTAAAAGAATTTAGAATACGTGGTGTAGAGACCAACATTCATTTTCTTCAAAATGTCATACAGACAGACGTTTTTAAAAATGGAGAGACCACAGTAAATTTCATTCAGAATACTCCTTCCCTATTTAATATTAAACTCTCTCAAGATC
GTACCTCTAAGGTTGTTCAGTTTCTGGCAGATGTGAGTGTAAATGGCAATAGTGATGTAAAACAAAAGATCCAAAGAAGGTATTTACAAACCCTAAGGTGCCACTTTACAGCCTGTCTGATCCGTTTCCAAAAGGGACAAAAGATATGCTTACAGAGCTAGGTCCGGAGGCATTTTGTAA
GTGGTTATTAGAGGAGAAAAAAATACACTTTACAGATACCACAATGCGTGACGCACACCAGTCTTTACTGGCAACGCGTATGCGCTCTTATGATATGCTCAAAGTAGCATCCAGTTTTGCAAAAAATCACCCTAACACCTTTAGTATGGAAGTATGGGGAGGTGCAACCTTTGATGTATGTA
TGCGATTCTTACACGAGAGCCCTTGGACACGTTTACGAGAGTTACGCAAGGCAGTACCTAACATCTTGTTCCAGATGCTCCTTCGTGGTTCTAATGGCGTGGGTTACAAAGCATATCCAGATAACCTTATAGAAAAGTTTGTAGAGAAGTCTTGGGAGAATGGTGTAGACCTCTTCCGTATT
TTTGACTCTCTTAACTGGGTAAAAGCAATGGAACCTAGTATACAATACGTACGTAATCTTACAGGAGGAATAGCCGAAGGTGCCATAAGTTATACGGGAGATATTCTAGATACTAAGCAAACTAAGTACAACCTCAAGTACTACACACAACTTGCAAAAGATCTAGAAAATGCTGGGGCAC
ATATGATTGCCATTAAGGATATGGCAGGATTATTAAAACCATATGCAGCGACAGAACTTGTTACAGCACTTAAGGACACTGTAAATGTACCTATACACTTACACACACACGACACGTCTTCTTTACAAACAGCTACTTACCTCAAGGCCATTGAGGCAAATGTAGATGTGGTAGATGTAGCT
CTAGGAGGACTTTCTGGATTAACTTCTCAACCTAACTTTAATGCTGTAGTAGAGATGCTTAAGTATCAAGATAGAGAACACCCATTTGAGATGGAAAAACTCAATCAGTTTTCTAATTATTGGGAAGATACGCGTGATCTCTACTACCCTTTTGAGTCTGACTTAAAGCCGGAACGGCAGAA
GTATTTCAGCACGAGATACCAGGTGGGCAGTACTCAAACTTGCGCCCTCAGGCAACTGCTCTAGGTCTTGCAGATAGATTTACAGAGGTAAAAAAGATGTATGCCGCTGTAAATAAAATGTTTGGCAACCTAGTAAAGGTTACGCCTAGCTCTAAGGTGGTAGGTGATATGGCCATTTTTA
TGGTGACCAATAACCTCACACCAGAAGATGTGATGGAGCGTGGTGATACCATTTCCTTCCCAGAGTCTGTTATAAACTTCTTTAAAGGAGATCTAGGTCAGCCAGTAGGTGGTTTCCCAAAAAAGCTACAAGCCATTATCTTAAAAAATAAAGAAGCTTATACAGATAGACCTAATGCACAC
CTAGCACCAGTAGATTTTGATAAGGAGTATAAAGTCTTTACAGAGACCTTCCAAAAAGGGTTTACCAGGCCACTAGAGATAGAAGATTTTCTATCATATATGCTCTACCCAAAGGTTTTTGAGCCAGCAGAAACTATAAGAAATACGGTAACCTAGCGCTCGTACCTACTAAGAATTT
CTTCTTTGGTATGAAACTAGGTGAAGAAACACTTATTGAGCTAGAACCTGGAAAAACGGTGATCATCAAGTTACTCTCTATTGGTATTCCTAATGATGAGGGAATGCGCACGGTATTCTTTAAAGTAAATGGAGAAAATCGCTTTGTAGAAGTACACGACACTTCCCTTAATATCAAAAAGG
TTGAAAACGTAAAAATTGATCCAGATAATGAAAACCACATAGGTGCTCCATTACAAGGCTCTCTTTACAAAGTGCTGGTCAAAAAAGGGCAAGAAGTCAAAGAAAATGACCCACTTTTTATAATAGAAGCTATGAAAATGGAAACTACCGTTGTTGCCTTTAAAGATGGTACAGTACAATCT
ATAATCGTGACAGATGGAACAATGGTGTCTCAAGATGATCTCGTTGTAACTCTTGAATAACCTGCAGGCTTAAGCCCCGCTTCGGCGGGGTTTTTTGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTAC
CCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGC
CAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGA
AGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAA
GCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTT
ATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATAT
TCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGAC
AAGGTGCTGATGCCGCTG

**Table S10.** DNA sequence of pMCL12

GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGT
AATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAG
TGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAG
CGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCC
CTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCT
TGAAGCACTTCACTGACACCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGTTCAAGCTACGGAATTGAGGCATTGCGCAGGTTTGGAGTGATAAAAGGCAGTTGGTTGACGGTGAAACGCGTATTAAAATGCCACCCTTTACACCCTGGTGGTGACGATCCCGTCCC
GCCCGGACCATTTGATACCAGAGAACACTAACGATGGATTCGCAACGCAATCTTTTAGTCATCGCTTTGCTGTTCGTGTCTTTCATGATCTGGCAAGCCTGGGAGCAGGATAAAAACCCGCAACCTCAGGCCCAACAGACCACGCAGACAACGACCACCGCAGCGGGTAGCGCCGCCGACC
AGGGCGTACCGGCCAGTGGCCAGGGGAAACTGATCTCGGTTAAGACCGACGTGCTTGATCTGACCATCAACACCCGTGGTGGTGATGTTGAGCAAGCTCTGCTGCCTGCTTACCCGAAAGAGCTGAACTCTACCCAGCCGTTCCAGCTGTTGGAAACTTCACCGCAGTTTATTTATCAGGC
ACAGAGCGGTCTGACCGGTCGTGATGGCCCGGATAACCCGGCTAACGGCCCGCGTCCGCTGTATAACGTTGAAAAAGACGCTTATGTGCTGGCTGAAGGTCAAAACGAACTGCAGGTGCCGATGACGTATACCGACGCGGCAGGCAACACGTTTACCAAAACGTTTGTCCTGAAACGTG
GTGATTACGCTGTCAACGTCAACTACAACGTGCAGAACGCTGGCGAGAAACCGCTGGAAATCTCCTCGTTTGGTCAGTTGAAGCAATCCATCACTCTGCCACCGCATCTCGATACCGGAAGCAGCAACTTCGCACTGCACACCTTCCGTGGCGCGGCGTACTCCACGCCTGACGAGAAGTA
TGAGAAATACAAGTTCGATACCATTGCCGATAACGAAAACCTGAACATCTCTTCGAAAGGTGGTTGGGTGGCGATGCTGCAACAGTATTTCGCGACGGCGTGGATCCCGCATAACGACGGTACCAACAACTTCTATACCGCTAATCTGGGTAACGGCATCGCCGCTATCGGCTATAAATCT
CAGCCGGTACTGGTTCAGCCTGGTCAGACTGGCGCGATGAACAGCACCCTGTGGGTTGGCCCGGAAATCCAGGACAAAATGGCAGCTGTTGCTCCGCACCTGGATCTGACCGTTGATTACGGTTGGTTGTGGTTCATCTCTCAGCCGCTGTTCAAACTGCTGAAATGGATCCATAGCTTTG
TGGGTAACTGGGGCTTCTCCATTATCATCATCACCTTTATCGTTCGTGGCATCATGTACCCGCTGACCAAAGCGCAGTACACCTCCATGGCGAAGATGCGTATGTTGCAGCCGAAGATTCAGGCAATGCGTGAGCGTCTGGGCGATGACAAACAGCGTATCAGCCAGGAAATGATGGCGCT
GTACAAAGCTGAGAAGGTTAACCCGCTGGGCGGCTGCTTCCCGCTGCTGATCCAGATGCCAATCTTCCTGGCGTTGTACTACATGCTGATGGGTTCCGTTGAACTGCGTCAGGCACCGTTTGCACTGTGGATCCACGACCTGTCGGCACAGGACCCGTACTACATCCTGCCGATCCTGATGG
GCGTAACGATGTTCTTCATTCAGAAGATGTCGCCGACCACAGTGACCGACCCGATGCAGCAGAAGATCATGACCTTTATGCCGGTCATCTTCACCGTGTTCTTCCTGTGGTTCCCGTCAGGTCTGGTGCTGTACTATATCGTCAGCAACCTGGTAACCATTATTCAGCAGCAGCTGATTTACC
GTGGTCTGGAAAAACGTGGCCTGCATAGCCGCGAGAAGAAAAAATCCTGATTCGGTGAGTTTTCGCTAAAATAAGGGCGGTCAGTTGACCGCCTTTTTTCTTTTCGTAGGGCGGATAAGCACCGCGCATCCGCCACACAAAGCAACAGGAACATCATGAGCGAGGTACCCATCGGGTAGT
TTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTGGGCCCCCCCTCGAGATGCCGCGTGGCCTGGAATTATTGATTGCTCAAACCATTTTGCAAGGCTTCGATGCTCAGTATGGTCGATTCCTCGAAGTGACCTCCGGTGCGCAGCAGCGTTTCGAACAGGCCGACTG
GCATGCTGTCCAGCAGGCGATGAAAAACCGTATCCATCTTTACGATCATCACGTTGGTCTGGTCGTGGAGCAACTGCGCTGCATTACTAACGGCCAAAGTACGGACGCGGCATTTTTACTACGTGTTAAAGAGCATTACACCCGGCTGTTGCCGGATTACCCGCGCTTCGAGATTGCGGAG
AGCTTTTTTAACTCCGTGTACTGTCGGTTATTTGACCACCGCTCGCTTACTCCCGAGCGGCTTTTTATCTTTAGCTCTCAGCCAGAGCGCCGCTTTCGTACCATTCCCCGCCCGCTGGCGAAAGACTTTCACCCCGATCACGGCTGGGAATCTCTACTGATGCGCGTTATCAGCGACCTACCGCT
GCGCCTGCGCTGGCAGAATAAAAGCCGTGACATCCATTACATTATTCGCCATCTGACGGAAACGCTGGGGACAGACAACCTCGCGGAAAGTCATTTACAGGTGGCGAACGAACTGTTTTACCGCAATAAAGCCGCCTGGCTGGTAGGCAAACTGATCACACCTTCCGGCACATTGCCATTT
TTGCTGCCGATCCACCAGACGGACGACGGCGAGTTATTTATTGATACCTGCCTGACGACGACCGCCGAAGCGAGCATTGTTTTTGGCTTTGCGCGTTCTTATTTTATGGTTTATGCGCCGCTGCCCGCAGCGCTGGTCGAGTGGCTACGGGAAATTCTGCCAGGTAAAACCACCGCTGAATT
GTATATGGCTATCGGCTGCCAGAAGCACGCCAAAACCGAAAGCTACCGCGAATATCTCGTTTATCTACAGGGCTGTAATGAGCAGTTCATTGAAGCGCCGGGTATTCGTGGAATGGTGATGTTGGTGTTTACGCTGCCGGGCTTTGATCGGGTATTCAAAGTCATCAAAGACAGGTTCGCG
CCGCAGAAAGAGATGTCTGCCGCTCACGTTCGTGCCTGCTATCAACTGGTGAAAGAGCACGATCGCGTGGGCCGAATGGCGGACACCCAGGAGTTTGAAAACTTTGTGCTGGAGAAGCGGCATATTTCCCCGGCATTAATGGAATTACTGCTTCAGGAAGCAGCGGAAAAAATCACCGAT
CTCGGCGAACAAATTGTGATTCGCCATCTTTATATTGAGCGGCGGATGGTGCCGCTCAATATCTGGCTGGAACAAGTGGAAGGTCAGCAGTTGCGCGACGCCATTGAAGAATACGGTAACGCTATTCGCCAGCTTGCCGCTGCTAACATTTTCCCTGGCGACATGCTGTTTAAAAACTTCG
GTGTCACCCGTCACGGGCGTGTGGTTTTTTATGATTACAATGAAATTTGCTACATGACGGAAGTGAATTTCCGCGACATCCCGCCGCCGCGCTATCCGGAAGACGAACTTGCCAGCGAACCGTGGTACAGCGTCTCGCCGGGCGATGTTTTCCCGGAAGAGTTTCGCCACTGGCTATGCGC
CGACCCGCGTATTGGTCCGCTGTTTGAAGAGATGCACGCCGACCTGTTCCGCGCTGATTACTGGCGCGCACTACAAAACCGCATACGTGAAGGGCATGTGGAAGATGTTTATGCGTATCGGCGCAGGCAAAGATTTAGCGTACGGTATGGGGAGATGCTTTTTTGAATGCATGTCGACCTT
AAGCCCCGCTTCGGCGGGGTTTTTTATCGATAAGCTTCATCGGGTAGTTTATTCTTGACAATTAAGTAGAGCCTGATATAATAGTTCAGTACTGTTACTAGTTCTAGAGCGGCCGCTACAAGCCAATCAAAATCCCCTAAGGAGGTAATACATGAAAATTAAAAAAGTCCTCGTCGCAAATCG
CGGGGAAATTGCTATTAGAGTACTACGTGCCTGTACAGAACTTAATATTGCAACCGTAGCTGTCTACACCTATGAAGATAGATACTCACAACATCGCAACAAAGCAGATGAGTCTTACCAGATAGGAGCAGATAATGAACCGCTCAAACCCTATTTAGATATGGATCAAATCATTGATCTAG
CAAAATCTAAAAATGTAGATGCCATTCACCCCGGCTACGGCTTCTTATCTGAAAACTCAGAATTTGCAAGAAGATGTGCAAAAAATGGCATTATATTTATTGGCCCTGAGCCAGAAGTAATGGATGCACTAGGTGATAAAATCACCGCAAAAAAAGTAGCACAATCTTGTGGAGTACCTATA
ATAGAAAGTAATAAAAAAGCCTTAGATTCACTAGAAACGGCCCTAAGTGAAGCTGGAAAAATAGGATATCCAGTAATGCTCAAAGCTGCCTCTGGTGGAGGTGGTCGCGGTATGCGTATTATACGTAAGGAAGAAGATTTACAGCTTAGTTTGACTCTGCACGTAATGAGGCGCTCAAC
TCTTTTGGAGATGATACAATGTTTTTAGAAAAATATGTAGAAGATCCAAAACATATAGAAGTACAAATAGTAGCAGACACACACGGTAATATAAGACACCTTTATGAACGTGATTGCTCTGTACAACGCAGGCACCAGAAGGTGGTAGAAATGGCTCCTTCATATAACCTATCACAAACGCT
ACGCGATAACCTTTATAAGTATGCTGTAGCTATTGCTACAGAGGTAAATTATAACAACATAGGTACCGTAGAGTTTCTTGTGGATGCAGATGATAATATTTACTTCATCGAGGTTAACCCACGTATACAAGTAGAGCACACAGTAACCGAAATGGTTACAGGTATTGACCTTGTAAAAACAC
AAATCTTTGTGGCTGGCGGTTATAAACTATCAGATAAGCAGATAAAGATTTATGGCCAAGAGTCTATAGGTACTTACGGCTTTGCCATACAGTGTAGACTTACCACAGAAGACCCTACTAATAAATTTACACCAGATTACGGAAAGATAACAACCTACCGTAGCGCCTCTGGTATGGGTATA
CGTCTAGATGCCGGTAGCATTTATCAAGGGTATAAAGTGAGTCCGTTTTTTGATTCGATGCTTGTAAAAGTATCTGCTCACGGTAGAACCCTGGATGGTGCAATACGTAAGATGGTGCGCGCCTTAAAAGAATTTAGAATACGTGGTGTAGAGACCAACATTCATTTTCTTCAAAATGTCAT
ACAGACAGACGTTTTTAAAAATGGAGAGACCACAGTAAATTTCATTCAGAATACTCCTTCCCTATTTAATATTAAACTCTCTCAAGATCGTACCTCTAAGGTTGTTCAGTTTCTGGCAGATGTGAGTGTAAATGGCAATAGTGATGTAAAAACAAAAGATCCAAAGAAGGTATTTACAAACCC
TAAGGTGCCACTTTACAGCCTGTCTGATCCGTTTCCAAAAGGGACAAAAGATATGCTTACAGAGCTAGGTCCGGAGGCATTTTGTAAGTGGTTATTAGAGGAGAAAAAAATACACTTTACAGATACCACAATGCGTGACGCACACCAGTCTTTACTGGCAACGCGTATGCGCTCTTATGATA
TGCTCAAAGTAGCATCCAGTTTTGCAAAAAATCACCCTAACACCTTTAGTATGGAAGTATGGGGAGGTGCAACCTTTGATGTATGTATGCGATTCTTACACGAGAGCCCTTGGACACGTTTACGAGAGTTACGCAAGGCAGTACCTAACATCTTGTTCCAGATGCTCCTTCGTGGTTCTAATG
GCGTGGGTTACAAAGCATATCCAGATAACCTTATAGAAAAGTTTGTAGAGAAGTCTTGGGAGAATGGTGTAGACCTCTTCCGTATTTTTGACTCTCTTAACTGGGTAAAAGCAATGGAACCTAGTATACAATACGTACGTAATCTTACAGGAGGAATAGCCGAAGGTGCCATAAGTTATAC
GGGAGATATTCTAGATACTAAGCAAACTAAGTACAACCTCAAGTACTACACACAACTTGCAAAAGATCTGAAAAATGCTGGGGCACATATGATTGCCATTAAGGATATGGCAGGATTATTAAAACCATATGCAGCGACAGAACTTGTTACAGCACTTAAGGACACTGTAAATGTACCTATA
CACTTACACACACACGACACGTCTTCTTTACAAACAGCTACTTACCTCAAGGCCATTGAGGCAAATGTAGATGTGGTAGATGTAGCTCTAGGAGGACTTTCTGGATTAACTTCTCAACCTAACTTTAATGCTGTAGTAGAGATGCTTAAGTATCAAGATAGAGAACACCCATTTGAGATGGAA
AAACTCAATCAGTTTTCTAATTATTGGGAAGATACGCGTGATCTCTACTACCCTTTTGAGTCTGGACTTAAAGCCGGAACGGCAGAAGTATTTCAGCACGAGATACCAGGTGGGCAGTACTCAAACTTGCGCCCTCAGGCAACTGCTCTAGGTCTTGCAGATAGATTTACAGAGGTAAAAAA
GATGTATGCCGCTGTAAATAAAATGTTTGGCAACCTAGTAAAGGTTACGCCTAGCTCTAAGGTGGTAGGTGATATGGCCATTTTTATGGTGACCAATAACCTCACACCAGAAGATGTGATGGAGCGTGGTGATACCATTTCCTTCCCAGAGTCTGTTATAAACTTCTTTAAAGGAGATCTAG
GTCAGCCAGTAGGTGGTTTCCCAAAAAAGCTACAAGCCATTATCTTAAAAAATAAAGAAGCTTATACAGATAGACCTAATGCACACCTAGCACCAGTAGATTTTGATAAGGAGTATAAAGTCTTTACAGAGACCTTCCAAAAAGGGTTTACCAGGCCACTAGAGATAGAAGATTTTCTATCA
TATATGCTCTACCCAAAGGTTTTTGAGCAAGCACACGAAAACTATAAGAAATACGGTAACCTAGCGCTCGTACCTACTAAGAATTTCTTCTTTGGTATGAAACTAGGTGAAGAAACACTTATTGAGCTAGAACCTGGAAAAACGGTGATCATCAAGTTACTCTCTATTGGTATTCCTAATGAT
GAGGGAATGCGCACGGTATTCTTTAAAGTAAATGGAGAAAATCGCTTTGTAGAAGTACACGACACTTCCCTTAATATCAAAAAGGTTGAAAACGTAAAAATTGATCCAGATAATGAAAACCACATAGGTGCTCCATTACAAAGTGCTGGTCAAAAAAGGGCAAGAAG
TCAAAGAAAATGACCCACTTTTTATAATAGAAGCTATGAAAATGGAAACTACCGTTGTTGCCTTTAAAGATGGTACAGTACAATCTATAATCGTGACAGATGGAACAATGGTGTCTCAAGATGATCTCGTTGTAACTCTTGAATAACCTGCAGGCTTAAGCCCCGCTTCGGCGGGGTTTTTTG
AGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTG
AATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGC
ACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGT
AAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAAT
GTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTT
ACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTT
CACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTG