Published in final edited form as:

Adv Exp Med Biol. 2018 ; 1080: 281–315. doi:10.1007/978-981-13-0854-3_12.

Regulatory tools for controlling gene expression in cyanobacteria

Gina Gordon^{1,2}Brian Pfleger^{1,2,*}

¹Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin, 53706, United States

²Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, Wisconsin, 53706, United States

Abstract

Cyanobacteria are attractive hosts for converting carbon dioxide and sunlight into desirable chemical products. To engineer these organisms and manipulate their metabolic pathways, the biotechnology community has had to develop genetic tools to control gene expression. Many native cyanobacterial promoters and related sequence elements have been used to regulate genes of interest and heterologous tools that use non-native small molecules to induce gene expression have been demonstrated. Overall, IPTG-based induction systems seem to be very leaky and have low fold induction ratios in cyanobacteria. A variety of other induction systems have been optimized to enable tighter control of gene expression. There are substantial differences in performance of many of these promoters and induction systems in cyanobacteria compared to *E. coli*, and we hypothesize that these differences are due to fundamental differences in physiology between organisms. This review is not intended to summarize all known products made in cyanobacteria and nor the performance (titer, rate, yield) of individual strains, but instead will focus on the genetic tools used to control gene expression the inherent aspects of cellular physiology that influence and control gene expression in cyanobacteria.

Keywords

Promoters; induction systems; termination; genome editing

Introduction

To meet growing societal needs and reduce the use of fossil fuels as chemical feedstocks, researchers are engineering photoautotrophs to create chemical products from low-cost, renewable inputs: carbon dioxide, water, and sunlight. Harnessing the power of photoautotrophs has substantial benefits compared to using heterotrophs because carbon dioxide is cheap, overabundant form of carbon that would not compete with edible sugars derived from crops¹. When selecting a photoautotrophic host, synthetic biologists are faced with two tradeoffs – production cost certainty and ease of genetic engineering. Modern

*To whom correspondence should be addressed. Tel: +1 608 890 1940; Fax: +1 608 262 5434; pflieger@engr.wisc.edu.

agricultural practices provide solid estimates of plant cultivation and harvesting costs, while the uncertainty of large-scale cyanobacteria cultivation remains a barrier to commercialization. Conversely, cyanobacteria can be genetically modified in the lab with mutation cycles measured in weeks rather than months for most crops. For this reason, metabolic engineering and synthetic biology in plants has progressed more slowly than in microbes and other heterotrophic hosts. That said, the growing number of fully sequenced plant genomes and the emerging ability to make precise insertions, deletions, and targeted mutations with Cas9 and similar nucleases will likely revolutionize the field of metabolic engineering in plants. Similar breakthroughs in cyanobacteria cultivation techniques will be needed to leverage the other physiological and processing advantages of cyanobacteria. These advantages include faster growth rates, greater salt tolerance, ability to grow on marginal lands, and ability to use waste streams as nutrients.

There is great diversity in cyanobacterial morphology, physiology, and environmental preferences. There are marine species, *Prochlorococcus* and *Synechococcus*, that are abundant in the ocean and could be responsible for as much as 25% of marine carbon fixation². There are freshwater cyanobacteria, as well as cyanobacteria found in glaciers, soil, deserts, and hot springs³. Different cyanobacteria species are cultivated as single cells or as multicellular aggregates; in salt or fresh water; in the presence or absence of reduced nitrogen (i.e. nitrogen fixing or non-nitrogen fixing). Given this wide diversity, researchers have selected a few model cyanobacteria to study and engineer in detail. *Synechocystis* sp. strain PCC 6803 is a unicellular, freshwater cyanobacteria that has been studied for decades as a model of photosynthesis⁴. Another freshwater cyanobacterium, *Synechococcus* sp. strain PCC 7942 has been extensively studied especially as a model of the circadian clock. There have been recent efforts to study and engineer *Synechococcus* sp. strain PCC 7002, a halotolerant, unicellular cyanobacteria, because of its faster doubling time compared to PCC 6803 and PCC 7942⁵. *Anabaena* (*Nostoc*) sp. strain PCC 7120 is a cyanobacteria capable of fixing nitrogen through the formation of heterocysts⁶, a trait that would reduce (if not eliminate) the need for providing reduced nitrogen to cultures. In the remainder of this chapter, we will largely focus on gene regulatory tools used in these model organisms (Table 1) because these species are where regulatory tools have been developed and deployed. Most regulatory tools were built to interrogate natural phenomena found in cyanobacteria or to engineer cells to produce valuable chemicals. The future will tell us whether existing tools will transfer to other novel cyanobacteria or if corresponding tools will need to be created to leverage advantageous traits in non-model species.

There are many successful examples of making native and foreign metabolites in cyanobacteria, and comprehensive lists of performance metrics (titer, rate, yield) can be found elsewhere^{7–10}. Unfortunately, titers and productivities lag significantly behind model heterotrophs engineered to produce the same compounds. It remains to be seen whether this discrepancy is due to differences in metabolism between photoautotrophs and heterotrophs, or arises from gaps in knowledge and engineering tools. While yeast, *E. coli*, and other heterotrophs have been the subject of metabolic engineering for several decades, a significantly smaller community of researchers has only recently pursued work with cyanobacteria. This difference can be seen in the standardized media and cultivation equipment used with heterotrophs and the diverse set-ups used to grow cyanobacteria.

Furthermore, culturing cyanobacteria requires optimization of many additional environmental variables including light source and intensity, timing of light/dark cycling, CO₂ concentration, and vessel geometries. Despite these obstacles, there have been a number of significant steps taken to increase production of target chemicals including enhancing flux through desired pathways, eliminating competing pathways, and developing tolerance to high product concentrations⁷. All three of these strategies are heavily reliant on the ability to manipulate gene expression.

The well-developed toolsets for gene expression combined with the fast generation times of bacteria has enabled significant process in engineering them to create non-native chemicals and protein products. There has been significant work to build and expand toolsets to control gene expression in cyanobacteria, because often tools developed in other organisms (*E. coli*) often do not behave similarly. We will cover many examples where directly using genetic tools from other organisms into cyanobacteria has yielded unpredictable and undesirable results. There are a great many factors and steps that facilitate gene expression (Figure 1), some of which may be unique and cyanobacteria-specific. Many of the successful gene expression tools in cyanobacteria have been created by systematically optimizing parameters (e.g. DNA and protein sequences) from underperforming embodiments. In this chapter, we will discuss several unexplored areas of cyanobacterial physiology (the transcriptional landscape, a highly over-represented repeat sequence, presence of multiple copies of the chromosome, as well as prevalence of small proteins), that could be playing significant roles in controlling gene expression and confounding transfer of existing regulatory tools. With future research in these areas we believe improved gene regulatory tools will be built and greatly accelerate the use of cyanobacteria as a chemical production platform.

Promoters

Endogenous Promoters

Many native promoters have been used to express heterologous genes in cyanobacteria and comprehensive lists can be found elsewhere⁸. There has been extensive study of many of the photosystem promoters^{11–13}, but these promoters have limited utility for many metabolic engineering applications because their activity is light-responsive. Although many native promoters have been tested on reporters or genes of interest there has been little systematic study of native promoters to determine relative strengths or sequence elements that could be useful in producing predictable expression of a heterologous gene.

A few native promoters from PCC 7002 have been used to compare expression levels from a common reporter. Potential promoters with a range of expression levels as measured by RNA-seq were chosen and expression of heterologous yellow fluorescence protein (YFP) was tested. There was no correlation between expression level of native sequence and experimental measured YFP levels, and most native promoters produced no more YFP than the background wild type fluorescence¹⁴. This suggests that native promoters cannot be simply fused to heterologous genes without additional design.

Similarly, native promoters from PCC 7942 that produced high levels of protein (measured by proteomics) were used to attempt to increase expression of limonene synthase for the

production of limonene. Varying levels of limonene were seen with different promoters, but only one native promoter produced as much limonene as the heterologous P_{trc} promoter with a synthetic RBS and none outperformed this strain. A heterologous pea *psbA* promoter produced about 13-fold more protein and produced the most limonene¹⁵. Many native cyanobacterial promoters have been used, but sometimes they contain unknown regulatory elements that have unintended effects on gene expression.

Genetic Instability

Using endogenous promoters and regulators can cause genetic instability. It's been suggested that genetic instability, where strains suddenly drop in productivity or performance, is much more widespread than is reported and that researchers may not be incentivized to investigate failed results¹⁶. One published example in PCC 7942 was an ethylene producing strain containing the ethylene forming enzyme (*efe*) that produced ethylene but looked yellow-green. Healthy, green/blue colonies soon appeared that no longer made ethylene, and these colonies were shown to have insertions at sites remarkably similar to the HIP1 sequence (see below) that caused a truncated protein¹⁷. This genomic instability of the ethylene-producing strain was not seen in PCC 6803 although this could be due to the use of a codon-optimized version of the gene¹⁸.

Another example of genetic instability was seen in a strain of PCC 7002 engineered to produce mannitol. Cultures would sometimes suddenly stop producing mannitol and start to grow at a faster growth rate closer to WT. One mutant was isolated that had an inactivating point mutation in one of the two heterologous enzymes¹⁹. Spontaneous inactivating mutations were also seen in a PCC 6803 strain that contained a gene to make lactic acid and also expressed a transhydrogenase²⁰. Mutations causing single amino acid changes that lowered enzyme activity was seen in a PCC 7942 strain engineered to produce isopropanol²¹.

The causes of these mutations and genetic instability is unknown and could be related to DNA repair and mutations caused by expressing a heterologous pathway. Another factor causing genetic instability in cyanobacteria is the maintenance of many copies of the chromosome (see Polyploidy section). There are countless examples of the difficulty and inability to obtain fully segregated strains, and variable productivities and phenotypes are seen when strains are not fully segregated. These observations motivate the development and use of more tightly-controlled gene expression regulation and highlight the advantages of using heterologous sequence elements that don't contain unknown native cyanobacterial control mechanisms.

Heterologous Promoters

Heterologous promoters have been tested to control gene expression to avoid issues with native regulation and potential strain instability due to repeated sequences. Widely used promoters in *E. coli* such as λP_L (BBa_R0040), λP_R (BBa_R0051), P_{Lac} (BBa_R0010) have been shown to have little to no detectable expression in PCC 6803²². Very little fluorescence could be detected from reporters expressed from P_{lac} , P_R , P_{trc10} , and P_{tet} in PCC 6803²². Changing elements such as the -10 element to the consensus TATAAT has

improved expression from heterologous promoters that originally worked well in *E. coli* but not in cyanobacteria²³. However, it's also been suggested that a fully consensus promoter is undesirable and may lead to low activity because of the large number of contacts between RNA polymerase and the promoter²⁴. Heterologous promoters may also perform differently in different cyanobacteria strains. Heterologous promoters P_{trc} , P_{LlacO1} , P_{conII} , P_{J23101} , and P_{J23119} were all tested in PCC 7942 and P_{conII} and P_{J23119} produced the highest expression levels²⁵.

Heterologous artificial promoters have been tested in cyanobacteria and show variable results. Members of the Anderson promoter library (designed in *E. coli*, part of the BioBrick collection from the iGEM Registry) has been tested in both PCC 7002²⁶ and in PCC 6803²⁷. In PCC 7002 there was relatively poor correlation when constructs were tested in both cyanobacteria and *E. coli* ($R^2 = 0.434^{26}$). Promoter libraries are a useful tool to get a large range of expression levels, but it is necessary to test them in a new host because they do not seem to perform similarly across organisms.

Differences in cyanobacterial transcription from *E. coli*

Heterologous promoters may not function as expected due to the fundamental differences in transcription. Cyanobacteria have a strong consensus -10 element but seem to lack consensus or motifs in the -35 region²⁸. It has been suggested that cyanobacteria could be reliant on sigma factors in the place of a consensus -35 element. A study of native promoters in MED4 also observed the conserved -10 element but lacked a consensus -35 element. The authors suggested that the -35 regions may be divided into conserved subclasses²⁹.

Cyanobacterial promoters have been divided into three groups based on the DNA sequence elements they contain and which sigma factors recognize them. Type I promoters (σ^{70}) contain both a -10 element (consensus 5'-TATAAT-3') and a -35 element (consensus 5'-TTGACA-3'). Type II promoters contain the -10 element but lack a -35 element, instead containing target sites for other transcriptional activators. An example of a type II promoter is *glnB* P2 in PCC 6803 that contains the -10 element and the GTAN₈TAC binding motif recognized by NtcA which mediates enhanced transcription under nitrogen-limiting conditions. Type III promoters do not have consensus -10 or -35 elements and most likely are involved in responses to various stresses and stimuli through binding of type III sigma factors³⁰.

Sigma factors are important proteins that facilitate binding of the core RNA polymerase with promoters to initiate transcription, and they are divided into three groups and recognize the corresponding type promoters. σ^{70} (Group 1) is the primary sigma factor in *E. coli* responsible for initiating transcription at housekeeping genes and cyanobacteria have a conserved homolog, *sigA*^{31,32}. Group 2 sigma factors are nonessential, and respond to environmental stimuli. Examples are the heat-shock responsive SigD, dark-responsive and nitrogen starvation-responsive SigB, and less well-studied SigC and SigE. Group 3 sigma factors have alternative structures and respond to stress and include SigF, SigG, SigH, and SigI. SigF has been implicated in control of pili genes but little else is known about others³³. The lack of knowledge of these important transcription factors as well as the coordinated

networks between them make predicting transcriptional levels based off of promoter sequence extremely difficult.

Differences in core transcription machinery may also be responsible for the observed differences in expression of promoters between *E. coli* and cyanobacteria. The cyanobacterial RNA polymerase (RNAP) is composed of different subunits than that of *E. coli*; instead of on β' subunit (RpoC) like in *E. coli* cyanobacteria have both γ (RpoC1) and β' (RpoC2)³³. Additionally, RpoC2 has large insertions between domains G and H. The reasons for this difference in architecture has not been explained or described in terms of gene expression control.

Other differences between cyanobacterial and *E. coli* RNAPs have been observed by testing purified RNAP complexes *in vitro*. RNAP from PCC 7942 and BP-1 both showed lower levels of abortive transcription and less miscincorporation than *E. coli* RNAP³⁴. Additionally, transcription of the same transcript was significantly slower in with cyanobacterial enzymes (taking 164 seconds to complete the transcript compared to 26 seconds). These cyanobacterial RNAPs also paused more frequently and cleaved its transcript faster when NTPs were absent. Mn^{2+} is typically thought to be toxic because RNAP tends to misincorporate noncognate bases, but cyanobacteria maintain approximately 2 orders of magnitude higher intracellular levels of Mn^{2+} than *E. coli* because of its key role in photosynthesis³⁵. It was proposed that cyanobacterial RNAP maintains high fidelity even in the presence of high Mn^{2+} levels with expense of slower elongation³⁴. If similar trends are also seen *in vivo*, these fundamental differences in transcription could have profound effects of gene expression and native mechanisms used to control gene expression.

Differences in RNAP-interacting factors may also be responsible for observed differences in transcription between cyanobacteria and other model organisms. The *E. coli* protein DksA has been shown to bind RNAP and help stimulate transcription from λP_R ³⁶. No homologs for DksA have been found in cyanobacteria. Additionally, ppGpp affects gene expression at many promoters including λP_R ³⁶. ppGpp has been shown to be a regulator in cyanobacteria just as it is in model heterotrophs like *E. coli*, but alternate signals can affect its synthesis. In *E. coli* ppGpp levels rise as part of the stringent response during stresses. Contrastingly, ppGpp has been shown to be involved to adaption to darkness in PCC 7942³⁷ and heterocyst formation in PCC 7120³⁸.

Many promoters, including P_{lac} , require cAMP receptor protein (Crp) for transcription. These promoters contain a Crp binding site and positioning of this binding site relative to the -10 element is important to ensure that Crp and RNAP are on the same side of the DNA and can interact to cause activation. The *E. coli* P_{lac} promoter has a Crp-binding site -61.5 bases from the transcription start site³⁹, which is 10.5 α -helical turns from the -10 element. A cAMP receptor protein has been identified in PCC 6803 (SYCRP1) and was shown to bind the consensus *E. coli* Crp sequence in the presence of cAMP⁴⁰. Several SYCRP1-dependent promoters have been identified as well as their putative Crp-binding sites, but not all have proper spacing to ensure Crp and RNAP localize to the same side of the DNA⁴¹. Study of these important transcription factors in cyanobacteria may reveal global differences in

specificity and function and help explain differential gene expression observed between organisms.

Global analyses have revealed intriguing aspects of the transcription landscape in cyanobacteria. Most of the genome seems to be transcribed with 88% having reads from RNA-seq align to either strand and 82% of the non-coding regions being transcribed in PCC 7942²⁸. Transcriptome reads mapping to almost the entire genome ($\geq 94\%$) was seen for PCC 6803, PCC 7002, and CCMP1986⁴². This shows remarkable similarity to plant chloroplasts where over 99% of the plant chloroplast genomes from rice, maize, *Arabidopsis*, and green algae *Chlamydomonas reinhardtii*⁴². The model for transcription in chloroplasts involves haphazard initiation and termination resulting in many variable precursor transcripts that are then processed to their final 5' and 3' ends by endo and exoribonucleases. To our knowledge there has been no in-depth study to see if a similar process could be occurring in cyanobacteria.

Other differences in transcription in cyanobacteria compared to *E. coli* include a distinct pause site approximately 63 nucleotides downstream of the transcription start site as observed by RNA pol ChIP data in PCC 7942²⁸. This is slightly different than has been reported in *E. coli* where the majority of RNAP is associated with the promoter⁴³. Additionally, a preference for an adenine was seen at the transcriptional start site in PCC 7942²⁸. These differences in both core RNAP machinery as well as factors that influence transcription initiation may help explain differences in promoter strength between *E. coli* and cyanobacteria.

Known Transcriptional Regulators

Global regulators play major roles in controlling gene expression, and there has been a small subset of global transcriptional regulators studied in cyanobacteria. One global transcriptional regulator, NtcA, has been shown to regulate nitrogen metabolism. NtcA belongs to the CAP (catabolite gene activator or cyclic AMP receptor protein) family and has been identified in nearly all cyanobacteria strains mentioned in this chapter⁴⁴. DNase footprinting revealed that NtcA bound to GTAN₈TAC which takes the place of the -35 element in NtcA-responsive promoters⁴⁵. Activation of these promoters is facilitated through an increase in the binding constant of RNAP and an increase in the rate constant for isomerization from the close to open complex⁴⁴. The consensus sequence for PCC 7120 NtcA was reported to be slightly different (TGTN₉ or 10ACA) and proposed to be both a positive and negative transcriptional regulator, like many other members of the CAP family. This PCC 7120 consensus sequence is also located upstream of the large subunit of RuBisCO, *rbcL*, suggesting it may be playing other regulatory roles.

Another transcriptional regulator involved in sensing nitrogen compounds, NtcB, is often located near NtcA and nitrate assimilation genes (*nir* operon)⁴⁶. In PCC 7120 NtcB has been shown to bind upstream of the *nir* operon and is a booster to enhance expression of the *nir* operon in combination with NtcA⁴⁷. There have been conflicting reports of conditions where NtcB activation is seen. In PCC 7942 NtcB activation of the *nir* operon is only seen in the presence of nitrite while in PCC 6803 and PCC 7120, activation is seen independently of the presence of nitrite⁴⁸.

Other transcriptional regulators have been shown to be involved in regulating genes involved in the CO₂ concentrating mechanism. CcmR and CmpR are both LysR family transcriptional regulators that sense and respond to CO₂ concentrations. Transcriptional activator, CmpR, has been identified in PCC 7942 and PCC 6803 and has been shown to bind to the upstream region of the *cmp* operon that encodes a high-affinity bicarbonate transporter⁴⁹. Ribulose-1,5-bisphosphate and 2-phosphoglycolate were able to enhance binding of CmpR to its targeting, suggesting that these metabolites are how this regulator senses CO₂ levels^{50,51}. The transcriptional repressor, CcmR, represses a suite of targets including *sbtA*, *bicA*, *porB*, as well as itself⁵². Its binding motif as well as novel regulon members have been identified through network analysis of PCC 7002 grown under many different conditions⁵³. α -ketoglutarate and NADP⁺ have been identified as co-repressors of CcmR⁵¹.

Other transcriptional regulators sense both nitrogen and carbon and mediate transcriptional responses. PCC 6803 has two cyanobacterial AbrB-like proteins (cyAbrBs); *cyabrB1*, which appears to have an overlapping or complementary regulon to NtcA⁵⁴, and *cyabrB2*, which regulates parts of the high affinity bicarbonate transport system and overlaps with CcmR and CmpR regulons⁵⁵.

Many transcriptional repressors respond to iron and the redox status of the cell. The SufR transcriptional repressor has been identified in PCC 6803, PCC 7002, PCC 7120, *N. punctiforme*, BP-1 as well as others. SufR represses transcription of the *suf* operon that encodes proteins involved in the assembly of iron-sulfur clusters. This repression is removed under conditions of oxidative stress or iron limitation⁵⁶. FurA is a global cyanobacterial regulator that binds a 19 to 23 bp consensus sequence in the presence of iron. It can act as either an activator or repress in PCC 7120⁵⁷ and similar results have been found for PCC 7942⁵⁸. PedR is a LuxR type transcriptional regulator that senses the redox state of the photosynthetic electron transport chain and facilitates gene expression change⁵⁹.

This is just a small subset of the major transcriptional regulators in cyanobacteria. There remain many uncharacterized transcriptional regulators, but elucidating both their sensing mechanism, targets, and functions is complicated by their many complementary and overlapping targets. Recently, 32 putative transcriptional regulators were studied in PCC 6803 by screening knockouts grown under a variety of conditions and using liquid chromatography-mass spectrometry based metabolomics⁶⁰. There are clearly many unknown factors that could be influencing expression of a gene of interest.

Induction/Repression Systems

Whether studying native functions or implementing synthetic biology strategies it is desirable to be able to turn on and off gene expression with desired timing and strength. Induction and repression systems are well-established in model heterotrophs and can respond to an environmental shift or presence/absence of a small molecule. Unfortunately, directly moving these systems into cyanobacteria has not been universally successful (discussed below). Therefore, significant efforts have been made to adapt and optimize them for use in cyanobacteria.

It is extremely difficult to compare induction systems both within and across cyanobacteria strains and species. Different reporters have been used, integrated at different chromosomal locations, with different selectable markers. Comparisons between different strains is difficult due to the different culturing conditions with varying media compositions, light intensities, temperatures, culture vessels and CO₂ concentrations. Most of the engineered genetic tools have been designed and optimized with fluorescent reporters as the output due to their ease of quantification. Most papers report fold-induction ranges between uninduced and induced strains, which gives little indication of overall ranges of expression. Additionally, these fold inductions take different forms: fluorescence of reporter proteins, protein quantification with western blots, transcript levels via quantitative PCR, measurements of product titers, or changes in cellular physiology.

A compilation of the fold induction data from all of the induction systems outlined in the following two sections can be seen in Figure 2. We've divided them into both transcription-based and translation-based induction systems. There is a wide range of induction ratios seen with various inducers and in various strains. Unfortunately, this comparison does not show overall expression levels due to the indirect measurements of induction in various publications. Even systems with low fold induction ratios can be extremely useful and may have tighter repression than systems with a larger dynamic range of induction.

Induction via transcription

There are many native cyanobacterial transcripts that have been shown to respond to environmental stimuli: high-light responsive *psbA* promoter, copper-regulated *petE* promoter, nitrate/nitrite-inducible *nirA* promoter, and the nickel-responsive *nrsA* promoter⁶¹. Some of these promoters have been used to induce expression of a gene of interest. A native PCC 6803 *cpcG2* promoter that is up-regulated in green-orange light and down-regulated in red light was used to create a green-light inducible lytic system in PCC 6803⁶². This light-responsive promoter was placed in front of a T4-derived lysis gene and was shown to induce cell lysis upon exposure to green light resulting in a reduced growth rate. Additionally, reporters driven from Ni²⁺-inducible P_{nrsB} and Co²⁺-inducible P_{coaT} showed 50 and 70-fold induction respectively, but have very limited ranges of inducer concentrations⁶³. A Zn²⁺ responsive system was built in PCC 7002 that utilizes the PCC 7942 SmtB repressor and SmtB-regulated promoter P_{smtA} and achieved 19-fold induction⁶⁴.

Systems that respond to these large environmental shifts or natural inducers have a limited utility for induction of proteins of interest because native targets would be affected as well. Often, it is ideal to have no other cellular or environmental changes besides the induction of your protein of interest, so cellular effects can be attributed to the protein of interest and not global shifts in regulation and metabolism. Efforts have been made to use non-native induction systems and optimized to achieve activity in cyanobacteria.

Directly moving these induction systems from other organisms has been unsuccessful and there has been significant difficulty with IPTG-inducible promoters. The widely used IPTG-inducible, P_{trc} did have high expression in PCC 6803, but only led to 1.6-fold induction when LacI was present and IPTG was added²². Leaky expression from the P_{trc} promoter was also seen without IPTG when a tagged protein was purified in PCC 6803⁶⁵. In PCC 7942 an

IPTG induction system using P_{trc} was sufficiently leaky to complement a mutant strain even in the absence of the inducer⁶⁶. Leaky expression of heterologous enzymes was sufficient for isobutanol production in PCC 6803 without inducer⁶⁷. However, there are other examples in PCC 7942 where 360-fold induction or higher was seen with P_{trc} ⁶⁸.

Many versions of IPTG inducible systems were very leaky in the absence of IPTG, resulting in low fold induction ratios. The P_{lac} (IPTG-inducible combination of *E. coli*'s P_{trp} and P_{lacUV5}) was inducible in PCC 7120, but induction was only 5-fold⁶⁹. Induction of heterologous genes in the 2,3-butanediol pathway in PCC 7942 using P_{LlacO1} was only 1.2 to 1.6-fold and showed very significant leaky expression in the absence of IPTG⁷⁰. Use of IPTG-inducible *trp-lac* promoter to control a sucrose permease gene in PCC 7942 showed 6-fold induction as measured by quantitative reverse transcriptase PCR⁷¹. A variant of P_{lac} , $P_{AllacO-1}$ that contains a second lac operator sequence, showed stronger repression than P_{trc} , but became leakier at higher cell densities¹⁸. However, there are examples where similar fold inductions (~2.5-fold) are seen in both *E. coli* and cyanobacteria⁷².

A systematic study varying the spacing and number of the lac operator sequences in PCC 6803 showed that there was weak repression when only a single lac operator sequence was used (P_{trc1O} and similar variants). However, when the second operator was added (P_{trc2O} and similar variants) there was very significant repression, as high as 400 times, when LacI was not present, but these strains could not be induced. They observed the same reliance on the 11 bp periodic placement of the operator sequences as seen in *E. coli*, ensuring LacI and DNA looping can occur on the same side of the DNA. However, very significant differences in repression and induction were observed, that remain unexplained⁷³.

There seems to be a trade-off between using heterologous IPTG-inducible systems from *E. coli* and native cyanobacterial induction systems. Different induction systems were used to drive expression of enzymes in ethylene metabolism in PCC 6803 and it was shown that P_{trc} produced high expression and ethylene levels, but was not tightly regulated like the same construct expressed in *E. coli*. Contrastingly native metal-inducible promoters (P_{petE} , P_{coa} , and P_{smt}) were tightly repressed, but had low expression even when induced¹⁸. Similarly, a native cyanobacterial promoter controlled by a metal-sensing repressor, P_{smt} , had the highest induction ratio in PCC 7942 when compared with P_{LtetO1} , $P_{ConII-ribo}$, P_{trc} , P_{LlacO1} , and P_{BAD} ²⁵.

Directly moving induction systems from *E. coli* has not always led to favorable expression levels or tight control of expression levels, so several induction systems have been optimized and modified for use in cyanobacteria. Two IPTG-inducible systems have been optimized for PCC 7002²⁶. Initial systems were optimized by changing the -10 elements, the spacing between the -10 element and the LacO operator sequence, as well as the spacing between the -35 and -10 element. Both constructs were titratable between 1 μ M – 10,000 μ M IPTG but had different ranges of expression. cLac143 had higher expression and was somewhat leaky in the absence of IPTG (~5% of the constitutive version), but had a 48-fold range of expression. cLac94 had a 38-fold range expression, but does not reach the same expression level when fully induced. These two constructs used the cyanobacterial phycocyanin truncated promoter (P_{cpt}) flanked by LacO operator sequences to drive expression of the

gene of interest. LacI expression was controlled by P_{MB2} , a synthetic promoter derived from BBa_J23119. Tighter repression was seen by using *lacI*W220F that has been shown to be less leaky possibly due to tighter DNA binding⁷⁴. Induction was also seen when P_{cpt} was replaced with P_{trc} , but a smaller dynamic range was achieved. These optimized IPTG-inducible systems have some of the largest fold induction ranges of those tested so far.

Anhydrotetracycline (aTc) induction systems have been developed for PCC 7002⁷⁵, PCC 7120⁷⁶, and PCC 6803²³. The aTc induction system in PCC 7002 had a 32-fold induction and was titratable with concentrations of aTc between 10 ng/mL and 1000 ng/mL aTc. Although the 32-fold induction is not as high as some other inducible systems, the tightness of this system has advantages compared to other leakier but higher expression systems. The aTc induction system was critical to achieve inducible repression using CRISPR interference in PCC 7002⁷⁷.

The aTc induction system in PCC 7120 achieve >1,200-fold induction and had a dynamic range with between 0 ng/mL and 2,000 ng/mL⁷⁶. There was a decrease in growth of PCC 7120 with 2000 ng/mL aTc, but the growth rate appeared to be unaffected with 200 ng/mL aTc. This system was adapted to respond to nitrate by placing the *nirA* nitrate-specific promoter in front of *tetR* and over 200-fold induction was seen in the absence of the inducer when cells were grown in nitrogen-free conditions in comparison to media supplemented with nitrate.

The aTc induction system for PCC 6803 was extensively optimized and achieved 230-fold induction under light-activated heterotrophic growth²³. Expression was titratable with aTc concentrations between 0 ng/mL - 4 ng/mL aTc. This induction system was created by systematically altering the -10 element sequence and the bases between the -10 element and the transcription start site. A guanine in the 2nd position downstream of the -10 element was shown to be important for high promoter strength, that might be caused by altered lifetime of the promoter-RNAP closed complex.

These systems had some of the largest fold induction ratios (Figure 2), but their use for long term induction is limited due to aTc's sensitivity to light. The dynamic range of induction in PCC 7002 dropped significantly after both 24 and 48 hours exposure to light⁷⁵. Induction persisted for at least 72 hours in PCC 7120, possibly because the cyanobacterial cells were absorbing light and protecting from aTc degradation⁷⁶. aTc has also been shown to be sensitive to temperature, with a half-life of only 15 hours at typical cyanobacterial cultivation temperature (37°C) compared to 30 hours 30°C⁷⁸. The lability of the inducer in the culture conditions is often forgotten and could be responsible for some of the different dynamics seen above.

Induction via translation

Induction can also occur at the level of translation, where the translation of a transcript can respond through a molecule that alters the occlusion or accessibility of the ribosome binding site. Theophylline-responsive riboswitches have been tested and used in a number of diverse cyanobacteria. A set of six riboswitches originally designed for and shown to be effective in a diverse group of Gram-negative and Gram-positive bacteria⁷⁹ have been tested in

cyanobacteria. In PCC 7942, riboswitch E* (a slight change in the RBS from E) showed 190-fold induction of a luciferase reporter while maintaining minimally leaky expression⁸⁰. Expression had a wide range of induction with concentrations between 0 and 2 mM theophylline. This riboswitch was successfully used to regulate circadian clock *kaiC* expression.

The same set of theophylline riboswitches was tested in a group of 4 diverse cyanobacteria strains (PCC 7942, BL0902, PCC 7120, and WHSyn). As was seen in the original diverse group of bacteria, various fold inductions, leakiness, and gene expression ranges were seen between the six riboswitches. Fold induction ratios in cyanobacteria seemed significantly lower than those reported for the diverse set of other bacteria with cyanobacteria fold induction ranges reaching maximally 30-fold⁶¹ whereas several of the previously tested bacterial strains reached 100 or 2000-fold⁷⁹.

It is thought that cyanobacterial small RNA (sRNA) play a large role in controlling gene expression. sRNAs are very prevalent with 371 reported in PCC 6803⁸¹, but only a few of these sRNA seem to be widely conserved (NsiR1, PsrR1, Yfr1, HliR1, PsiR1, and PsiR2)⁸¹. The lack of development of sRNA tools for cyanobacteria is surprising, with only two reports of synthetic sRNA's used to mediate gene expression. A trans-acting sRNA system based on the *E. coli* IS10 RNA-IN/OUT system was adapted for use in PCC 7002 and achieved 70% attenuation⁷⁵, slightly less than the 90% attenuation in *E. coli*⁸². sRNA control has also been implemented in cyanobacteria through the use of riboregulators. Riboregulators consist of a *cis*-repressed mRNA (crRNA) that forms a stem loop that occludes the RBS. A *trans*-activating RNA (taRNA) is complementary to part of the stem loop and hybridizes with the crRNA and enhances ribosome binding. By driving the crRNA from the *trc* promoter and the taRNA from the Ni²⁺ inducible promoter *PnrsB*, they saw 13-fold induction of the GFP reporter upon the addition of Ni²⁺ in PCC 6803⁸³. Induction systems based on the level of translation have proved to be a useful tool in cyanobacteria and provide an avenue for gene expression control independent of the complex and unstudied aspects of cyanobacterial transcription machinery.

Ribosome bindings sites

The ribosome binding site (RBS) is a major factor controlling the amount of translation of a given gene. The complementarity of the RBS to the 16S rRNA in the ribosome is a major determinate of protein expression as well as the accessibility of this site given the secondary structure of the surrounding RNA sequence. Most cyanobacteria mentioned here (PCC 7002, PCC 6803, PCC 7942, PCC 7120, and PCC 6301) have the portion of the 16S rRNA that pairs with messenger RNA with a sequence of ACCTCCTTT which would suggest optimal pairing with mRNA with a central region of the RBS being AGGAGG (the consensus prokaryotic sequence).

Thermodynamic models have been developed to predict translation efficiency and have been quite successful in both *E. coli* and other organisms⁸⁴. The most widely used software is the RBS Calculator⁸⁵, but other translational rate calculators include the RBS Designer⁸⁶ and UTR Designer⁸⁷. These calculators use statistical thermodynamic models that calculate free energies for interactions between the mRNA and rRNA and utilizes RNA folding software to

predict mRNA secondary structure. These calculators are capable of both designing sequences for a specified translation rate (forward engineering) and predicting the location and strength of a given RBS (reverse engineering). Each of these calculators tested predicted translation initiation rates and observed fluorescence or β -galactosidase activity of reporters in *E. coli* and saw good correlations using the forward engineering mode ($R^2 = 0.84$ for the RBS calculator⁸⁵, $R^2 = 0.87$ for UTR Designer⁸⁸, and $R^2 = 0.87$ for the RBS Designer⁸⁹). These calculators have been very useful and enabling, but there are limitations including the effects of antisense RNA, RNA binding proteins, translational coupling, or differences in translational elongation⁹⁰.

There are many examples showing how manipulation of the RBS has enabled desirable changes in cyanobacteria protein expression. Using an optimized RBS in PCC 7120, L0908, PCC 6803, and PCC 7942 increased expressed of both CFP and YFP by several fold⁹¹. Although there are limited examples, RBS predictions do not seem to be as reliable in cyanobacteria as they are in *E. coli*. Fluorescence from a 11-member library of RBS' in front of a YFP reporter was measured and compared with the predicted translation initiation rate using the RBS calculator, UTR Designer, and RBS Designer²⁶. Once an outlier that had a large increase in transcript level compared to all other members was removed, R^2 values were respectively 0.74, 0.78, and 0.04. Altogether, RBS design software is not perfect, but is an extremely useful tool to both study native regulation and achieve desirable expression levels of heterologous proteins.

CRISPR interference

CRISPR interference (CRISPRi) has proved to be a powerful method to repress native and heterologous gene repression in bacteria. CRISPRi utilizes the native RNA based defense system that bacteria use to target invading nucleic acids⁹². The system most widely used is derived from the *Streptococcus pyogenes* type II system that utilizes a single nuclease protein (Cas9) and a small guide RNA (sgRNA) complementary to the target sequence. CRISPRi utilizes a deactivated Cas9 (dCas9) that has point mutations in its active sites so it can no longer make double-stranded breaks in DNA. Genes targeted with CRISPR interference are thought to be repressed because of tight binding of the dCas9/sgRNA complex to the DNA that causes a steric clash, dislodging RNAP⁹³. This heterologous system has been implemented in both PCC 7002, PCC 6803, and PCC 7942 and has been used to re-direct flux toward desired products. Several of these cyanobacteria have a native CRISPR system or multiple systems, but there has been little exploration in their function or mention of cross-talk with CRISPRi.

In PCC 6803 a number of constitutive and inducible promoters were tested to drive expression of dCas9 and sgRNA and their efficiencies were measured by the output of reporter green fluorescent protein (GFP)⁹⁴. They observed significant repression of GFP whenever both dCas9 and the sgRNA was present, and the best combination was using the aTc-inducible P_{L22} . They observed 96% of the maximal GFP fluorescence without aTc, and only 6% when aTc was added, but there was still significant repression with intermediate levels of aTc. Other promoters caused maximal repression (>99%) even without the presence of the inducer, highlighting that low levels of expression are necessary to achieve

inducible repression. They targeted native *phaE* for repression and saw no PHB accumulation under nitrogen-starvation conditions which normally accumulates under these conditions and functions as a carbon and redundant storage polymer. They also targeted *glgC*, an enzyme involved in glycogen synthesis and showed that under nitrogen starvation the *glgC* repression strain only accumulated 25% of the glycogen compared to WT. They showed efficient repression of four targets at the same time. Repression of one target in PCC 6803 (*slr0091*) was only 2-fold whereas other targets were about 10-fold. This weak repression could have been because this gene is very weakly transcribed than the others, or because it is part of an operon.

An inducible and titratable system for CRISPRi was made in PCC 7002⁷⁷. Repression was seen immediately whenever dCas9 and the sgRNA was present, and the promoters and ribosome binding sites were changed to attempt to relieve this repression in the absence of the inducer. The system with the best titratable range had both dCas9 and the sgRNA under the control of the aTc-inducible promoter, P_{EZtet} ⁷⁵. While this system only showed ~70% of the maximal protein expression without the inducer, it was titratable with intermediate levels of aTc and only ~10% of the maximal protein expression with maximum aTc concentrations. Significant repression of targets was seen even when the promoter of dCas9 was removed and when the sgRNA was truncated down to only 15 bases. Native phycobilisome (*cpcBACDEF*) and carboxysome genes (*ccmK₁K₂LMN*) were also targeted and showed characteristic decreases in absorbance at 635nm and requirement for high CO₂ concentrations for growth respectively. This repression scheme was used to increase flux towards lactate by targeting glutamine synthetase (*glnA*) and over 2-fold more lactate was made in this strain than the baseline engineered strain. Interestingly, increased lactate production was also seen in a strain that contained the sgRNA targeting *glnA* but lacked dCas9.

A CRISPRi system was built in PCC 7942 to target native genes and increase succinate production. Even when dCas9 was under control of inducible P_{smt} leaky expression was sufficient to repress targets in the absence of the inducer. Repression of *glgC*, *sdhA*, and *sdhB* (separately) using CRISPRi increased succinate production by ~12.5-fold²⁵.

These CRISPRi systems have enabled both increased titers of desired products as well as provided platforms to manipulate and study native gene regulation. Combined, these three systems show that CRISPRi is extremely effective at repressing gene expression, stable (effective for maximum time tested in PCC 7942: 21 days²⁵), and reversible^{77,94}. These studies also show that sgRNA's can efficiently target both the template and non-template strand, but most effective repression is seen when the target is near the transcription start site or near the beginning of the coding sequence. One remaining challenge is the fact that many bacterial and cyanobacteria genes are located within operons, limiting the application of CRISPRi. Nonetheless, CRISPRi is a very promising tool, enabling regulation of gene expression in *trans*.

RNA stability and degradation

Of the many steps in the central dogma of biology, RNA turnover is the least considered when designing gene expression modules. In cyanobacteria, RNA turnover is rapid and may

play a substantial role in determining gene expression levels. Global half-lives of transcripts have been determined in *Prochlorococcus* MED4 using microarrays, and the median half-life was 2.4 minutes⁹⁵. The observation that RNA is very labile and that RNA is turned over rapidly in bacteria has been long observed, but the half life reported for MED4 is on the shorter end compared with other recent global RNA studies (*B. subtilis* ~ 5 minutes⁹⁶, *Bacillus cereus* 2.4 minutes⁹⁷, *E. coli* 6.8 minutes⁹⁸; 4.7 minutes⁹⁹; 2.5 minutes¹⁰⁰, *Chlamydia trachomatis trachoma* 15–17 minutes¹⁰¹, *Mycobacterium tuberculosis* 9.5 minutes¹⁰²). The reason for this especially rapid turnover remains unexplained although it has been suggested that RNA turnover could be linked with cell doubling time. This seems unlikely when the doubling times and average/median global transcript turnover is compared⁹⁵.

This rapid turnover of RNA in cyanobacteria could be a result of its somewhat unique composition of ribonucleases (RNases). Cyanobacteria have homologs of both RNase E (endoribonuclease that often initiates turnover in *E. coli*) and RNase J (5'→3' exoribonuclease responsible for initiation of mRNA degradation in *B. subtilis*). Both RNase E and RNase J appear to be essential in PCC 7002 because attempts to delete either were unsuccessful¹⁰³. This combination of RNase E and J exists in several other bacteria including *Sinorhizobium meliloti*, *Mycobacterium smegmatis*, *Rhodobacter sphaeroides*, and *Streptomyces coelicolor*. Studies in these organisms have shown that both RNase E and RNase J play important roles in rRNA maturation^{104,105} as well as in the production of antibiotics¹⁰⁶ and SAM metabolism¹⁰⁷. A global model of RNA degradation in organisms containing both RNase E and J has yet to be proposed and universal tools for modulating RNA turnover are not available.

Another difference between the cyanobacteria RNA degradation machinery and other more well-characterized systems is the apparent lack of a 5' pyrophosphohydrolase, RppH¹⁰³. RppH is part of the Nudix (nucleoside diphosphates linked to x) protein family that removes pyrophosphate from the 5' end of transcripts leaving a monophosphate¹⁰⁸. This processing event affects transcript degradation because both RNase E and RNase J can sense the 5' of a transcript and cleave more efficiently in the presence of the 5' monophosphate^{109–111}. While no proteins with 5' pyrophosphohydrolase activity have been identified in cyanobacteria, RppH could still be present and be an important element controlling transcript degradation. All RppH proteins identified so far have been part of the Nudix family, but it is difficult to determine the substrates of these protein family members based on amino acid sequence. RppH homologs have been identified previously by purifying these proteins and testing their activities *in vitro*¹¹². Cyanobacteria possess many proteins annotated as members of the Nudix family and it is yet to be seen if one may have the activity of RppH. Recent reports show that RppH may not be the only enzyme involved in the processing of the 5' end, and that a yet-identified protein may be responsible of initial cleavage from a 5' triphosphate to a 5' diphosphate and serve as the precursor for RppH¹¹³. Given that ribonucleases play a substantial role in regulating gene expression, it is important to identify the full complement of RNA processing enzymes in cyanobacteria.

Transcriptional Termination

Transcription termination has not been widely studied in cyanobacteria. Observations that the cyanobacterial genome is nearly fully transcribed, points to potential differences in transcriptional termination from other model hosts. In bacteria, termination is either rho-dependent, where the protein Rho recognizes a target sequence and dissociates RNAP, or rho-independent where an RNA structure in the transcript destabilizes RNA-RNAP interactions. In rho-dependent termination Rho recognizes a Rho-utilization site (RUT) in recently transcribed RNA, wraps the RNA inside the Rho hexamer, moves along the transcript, and is thought to pull the RNA out of the RNAP active site. The site of termination is not well defined and can be as far as 120 bases from the RUT site¹¹⁴. It is unlikely that cyanobacteria use rho-independent termination because PCC 7942, PCC 6803, and PCC 7002 do not have an obvious homolog of Rho²⁸. Many assume that cyanobacteria rely on rho-independent termination also known as intrinsic termination. Transcripts that are terminated via intrinsic termination have a 3' secondary structure that is composed of a GC-rich palindromic sequence followed by stretch of T's¹¹⁵. The stretch of T's is thought to induced transcriptional pausing, allowing time for the hairpin to form which in turn dislodges key interactions between three nucleic acid binding sites in RNAP (the RNA:DNA T-A hybrid, RNA-binding site of just added ribonucleotides, and upstream RNA-binding site of RNA exiting RNAP)¹¹⁴. Intrinsic termination structures can be found for almost 50% of protein-coding genes in *E. coli*¹¹⁶.

It was originally presumed that cyanobacteria must use rho-independent termination because of the lack of a Rho homolog, but genome wide predictions of hairpin formation near stop codons showed that termination in cyanobacteria may not involve hairpin formation. In *E. coli* there is a significant sharp drop in free energy approximately 30 bases downstream of stop codons indicating the presence of a hairpin. Genome-wide free energy calculations were made for many organisms including PCC 6803, and no sharp dip in free energy was seen¹¹⁷. The lack of dips in free energy values was also seen in several other genomes including *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Helicobacter pylori*, *Borrelia burgdorferi*, *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum*, *Aquifex aeolicus*, *Pyrococcus horkoshii*, *Mycobacterium tuberculosis*, and *Treponema pallidum*¹¹⁷. Dips in the free energy prediction near the 3' end of transcripts have, however, were seen in PCC 7942²⁸.

Transcriptome studies suggest that cyanobacteria transcribe a large portion of non-coding regions as well as sequences within coding sequences, but on the opposite strand⁴². Cyanobacteria have intriguing parallels in terms of their transcriptional landscapes when compared with plants chloroplasts. It is a well-established in chloroplasts that transcription is very inefficient, and it has been proposed that transcription initiation and elongation are very haphazard, creating a wide array of varying length transcripts⁴². Many transcripts in chloroplasts have been shown to be variable in length with a mixture of 3' extensions¹¹⁸. Transcripts are thought to be processed by proteins to have the proper 5' and 3' ends. This phenomenon has been analyzed globally using RNA-sequencing and also been studied on an individual transcript basis. Similar results have been shown in plant

mitochondria where variable 3' extensions are trimmed back to stem loop structures by exonucleases¹¹⁹.

It's been observed that many chloroplast genes have inverted repeats at the 3' end of the transcript (including *psbA*, *rbcL*, *petD*, and *atpE*)¹¹⁸. These 3' inverted repeats have been shown not to cause termination like 3' secondary structure does on many bacterial transcripts, but instead serve as processing sites. Without these 3' inverted repeats these transcripts are rapidly degraded¹¹⁸.

Transcriptional termination can play a role in controlling gene expression of genes that are located in operons. Terminators are not necessarily efficient and can allow various amounts of read-through. In PCC 7002 the phycocyanin operon (*cpcBCADEF*) encodes six genes that encode structural subunits of the phycobilisome, linker proteins and assembly enzymes, and has been shown to make eight to ten different length transcripts¹²⁰ (Figure 3). There are predicted stable stem loop structures at several locations within this transcript that could serve as terminators or processing sites. Support for their role as inefficient terminators includes that only 1–2% of the transcripts include *cpcE* and *cpcF* located at the 3' end of the operon¹²⁰. The structural components of the phycobilisome are encoded in the first two genes of the operon, followed by two linker proteins, and finally enzymatic lyases for attachment and their decrease in expression levels seems to match the stoichiometry needed for phycobilisome function²⁸.

Similar regulation of the phycobilisome operon is seen in PCC 7120, except with slight differences in operon organization and the number of different transcripts made. The PCC 7120 operon contains ten open reading frames (*cpcBACDEFG1G2G3G4*), and several different transcripts are seen^{121,122}. The major transcript includes *cpcBA*, but full-length transcripts as well as partial transcripts have also been observed, including some that suggest an internal *cpcC* promoter. Although the presence of processing events could not be ruled out, it was suggested that these transcripts arose through partial termination¹²¹. Most of the locations where partial transcriptional termination is occurring are also predicted to form strong secondary structures. This partial read-through seems to be controlling expression levels of important parts of this light harvesting complex.

Partial termination may be playing a large role in regulating gene expression from the *cpc* operon, an operon extensively studied because of its important role in light harvesting and photosynthesis. The need to for analysis of transcriptional termination in cyanobacteria has been mentioned before¹²³, and perhaps difference in transcriptional termination could play major roles in controlling gene expression.

Advances in Genome Editing

One significant advantage of working in cyanobacteria is that many of the commonly used strains are naturally competent and readily take up foreign DNA (PCC6803¹²⁴, PCC 7002¹²⁵, PCC 7942¹²⁶, and others). DNA-uptake in cyanobacteria is thought to involve type IV pilus apparatus that may be responsible for DNA translocation as a single strand¹²⁷. Many established protocols make use of this natural competency and involve incubating cyanobacteria cells with a plasmid or linear DNA piece containing regions homologous to

the chromosome prior to plating on media with a selective pressure corresponding to a selection marker in the foreign DNA.

Traditionally, genomic integrations have been made by providing a plasmid or PCR product that contains the desired sequence surrounded by homology arms. Higher efficiency has been seen with longer homology arms, but they can be as short as 250 bp¹⁴. Several studies have examined the length relationship between homology arm length and integration efficiency, and generally 500–1000 bases of homology are used.

Integrating constructs onto the genome may be more stable than maintaining them on a plasmid, but deciding the location of integration can be challenging. Ideally genes could be inserted in neutral sites that have no effect on physiology when grown in any conceivable environmental condition. Historically in PCC 7002 *desB* was used as an integration site¹⁴, but its expression at low temperatures and role in lipid desaturation¹²⁸ make it an unattractive candidate. Many have used *glpK*, encoding glycerol kinase, as a neutral insertion site because of its original annotation as a pseudogene. Re-sequencing of PCC 7002 showed that *glpK* is intact and an error was made in the original sequencing¹²⁹. Genomic integrations have also been introduced between A0935 and A0936 (termed NS1)¹⁴. Given the abundance of transcription from the cyanobacteria genome, it may not be possible to identify a truly neutral site. That said, systematic studies of gene expression around the genome are not available for all cyanobacteria and many other hosts, a knowledge gap that could provide immediate help in designing engineered strains.

Several methods have been developed to obtain markerless genome edits, but many of these systems require multiple transformation and homologous recombination steps. Levansucrase, encoded by *sacB*, confers sensitivity to sucrose, is one of the most popular counterselectable markers and has been adapted for uses in many organisms¹³⁰, including cyanobacteria. *sacB* was used as early as 1990 in PCC 7120 by integrating a selectable plasmid containing *sacB* onto the chromosome with a single crossover event and then selecting for a replacement event with the desired sequence¹³¹. Unfortunately this technique produced a significant amount of spontaneous mutants as well as a mixture of five different insertion sequences at the target locus. Recently, a faster *sacB* system has been developed in PCC 6803 that utilizes a single vector containing kanamycin resistance gene and *sacB* nested within a split gene of interest with partial homology regions. This method requires two homologous recombination events: (1) insertion of the entire cassette and (2) recombination fusing the gene of interest and excising the kanamycin resistance-*sacB* region, but the time needed to produce mutant strains is greatly reduced without the need for a second conjugation or transformation step¹³². Another one of the most popular markerless counterselectable markers, the *rsp12* (*strA*) system that relies on the resistance and sensitivity of different alleles to streptomycin, has been used in PCC 7942¹³³. This system unfortunately requires working in a mutant *rsp12* background. A novel counter selection system utilizing the sensitivity to acrylic acid in the presence of native *acsA* gene has been developed in PCC 7002¹³⁴.

The adaptation of the CRISPR system as a tool for genome editing has been an enabling technology in bacteria and higher organisms. CRISPR Cas9 systems are useful in

cyanobacteria to allow genomic integrations or changes without the use of selectable markers and can be used to alleviate problems with segregation. One of the first examples of a CRISPR Cas9 genome editing system was in UTEX 2973¹³⁵. The system was based on one developed for use in *Streptomyces lividans* that uses a plasmid that is not functional at 34°C or higher. This plasmid was needed because Cas9 was shown to be toxic in UTEX 2973 when cells were transformed with a replicating RSF1010-based plasmid. This toxicity was still seen when the promoter and ribosome binding site of Cas9 was removed. Growing cells at a non-permissive temperature (38°C) while forcing them to maintain the plasmid due to the presence of the antibiotic, appeared to allow enough transient expression for genome editing and genome edits appeared to be segregated in the first patch. The system is all contained on a single plasmid encoding Cas9, the gRNA, and repair template and can be introduced into cyanobacteria with triparental mating with a conjugal and helper plasmid. The editing plasmid could be cured from the edited strain by re-streaking without the antibiotic, but unfortunately this took 10 rounds of patching.

Another CRISPR Cas9 system was used in cyanobacteria was in PCC 7942. Cas9 and repair templates were provided on two separate plasmids and they saw successful integration at higher efficiencies than traditional transformations and found that they could use shorter homology arms (400bp vs 700 bp)¹³⁶. They found that Cas9 cleavage seemed to help with segregation, but strains were still not segregated until 3 rounds of passaging on the antibiotic. Cas9 was expressed from a plasmid that appeared to only be transiently present because it was not detected after 9 days. Notably, Cas9, and tracrRNA, crRNA were all expressed from native *Streptococcus pyogenes* promoters. This CRISPR Cas9 system was effective, but genomic edits still contained an antibiotic resistance marker, limiting its utility to make some precise edits.

Cpf1, a type V CRISPR nuclease, has also been used for genome editing in cyanobacteria. The Cpf1 system differs from type II Cas9 CRISPR system in that Cpf1 processes pre-crRNA¹³⁷ and introduces a staggered cut 17 bases from the PAM, and the PAM (TTN) is located 5' of the crRNA¹³⁸. Cpf1 appeared less toxic than Cas9 because more colonies were recovered even when compared to Cas9 that lacked a promoter. Successful editing events were shown in UTEX 2973, PCC 6803, and PCC 7120 using a RSF1010 based plasmid containing Cpf1, gRNA, and repair template¹³⁹. Only ~20% of the initial colonies were edited and segregated, but subsequent patching on the antibiotic increased the recovery of segregated mutants. They showed markerless point mutations, knock-outs, and insertions. Continued advancements to improve the efficiency and minimize the time needed to make genome edits will help accelerate the development of gene regulatory tools.

Use of plasmids

In other model organisms, plasmids are used to express heterologous protein. Their use is much less widespread in cyanobacteria both because of the lack of characterized plasmids that can replicate in cyanobacteria as well as the difficulties introducing plasmids into the cell. The natural competence of cyanobacteria is advantageous for introducing foreign DNA with selectable markers, but unfortunately is not useful for introducing plasmids. The mechanism of DNA uptake and natural competence in cyanobacteria is relatively unknown,

but likely contains core DNA uptake genes as well as hundreds of regulatory genes as in other model organisms used to study natural competence¹⁴⁰. DNA uptake is thought to involve an extendable (and potentially retractable) type IV pilus that translocates single-stranded DNA that is then decorated with protective proteins and then acted on by recombinases. This process imports only single-stranded DNA, preventing its utility for transformation of plasmids.

Both integrative and replicative plasmids have been used in cyanobacteria, but their use lags significantly behind their use and utility in other organisms. There has been a number of shuttle plasmids built that contain both *E. coli* and cyanobacteria-based origins⁸. *E. coli* origins are more standard but only a limited number of cyanobacterial origins have been used: pDU1 in PCC7120, pDC1 in *N. punctiforme*, pUH24 and pANS in PCC 7942, pAQ1 in PCC 7002, pBA1 in PCC 6301, and pFDA in *Fremyella diplosiphon* UTEX481^{8,91}. In most of these cases native origins have been used and re-introduced on a new vector in the same host. However, pDU1 was originally isolated from *Nostoc* PCC 7524 but shuttle vectors containing pDU1 have shown to replicate in many other strains of cyanobacteria including PCC 7120¹⁴¹, *Anabaena* M-131¹⁴¹, *Fischerella muscicola* PCC 7414¹⁴², *Chlorogloeopsis fritschii* PCC 6912¹⁴², *Chroococcidiopsis* spp. Str. 029, 057, and 123¹⁴³, and *Oscillatoria* MKU 277¹⁴⁴. There is a significant lack of understanding of the maintenance, copy number control, and replication mechanisms of cyanobacterial plasmids and if these aspects are similar or different from plasmids native to other bacterial species.

The most widely used plasmid in cyanobacteria is broad host range plasmid RSF1010⁸. RSF1010 is a broad host range plasmid in the IncQ compatibility group that can replicate in over 30 Gram-negative species¹⁴⁵. This plasmid on its own is not transferable via conjugation, but it contains the necessary elements (*oriT* and *mob* mobilization genes) to be transferred in combination with a 'helper plasmid.' Conjugation protocols are standard in many strains of cyanobacteria and typically consist of biparental or triparental matings^{22,146,147}. Plasmid based expression has significant advantages, but there are not many widely used or characterized plasmids that replicate in many cyanobacterial hosts.

Standardization of genetic tools

One of the reasons gene expression regulatory tools have only been built in tested in a few strains is because of the lack of standardization of cyanobacteria genetic tools. There have been significant efforts to standardize genetic tools in synthetic biology. The BioBrick 'Registry of Standard Biological Parts' contains thousands of promoters, ribosome bindings sites, terminators, coding sequences, and origins that are surrounded by standardized restriction enzyme sites for simple assembly of parts. There have been a few cyanobacterial plasmids built using the BioBrick system but they have not gained much popularity. The plasmid, pPMQAK1, was built by adding the BioBrick multiple cloning site and antibiotic resistant cassettes to the backbone of the broad-host range plasmid RSF1010. The plasmid was shown to replicate in and *E. coli*, PCC 6803, PCC 7120, and *N. punctiforme*²².

A broad-host-range vector system based off of RSF1010 was built with and used to create cyanobacterial shuttle vectors⁹¹. This system utilizes GC-adaptors flanked by restriction enzyme sites, so that elements can be digested out of donor plasmids, purified, and

assembled using Gibson assembly. Four different sites with overlapping regions were used that could be used for an *E. coli* origin, cyanobacterial origin, antibiotic resistance marker, and expression cassette for the gene of interest. These elements can also be replaced with homology regions to direct homologous recombination onto the chromosome. This design scheme and sequence elements have been integrated into a web service, CYANO-VECTOR (<http://golden.ucsd.edu/CyanoVECTOR/>) that can be used to design and export GenBank files.

The lack of standardization of parts for gene expression in cyanobacteria likely stems from the widespread practice of integrating cassettes onto the chromosome of cyanobacteria. With the ease of Gibson assembly¹⁴⁸, most labs design and assemble custom plasmids. Genomic integration is a popular method of expressing genes of interest because transformation of plasmids to date has been achieved through conjugation with helper plasmids which takes approximately the same amount of time and reagents as using natural competence and homologous recombination. An effort to make gene regulatory tools accessible across strains and labs would greatly enhance current tool building and engineering efforts.

Unexplored Physiology

The development of predictable gene regulatory tools in cyanobacteria may be limited by unexplored physiology. There are several cyanobacteria-specific elements that remain relatively unexplained that could be influencing gene expression and could explain why directly moving elements from other organisms has been somewhat unsuccessful. We highlight three aspects of cellular physiology that may be playing significant roles in influencing gene expression.

Highly iterative palindrome 1 (HIP1)

The extreme prevalence of one sequence in cyanobacterial genomes remains unexplained. The palindrome 5'-GCGATCGC-3' or the extended version, 5'-GGCGATCGCC-3', is extremely overrepresented in many cyanobacterial genomes and has been named highly iterative palindrome 1 (HIP1). This sequence was first described in PCC 6301 in 1993 as the two flanking sequences surrounding a deletion event that conferred a tolerance to cadmium¹⁴⁹. Since then the overrepresentation of this sequence has been discovered in many cyanobacterial strains and has been shown to have been gained once during evolution and then lost in marine cyanobacteria¹⁵⁰.

HIP1 has been shown to propagate through nucleotide substitutions. Comparisons of several conserved regions between *E. coli* and PCC 7942 show that *E. coli* contains partial HIP1 sequences (on average 5.2 of the 8 nt) that require minimal changes to the HIP1 sequence that result in synonymous or conservative amino acid changes¹⁵¹. Deletions flanked by HIP1 sequences have been observed many times. Direct tests comparing deletion frequencies in *E. coli* and PCC 7942 show that excision frequencies surrounded by HIP1 sequences are 2-fold higher than those surrounded by control palindromic sequences. Additionally, the frequency of excision was similar in both *E. coli* and PCC 7942, suggesting that this phenomenon is not due to a cyanobacterial-specific recombinase¹⁵¹. It has been suggested that these deletion events could be due to HIP1 sequences forming hairpins or other structures that

result in replication slippage and illegitimate replication. A similar hypothesis has been suggested for repeated sequences found in radiation resistant *Deinococcus radiodurans*, that is thought to survive double-stranded breaks through extensive recombination of its many chromosomes¹⁵².

It's possible that HIP1 could be involved in gene expression control, but potential mechanisms and regulatory partners remain elusive. Comparative genomics of many cyanobacterial genomes show that is not preferentially located in coding or noncoding regions and has no preference for reading frame. Additionally, attempts to see trends between HIP1 prevalence and location within a transcript, GC content, codon usage, and transcript abundance have been unsuccessful¹⁵³. So far, no HIP1-specific proteins or complexes have been identified via electrophoretic mobility-shift assays¹⁵¹. We urge the cyanobacterial community to be cautious of DNA motif finders such as MEME¹⁵⁴ because they will most likely find HIP1 enrichment no matter what native cyanobacterial sequences are entered. The overabundance of the HIP1 sequence remains unexplained and could be influencing gene expression or other processes that have large implications for growth and study of cyanobacteria.

Polyploidy

It has long been reported that many strains of cyanobacteria contain multiple copies of its chromosome, but it's implications on gene expression control is often ignored. PCC 7942 and WH7803 have been shown to be oligoploid containing 3–4 copies of its chromosome per cell irrespective of growth phase¹⁵⁵ or light-dark/continuous light cycles¹⁵⁶. Other strains such as PCC 6803 have been shown to be highly polyploid with as many as 218 copies of its chromosome per cell in exponential phase. In PCC 6803 the number of copies of its chromosome changes and only ~ 60 copies per cell are present in linear and stationary phase. Previous studies report only 12 copies of the chromosome in PCC 6803¹⁵⁷. PCC 6803 has the highest ploidy level of any bacterium besides the very large *Epulopiscum sp.* that contains tens of thousands copies of its chromosome¹⁵⁵. Discrepancies in chromosomal as well as plasmid copy number may come from lysis efficiency issues as well as the method of detection (PCR vs total DNA quantification). There has been also been mention that these chromosome copy number discrepancies could be due to cultivation of these strains in many labs over the past 20 years that could be resulting in mutations.

The ploidy of cyanobacteria could confer resistance to UV radiation; both PCC 6803 and PCC 7942 were more resistant to UV radiation than *E. coli*. Additionally, PCC 6803 (highly polyploid) is much more resistant to X-ray irradiation than PCC 7942 (oligoploid)¹⁵⁸. The ploidy of cyanobacteria leaves many unanswered questions regarding gene expression. It is unknown when genes are expressed from all chromosomes or selectively from different chromosome copies. The polyploid nature of cyanobacteria has unknown implications on gene expression.

The presence of multiple copies of the genome is challenging for the introduction of foreign genes and making targeted mutations. It has only been recently appreciated that it is imperative to show segregation of strains, to ensure predictable and reliable expression patterns and production results. In some early papers using cyanobacteria to create products,

there is no mention of segregation while in others the authors acknowledge that the strains were indeed not segregated, making it almost impossible to re-create these strains and verify these results. PCR, using primers flanking the inserted gene of interest, is often used to demonstrate segregation of strains, but this is not always sufficient to demonstrate segregation. There have been cases where the PCR product made with flanking primers indicates the proper mutated size, but it is still possible to amplify the native WT sequence using an internal primer¹⁰³. A combination of both internal and external primers is necessary to demonstrate segregation. The presence of multiple copies of the genome in cyanobacteria is challenging, and the cyanobacteria community must remain vigilant to ensure thorough and meticulous strain construction.

Role of small proteins

Small proteins often escape annotation by automated pipeline systems despite their important functions, and many of these small proteins have important roles in photosynthesis. These small proteins are sometime known as sprteins, SPs, or μ -proteins with arbitrary cutoffs ranging from < 100, <80, and < 50 amino acids. These small proteins have been implicated in many processes such as being secreted as signals, serve as toxins, membrane components, chaperones of metals or nucleic acids, regulators, or stabilizing factors of large protein complexes¹⁵⁹.

Issues with small protein annotation likely arise because short DNA sequences are likely to contain both start and stop codons. Random DNA sequences have been modelled in yeast, and the frequencies that different length sequences would contain ORFs showed that a large number of short ORFs could be occurring by chance¹⁶⁰. Due to this, most gene annotation algorithms have set a cutoff for ORFs of greater than 100 amino acids to lower then number of false positive annotations¹⁶¹.

Only 80.8% of the entire PCC 6803 predicted proteome has been detected even when the data was combined from over 50 independent proteomic studies¹⁶². Hydrophobicity as well as low or condition-specific expression (especially proteins encoded on plasmids) may cause the inability to detect expression¹⁶², but a major class of proteins that is not-represented is small proteins. Proteomic approaches fail to identify many small proteins for several reasons. Small proteins contain few or no trypsin cleavage sites, necessary for many mass spectrometry analyses. Additionally, many small proteins be lost in the filtering process where two positively charged peptides are necessary to match to the database. Modifications such as phosphorylation, acetylation, glycosylation, and oxidation may prevent detection¹⁶² (all of which have been observed in PCC 6803).

Many short proteins have been shown to play important roles in photosynthesis including 19 proteins < 50 amino acids that interact with photosystem II (*psbM*, *psbT*, *psbI*, *psbL*, *psb30*, *psbJ*, *psbM*, *psbX*, *psbY*, *psbN*, *psbF*, and *psbK*¹⁶³). Other small proteins are involved in photosystem I (*psaM*, *psaJ*, and *psaI*) as well as electron transport (*petL*, *petN*, *petM*, and *petG*). The smallest protein currently annotation is *petN* (29 amino acids)¹⁵⁹. The presence and expression of many of these small genes was completely unknown and only found through purification of photosystem II complexes. Purification of PSII from PCC 6803 revealed a total of 31 distinct proteins, 12 of which were less than 8kDa, and 5 of which

were novel¹⁶³. Six of these small proteins remained unidentified, highlighting the challenges of both predicting and studying these small proteins. The important role these small proteins play is highlighted by the fact that many are conserved in cyanobacteria, algae and plants.

There have been attempts to predict the presence and expression of small proteins in cyanobacteria by examining transcriptomic data and comparing the conservation of possible coding sequences across different species and organisms. 146 small proteins were predicted to be shared in both PCC 6803 and PCC 6714 and a smaller subset of 42 proteins shared between both cyanobacteria strains and *Arabidopsis thaliana*, and 29 proteins shared between the cyanobacteria strains and *E. coli*¹⁶⁴. Expression of a test set of 5 of these genes were confirmed with immunoblotting. Two of these small proteins were originally thought to be non-coding RNAs¹⁶⁴. Challenges to discovering the expression of these small proteins include the time it takes to make markerless mutations such as a C-terminal tag as well as their regulated expression under specific environmental conditions. There have been a number of short proteins that have been found to play important roles in photosynthesis, and we postulate that other unknown and uncharacterized sequences may play large roles in cellular physiology.

Conclusions

There have been significant advances in gene regulatory tools in cyanobacteria, but lack of knowledge of basic cellular processes as well as cyanobacteria-specific physiology may still play a large role in controlling heterologous gene expression. Endogenous cyanobacteria promoters have been used to drive gene expression as well as heterologous promoters. Adapting native sequences for expression of heterologous transcripts can cause problems with genetic instability. Many groups have tried using promoters widely-used in *E. coli*, but the majority of them are leaky and have low fold induction values, especially IPTG-based systems. Induction systems have been used both on the level of transcription and translation with varying results. It is difficult to compare induction data and measures of tightness of repression in different publications, and currently there is no method to compare overall expression levels between publications and strains. The CRISPR system has expanded the collection of cyanobacteria gene regulatory tools both in use as a repression system (CRISPRi) and for genome editing.

Gene expression tools often do not operate as expected in cyanobacteria, and we explore several relatively under-studied areas relating to cyanobacteria physiology that could be causing these inconsistencies. It's been observed that most of the cyanobacterial genome is transcribed, but implications for gene expression and gene expression tools remains unknown. Additionally, there has been extremely little study of cyanobacterial transcription termination. There seem to be intriguing similarities in the transcriptional landscape between cyanobacteria and plant chloroplasts, and continued study in this area might be quite fruitful. We conclude with three topics that are cyanobacteria-specific that may be influencing gene expression: the unknown role of the highly prevalent repeat sequence (HIP1), the maintenance of many copies of chromosome, and the role and prevalence of small proteins.

Future Trends

The ability to do metabolic and synthetic biology in cyanobacteria is heavily reliant on the quality and availability of gene expression regulatory tools. We suspect that the cyanobacterial community will continue to develop novel and more dependable gene expression regulatory systems that will enable rapid acceleration of the use of cyanobacteria for biotechnology purposes.

Due to the transcription landscape in cyanobacteria, it is likely that tools to control gene expression on the level of translation may allow stricter control. Leaky expression of genes of interest is a phenomenon seen across many cyanobacteria species. There is significant need for a tightly regulated induction system that has little to no leaky expression.

There are many intriguing aspects of cyanobacterial physiology that may play substantial roles in controlling gene expression. We hope researchers will draw parallels from studies of plant chloroplasts and help unravel the complex gene regulatory networks in cyanobacteria.

References

- Desai SH & Atsumi S Photosynthetic approaches to chemical biotechnology. *Curr. Opin. Biotechnol* 24, 1031–1036 (2013). [PubMed: 23578466]
- Flombaum P et al. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Pnas* 110, 9824–9829 (2013). [PubMed: 23703908]
- Whitton B & Potts M Introduction to the Cyanobacteria. in *The Ecology of Cyanobacteria* 1–11 (Kluwer Academic Publishers, 2000). doi:10.1007/s13398-014-0173-7.2
- Ikeuchi M & Tabata S *Synechocystis* sp. PCC 6803 - a useful tool in the study of the genetics of cyanobacteria. *Photosynth. Res* 70, 73–83 (2001). [PubMed: 16228363]
- Bernstein HC Unlocking the Constraints of Cyanobacterial Productivity: *MBio* 7, 1–10 (2016).
- Kumar K, Mella-Herrera RA & Golden JW Cyanobacterial heterocysts. *Cold Spring Harb. Perspect. Biol* 2, 1–19 (2010).
- Nozzi NE, Oliver JWK & Atsumi S Cyanobacteria as a Platform for Biofuel Production. *Front. Bioeng. Biotechnol* 1, 1–6 (2013). [PubMed: 25023785]
- Wang B, Wang J & Meldrum DR Application of synthetic biology in cyanobacteria and algae. *Front. Microbiol* 3, 1–15 (2012). [PubMed: 22275914]
- Case AE & Atsumi S Cyanobacterial chemical production. *J. Biotechnol* 231, 106–114 (2016). [PubMed: 27238233]
- Lai MC & Lan EI Advances in metabolic engineering of cyanobacteria for photosynthetic biochemical production. *Metabolites* 5, 636–658 (2015). [PubMed: 26516923]
- Nair U, Thomas C & Golden SS Functional Elements of the Strong *psbAI* Promoter of *Synechococcus elongatus* PCC 7942. *J. Bacteriol* 183, 1740–1747 (2001). [PubMed: 11160106]
- Mulo P, Sakurai I & Aro EM Strategies for *psbA* gene expression in cyanobacteria, green algae and higher plants: From transcription to PSII repair. *Biochim. Biophys. Acta - Bioenerg* 1817, 247–257 (2012).
- Takahashi T, Nakai N, Muramatsu M & Hihara Y Role of multiple *HLR1* sequences in the regulation of the dual promoters of the *psaAB* genes in *Synechocystis* sp. PCC 6803. *J. Bacteriol* 192, 4031–4036 (2010). [PubMed: 20511509]
- Ruffing AM, Jensen TJ & Strickland LM Genetic tools for advancement of *Synechococcus* sp. PCC 7002 as a cyanobacterial chassis. *Microb. Cell Fact* 15, 190 (2016). [PubMed: 27832791]
- Wang X et al. Enhanced limonene production in cyanobacteria reveals photosynthesis limitations. *Proc. Natl. Acad. Sci. U. S. A* 113, 201613340 (2016).
- Jones PR Genetic Instability in Cyanobacteria - An Elephant in the Room? *Front. Bioeng. Biotechnol* 2, 1–5 (2014). [PubMed: 25152877]

17. Takahama K, Matsuoka M, Nagahama K & Ogawa T Construction and analysis of a recombinant cyanobacterium expressing a chromosomally inserted gene for an ethylene-forming enzyme at the *psbAI* locus. *J. Biosci. Bioeng* 95, 302–305 (2003). [PubMed: 16233410]
18. Guerrero F, Carbonell V, Cossu M, Correddu D & Jones PR Ethylene Synthesis and Regulated Expression of Recombinant Protein in *Synechocystis* sp. PCC 6803. *PLoS One* 7, (2012).
19. Jacobsen JH & Frigaard NU Engineering of photosynthetic mannitol biosynthesis from CO₂ in a cyanobacterium. *Metab. Eng* 21, 60–70 (2014). [PubMed: 24269997]
20. Angermayr SA, Paszota M & Hellingwerf KJ Engineering a cyanobacterial cell factory for production of lactic acid. *Appl. Environ. Microbiol* 78, 7098–7106 (2012). [PubMed: 22865063]
21. Kusakabe T et al. Engineering a synthetic pathway in cyanobacteria for isopropanol production directly from carbon dioxide and light. *Metab. Eng* 20, 101–108 (2013). [PubMed: 24076145]
22. Huang HH, Camsund D, Lindblad P & Heidorn T Design and characterization of molecular tools for a synthetic biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res* 38, 2577–2593 (2010). [PubMed: 20236988]
23. Huang H-H & Lindblad P Wide-dynamic-range promoters engineered for cyanobacteria. *J. Biol. Eng* 7, 10 (2013). [PubMed: 23607865]
24. Hook-Barnard IG & Hinton DM Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul Syst Bio* 1, 275–293 (2007).
25. Huang CH et al. CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium *S. elongatus* PCC 7942. *Microb. Cell Fact* 15, 196 (2016). [PubMed: 27846887]
26. Markley AL, Begemann MB, Clarke RE, Gordon GC & Pflieger BF Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium *Synechococcus* sp. strain PCC 7002. *ACS Synth. Biol* 4, 595–603 (2015). [PubMed: 25216157]
27. Camsund D, Heidorn T & Lindblad P Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium. *J. Biol. Eng* 8, 4 (2014). [PubMed: 24467947]
28. Vijayan V, Jain IH & O'Shea EK A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12, R47 (2011). [PubMed: 21612627]
29. Vogel J, Axmann IM, Herzel H & Hess WR Experimental and computation analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. *Nucleic Acids Res* 31, 2890–2899 (2003). [PubMed: 12771216]
30. Camsund D & Lindblad P Engineered Transcriptional Systems for Cyanobacterial Biotechnology. *Front. Bioeng. Biotechnol* 2, 1–9 (2014). [PubMed: 25152877]
31. Caslake LF & Bryant DA The *sigA* gene encoding the major sigma-factor of RNA polymerase from the marine cyanobacterium *Synechococcus* sp. strain PCC-7002: Cloning and characterization. *Microbiol* 142, 347–357 (1996).
32. Brahamsha B & Haselkorn R Isolation and Characterization of the Gene Encoding the Principal Sigma Factor of the Vegetative Cell RNA Polymerase from the Cyanobacterium *Anabaena* sp. Strain PCC 7120. *J. Bacteriol* 173, 2442–2450 (1991). [PubMed: 1901566]
33. Imamura S & Asayama M Sigma factors for cyanobacterial transcription. *Gene Regul. Syst. Biol* 3, 65–87 (2009).
34. Imashimizu M, Tanaka K & Shimamoto N Comparative Study of Cyanobacterial and *E. coli* RNA Polymerases: Misincorporation, Abortive Transcription, and Dependence on Divalent Cations. *Genet. Res. Int* 2011, 1–11 (2011).
35. Keren N, Kidd MJ, Penner-Hahn JE & Pakrasi HB A light-dependent mechanism for massive accumulation of manganese in the photosynthetic bacterium *Synechocystis* sp. PCC 6803. *Biochemistry* 41, 15085–15092 (2002). [PubMed: 12475258]
36. Łyze R, Kochanowska M, Węgrzyn G & Szalewska-Pałasz A Transcription from bacteriophage λ pR promoter is regulated independently and antagonistically by DksA and ppGpp. *Nucleic Acids Res* 37, 6655–6664 (2009). [PubMed: 19759216]
37. Hood RD, Higgins SA, Flamholz A, Nichols RJ & Savage DF The stringent response regulates adaptation to darkness in the cyanobacterium *Synechococcus elongatus*. *Proc. Natl. Acad. Sci* 201524915 (2016). doi:10.1073/pnas.1524915113

38. Zhang SR, Lin GM, Chen WL, Wang L & Zhang CC PpGpp metabolism is involved in heterocyst development in the Cyanobacterium *Anabaena* sp. Strain PCC 7120. *J. Bacteriol* 195, 4536–4544 (2013). [PubMed: 23935047]
39. Busby S & Ebright RH Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol* 293, 199–213 (1999). [PubMed: 10550204]
40. Yoshimura H, Hisabori T, Yanagisawa S & Ohmori M Identification and characterization of a novel cAMP receptor protein in the cyanobacterium *Synechocystis* sp. PCC 6803. *J Biol Chem* 275, 6241–6245 (2000). [PubMed: 10692419]
41. Hedger J et al. Illumination stimulates cAMP receptor protein-dependent transcriptional activation from regulatory regions containing class I and class II promoter elements in *Synechocystis* sp. PCC 6803. *Microbiology* 155, 2994–3004 (2009). [PubMed: 19542007]
42. Shi C et al. Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Sci. Rep* 6, 30135 (2016). [PubMed: 27456469]
43. Reppas NB, Wade JT, Church GM & Struhl K The Transition between Transcriptional Initiation and Elongation in *E. coli* Is Highly Variable and Often Rate Limiting. *Mol. Cell* 24, 747–757 (2006). [PubMed: 17157257]
44. Herrero A et al. Nitrogen Control in Cyanobacteria. *J. Bacteriol* 183, 411–425 (2001). [PubMed: 11133933]
45. Luque I, Flores E & Herrero A Molecular mechanism for the operation of nitrogen control in cyanobacteria. *EMBO J* 13, 2862–9 (1994). [PubMed: 8026471]
46. Flores E, Frías JE, Rubio LM & Herrero A Photosynthetic nitrate assimilation in cyanobacteria. *Photosynth. Res* 83, 117–133 (2005). [PubMed: 16143847]
47. Frías JE, Flores E & Herrero A Activation of the *Anabaena* nir operon promoter requires both NtcA (CAP family) and NtcB (LysR family) transcription factors. *Mol. Microbiol* 38, 613–625 (2000). [PubMed: 11069684]
48. Aichi M, Takatani N & Omata T Role of NtcB in Activation of Nitrate Assimilation Genes in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. *J. Bacteriol* 183, 5840–5847 (2001). [PubMed: 11566981]
49. Omata T, Gohta S, Takahashi Y, Harano Y & Maeda SI Involvement of a CbbR homolog in low CO₂-induced activation of the bicarbonate transporter operon in cyanobacteria. *J. Bacteriol* 183, 1891–1898 (2001). [PubMed: 11222586]
50. Nishimura T et al. Mechanism of low CO₂-induced activation of the cmp bicarbonate transporter operon by a LysR family protein in the cyanobacterium *Synechococcus elongatus* strain PCC 7942. *Mol. Microbiol* 68, 98–109 (2008). [PubMed: 18312274]
51. Daley SME, Kappell AD, Carrick MJ & Burnap RL Regulation of the cyanobacterial CO₂-concentrating mechanism involves internal sensing of NADP⁺ and α -ketoglutarate levels by transcription factor CcmR. *PLoS One* 7, 1–10 (2012).
52. Woodger FJ, Bryant D.a & Price GD. Transcriptional regulation of the CO₂-concentrating mechanism in a euryhaline, coastal marine cyanobacterium, *Synechococcus* sp. Strain PCC 7002: role of NdhR/CcmR. *J. Bacteriol* 189, 3335–47 (2007). [PubMed: 17307862]
53. McClure RS et al. Network analysis of transcriptomics expands regulatory landscapes in *Synechococcus* sp. PCC 7002. *Nucleic Acids Res* 44, 8810–8825 (2016). [PubMed: 27568004]
54. Ishii A & Hihara Y An AbrB-Like Transcriptional Regulator, Sll0822, Is Essential for the Activation of Nitrogen-Regulated Genes in *Synechocystis* sp. PCC 6803. *Plant Physiol* 148, 660–670 (2008). [PubMed: 18667724]
55. Orf I et al. CyAbrB2 contributes to the transcriptional regulation of low CO₂ acclimation in *Synechocystis* sp. PCC 6803. *Plant Cell Physiol* 57, 2232–2243 (2016). [PubMed: 27638927]
56. Wang T et al. The sufR Gene (sll0088 in *Synechocystis* sp. Strain PCC 6803) Functions as a Repressor of the sufBCDS Operon in Iron-Sulfur Cluster Biogenesis in Cyanobacteria. *J. Bacteriol* 186, 956–967 (2004). [PubMed: 14761990]
57. González A, Angarica VE, Sancho J & Fillat MF The FurA regulon in *Anabaena* sp. PCC 7120: In silico prediction and experimental validation of novel target genes. *Nucleic Acids Res* 42, 4833–4846 (2014). [PubMed: 24503250]

58. Ghassemian M & Straus NA Fur regulates the expression of iron-stress genes. *Microbiology* 142, 1469–1476 (2017).
59. Horiuchi M et al. The PedR transcriptional regulator interacts with thioredoxin to connect photosynthesis with gene expression in cyanobacteria. *Biochem. J* 431, 135–140 (2010). [PubMed: 20662766]
60. Shi M, Zhang X, Pei G, Chen L & Zhang W Functional diversity of transcriptional regulators in the cyanobacterium *Synechocystis* sp. PCC 6803. *Front. Microbiol* 8, 1–10 (2017). [PubMed: 28197127]
61. Ma AT, Schmidt CM & Golden JW Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. *Appl. Environ. Microbiol* 80, 6704–6713 (2014). [PubMed: 25149516]
62. Miyake K et al. A green-light inducible lytic system for cyanobacterial cells. *Biotechnol. Biofuels* 7, 56 (2014). [PubMed: 24713090]
63. Peca L, Kós PB & Vass I Characterization of the activity of heavy metal-responsive promoters in the cyanobacterium *Synechocystis* PCC 6803. *Acta Biol. Hung* 58 Suppl, 11–22 (2007).
64. Perez AA et al. Zn²⁺-Inducible Expression Platform for *Synechococcus* sp. strain PCC 7002 Based on the *smtA* Promoter/Operator and *smtB* Repressor. *Appl. Environ. Microbiol* 83, 1–14 (2017).
65. Ng WO, Zentella R, Wang Y, Taylor JS & Pakrasi HB PhrA, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase. *Arch. Microbiol* 173, 412–417 (2000). [PubMed: 10896222]
66. Mutsuda M, Michel KP, Zhang X, Montgomery BL & Golden SS Biochemical properties of CikA, an unusual phytochrome-like histidine protein kinase that resets the circadian clock in *Synechococcus elongatus* PCC 7942. *J. Biol. Chem* 278, 19102–19110 (2003). [PubMed: 12626498]
67. Varman AM, Xiao Y, Pakrasi HB & Tang YJ Metabolic engineering of *Synechocystis* sp. Strain PCC 6803 for isobutanol production. *Appl. Environ. Microbiol* 79, 908–914 (2013). [PubMed: 23183979]
68. Geerts D, Bovy A, Devrieze G, Borrias M & Weisbeek P Inducible expression of heterologous genes targeted to a chromosomal platform in the cyanobacterium *Synechococcus* sp. PCC 7942. *Microbiol. Uk* 141, 831–841 (1995).
69. Elhai J Strong and regulated promoters in the cyanobacterium *Anabaena* PCC 7120. *FEMS Microbiol. Lett* 114, 179–184 (1993). [PubMed: 8282186]
70. Oliver JWK, Machado IMP, Yoneda H & Atsumi S Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc. Natl. Acad. Sci* 110, 1249–1254 (2013). [PubMed: 23297225]
71. Ducat DC, Avelar-Rivas JA, Way JC & Silvera PA Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol* 78, 2660–2668 (2012). [PubMed: 22307292]
72. Nozzi NE & Atsumi S Genome Engineering of the 2,3-Butanediol Biosynthetic Pathway for Tight Regulation in Cyanobacteria. *ACS Synth. Biol* 4, 1197–1204 (2015). [PubMed: 25974153]
73. Camsund D & Lindblad P Engineered transcriptional systems for cyanobacterial biotechnology. *Front. Bioeng. Biotechnol* 2, 40 (2014). [PubMed: 25325057]
74. Gatti-Lafronconi P, Dijkman WP, Devenish SR & Hollfelder F A single mutation in the core domain of the lac repressor reduces leakiness. *Microb. Cell Fact* 12, 67 (2013). [PubMed: 23834731]
75. Zess EK, Begemann MB & Pflieger BF Construction of new synthetic biology tools for the control of gene expression in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Biotechnol. Bioeng* 113, 424–432 (2016). [PubMed: 26192329]
76. Higo A, Isu A, Fukaya Y & Hisabori T Efficient gene induction and endogenous gene repression systems for the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Plant Cell Physiol* 57, 387–396 (2016). [PubMed: 26684202]
77. Gordon GC et al. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metab. Eng* 38, 170–179 (2016). [PubMed: 27481676]

78. Politi N et al. Half-life measurements of chemical inducers for recombinant gene expression. *J. Biol. Eng* 8, 5 (2014). [PubMed: 24485151]
79. Topp S et al. Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl. Environ. Microbiol* 76, 7881–7884 (2010). [PubMed: 20935124]
80. Nakahira Y, Ogawa A, Asano H, Oyama T & Tozawa Y Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Cell Physiol* 54, 1724–1735 (2013). [PubMed: 23969558]
81. Kopf M & Hess WR Regulatory RNAs in photosynthetic cyanobacteria. *FEMS Microbiol. Rev* 39, 301–315 (2015). [PubMed: 25934122]
82. Mutalik VK, Qi L, Guimaraes JC, Lucks JB & Arkin AP Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol* 8, 447–454 (2012). [PubMed: 22446835]
83. Abe K et al. Design of riboregulators for control of cyanobacterial (*Synechocystis*) protein expression. *Biotechnol. Lett* 36, 287–294 (2014). [PubMed: 24068508]
84. Espah Borujeni A, Channarasappa AS & Salis HM 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 (2014). [PubMed: 24234441]
85. Salis HM, Mirsky EA & Voigt CA Automated Design of Synthetic Ribosome Binding Sites to Precisely Control Protein Expression. *Nat Biotechnol* 27, 946–950 (2010).
86. Na D & Lee D RBSDesigner: software for designing synthetic ribosome binding sites that yield a desired level of protein expression. *Bioinformatics* 3, 1–2 (2010).
87. Seo SW et al. Predictive combinatorial design of mRNA translation initiation regions for systematic optimization of gene expression levels. *Sci. Rep* 4, 4515 (2014). [PubMed: 24682040]
88. Seo SW et al. Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng* 15, 67–74 (2013). [PubMed: 23164579]
89. Na D, Lee S & Lee D Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Syst. Biol* 4, 71 (2010). [PubMed: 20504310]
90. Reeve B, Hargest T, Gilbert C & Ellis T Predicting Translation Initiation Rates for Designing Synthetic Biology. *Front. Bioeng. Biotechnol* 2, 1–6 (2014). [PubMed: 25152877]
91. Taton A et al. Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Res* 42, 1–16 (2014). [PubMed: 24376271]
92. Qi LS et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183 (2013). [PubMed: 23452860]
93. Gilbert LA et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451 (2013). [PubMed: 23849981]
94. Yao L, Cengic I, Anfelt J & Hudson EP Multiple Gene Repression in Cyanobacteria Using CRISPRi. *ACS Synth. Biol* 5, 207–212 (2016). [PubMed: 26689101]
95. Steglich C et al. Short RNA half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biol* 11, R54 (2010). [PubMed: 20482874]
96. Hambræus G, Von Wachenfeldt C & Hederstedt L Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Genet. Genomics* 269, 706–714 (2003). [PubMed: 12884008]
97. Kristoffersen SM et al. Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium. *Genome Biol* 13, R30 (2012). [PubMed: 22537947]
98. Selinger DW, Saxena RM, Cheung KJ, Church GM & Rosenow C Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res* 13, 216–223 (2003). [PubMed: 12566399]
99. Bernstein J.a, Khodursky AB, Lin P-H, Lin-Chao S & Cohen SN. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A* 99, 9697–702 (2002). [PubMed: 12119387]

100. Chen H, Shiroguchi K, Ge H & Xie XS Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol* 11, 781–781 (2015). [PubMed: 25583150]
101. Ferreira R, Borges V, Borrego MJ & Gomes JP Global survey of mRNA levels and decay rates of *Chlamydia trachomatis trachoma* and *lymphogranuloma venereum biovars*. *Heliyon* 3, e00364 (2017). [PubMed: 28795162]
102. Rustad TR et al. Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic Acids Res* 41, 509–517 (2013). [PubMed: 23125364]
103. Cameron JC, Gordon GC & Pflieger BF Genetic and genomic analysis of RNases in model cyanobacteria. *Photosynth. Res* 126, 171–183 (2015). [PubMed: 25595545]
104. Taverniti V, Forti F, Ghisotti D & Putzer H *Mycobacterium smegmatis* RNase J is a 5′–3′ exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol. Microbiol* 82, 1260–1276 (2011). [PubMed: 22014150]
105. Rische T & Klug G The ordered processing of intervening sequences in 23S rRNA of *Rhodobacter sphaeroides* requires RNase J. *RNA Biol* 9, 343–350 (2012). [PubMed: 22336705]
106. Bralley P, Aseem M & Jones GH SCO5745, a Bifunctional RNase J Ortholog, affects antibiotic production in *Streptomyces coelicolor*. *J. Bacteriol* 196, 1197–1205 (2014). [PubMed: 24415725]
107. Baumgardt K et al. RNase E and RNase J are needed for S-adenosylmethionine homeostasis in *Sinorhizobium meliloti*. *Microbiology* 163, 570–583 (2017). [PubMed: 28141492]
108. Deana A, Celesnik H & Belasco JG The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. *Nature* 451, 355–358 (2008). [PubMed: 18202662]
109. Mackie G a. Ribonuclease E is a 5′-end-dependent endonuclease. *Nature* 395, 720–723 (1998). [PubMed: 9790196]
110. Callaghan AJ et al. Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* 437, 1187–91 (2005). [PubMed: 16237448]
111. Pei XY, Bralley P, Jones GH & Luisi BF Linkage of catalysis and 5′ end recognition in ribonuclease RNase J. *Nucleic Acids Res* 43, 8066–8076 (2015). [PubMed: 26253740]
112. Bischler T et al. Identification of the RNA pyrophosphohydrolase RppH of *Helicobacter pylori* and global analysis of its RNA targets. *J. Biol. Chem* 292, 1934–1950 (2017). [PubMed: 27974459]
113. Luciano DJ, Vasilyev N, Richards J, Serganov A & Belasco JG A Novel RNA Phosphorylation State Enables 5′ End-Dependent Degradation in *Escherichia coli*. *Mol. Cell* 1–11 (2017). doi: 10.1016/j.molcel.2017.05.035
114. Nudler E & Gottesman ME Transcription termination and anti-termination in *E. coli*. *Genes to Cells* 7, 755–768 (2002). [PubMed: 12167155]
115. Carafa Y.d. A., Brody E & Thermes C. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol* 216, 835–858 (1990). [PubMed: 1702475]
116. Lesnik EA et al. Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res* 29, 3583–94 (2001). [PubMed: 11522828]
117. Washio T, Sasayama J & Tomita M Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res* 26, 5456–63 (1998). [PubMed: 9826772]
118. Stern DB & Grissem W Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* 51, 1145–1157 (1987). [PubMed: 3690662]
119. Dombrowski S, Brennicke A & Binder S 3′-inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *EMBO J* 16, 5069–5076 (1997). [PubMed: 9305648]
120. Zou J, Gasparich GE, Stirewalt VL, Lorimier R. de & Bryant DA The *cpcE* and *cpcF* Genes of *Synechococcus* sp. PCC 7002. *J. Biol. Chem* 267, 16138–16145 (1999).
121. Belknap WR & Haselkorn R Cloning and light regulation of expression of the phycocyanin operon of the cyanobacterium *Anabaena*. *EMBO J* 6, 871–884 (1987). [PubMed: 3109890]

122. Grossman AR, Schaefer MR, Chiang GG & Collier JL The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol. Rev* 57, 725–49 (1993). [PubMed: 8246846]
123. Ramey CJ, Baron-Sola A, Aucoin HR & Boyle NR Genome Engineering in Cyanobacteria: Where We Are and Where We Need to Go. *ACS Synth. Biol* 4, 1186–1196 (2015). [PubMed: 25985322]
124. Barten R & Lill H DNA-uptake in the naturally competent cyanobacterium, *Synechocystis* sp. PCC 6803. *FEMS Microbiol. Lett* 129, 83–88 (2017).
125. Stevens SE & Porter RD Transformation in *Agmenellum quadruplicatum*. *Proc. Natl. Acad. Sci* 77, 6052–6056 (1980). [PubMed: 16592896]
126. Golden SS & Sherman LA Optimal conditions for genetic transformation of the cyanobacterium *Anacystis nidulans* R2. *J. Bacteriol* 158, 36–42 (1984). [PubMed: 6425267]
127. Schuergers N & Wilde A Appendages of the Cyanobacterial Cell. *Life* 5, 700–715 (2015). [PubMed: 25749611]
128. Sakamoto T & Bryant D.a Temperature-regulated mRNA accumulation and stabilization for fatty acid desaturase genes in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Mol. Microbiol* 23, 1281–92 (1997). [PubMed: 9106218]
129. Peterson ES et al. VESPA: software to facilitate genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data. *BMC Genomics* 13, 131 (2012). [PubMed: 22480257]
130. Reytrat J, Pelicic V & Gicquel B Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis. *Infect. Immun* 66, 4011–4017 (1998). [PubMed: 9712740]
131. Cai Y & Wolk CP Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. *J. Bacteriol* 172, 3138–3145 (1990). [PubMed: 2160938]
132. Viola S, Rühle T & Leister D A single vector-based strategy for marker-less gene replacement in *Synechocystis* sp. PCC 6803. *Microb. Cell Fact* 13, 4 (2014). [PubMed: 24401024]
133. Matsuoka M, Takahama K & Ogawa T Gene replacement in cyanobacteria mediated by a dominant streptomycin-sensitive *rps12* gene that allows selection of mutants free from drug resistance markers. *Culture* 2077–2087 (2001).
134. Begemann MB et al. An Organic Acid Based Counter Selection System for Cyanobacteria. *PLoS One* 8, 1–12 (2013).
135. Wendt KE, Ungerer J, Cobb RE, Zhao H & Pakrasi HB CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Microb. Cell Fact* 15, 115 (2016). [PubMed: 27339038]
136. Li H et al. CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metab. Eng* 38, 293–302 (2016). [PubMed: 27693320]
137. Fonfara I, Richter H, Bratovi M, Le Rhun A & Charpentier E The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517–521 (2016). [PubMed: 27096362]
138. Zetsche B et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163, 759–771 (2015). [PubMed: 26422227]
139. Ungerer J & Pakrasi HB Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Sci. Rep* 6, 39681 (2016). [PubMed: 28000776]
140. Blokesch M Natural competence for transformation. *Curr. Biol* 26, R1126–R1130 (2016). [PubMed: 27825443]
141. Wolk CP, Vonshak A, Kehoe P & Elhai J Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. *Microbiology* 81, 1561–1565 (1984).
142. Stucken K, Ilhan J, Roettger M, Dagan T & Martin WF Transformation and conjugal transfer of foreign genes into the filamentous multicellular cyanobacteria (subsection V) *Fischerella* and *Chlorogloeopsis*. *Curr. Microbiol* 65, 552–560 (2012). [PubMed: 22833222]
143. Billi D, Friedmann EI, Helm RF & Potts M Gene transfer to the desiccation-tolerant cyanobacterium *Chroococcidiopsis*. *J. Bacteriol* 183, 2298–2305 (2001). [PubMed: 11244070]

144. Ravindran CRM, Suguna S & Shanmugasundaram S Electroporation as a tool to transfer the plasmid pRL489 in *Oscillatoria* MKU 277. *J. Microbiol. Methods* 66, 174–176 (2006). [PubMed: 16406130]
145. Frey J & Bagdasarian M The Molecular Biology of IncQ Plasmids. in *Promiscuous Plasmids of Gram-Negative Bacteria* 79–91 (1989).
146. Brahamsha B A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Appl. Environ. Microbiol* 62, 1747–1751 (1996). [PubMed: 8633873]
147. Taton A et al. Gene transfer in *Leptolyngbya* sp. strain BL0902, a cyanobacterium suitable for production of biomass and bioproducts. *PLoS One* 7, (2012).
148. Gibson DG et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345 (2009). [PubMed: 19363495]
149. Gupta A, Morby AP, Turner JS, Whitton BA & Robinson NJ Deletion within the metallothionein locus of cadmium-tolerant *Synechococcus* PCC 6301 involving a highly iterated palindrome (HIP1). *Mol. Microbiol* 7, 189–195 (1993). [PubMed: 8446026]
150. Delaye L & Moya A Abundance and distribution of the highly iterated palindrome 1 (HIP1) among prokaryotes. *Mob. Genet. Elements* 1, 159–168 (2011). [PubMed: 22312590]
151. Robinson PJ, Cranenburgh RM, Head IM & Robinson NJ HIP1 propagates in cyanobacterial DNA via nucleotide substitutions but promotes excision at similar frequencies in *Escherichia coli* and *Synechococcus* PCC 7942. *Mol. Microbiol* 24, 181–189 (1997). [PubMed: 9140975]
152. Minton K DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. *Mol. Microbiol* 13, 9–15 (1994). [PubMed: 7984097]
153. Xu M Comparative Genomics Reveals Forces Driving the Evolution of Highly Iterated Palindrome-1 (HIP1) in Cyanobacteria. 1, (Carnegie Mellon University, 2015).
154. Bailey TL & Elkan C Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers. *Proc. Second Int. Conf. Intell. Syst. Mol. Biol* 28–36 (1994). doi:citeulike-article-id:878292
155. Griesse M, Lange C & Soppa J Ploidy in cyanobacteria. *FEMS Microbiol. Lett* 323, 124–131 (2011). [PubMed: 22092711]
156. Mori T, Binder B & Johnson CH Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl. Acad. Sci. U. S. A* 93, 10183–10188 (1996). [PubMed: 8816773]
157. Labarre J, Chauvat F & Thuriaux P Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. *J. Bacteriol* 171, 3449–3457 (1989). [PubMed: 2498291]
158. Domain F, Houot L, Chauvat F & Cassier-Chauvat C Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation. *Mol. Microbiol* 53, 65–80 (2004). [PubMed: 15225304]
159. Hobbs EC, Fontaine F, Yin X & Storz G An expanding universe of small proteins. *Curr. Opin. Microbiol* 14, 167–173 (2011). [PubMed: 21342783]
160. Fickett JW ORFs and Genes: How Strong a Connection? *J. Comput. Biol* 2, 117–123 (1995). [PubMed: 7497114]
161. Su M, Ling Y, Yu J, Wu J & Xiao J Small proteins: Untapped area of potential biological importance. *Front. Genet* 4, 1–9 (2013).
162. Gao L et al. Toward the complete proteome of *Synechocystis* sp. PCC 6803. *Photosynth. Res* 126, 203–219 (2015). [PubMed: 25862646]
163. Kashino Y et al. Proteomic analysis of a highly active Photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* 41, 8004–8012. (2002). [PubMed: 12069591]
164. Baumgartner D, Kopf M, Klähn S, Steglich C & Hess WR Small proteins in cyanobacteria provide a paradigm for the functional analysis of the bacterial micro-proteome. *BMC Microbiol* 16, 285 (2016). [PubMed: 27894276]

Figure 1.
Overview of key factors governing gene expression.

Figure 2. Fold induction ratios from inducible systems in cyanobacteria. We've color-coded bars by inducer (IPTG, aTc, metals, and theophylline), and separated them by induction mechanism (transcription vs translation). The organism and citation can be found below the bar.

Figure 3. Organization and transcripts observed from *cpc* operons in PCC 7002 and PCC 7120. Genes are color-coded for their functions (red: structural components, blue: linker proteins, and green: attachment enzymes). Promoters are marked as well as regions that predicted to form significant secondary structures. Observed transcripts are marked below with the thickness indicating relative abundance.

Table 1.

Cyanobacteria strains mentioned in this chapter

Cyanobacterial Strain Name	Abbreviation
<i>Anabaena (Nostoc) sp.</i> PCC 7120	PCC 7120
<i>Gleobacter violaceus</i>	<i>G. violaceus</i>
<i>Leptolyngbya sp.</i> strain BL0902	BL0902
<i>Nostoc punctiforme</i>	<i>N. punctiforme</i>
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. CCMP1986	CCMP1986
<i>Prochlorococcus sp.</i> MED4	MED4
<i>Synechococcus</i> PCC 6301	PCC 6301
<i>Synechococcus</i> WH7803	WH7803
<i>Synechococcus elongates</i> PCC 7942	PCC 7942
<i>Synechococcus elongatus</i> UTEX 2973	UTEX 2973
<i>Synechococcus sp.</i> strain PCC 7002	PCC 7002
<i>Synechocystis sp.</i> strain PCC 6714	PCC 6714
<i>Synechocystis sp.</i> strain PCC 6803	PCC 6803
<i>Synechocystis sp.</i> strain WHSyn	WHSyn
<i>Thermosynechococcus elongatus</i> BP-1	BP-1