



# Transcription control engineering and applications in synthetic biology



Michael D. Engstrom<sup>a, b, 1</sup>, Brian F. Pfleger<sup>b, c, \*</sup>

<sup>a</sup> Genetics-Biotechnology Center, University of Wisconsin-Madison School of Medicine and Public Health, USA

<sup>b</sup> Department of Chemical and Biological Engineering, University of Wisconsin-Madison College of Engineering, USA

<sup>c</sup> Microbiology Doctoral Training Program, University of Wisconsin-Madison, USA

## ARTICLE INFO

### Article history:

Received 28 July 2017

Received in revised form

26 September 2017

Accepted 26 September 2017

### Keywords:

Promoter

Transcription

Synthetic biology

Bacteria

Termination

Metabolic engineering

## ABSTRACT

In synthetic biology, researchers assemble biological components in new ways to produce systems with practical applications. One of these practical applications is control of the flow of genetic information (from nucleic acid to protein), a.k.a. gene regulation. Regulation is critical for optimizing protein (and therefore activity) levels and the subsequent levels of metabolites and other cellular properties. The central dogma of molecular biology posits that information flow commences with transcription, and accordingly, regulatory tools targeting transcription have received the most attention in synthetic biology. In this mini-review, we highlight many past successes and summarize the lessons learned in developing tools for controlling transcription. In particular, we focus on engineering studies where promoters and transcription terminators (*cis*-factors) were directly engineered and/or isolated from DNA libraries. We also review several well-characterized transcription regulators (*trans*-factors), giving examples of how *cis*- and *trans*-acting factors have been combined to create digital and analogue switches for regulating transcription in response to various signals. Last, we provide examples of how engineered transcription control systems have been used in metabolic engineering and more complicated genetic circuits. While most of our mini-review focuses on the well-characterized bacterium *Escherichia coli*, we also provide several examples of the use of transcription control engineering in non-model organisms. Similar approaches have been applied outside the bacterial kingdom indicating that the lessons learned from bacterial studies may be generalized for other organisms.

© 2017 Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1. Introduction .....	177
2. Prokaryotic promoters, transcription initiation, termination, and tools .....	178
2.1. Elements of promoter recognition by RNA polymerases .....	178
2.2. Promoter libraries .....	179
2.3. Transcription terminators .....	180
2.4. Terminator libraries .....	181
3. Transcription control through transcription factor tools .....	181
3.1. CI and Cro .....	182
3.2. CAP (catabolite activator protein) .....	182
3.3. LacI .....	182
3.4. AraC .....	184

\* Corresponding author. 3629 Engineering Hall, 1415 Engineering Dr. Madison, WI 53706, USA.

E-mail address: [pfleger@engr.wisc.edu](mailto:pfleger@engr.wisc.edu) (B.F. Pfleger).

Peer review under responsibility of KeAi Communications Co., Ltd.

<sup>1</sup> Present address: 2730 Engineering Hall, 1415 Engineering Dr. Madison, WI 53706, USA.

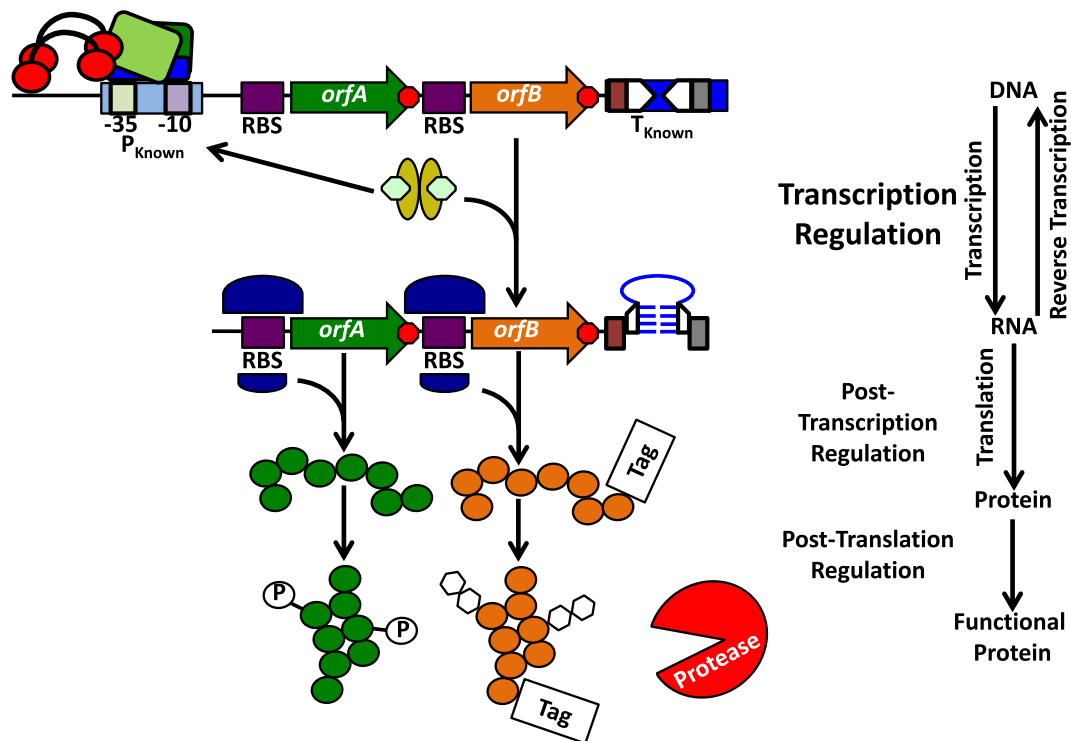
3.5.	TetR .....	184
3.6.	LuxR .....	184
3.7.	TALEs (transcription activator-like effectors) .....	184
3.8.	CRISPRi (clustered regularly interspaced short palindromic repeats interference) .....	185
3.9.	Two-component regulatory systems .....	185
4.	Conclusion .....	186
	Acknowledgements .....	186
	References .....	186

## 1. Introduction

A central tenet of synthetic biology is the ability to assemble a practical system from biological components. In most cases, the components are proteins synthesized from genetic information encoded in DNA. One of the challenges faced by synthetic biologists is designing the appropriate genetic sequence to control protein expression at the level that optimizes system function. The central dogma of molecular biology (Fig. 1) describes the flow of genetic information from nucleic acid polymers (DNA and RNA) to amino acid polymers (proteins) [1]. Transcription is the process during which information encoded in DNA is converted to a transient carrier of information, mRNA. Translation is the process in which the mRNA code is converted to an amino acid polymer by

ribosomes, themselves RNA catalysts (ribozymes) [2–5]. The flow of genetic information terminates with the production of a protein, but often post-translational modifications and processes (see below) are needed to produce the final functional component. To optimize the abundance and other properties of the protein and the performance of the ultimate application, the rate and timing of each of these processes can be controlled by both *cis*-acting (within or adjacent to the protein coding sequence) and *trans*-acting (separate from the protein coding sequence) components.

Transcriptional control is a prevalent mode of regulating protein production because it is the first step in the process. Accordingly, it has received the most attention in terms of tool development and engineering applications. Transcriptional regulation allows a cell to allocate its valuable resources towards production of proteins when



**Fig. 1. Overview of the central dogma of molecular biology.** An operon encoding *orfA* and *orfB* is under the control of the indicated promoter ( $P_{\text{Known}}$ ), which is engaged by the RNA polymerase holoenzyme (green and blue boxes, red circles, and black arcs). Sequences that will be transcribed into the ribosome binding sites (RBSs) for *orfA* and *orfB* (purple boxes) and the transcription terminator ( $T_{\text{Known}}$ ; two inverted repeats [inverted arrowheads] flanked by a poly A-tract [red box] and poly U-tract [grey box]) are indicated. Red octagons represent translation stop codons. At the level of **transcription regulation**, alterations to the promoter and terminator sequences, the presence of transcription activators and repressors (depicted as golden ovals with green hexagon ligands bound), and nucleoid/supercoiling states may all influence *orfA* and *orfB* transcription (not shown). Through the process of transcription itself, the associated *orfA* and *orfB* mRNA is produced, which may in turn be bound by ribosomes (depicted as two variable size purple half circles) at the RBS. To terminate transcription, the depicted terminator sequence forms a stem-loop (blue), mediated by base pairing/stacking with the complementary sequences of the former inverted repeats (half arrowheads), and mRNA synthesis ceases at this site (poly A- and poly U-tracts are indicated as above). Modifying RBS strength, RNA riboswitches, aptazymes (aptamer ribozymes), and RNA stability may each contribute to **post-transcription regulation** (not shown). Folding of a translated protein (green and orange circles) with the assistance of folding chaperones (not shown), enzymes that promote post-translational modifications such as phosphorylation (P in a white circle) or glycosylation (white hexagons), the presence of proteolytic enzymes (indicated in the figure and potentially targeted to a protein through a specific motif [Tag]) represent methods of **post-translation regulation**.

these proteins would increase fitness in a given environment. Furthermore, transcriptional regulation also provides a method for the cell to control the levels/stoichiometry of enzymes to avoid futile cycles, underutilized enzymes, build-up of undesired/toxic metabolites, and assembly of macromolecular structures. Over the last several decades, researchers have assembled methodologies for creating tools that enable fine-tuning of gene expression without the evolution time required to optimize natural systems. These tools have been applied to many organisms and heterologous systems. Indeed, fine tuning the levels of enzymes and other proteins has proven to be essential to many applications in metabolic engineering [6–15] and common synthetic biology systems such as toggle switches and Boolean logic circuits [16–23], cellular memory and calculation elements [24–27], and regulatory oscillators [13,20,28]. Promoter and terminator libraries [12,29–40] have enabled many of these applications.

This mini-review will describe elements of transcriptional control and applications for engineered systems, but many control topics are beyond its scope. Although we will focus primarily on well-characterized examples of transcriptional control in model bacteria (e.g., *Escherichia coli*), we acknowledge that these transcriptional control methods are being increasingly extended to non-model organisms and eukaryotes. We are confident that the approaches developed for *E. coli* and other simple systems will lead to the creation of transcription tools for these organisms and find use in both metabolic engineering and synthetic biology applications. While this review will focus on transcription control, there are additional layers of control over protein expression at the post-transcriptional (e.g., riboswitches/toehold RNA switches [41–44], RNA splicing/ribozymes [9,45–47], and at the level of modifying ribosome binding site [RBS] strength [48,49]) and post-translational levels (e.g., proteolytic activation/inactivation [50–56], glycosylation [57–60], phosphopantetheine addition [61–63], phosphorylation [64–68], quaternary structure formation [69–74], protease tags [75], and folding chaperones [76]). However, these are methods beyond the intended scope of this mini-review. Likewise, most genes in bacteria are arranged in operons [77–81]. In addition to understanding promoter strength and transcription termination (discussed here, but in a different context), engineering synthetic operons requires careful RBS design and positioning [48,49]; an understanding of gene positioning, order, and orientation [78–81]; an understanding of RNA stability and structure [82–84]; and an understanding of intergenic DNA and RNA sequences (related to the other operon engineering considerations) [29,30,85]. These too are all beyond the scope of this mini-review. Please note that transcription and translation are coupled in prokaryotes (i.e., simultaneously occurring in the same location). Intriguingly, once these processes are decoupled, transcripts become rapidly inactivated through such mechanisms as alterations to the RNA structure, RNA degradation, and forming RNA-DNA R-Loops [86–91]. Furthermore, the bacterial nucleoid structure and DNA supercoiling surely play an overwhelming [92–98], but still poorly understood, function during transcription regulation. Nevertheless, transcription-translation decoupling, DNA structure, and supercoiling are all beyond the intentional scope of this mini-review.

## 2. Prokaryotic promoters, transcription initiation, termination, and tools

A logical point to begin any discussion of transcription regulation is at the level of the machinery and DNA/RNA sequence elements that facilitate mRNA transcript production. In particular, the machinery referred to here includes bacterial RNA polymerase (RNAP) holoenzymes. The DNA/RNA sequence elements referred to

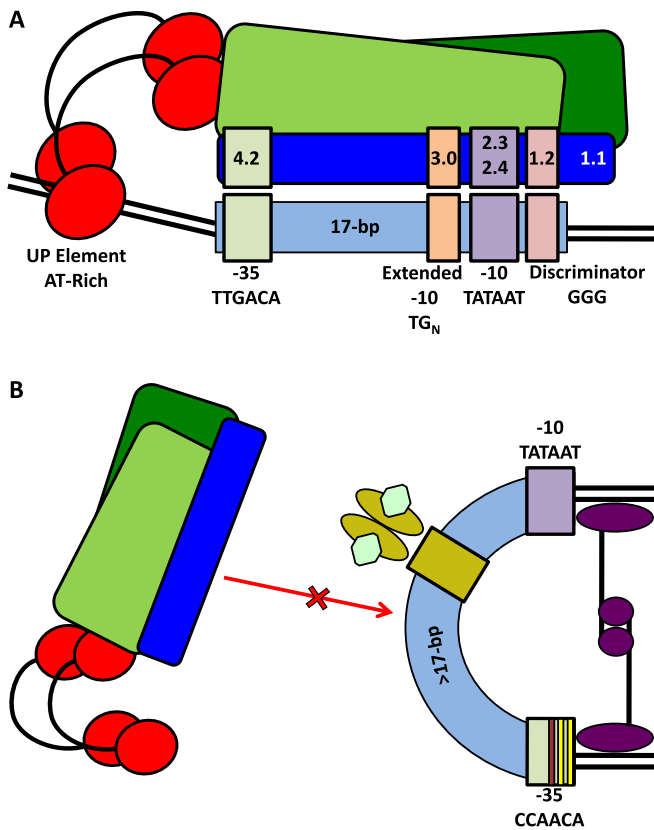
above include promoters and terminators. Remarkably, the components of the RNAP holoenzyme,  $\alpha_2\beta\beta'\omega\sigma$ , are conserved among bacteria [74,99,100] and have similar characteristics to eukaryotic and viral RNA polymerases [70,73,99,100]. Additionally, RNAP holoenzymes have different activities toward different promoters and terminators [16,35,36,101–108], which motivates promoter and terminator library development.

### 2.1. Elements of promoter recognition by RNA polymerases

Among the well-characterized prokaryotic promoters are those promoters recognized by  $\sigma^{70}$  (sigma-70) and other group 1 sigma factors. Indeed, group 1 sigma factors represent those factors that drive the expression of genes necessary for cell viability and other housekeeping activities [109–111]. In *E. coli*,  $\sigma^{70}$  is the principal sigma factor ( $\sigma^A$  is the principal sigma factor in other prokaryotes) [109–111]. Although there are many additional sigma factors that serve important functions under a variety of different cellular conditions [109–113], we will primarily focus on  $\sigma^{70}$ - or  $\sigma^A$ -regulated promoters.

$\sigma^{70}$ , in complex with RNAP, recognizes key promoter sequences during early events of transcription initiation (for additional descriptions of transcription initiation and the anatomy of RNAP holoenzyme-promoter interactions please see Fig. 2A, Fig. 2B, and references [70–72,110,111,114–121]). This sigma factor is further sub-divided into four regions; elements of sub-region 4 (sub-region 4.2) and 2 (sub-regions 2.3 and 2.4) are involved in recognizing the –35 and –10 regions of a typical  $\sigma^{70}$  promoter, respectively [70–72,109–111,114,115,120,121]. The –35 sequence has the consensus sequence TTGACA, while the –10 sequence has the consensus sequence (TATAAT) [121–123]. An additional TG<sub>N</sub> sequence upstream of the –10 sequence (the extended –10 sequence) is present in some bacterial promoters and is recognized by the 3.0 sub-region of  $\sigma^{70}$  [70–72,111,114,115,120,121]. The optimal spacing between the end of the –35 sequences and the beginning of the –10 sequences is 17-bp (base pairs), which is consistent with sub-regions 2, 3, and 4 positioning when  $\sigma^{70}$  is docked into the RNAP holoenzyme [70–72,109–111,114,120–123]. In addition, a 6- to 8-bp variable region (recognized by the 1.2 sub-region of  $\sigma^{70}$ ) immediately downstream of the –10 sequence (the discriminator) also serves a function during early transcription initiation [71,72,110,111,114,117–121].

Optimal spacing and function of promoter sequences reflects the position of individual components in the RNAP holoenzyme both during promoter recognition and subsequent transcription initiation events. Briefly, during the early stages of promoter recognition, the RNAP holoenzyme  $\alpha_2$  subunits CTDs (carboxyl-terminal domains) bind to the promoter UP elements (an AT-rich sequence [114,121,124,125]), the  $\sigma^{70}$  sub-region 4.2 (proximal to the  $\alpha_2$  subunits CTDs) associates with the promoter –35 sequence, and sub-regions 2.3 and 2.4 associate with the –10 sequence [70–72,109–111,114,115,120,121]. A linker region between  $\sigma^{70}$  sub-region 4.2 and sub-regions 2.3, 2.4, and 3.0 accounts for the spacing between the –35 and –10 sequences (i.e., the linker is in an extended conformation when  $\sigma^{70}$  is docked into the RNAP holoenzyme) [70–72,109–111,114,120,121]. The clamp and jaw formed by the  $\beta$  and  $\beta'$  subunits of the RNAP holoenzyme, respectively, are in an open conformation, while promoter DNA at this point is still in a closed duplex (referred to as the closed complex) [71,72,114,126]. After initial contact with a promoter, DNA is bent around the RNAP holoenzyme; first the upstream portion of DNA proximal to the  $\alpha_2$  subunits CTDs is bent, followed by bending of the region immediately upstream of the –10 sequence through the –10 sequence itself [70–72,114]. During this bending,  $\sigma^{70}$  sub-region 1.2 is in contact with the discriminator, base stacks in duplex DNA are



**Fig. 2. A representation of RNA polymerase holoenzyme-promoter interactions.** (A) RNAP  $\beta$  and  $\beta'$  subunits are indicated by green boxes. The RNAP  $\alpha$  subunits (red circles and black arcs) are in contact with the bent upstream UP element.  $\sigma^{70}$  (dark blue box) sub-region 4.2 (pale green), 3.0 (pale orange), 2.3/2.4 (pale purple), 1.2 (pale pink), and 1.1 are associated with the indicated  $-35$  sequence (pale green), extended  $-10$  sequence (pale orange),  $-10$  sequence (pale purple), discriminator (pale pink), and double stranded DNA, respectively. In this case, promoter elements are separated by 17-bp. (B) RNAP holoenzyme promoter binding is obstructed (red arrow and a red cross) by DNA bending mediated by nucleoid structuring factors or transcription factors (purple ovals, circles, and lines), other transcription regulators (golden ovals with green hexagon ligands bound) bound to operator sequences (gold box), spacing greater than 17-bp between the  $-35$  and  $-10$  sequences, and a  $-35$  sequence deviating from the canonical  $-35$  sequence (yellow and red bars). These figures are modified from the work by Browning and Busby [121].

disrupted in the AT-rich  $-10$  sequence, the base at position  $-11$  (usually an A) is flipped into  $\sigma^{70}$  sub-region 2.3, and the  $-11/-10$  nucleotide through the transcription initiation site (+1) (usually initiated with a purine [122,127]) and adjacent bases (+2/+3) of duplex DNA are eventually dehybridized (*i.e.*, this is step is kinetically limiting) forming the open complex [70–72,111,114,115, 117,120,121,128]. A separation between the clamp and jaw of RNAP, where central  $Mg^{2+}$  ions are coordinated by amino acids found in the  $\beta'$  subunit and water molecules coordinated by amino acids found in both the  $\beta'$  and  $\beta$  subunits [70–72,74,114], forms the active site cleft. The single stranded transcriptional start site (on the template strand), typically seven- or eight-nucleotides downstream of the  $-10$  sequence [70–72,110,114,120], is positioned into the active site cleft, and isomerization events result in a tightening (closing) of the RNAP jaw and clamp [71,72,114,126]. Intriguingly, recent work shows that the RNAP jaw and clamp may transiently close during initial promoter recognition and reopen before the final tightening above [126]. Discriminator sequence-RNAP holoenzyme contacts presumably assist in positioning the template strand (*i.e.*, open DNA strands may be scrunched into this cleft with different DNA-RNAP holoenzyme interactions leading to

transcription initiation sites different from the typical seven- or eight- nucleotides downstream of the  $-10$  sequence noted above) [71,72,114,116,119,120].

The discriminator also serves a function in mediating open complex stability and may influence length of the initially transcribed RNA polymer at the time of promoter escape. As noted above, there is no consensus sequence for the discriminator region [116,118]. Therefore, it instead seems that interactions between open DNA and  $\sigma^{70}$  sub-region 1.2, the presence of G/(G/C) bases at positions  $-6/-5$ , interactions between DNA and  $\sigma^{70}$  sub-region 1.1, continued interactions between  $\sigma^{70}$  and the aforementioned promoter elements above, spacing of these promoter elements, and other interactions between open- and closed-form DNA and RNAP holoenzyme subunits all likely contribute to recognition of the discriminator sequence [70–72,110,111,114–121]. It has recently been demonstrated that open complexes at certain promoters have discriminator regions that mediate stable open complexes and increased length of abortive, and presumably initial escape transcripts, while other promoters with discriminators that mediate unstable open complexes have shorter length abortive and initial escape transcripts [116]. The reason for these differences is likely accounted for by weaker discriminator sequence interactions with the RNAP holoenzyme along with needing to prescrunch these sequences into the RNAP active site cleft to make any productive interactions [70,116,118,119]. Thus, relieving stress buildup resulting from scrunching open complex DNA into the active site cleft may be a significant driver of RNAP promoter escape [116,119,129,130].

## 2.2. Promoter libraries

RNAP holoenzyme activity can vary over a wide range depending on the sequence of a series of promoters [16,36,101–105]. This correlation has been used to tune control of gene expression at the level of promoter recognition and transcription initiation. Constitutive promoters under the control of  $\sigma^{70}$ , in contrast to non-constitutive inducible and repressible promoters (see the Transcription Factors section below), usually direct comparable levels of gene expression independent of environmental and regulator context. Although libraries of characterized non-constitutive promoters exist [33,34,38–40] (<http://parts.igem.org/Promoters/Catalog/Ecoli/Multiple>, <http://parts.igem.org/Promoters/Catalog/Ecoli/Repressible>, <http://parts.igem.org/Promoters/Catalog/Ecoli/Positive>), this mini-review will primarily focus on constitutive promoter libraries (*e.g.*, <http://parts.igem.org/Promoters/Catalog/Ecoli/Constitutive>). These constitutive promoters are often characterized as being weak (yielding low levels of reporter gene expression) or strong (yielding high levels of reporter gene expression). Although weak promoters often deviate from the  $-35$  and  $-10$  consensus sequences and typical element spacing (strong promoters closely match these consensus sequences and typical element spacing), considerable variability may still exist between promoters with identical or nearly identical  $-35$  and  $-10$  elements [32,37,101,102,105,131]. In addition, interactions between the 5' UTRs (untranslated regions) and identical promoters contribute to gene expression variability, suggesting that promoter context is as important as the  $-35$  and  $-10$  sequences [29,30].

Promoter libraries are usually generated either through synthesizing putative constitutive promoter sequences [29,33,37], randomly mutating the promoter sequence [12,34], or generating libraries of putative promoter inserts [29,30,36]. For instance, Jensen et al. [37] demonstrated that a synthetic promoter sequence library originally engineered for *Lactococcus lactis* (see below) and inserted upstream of *lacZ* (encoding  $\beta$ -galactosidase), could be used in *E. coli* to titrate *lacZ* expression over four different orders of



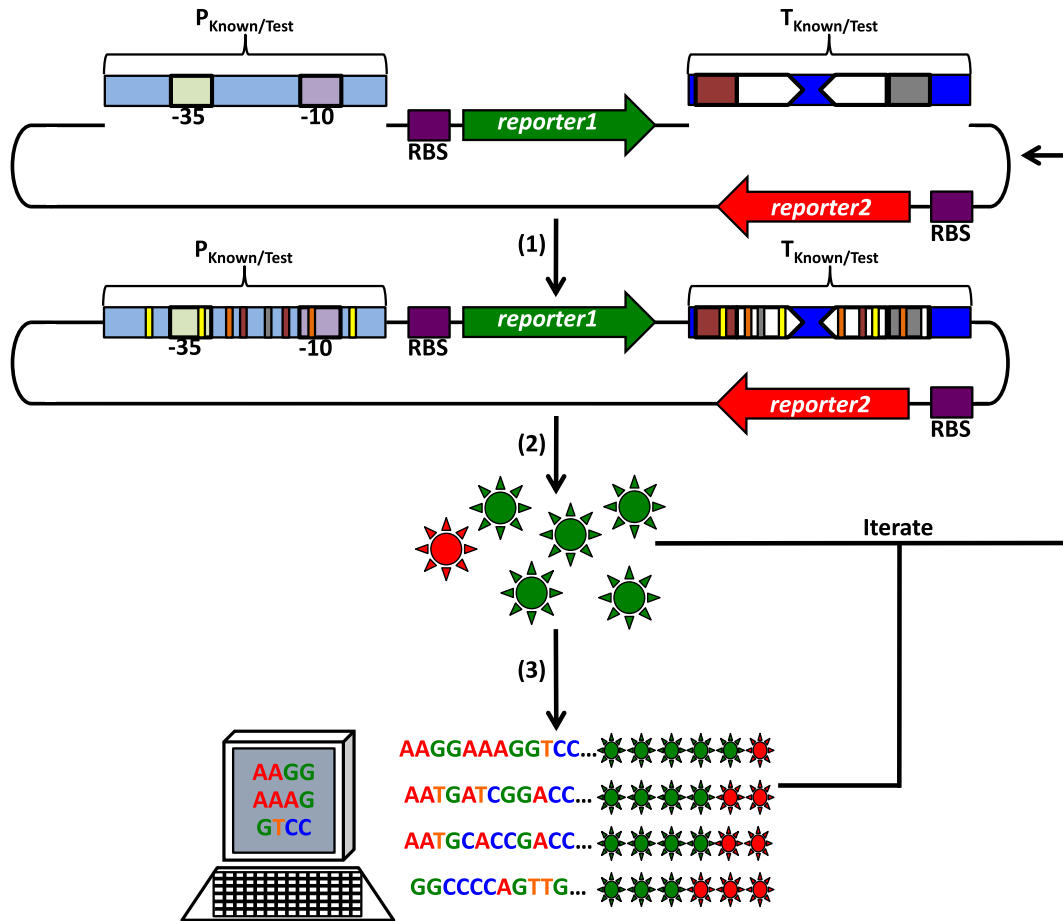
magnitude. This library was synthesized with changes primarily to regions flanking the  $-35$  and  $-10$  sequences, with a few  $-35$  and  $-10$  sequence variants. As one may expect, the strongest promoters in this collection have the  $-35$  and  $-10$  consensus sequence, along with optimal spacing. However many of the weaker promoters in this collection also match these elements and some strong promoters do not. The reasons for these deviations remain unclear, but may result from complex promoter-RNAP-holenzyme interactions, DNA supercoiling, and 3D DNA architecture noted in previous sections above. Indeed, Mutalik et al. [30] developed an analysis of variance (ANOVA) tool to quantify transcription variability arising from different promoter and 5' UTR couplings. In other work by Mutalik et al. [29], these researchers subsequently leveraged known and random promoters, coupled to 5' UTR, monocistron, and bicistron variants, to obtain transcription and translation initiation constructs that promote reliable reporter gene expression over a 1000-fold dynamic range. In the work by Alper et al. [12] the  $P_{\text{Ltet-O1}}$  promoter was randomly mutated, inserted upstream of a *gfp* (encoding green fluorescent protein) reporter gene, and a library of 22 promoter mutants were selected for further characterization based on high reporter expression across the entire population of *E. coli* harboring these constructs. This promoter library had a 196-fold range in GFP production among member constructs. The *cat* (chloramphenicol acetyltransferase) gene was also swapped for *gfp* in this promoter library, the minimum inhibitory concentrations (MICs) of chloramphenicol were measured, and these MICs were correlated with promoter strength. This swap not only demonstrates that gene expression from these library promoters was constitutive, but that these promoters were modular. Leveraging this modularity, Alper et al. inserted library promoters upstream of *ppc* (encoding phosphoenolpyruvate carboxylase) and *dxs* (encoding deoxy-xylulose-5-P synthase). Biomass yields and lycopene production were to varying degrees associated with promoter strength. The Anderson promoter library (<http://parts.igem.org/Promoters/Catalog/Anderson>) is a tool composed of 20 promoters for titrating gene expression over three different orders of magnitude. This library is composed primarily of  $-35$  and  $-10$  sequence variants. Finally, the Weizmann Institute of Science promoter library [36] features nearly 2000 unique members for titratable and dynamic expression of a *gfp* reporter. This library was engineered using random inserts of genomic DNA cloned upstream of a previously promoterless reporter gene, a technique often followed when promoter sequences or regulation profiles are unknown [36,132–137]. Analyzing transcriptome data for putative transcription start sites or conserved upstream sequences are additional methods to identify promoter elements [127,138–141].

Promoter libraries have been developed and implemented in non-model organisms. As alluded to above, Jensen et al. [37] synthesized an *Lc. lactis* library using the known consensus sequences of the  $-35$  and  $-10$  regions along with variable sequence adjacent motifs. This library yielded titratable *lacZ* expression over four different orders of magnitude in this organism. Unlike *E. coli* above, in terms of *lacZ* reporter expression, *Lc. lactis* is less tolerant to changes in the  $-35$  and  $-10$  sequences. However, as was the case with *E. coli*, the spacer region and additional promoter elements also contribute to titratable *lacZ* expression in *Lc. lactis*. As was the case for *Lc. lactis* [37], to engineer a *Corynebacterium glutamicum* library Rytter et al. [33] took consensus sequences for the *C. glutamicum*  $-10$  region (TGNGNTAYAATGG), the *E. coli*  $-35$  region with and without an AT-rich region upstream, and primarily changed the spacer between these regions to titrate *lacZ* expression over three different orders of magnitude. Finally, Markley et al. [34] used mutagenesis of a truncated version of the well-characterized *Synechococcus elongatus* strain PCC 7002 promoter  $P_{\text{CpeB}}$  to

engineer a library of 29 unique members to control expression of a *yfp* (encoding yellow fluorescent protein) reporter gene. This library yielded titratable expression of *yfp* over three orders of magnitude, with a mutation changing the original  $P_{\text{CpeB}}$   $-10$  sequence (TATAAA) to the *E. coli*  $-10$  consensus being responsible for much of the increased *yfp* expression of the top performing library member. Fig. 3 provides a roadmap for engineering promoter libraries, which may be a useful guide for further extending promoter libraries in other non-model organisms.

### 2.3. Transcription terminators

In *E. coli*, transcription primarily terminates either with the assistance of Rho helicase (Rho-dependent termination) or through innate elements of the RNA transcript itself (Rho-independent termination) [142–144]. Engineering termination elements becomes especially important for many transcriptional control systems to avoid polar effects associated with transcription. Many of these polar effects may include transcription unit desegmentation (*i.e.*, circumventing transcription regulation of downstream genes) [35,145–152], increased translation of upstream genes by prolonging transcription-translation coupling along with other alterations to translation mechanics (*i.e.*, translational coupling) [91,144,150,153–156], potentially arresting downstream transcription through R-Loop formation [144,157–159], altering DNA supercoiling [160], and influencing plasmid copy number [161]. Multiple terminators present in a single cell must be designed such that homologous recombination will not delete or invert intervening sequences between these constructs, underscoring the need to have a broad understanding of a variety of different terminators. To further characterize transcriptional termination [142–144]. As a hexamer complex, Rho associates with RNA, recognizes a C-rich and G-poor RNA *rut* (Rho utilization) sequence, allows RNA to threaded through an open region in the Rho complex, and subsequently closes around the associated RNA [142–144]. Signals that are still poorly understood pause the procession of RNAP-mediated RNA polymer elongation or promote RNAP backtracking (albeit, nucleoid structure may serve a function perturbing RNAP elongation dynamics along with promoting RNAP backtracking) [162,163], and Rho, coupled to ATP hydrolysis, proceeds wrapping RNA around the hexamer complex [142,143] and catches up to this previously paused RNAP (*i.e.*, resuming transcript elongation may be a requirement for Rho-mediated termination [143,164]). Subsequently, Rho disrupts the RNA-DNA duplex inside of RNAP and transcription terminates after dissociation of the RNA-DNA duplex and rehybridization of the DNA-DNA duplex [142,143]. In contrast to Rho-dependent termination, Rho-independent termination events usually occur when an RNA stem-loop forms in the RNAP exit channel and a poly U-tract downstream of this stem loop is dehybridized from the associated DNA poly A-tract [35,142,144,165]. In particular, pausing at this poly U-tract allows the nucleation of a GC-rich stem [35,142,144]. An additional poly A-tract is sometimes present upstream of some terminator stem-loops [35,142,144], but the precise function for this sequence remains unclear (*e.g.*, this sequence may be the poly U-tract for transcription directed from the opposite strand of DNA, promote extension of the terminator stem into the poly U-tract, etc). Well known Rho-independent transcription terminators include the  $T_{\text{RimB1}}$  and  $T_{\text{RimB2}}$  terminators [166–168], the  $T_{\text{TE}}$  T7 phage terminator [167,168], and the  $\lambda$  phage  $T_{\text{O}}$  terminator [169–171]. To further characterize transcriptional termination and solve the issue of integrating multiple terminators in tandem without promoting homologous recombination, additional libraries of known and tested Rho-independent terminators have been developed (<http://parts.igem.org/Terminators/Catalog>) [35]. As was the case for



**Fig. 3. A procedure to generate promoter and terminator libraries.** (1) To test putative promoters as sites of transcription initiation, known promoters ( $P_{\text{Known}}$ ) or random sequences ( $P_{\text{Test}}$ ) may be inserted upstream of a previously promoterless ribosome binding site (RBS)-*reporter1* gene translational fusion (purple box with a green arrow). These promoters may be constitutive (shown) or subject to transcription regulation by activators and repressors (not shown). Pale green and pale purple boxes indicate the  $-35$  and  $-10$  promoter sequences, respectively. To test putative transcription terminators, known terminators ( $T_{\text{Known}}$ ) or random sequences ( $T_{\text{Test}}$ ) may be inserted upstream of a promoterless RBS-*reporter2* translational fusion (purple box with a red arrow) and downstream of the  $P_{\text{Known/Test}}$ -*reporter1* transcription fusion above. Additional known or random terminators may be inserted in tandem, and a known terminator may be inserted downstream of *reporter2* to ensure termination of transcription after RNAP reads-through this gene (not shown). Inverted white arrows represent the sequence forming a terminator stem-loop (inverted repeats), and red and grey boxes represent poly A- and poly U-tracts, respectively. (2) To introduce variability in promoter and terminator libraries (thin yellow, orange, red, grey, and white bars) these *cis*-regulatory elements may be derived from randomly sheared DNA, synthesized on demand with desired sequences, subjected to mutagenic PCR (polymerase chain reaction), or other similar methods. These *cis*-regulatory element constructs may be introduced into an organism of interest, reporter gene expression may be assessed for each library member (green and red suns), and this procedure may be iterated (starting with new random/mutated sequences or using library member sequences as inputs for further sequence variation generation) to achieve a desired result or a wider expression variability profile. (3) To identify sequence characteristics that may be responsible for reporter gene expression level differences (rows of green [gene expression from promoters] and red [gene expression from transcription read-through] suns) among library member sequences (indicated), these constructs should be sequenced and compared with other members or known promoter sequences (iterated as desired).

promoter libraries above, RNAP also has a wide range of activities toward intrinsic terminators [35,106–108].

#### 2.4. Terminator libraries

In the work by Chen et al. [35], 582 natural and synthetic Rho-independent terminators were characterized and a biophysical model of terminator strength was developed. This group used an experimental setup where a single promoter,  $P_{\text{BAD}}$  (see below), drove expression of a *gfp* reporter gene. Downstream of *gfp*, putative terminators were positioned to terminate expression 5' of a promoterless *rfp* (red fluorescent protein) reporter gene (i.e., *rfp* will only be expressed if RNAP reads through these tested terminators). Terminator strengths were determined by taking the ratio of *gfp* to *rfp* expression levels, normalized to the same ratio for a terminator-less RNA molecule, and varied over three different orders of magnitude. As was previously known for intrinsic terminators, the presence of a poly A-tract upstream (5') of a GC-rich

terminator stem, a longer terminator stem (approximately 8-bp), a stable unpaired loop region, and a poly U-tract downstream (3') of the stem were all characteristics associated with strong terminators studied by this group. Free energy changes associated with the formation of all of these features, thus, went into the design of a biophysical model for predicting terminator strength. In addition, through identifying a variety of strong terminators in their library and model, Chen et al. were able to predictably terminate transcription with reduced homologous recombination-mediated changes to these terminator constructs. Fig. 3, in addition to describing promoter library generation, describes building terminator libraries.

### 3. Transcription control through transcription factor tools

As described above, promoters can often be classified as weak or strong. However, *trans*-acting factors, referred to as transcription factors, have the potential to invert these promoter strength

metrics (*i.e.*, a weak promoter could be transformed into a strong promoter, and a strong promoter into a weak one) (Fig. 4A). Factors that transform a weak promoter into a strong one are referred to as transcription activators; the factors that transform a strong promoter into a weak one are referred to as transcription repressors. Promoters with titratable (linear/dynamic) activity toward these transcription factors are generally used for analogue-type computation (Fig. 4A and B) [25]. Promoters that strongly respond to transcription activators or repressors (often with feedforward and feedback loops), resulting in a step-change in promoter strength, are generally used for digital-type gene circuits [16,18,19,25] (Fig. 4A and C). DNA binding by many of these transcription factors is affected by environmental conditions, such as the presence of small metabolites and other ligands. In this way, small molecules can be used as switches of transcriptional control. Of particular note, in terms of digital transcription control design, Nielson et al. [23] have developed the automated system Cello (<http://www.cellocad.org/>). The Cello system uses the Verilog design language, which can be converted into DNA sequences by this system, and reliable and modular gene circuit components to automate gene circuit engineering. In the following sections, we highlight some of the more commonly used transcription factors and how they have been used to control gene expression by synthetic biologists.

### 3.1. *CI* and *Cro*

Together, *CI* and *Cro* are transcription factors that naturally function as a toggle switch and have been utilized in a variety of synthetic regulatory systems. In nature, *Cro* binds  $P_{RM}$  and represses *cl*, and *CI* binds  $P_R$  and represses *cro* [172–176]. This toggle switch controls the fate of a cell infected with bacteriophage  $\lambda$ . If the *Cro* binding state is predominate a lytic infection with this phage occurs, but if the *CI* state predominates the phage may enter into the lysogenic cycle [172–176]. The *Cro* and *CI* binding sites have the same consensus sequence (TATCACGGCGGTGATA), but each repressor has different affinities for variants of this sequence, explaining the switch like behavior [172–178]. The *Cro* and *CI* binding sites overlap the respective promoters of each repressor, which suggests that each blocks transcription by sterically interfering with RNAP holoenzyme promoter binding (potentially through DNA looping) [115,121,173–175,179]. In synthetic biology, *CI* was one of three repressors tested in early toggle switches development by Gardner et al. [16]. Gardner et al. concluded that the *CI*857 temperature-sensitive variant (functional as a repressor at 30 °C, but not at 42 °C [180]) is an extremely efficient repressor at  $P_{LS1con}$ . Kotula et al. [24] constructed a *Cro* and *CI* memory element based on the natural  $P_R$  and  $P_{RM}$  promoters. This group used a stable *CI*-ON state (resulting in *lacZ* reporter gene repression) coupled to TetR (see below) repression of *cro*. Adding aTc (anhydrotetracycline) derepressed *cro*, which in turn repressed *cl* and generated a stable *Cro*-ON state (resulting in *lacZ* reporter gene expression). Using a mouse model, where mice were treated with a strain of *E. coli* harboring this memory element and with or without aTc, Kotula et al. also demonstrated that this memory element is a useful *in vivo* reporter system. Finally, Tamsir et al. [18] used *CI*, CAP (catabolite activator protein), AraC, and TetR (see below) during their construction of all 16 two-input Boolean logic gates.

### 3.2. CAP (catabolite activator protein)

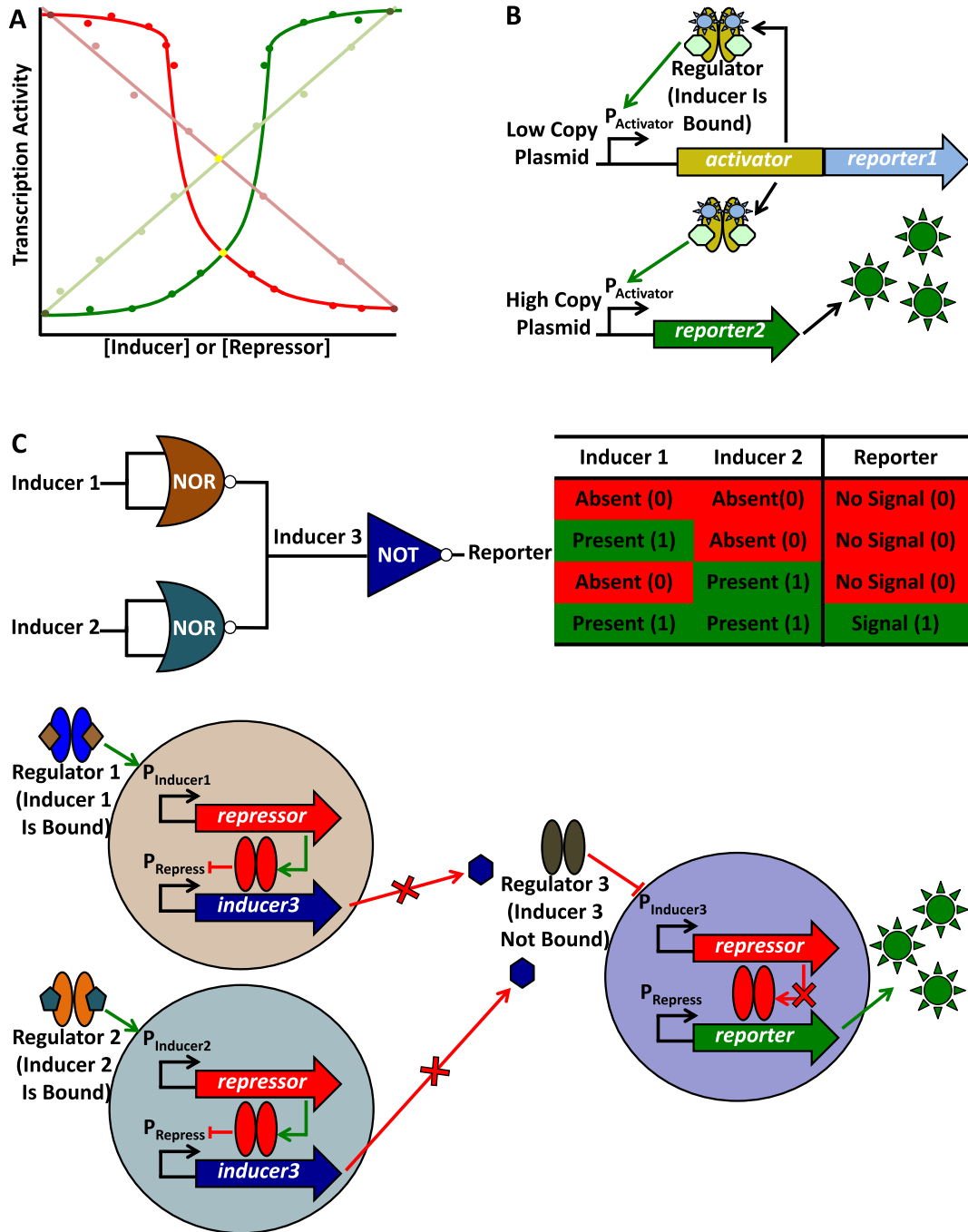
CAP serves a key regulatory function in catabolite repression and has been leveraged in synthetic systems to control gene expression. When CAP binds cAMP (cyclic AMP) it has a higher affinity for its cognate binding sites (AATGTGATCTAGATCACATTT), which are often upstream of operons related to carbohydrate

metabolism [181–184]. Indeed, the  $\alpha$ CTD subunits of the RNAP holoenzyme make contacts with CAP, which improves promoter recognition by the holoenzyme [115,181,185,186]. In terms of catabolite repression, adenylyl cyclase activity and cAMP synthesis is modulated by the availability of glucose in the extracellular milieu [184]. When glucose levels are high, adenylyl cyclase activity is low, and when glucose levels are low, adenylyl cyclase activity is high [184]. Synthetic systems based on promoters that are subject to catabolite repression (*e.g.*,  $P_{BAD}$  [found on vector pBAD and others],  $P_{LacUV5}$  expression systems, etc) are often used when strong negative regulation of genes is desired (*i.e.*, for auto-induction systems that will only produce desired proteins when glucose is depleted from the extracellular milieu) [187,188]. Daniel et al. [25], additionally, used  $P_{BAD}$  to engineer transcription systems capable of performing analogue computation. Other components of these transcription control systems include *LacI*, AraC, and LuxR (see below).

### 3.3. *LacI*

Describing *LacI*-mediated repression of the *E. coli lac* operon represents one of the earliest models of gene regulation [77,189]. Naturally, *LacI* dimers bind the operator sequence (TGAATTGTGACGGATAACAATT) in the absence of the small molecule allolactose (or isopropyl  $\beta$ -D-1-thiogalactopyranoside [IPTG]) [77,189–194]. Occupancy at these operator sequences obstructs RNAP holoenzyme binding to  $P_{Lac}$  (a weak promoter without additional CAP-mediated activation) [121,189,192,195–197]. In addition, *LacI*-mediated DNA looping (potentially assisted by nucleoid structuring factors and DNA supercoiling) may obstruct RNAP holoenzyme- $P_{Lac}$  association [115,121,179,189,195,196,198]. When present, allolactose or IPTG occupy ligand-binding pockets on *LacI*, which results in a conformational change of this regulator and reduces its affinity for operator binding sequences [77,189,191,193,194]. Numerous regulation tools (noted above and below) have been based off of *LacI*-mediated repression.

*LacI*-mediated repression is a tool used for engineering a variety of genetic circuits. To explore intrinsic and extrinsic noise in gene expression, Elowitz et al. [199] used *LacI*-mediated repression of reporter genes *yfp* and *cfp* (encoding the cyan fluorescent protein) placed nearly equidistant from the origin of replication on opposite sides of the *E. coli* chromosome. This group identified noisy reporter gene expression when *LacI* repressed these reporter genes. However, they found that varying levels of IPTG affected intrinsic and extrinsic reporter gene noise differently. Increasing IPTG levels decreased intrinsic noise, and extrinsic noise initially increased then decreased. *LacI* was used in addition to *CI* (see above) and TetR (see below) repressors, to construct toggle switches used by Gardner et al. [16]. In this work, *LacI* repressed *cl* or *tetR* expression in addition to a *gfp* reporter gene expressed from  $P_{Trc-2}$  (a strong synthetic promoter generated by fusing  $P_{Trp}$  and  $P_{Lac}$ —containing a single *LacI* operator sequence). This approximates a step-change in reporter expression, where derepression by addition of IPTG is supplemented by *cl* expression and subsequent *CI*-mediated *lacI* repression at  $P_{LS1con}$ . Earlier work has established that *LacI*-mediated regulation is subject to the all or nothing phenomenon (a step-change as opposed to a titration in gene expression) [200,201]. Indeed, suboptimal induction with IPTG leads to a heterogeneous response at the single cell level (*i.e.*, some cells express high levels of *lacZ* while others do not), but at the cell population level may appear as though the cells have uniform or expression levels similar to a fully induced system [201,202]. Nevertheless, other groups have used systems where it was possible to titrate IPTG-mediated derepression [25,200,202]. For example, titratable expression is also sometimes achieved when *lacY* (the lactose and



**Fig. 4. An overview of transcription regulation following analogue and digital expression dynamics.** (A) A hypothetical transcription activity dynamics plot is depicted. Faint green and red lines represent transcription activity (activation and repression, respectively) subject to linear/analogue dynamics. Dark green and red curves represent activation and repression subject to step-change/digital dynamics, respectively. Individual points represent one hypothetical measurement of transcription activity. Concentrations of hypothetical activator or repressor ligands increase from left to right on the horizontal axis, and transcription activity increases from the bottom to the top of the vertical axis. (B) An example of a system subject to linear/analogue dynamics. This model is modified from the work of Daniel et al. [25]. Activator (gold ovals) fused to Reporter 1 (blue suns) is produced from a gene expressed from a low copy plasmid and subject to a positive feedback (auto-activation) loop at  $P_{Activator}$  (green arrow). Activator (bound to inducer) also binds target sites on a high copy number plasmid activating *reporter2*, again at  $P_{Activator}$ , and simultaneously titrates Activator from the low copy plasmid above. Based on inducer concentration, this model allows for log-linear titration of Reporter 2 (green suns) levels [25]. (C) A digital AND logic gate (engineered from two NOR gates and a NOT gate) is symbolically represented along with the associated truth table. The depicted model is modified from the work of Tamsir et al. [18]. This circuit is distributed among three co-cultured cells (pale brown, blue, and purple circles), and the scenario depicted represents the ON output state, when Reporter (green suns) is produced (1 and a green box in the truth table). Both inducer 1 (brown diamond) and inducer 2 (blue pentagon) are present (1 and a green box in the truth table) and bound to Regulators 1 and 2, respectively. These regulators bind  $P_{Inducer1}$  and  $P_{Inducer2}$ , respectively, activating *repressor* (green arrow). Repressor is subsequently produced and represses *inducer3* (red crossbar), the gene encoding the producer of inducer 3 (purple hexagon), at  $P_{Repress}$ . Inducer 3 is not secreted (red arrow with a cross), and Regulator 3 (without inducer 3 bound) binds  $P_{Inducer3}$  repressing *repressor*. Subsequently, Repressor does not repress *reporter*, at  $P_{Repress}$ , leading to *reporter* expression and Reporter production. In the absence of either or both inducers 1 and 2 (0 and red boxes in the truth table) *inducer3* is not repressed, which leads to inducer 3 secretion. Regulator 3 (with inducer 3 bound) activates *repressor*, at  $P_{Repress}$ . Repressor subsequently represses *reporter*, blocking Reporter production (0 and a red box in the truth table).



IPTG importer) is deleted from the background of the *E. coli* host strain [200,202].

### 3.4. AraC

AraC-mediated regulation of the *ara* operon is another well-characterized system of transcriptional control. In nature, AraC is both a repressor and activator depending on L-arabinose-mediated isomerizations of this regulator [203,204]. Indeed, when L-arabinose is absent in the extracellular milieu, AraC dimers bind promoter proximal and upstream distal sites in an outstretched configuration, which bends/loops DNA and obstructs RNAP holoenzyme promoter binding [115,121,204]. AraC recognizes an AT-rich sequence (usually 6-bp long) flanked by the sequence TRTGGAC on the upstream side and GCTA on the downstream side [203]. In the presence of L-arabinose, AraC binds this metabolite and isomerizes to a collapsed configuration [204]; AraC and CAP ( $P_{Ara}$  is a weak promoter without CAP activation) occupy binding sites proximal to  $P_{Ara}$  and  $P_{AraC}$  and induce RNAP holoenzyme promoter binding and open complex formation [204]. This may be achieved by locally bending DNA such that the  $\alpha$ CTDs of the RNA holoenzyme can still contact CAP despite the presence of AraC dimers between these two elements [115,204].

AraC has been mobilized to engineer a variety of systems under the transcriptional control of this regulator. Many of these engineered systems promote both digital/step-change dynamics [18,23] and analogue/linear dynamics [25,205]. It is intriguing to note, however, that for AraC-mediated induction at subsaturating levels of L-arabinose (intermediate induction) a dichotomy exists at the single cell level [205,206]. Thus, a subset of cells in the population express a gene of interest from  $P_{BAD}$ , while the remaining cells do not express this gene of interest. This gives the false impression that all cells in the population express the gene of interest at the same intermediate level [205,206]. As was the case for LacI above, however, systems can be engineered where AraC-mediated regulation can be titratable. For example, titratable expression may be achieved when arabinose importers (e.g. *araE*) are constitutively expressed (natively, expression is linked to the presence of arabinose) from the *E. coli* chromosome [205].

### 3.5. TetR

TetR-mediated negative regulation of  $P_{Tet}$  is another model of tight *E. coli* gene repression. In its natural context, TetR binds as a dimer to the operator sequence (TCTATCATTGATAGG), which in turn obstructs RNAP holoenzyme-promoter interactions [115,207–209]. In the presence of tetracycline (or the analogue aTc) this small molecule binds a tetracycline binding pocket in each dimer, which in turn causes a conformational change in this regulator and reduces its affinity for the above operator sequence [103,207–210]. Regulation by TetR at cognate promoters is incredibly tight, with as high as a 5000-fold difference between aTc induced and uninduced expression from some promoters [103].

TetR-mediated regulation has been leveraged to engineer synthetic regulatory elements. When constructing all two-input Boolean logic gates using site-specific recombinases to invert promoters or terminators controlling reporter gene expression, Siuti et al. [19] used TetR- and aTc-mediated regulation to control one of two site-specific recombinases (the other recombinase being controlled by a small regulatory RNA and LuxR- and AHL (acyl-homoserine lactone)-mediated regulation at  $P_{Lux}$  [see below]). TetR, CAP, LacI, and AraC (discussed above) are also useful components for automated logic gate assembly by the Cello design system [23]. Kotula et al. [24], as noted above, used TetR- and aTc-mediated regulation as a trigger element to regulate *cro* expression.

Elowitz and Leibler [28] used TetR, along with LacI and CI (see above), to construct a synthetic repressilator. As the name of this construct suggests, Elowitz and Leibler described the synchronization and subsequent innate oscillatory character of reporter and regulator protein production and turnover.

### 3.6. LuxR

Although best understood for its functions during the quorum sensing phenomenon [211–214], AHL-mediated regulation of LuxR is also associated with a variety of engineered transcription control systems. In nature, LuxI synthesizes the auto-inducer AHL, which in turn binds LuxR once sufficiently high levels of AHL are present in the cellular milieu (i.e., AHL binds reversibly and weakly to LuxR at low AHL levels) [211–215]. AHL-bound LuxR dimers recognize a LuxR box (ACCTGTAGGATCGTACAGGT) upstream of the extended –10 sequence of  $P_{Lux}$  [211–217]. In the case of  $P_{Lux}$ , LuxR dimers make contact with the  $\alpha$ CTD subunits of the RNAP holoenzyme, which in turn facilitates promoter recognition (of the weak  $P_{Lux}$ ) [216]. In synthetic biology, leveraging the natural functions of AHL and LuxR in quorum sensing allowed Basu et al. [218] to engineer a system where cells could be induced (aTc-mediated derepression of  $P_{Ltet-O1}$ ) to produce AHL (sender cells) that in turn binds LuxR in receiver cells leading to a pulse of *gfp* expression (under control of  $P_{Lux}$ ). Due to limited diffusion of AHL in solid media, follow-up work [219] extended LuxR-mediated regulation to demonstrate that band patterns of fluorescent reporter production could form when undifferentiated mixtures of receiver cells (harboring a complex mixture of plasmids with varying copy numbers encoding CI, LuxR variants, LacI variants, promoters responsive to these regulators, and *gfp* or *dsred* reporters) were plated with AHL-sender cells. Thus, LuxR- and AHL-mediated regulation could be useful for engineering auto-induction systems [220,221] to simplify protein production and assist in cross-strain and cross-species communication in microbial consortia [222–226].

### 3.7. TALEs (transcription activator-like effectors)

Unlike the previous regulators above, which recognize specific DNA sequences, TALEs (transcription activator-like effectors) have a modular structure and may be naturally or artificially engineered to recognize nearly any specific DNA sequence. Thus, these regulators are especially useful tools for engineering transcription control regulation at any sequence on demand (i.e., where no known regulators are able to bind to desired target sequences or known regulators have been used elsewhere in the same construct) [227]. Although originally described as plant transcriptional activators secreted by prokaryotic plant pathogens [228–232], TALEs have been well-studied as modular DNA binding proteins. TALEs are large (>120 kDa) DNA binding proteins with N- (amino-) and C-termini containing secretion and nuclear localization signals, respectively, and a highly repetitive central helical region [228,229,231–234]. The internal helical region typically contains 15.5–19.5 near perfect repeats of 34 amino acids, which accounts for most of the size of these regulators and imparts the overwhelming majority of DNA sequence-specific binding activity [228,231–238]. Indeed, Boch et al. [235] and Moscou and Bogdanov [236] cracked the TALE DNA binding specificity code, where residues 12/13 of the 34 amino acid repeat impart this nucleotide binding specificity (H/D=C, N/G=T, N/I=A, N/S=N, N/N=R, N/—=Y, H/G=T, H/A=V, N/D=C, N/K=G, H/—=T, H/I=C, H/N=G, N/A=G, and I/G=T). Thus, shuffling different repeat regions imparts TALEs with new DNA binding sequence specificities [21,237,239–245], a testament to the TALE modularity. The region immediately N-terminal

to the central repeat region is required for initial nonspecific DNA binding [231,233,246–248]. In this initial event, the TALE associates loosely with DNA in an extended conformation and begins a search for its cognate binding sequence through 1D diffusion [246]. Upon recognition of its cognate binding sequence, the extended TALE conformation contracts, and the helical repeats tightly wrap around the major groove of this binding sequence [246]. In addition, the C-terminal domain of natural TALEs may contain regions that interact with transcription factors or may be fused to other regulatory proteins [21,228,230,232,240,243–245].

TALEs have been mobilized as tools to control gene expression in synthetic biology. For the most part, TALE tools have been implemented in eukaryotes [21,237,239,240,243–245]. However, Politz et al. [242] demonstrated that a TALE recognizing the *lacI* operator sequence could effectively block transcription initiation and elongation. Expanding on this work and to demonstrate that TALE-mediated repression in *E. coli* could be reversed with the addition of small molecule ligands, Copeland et al. [241] used TALEs with TEV protease [249,250] cleavage sequences and a *tev* gene expression construct under the control of  $P_{Tet}$  (aTc inducible). Upon TetR depression of *tev* (by the addition of aTc) and subsequent TEV-mediated TALE cleavage, a formerly TALE-repressed *gfp* or *mcherry* reporter could be subsequently expressed. This response to a small ligand (aTc) had similar dynamics to IPTG-mediated *lacI* derepression of a *mcherry* reporter gene, despite being an indirect response. In addition, by changing the binding specificity of the repeat region inside of the TALE scaffold used above, Copeland et al. demonstrated multiplexed gene repression at two *E. coli* chromosomal loci, the *lysA* and *sucA* promoters.

### 3.8. CRISPRi (clustered regularly interspaced short palindromic repeats interference)

The CRISPR (clustered regularly interspaced short palindromic repeats) system is believed to have evolved as a defensive system against foreign DNA [251–255]. Spacers intervening between inverted repeats (form stem-loops in RNA transcripts) contain nucleic acid sequences of bacteriophages and other mobile genetic elements [251–255]. In the case of Type II CRISPR systems, the spacer locus is transcribed as a single RNA transcript, a tracrRNA binds adjacent to the spacers (in the stem-loop), and RNase III-mediated processing generates mature guide crRNAs (containing the spacer nucleic acid sequences above) and tracrRNAs [256,257]. These RNAs are subsequently loaded into Cas9 forming a mature endonuclease complex that will cleave DNAs in a sequence specific manner (directed by the guide crRNAs above) [257–263]. An additional sequence motif, referred to as the PAM (protospacer adjacent motif), also serves functions during target sequence recognition and subsequent DNA-RNA hybridization [257–263]. In the case of the commonly used Cas9 variant from *Streptococcus pyogenes*, NGG (the reverse complement of which occurs upstream of the DNA sequence hybridized to the RNA spacer sequence) is the PAM utilized by this enzyme [261,262]. Although the catalytically active Cas9 has been extensively used for recombination-mediated genetic engineering [261,262,264–266], catalytically inactive Cas9 (dCas9) also has important regulatory functions [6–8,267–273].

Modifications to the CRISPR system promote sequence-specific transcriptional regulation, CRISPRi (CRISPR interference). Indeed, CRISPR spacer sequences and dCas9 can be directed to nearly any DNA sequence, a similar phenomenon to what was described for TALEs above. Thus, as was the case for TALEs, CRISPRi is an especially useful tool for targeting regulatory proteins (e.g., dCas9) to any DNA sequence on demand [227]. For example, Qi et al. [268] used a dCas9 variant to block transcription initiation and elongation using sgRNAs (single guide RNAs) to various targets in the

promoters and coding sequences of variant *rfp* and *gfp* reporter genes. Blocking transcription elongation was the most efficient method to decrease expression of the tested reporter genes. To demonstrate that CRISPRi systems could be used to probe a complex regulatory network, Qi et al. also used sgRNAs directed against genes in the *lac* operon and involved in catabolite repression (see above) to knockdown expression of tested genes. Bikard et al. [269] used a dCas9 variant to block transcription initiation and elongation by targeting the *gfp* reporter open reading frame, the reporter *gfp* promoter, and the *gfp* reporter ribosome binding site using crRNA guides directed against these sequences. This same group also fused the  $\omega$  subunit of the RNAP core to dCas9 and demonstrated that transcription activation could be promoted by crRNA guides directed at sequences 43 to 59 base pairs upstream of the –35 sequence of a weak promoter. To successfully minimize acetate production in *E. coli*, Fernandez-Rodríguez et al. [14] used dCas9, a series of light-responsive two-component systems (see below), and sgRNAs directed against *poxB*, *pta*, and *ackA* (all involved in acetate production [274]). This demonstrates that dCas9 is also useful for controlling transcription of genes that encode factors controlling flux through a metabolic pathway. Through fusing activators and repressors to dCas9, others have also shown (for the most part in eukaryotes) that these hybrid proteins can mediate transcription activation and repression [270–273].

In non-model organisms, CRISPRi can also be used to regulate gene expression. Indeed, Yao et al. [6] used CRISPRi to first regulate *gfp* reporter gene expression in the cyanobacterium *Synechocystis* sp. PCC 6803. This group then used CRISPRi to repress expression of *phaE* (involved in polyhydroxybutyrate synthesis [275]), *glgC* (involved in generating glycogen storage molecules [276]), and a variety of aldehyde dehydrogenases and reductases (involved in depleting aldehyde pools available for alkane production [277]). Gordon et al. [8] used CRISPRi in *S. elongatus* strain PCC 7002 to tune expression of the *cpc* operon (involved in phycocyanin antennae synthesis) by using sgRNAs directed against *cpcB*. They also used sgRNAs directed against *glnA* (depletes  $\alpha$ -ketoglutarate pools by converting this metabolite to glutamine, which promotes ammonium [preferred nitrogen source of cyanobacteria] assimilation and is involved in signaling carbon/nitrogen balance [278]) to moderately increase  $\alpha$ -ketoglutarate levels, activate NtcA (a catabolic gene regulator activated by reduced ammonium assimilation and the modest increase in  $\alpha$ -ketoglutarate levels [278–280]), and subsequently increase lactate levels (driven by NtcA-mediated enhanced flux to pyruvate from fixed carbon [279]). In the same pursuit, to redirect carbon flow toward succinate accumulation, Huang et al. [7] targeted *glgC*, *sdhA*, and *sdhB* (blocks production of fumarate from succinate [281,282]) with sgRNAs directed against these genes. Thus, CRISPRi is a useful technology for rationally routing metabolites for the purposes of metabolic engineering.

### 3.9. Two-component regulatory systems

Unlike the one-component systems described above, two-component systems perturb gene expression in response to signals that are often located outside of the cytoplasm. In general, a sensor kinase binds a ligand (or in some cases low ligand concentration shifts the receptor-ligand complex equilibrium toward releasing a ligand). In turn, this ligand binding (or disengagement) stimulates autokinase activity (often at specific histidine residues) of the sensor kinase [64–66]. The specific response regulator for the activated sensor kinase (crosstalk is generally avoided in prokaryotic systems [64,283–285]) is subsequently phosphorylated (usually at specific aspartate residues), undergoes conformational changes, and subsequently mediates differential gene regulation (direct or indirect activation or repression) at regulator-specific

promoters [64–66].

Two-component systems have been used for regulation of synthetic transcription systems. Anderson et al. [22] used the PhoPQ system (PhoQ kinase is inactive under high  $Mg^{2+}$  concentrations [64,65]), the cognate  $P_{MgR}$ ,  $P_{Lux}$  and  $P_{Sal}$  (positively regulated by NahR in the presence of salicylate [286]), and *gfp* and *inv* (Invasin) [287] reporters to develop an AND digital logic gate. Intriguingly, this logic gate can be used to program bacterial cell invasion of mammalian cells of interest. Using a chimeric receptor (composed of a Cph1 photoreceptor [288–290] and the EnvZ histidine kinase domain [291]), Levskaya et al. [292] developed a sensor that would be inactivated in the presence of red light (i.e., OmpR would not be phosphorylated by EnvZ [291,293] to drive expression of a *lacZ* reporter from  $P_{OmpC}$ ). In an additional impressive feat, this same chimeric light receptor above, LuxR, LuxI, CI, and a *lacZ* reporter were used by Tabor et al. [294] to build a system that could recognize the edge of projected shapes. Recently, to engineer *E. coli* capable of distinguishing between red, green, and blue light, Fernandez-Rodriguez et al. [14] used the above chimeric receptor (switched on by infrared light and off by red light, switching the OmpR response regulator off or on, respectively); the CcaS receptor (switched on by green light and off by far-red light, switching the CcaR response regulator on or off, respectively [295,296]); the YF1 receptor (switched off by blue light, which switches off response regulator FixJ [297]); CI and PhlF regulators; a T7 RNAP core enzyme; supplemental sigma factors for this T7 RNAP core; promoters recognized by each regulator or T7 RNAP holoenzyme; and *bFMO*, *lacZ*, and *gusA* reporter genes. These complex circuits allowed this group to replicate several colored images projected onto *E. coli* cultures harboring these constructs.

#### 4. Conclusion

In this mini-review we have discussed several features of the natural biology of transcriptional regulation and several systems that have been used to control transcriptional regulation in model bacteria. Engineering control at this level is especially valuable for titrating protein (and therefore activity) levels, which leads to titratability of metabolites of interest, and engineering a plethora of gene circuits. Indeed, as outlined by the central dogma of molecular biology, transcriptional regulation represents the first level that the flow of genetic information from nucleic acid polymers (DNA and RNA) to amino acid polymers (proteins) can be regulated. Transcriptional regulation, as discussed in detail above, is mediated by controlling promoter strength, positioning and strength of terminators, and by transcriptional regulators. These methods of engineering transcription control can be readily extrapolated to controlling metabolic flux and control engineering in non-model organisms and eukaryotes. Nevertheless, there are still additional layers of post-transcriptional and post-translational control not covered by this mini-review. Furthermore, translation is reduced once transcription of an mRNA ceases. However, this phenomenon and the global effects of nucleoid structure and DNA supercoiling on gene transcription and translation (presumably at the apex of genetic information flow) remain poorly understood. Thus, along with combining the well-characterized regulatory systems discussed above in novel and interesting ways (potentially for practical applications or improved predictability of engineered constructs), exploring the use of novel regulatory systems and expanding control engineering to non-model organisms and eukaryotes remain areas of active research.

#### Acknowledgements

MDE was supported by an NHGRI training grant to the Genomic

Sciences Training Program 5T32-HG002760. BFP was supported by a grant from the National Science Foundation (EFRI-1240268).

#### References

- [1] Crick F. Central dogma of molecular biology. *Nature* 1970;227:561–3.
- [2] Cech TR. Structural biology. The ribosome is a ribozyme. *Science* 2000;289:878–9.
- [3] Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 2000;289:905–20.
- [4] Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science* 2000;289:920–30.
- [5] Muth GW, Ortleva-Donnelly L, Strobel SA. A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science* 2000;289:947–50.
- [6] Yao L, Cengic I, Anfelt J, Hudson EP. Multiple gene repression in cyanobacteria using CRISPRi. *ACS Synth Biol* 2016;5:207–12.
- [7] Huang CH, Shen CR, Li H, Sung LY, Wu MY, Hu YC. CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium *S. elongatus* PCC 7942. *Microb Cell Fact* 2016;15:196.
- [8] Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, Pfeleger BF. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metab Eng* 2016;38:170–9.
- [9] Stevens JT, Carothers JM. Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synth Biol* 2015;4:107–15.
- [10] Farmer WR, Liao JC. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat Biotechnol* 2000;18:533–7.
- [11] Zhang F, Carothers JM, Keasling JD. Design of a dynamic sensor–regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol* 2012;30:354–9.
- [12] Alper H, Fischer C, Nevoigt E, Stephanopoulos G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U. S. A* 2005;102:12678–83.
- [13] Fung E, Wong WW, Suen JK, Bulter T, Lee SG, Liao JC. A synthetic gene-metabolic oscillator. *Nature* 2005;435:118–22.
- [14] Fernandez-Rodriguez J, Moser F, Song M, Voigt CA. Engineering RGB color vision into *Escherichia coli*. *Nat Chem Biol* 2017;13:706–8.
- [15] Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, et al. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab Eng* 2008;10:305–11.
- [16] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 2000;403:339–42.
- [17] Purcell O, Lu TK. Synthetic analog and digital circuits for cellular computation and memory. *Curr Opin Biotechnol* 2014;29:146–55.
- [18] Tamsir A, Tabor JJ, Voigt CA. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 2011;469:212–5.
- [19] Siuti P, Yazbek J, Lu TK. Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol* 2013;31:448–52.
- [20] Atkinson MR, Savageau MA, Myers JT, Ninfa AJ. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 2003;113:597–607.
- [21] Gaber R, Lebar T, Majerle A, Šter B, Dobnikar A, Benčina M, et al. Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat Chem Biol* 2014;10:203–8.
- [22] Anderson JC, Voigt CA, Arkin AP. Environmental signal integration by a modular AND gate. *Mol Syst Biol* 2007;3:133.
- [23] Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. *Science* 2016;352. aac7341.
- [24] Kotula JW, Kerns SJ, Shakel LA, Siraj L, Collins JJ, Way JC, et al. Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Natl Acad Sci U. S. A* 2014;111:4838–43.
- [25] Daniel R, Rubens JR, Sarpeshkar R, Lu TK. Synthetic analog computation in living cells. *Nature* 2013;497:619–23.
- [26] Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. Synthetic gene networks that count. *Science* 2009;324:1199–202.
- [27] Farzadfard F, Lu TK. Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* 2014;346:1256272.
- [28] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 2000;403:335–8.
- [29] Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai QA, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 2013;10:354–60.
- [30] Mutalik VK, Guimaraes JC, Cambray G, Mai QA, Christoffersen MJ, Martin L, et al. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat Methods* 2013;10:347–53.
- [31] Hammer K, Mijakovic I, Jensen PR. Synthetic promoter libraries—tuning of gene expression. *Trends Biotechnol* 2006;24:53–5.
- [32] Jensen PR, Hammer K. Artificial promoters for metabolic optimization. *Biotechnol Bioeng* 1998;58:191–5.
- [33] Rytter JV, Helmark S, Chen J, Lezyk MJ, Solem C, Jensen PR. Synthetic



- promoter libraries for *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 2014;98:2617–23.
- [34] Markley AL, Begemann MB, Clarke RE, Gordon GC, Pfeleger BF. Synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechococcus* sp. strain PCC 7002. ACS Synth Biol 2015;4:595–603.
- [35] Chen YJ, Liu P, Nielsen AA, Brophy JA, Clancy K, Peterson T, et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. Nat Methods 2013;10:659–64.
- [36] Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat Methods 2006;3:623–8.
- [37] Jensen PR, Hammer K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 1998;64:82–7.
- [38] Huang HH, Lindblad P. Wide-dynamic-range promoters engineered for cyanobacteria. J Biol Eng 2013;7:10.
- [39] Camsund D, Heidorn T, Lindblad P. Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium. J Biol Eng 2014;8:4.
- [40] Zucca S, Pasotti L, Politi N, Casanova M, Mazzini G, Cusella De Angelis MG, et al. Multi-faceted characterization of a novel LuxR-repressible promoter library for *Escherichia coli*. PLoS One 2015;10:e0126264.
- [41] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.
- [42] Green AA, Silver PA, Collins JJ, Yin P. Toehold switches: de-novo-designed regulators of gene expression. Cell 2014;159:925–39.
- [43] Winkler WC, Breaker RR. Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 2005;59:487–517.
- [44] Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, et al. Tandem riboswitch architectures exhibit complex gene control functions. Science 2006;314:300–4.
- [45] Perriman R, Ares Jr M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. RNA 1998;4:1047–54.
- [46] Belfort M, Pedersen-Lane J, West D, Ehrenman K, Maley G, Chu F, et al. Processing of the intron-containing thymidylate synthase (*td*) gene of phage T4 is at the RNA level. Cell 1985;41:375–82.
- [47] Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell 1982;31:147–57.
- [48] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009;27:946–50.
- [49] Espah Borujeni A, Salis HM. Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. J Am Chem Soc 2016;138:7016–23.
- [50] Alba BM, Leeds JA, Onufryk C, Lu CZ, Gross CA, DegS and YaeL participate sequentially in the cleavage of RseA to activate the  $\sigma^E$ -dependent extracytoplasmic stress response. Genes Dev 2002;16:2156–68.
- [51] Kanehara K, Ito K, Akiyama Y. YaeL (EcfE) activates the  $\sigma^E$  pathway of stress response through a site-2 cleavage of anti- $\sigma^E$ , RseA. Genes Dev 2002;16:2147–55.
- [52] Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell 2000;100:391–8.
- [53] Stragier P, Losick R. Molecular genetics of sporulation in *Bacillus subtilis*. Annu Rev Genet 1996;30:297–341.
- [54] LaBell TL, Trempey JE, Haldenwang WG. Sporulation-specific  $\sigma$  factor  $\sigma^{29}$  of *Bacillus subtilis* is synthesized from a precursor protein, P<sup>31</sup>. Proc Natl Acad Sci U. S. A 1987;84:1784–8.
- [55] Christensen SK, Maenhaut-Michel G, Mine N, Gottesman S, Gerdes K, Van Melderen L. Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the yefM-yoeB toxin-antitoxin system. Mol Microbiol 2004;51:1705–17.
- [56] Gottesman S. Regulation by proteolysis: developmental switches. Curr Opin Microbiol 1999;2:142–7.
- [57] Nothaft H, Szymanski CM. Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol 2010;8:765–78.
- [58] Mescher MF, Strominger JL. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarum*. J Biol Chem 1976;251:2005–14.
- [59] Sleytr UB. Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. Nature 1975;257:400–2.
- [60] Sleytr UB, Thorne KJ. Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. J Bacteriol 1976;126:377–83.
- [61] Zhang YM, Frank MW, Virga KG, Lee RE, Rock CO, Jackowski S. Acyl carrier protein is a cellular target for the antibacterial action of the pantothenamide class of pantothenate antimetabolites. J Biol Chem 2004;279:50969–75.
- [62] Clifton G, Bryant SR, Skinner CG. N<sup>1</sup>-(substituted) pantothenamides, antimetabolites of pantothenic acid. Arch Biochem Biophys 1970;137:523–8.
- [63] Magnuson K, Jackowski S, Rock CO, Cronan Jr JE. Regulation of fatty acid biosynthesis in *Escherichia coli*. Microbiol Rev 1993;57:522–42.
- [64] Mitrophanov AY, Groisman EA. Signal integration in bacterial two-component regulatory systems. Genes Dev 2008;22:2601–11.
- [65] Groisman EA. The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol 2001;183:1835–42.
- [66] Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annu Rev Biochem 2000;69:183–215.
- [67] Adcock JB, Ninfa AJ, Stock AM. Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol Rev 1989;53:450–90.
- [68] Garnak M, Reeves HC. Phosphorylation of isocitrate dehydrogenase of *Escherichia coli*. Science 1979;203(4385):1111–2.
- [69] Klotz IM, Langerman NR, Darnall DW. Quaternary structure of proteins. Annu Rev Biochem 1970;39:25–62.
- [70] Murakami KS, Darst SA. Bacterial RNA polymerases: the whole story. Curr Opin Struct Biol 2003;13:31–9.
- [71] Saeclyev RM, Record Jr MT, deHaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol 2011;412:754–71.
- [72] Haugen SP, Ross W, Gourse RL. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol 2008;6:507–19.
- [73] Ebright RH. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol 2000;304:687–98.
- [74] Vassylyev DG, Sekine S, Laptchenko O, Lee J, Vassylyeva MN, Borukhov S, et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. Nature 2002;417:712–9.
- [75] Farrell CM, Grossman AD, Sauer RT. Cytoplasmic degradation of ssrA-tagged proteins. Mol Microbiol 2005;57(6):1750–61.
- [76] Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem 2013;82:323–55.
- [77] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 1961;3:318–56.
- [78] Price MN, Arkin AP, Alm EJ. The life-cycle of operons. PLoS Genet 2006;2:e96.
- [79] Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 2001;11:356–72.
- [80] Ermolaeva MD, White O, Salzberg SL. Prediction of operons in microbial genomes. Nucleic Acids Res 2001;29:1216–21.
- [81] Price MN, Huang KH, Alm EJ, Arkin AP. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucleic Acids Res 2005;33:880–92.
- [82] 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 2003;13:216–23.
- [83] Emory SA, Bouvet P, Belasco JG. A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. Genes Dev 1992;6:135–48.
- [84] Watters KE, Abbott TR, Lucks JB. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res 2016;44:e12.
- [85] Pfeleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 2006;24:1027–32.
- [86] Hansen MM, Ventosa Rosquelles M, Yelleswarapu M, Maas RJ, van Vugt-Jonker AJ, Heus HA, et al. Protein synthesis in coupled and uncoupled cell-free prokaryotic gene expression systems. ACS Synth Biol 2016;5:1433–40.
- [87] lost I, Dreyfus M. The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. EMBO J 1995;14:3252–61.
- [88] Lewicki BT, Margus T, Remme J, Nierhaus KH. Coupling of rRNA transcription and ribosomal assembly in vivo. Formation of active ribosomal subunits in *Escherichia coli* requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. J Mol Biol 1993;231:581–93.
- [89] Massé E, Drolet M. *Escherichia coli* DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. J Biol Chem 1999;274:16659–64.
- [90] Iskakova MB, Szaflarski W, Dreyfus M, Remme J, Nierhaus KH. Troubleshooting coupled in vitro transcription-translation system derived from *Escherichia coli* cells: synthesis of high-yield fully active proteins. Nucleic Acids Res 2006;34:e135.
- [91] Lim HN, Lee Y, Hussein R. Fundamental relationship between operon organization and gene expression. Proc Natl Acad Sci U. S. A 2011;108:10626–31.
- [92] Wolfe AJ. The acetate switch. Microbiol Mol Biol Rev 2005;69:12–50.
- [93] Mogil LS, Becker NA, Maher 3rd LJ. Supercoiling effects on short-range DNA looping in *E. coli*. PLoS One 2016;11:e0165306.
- [94] Wang Q, Wu J, Friedberg D, Plakto J, Calvo JM. Regulation of the *Escherichia coli trp* gene. J Bacteriol 1994;176:1831–9.
- [95] Travers A, Muskhelishvili G. DNA supercoiling - a global transcriptional regulator for enterobacterial growth? Nat Rev Microbiol 2005;3:157–69.
- [96] Kelly A, Goldberg MD, Carroll RK, Danino V, Hinton JC, Dorman CJ. A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar Typhimurium. Microbiology 2004;150:2037–53.
- [97] Hommais F, Krin E, Laurent-Winter C, Soutourina O, Malpertuy A, Le Caer JP, et al. Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. Mol Microbiol 2001;40(1):20–36.
- [98] Higgins CF, Dorman CJ, Stirling DA, Waddell L, Booth IR, May G, et al. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. Cell 1988;52:569–84.
- [99] Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the



- three domains of life. *Nat Rev Microbiol* 2011;9:85–98.
- [100] Werner F. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol* 2008;16:247–50.
- [101] Deuschle U, Kammerer W, Gentz R, Bujard H. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J* 1986;5:2987–94.
- [102] Mulligan ME, Hawley DK, Enriken R, McClure WR. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. *Nucleic Acids Res* 1984;12:789–800.
- [103] Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-12 regulatory elements. *Nucleic Acids Res* 1997;25(6):1203–10.
- [104] Ross W, Aiyar SE, Salomon J, Gourse RL. *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *J Bacteriol* 1998;180:5375–83.
- [105] de Boer HA, Comstock LJ, Vasser M. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci U. S. A* 1983;80:21–5.
- [106] d'Aubenton Carafa Y, Brody E, Thermes C. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J Mol Biol* 1990;216:835–58.
- [107] Yager TD, von Hippel PH. A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry* 1991;30:1097–118.
- [108] von Hippel PH, Yager TD. Transcript elongation and termination are competitive kinetic processes. *Proc Natl Acad Sci U. S. A* 1991;88:2307–11.
- [109] Paget MS, Helmann JD. The  $\sigma^{70}$  family of sigma factors. *Genome Biol* 2003;4:203.
- [110] Feklistov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* 2014;68:357–76.
- [111] Österberg S, del Peso-Santos T, Shingler V. Regulation of alternative sigma factor use. *Annu Rev Microbiol* 2011;65:37–55.
- [112] Kazmierczak MJ, Wiedmann M, Boor KJ. Alternative sigma factors and their roles in bacterial virulence. *Microbiol Mol Biol Rev* 2005;69:527–43.
- [113] Errington J. *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. *Microbiol Rev* 1993;57:1–33.
- [114] Ruff EF, Record Jr MT, Artsimovitch I. Initial events in bacterial transcription initiation. *Biomolecules* 2015;5:1035–62.
- [115] Browning DF, Busby SJ. The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2004;2:57–65.
- [116] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc Natl Acad Sci U. S. A* 2017;114:E3032–40.
- [117] Ruff EF, Drennan AC, Capp MW, Poulos MA, Artsimovitch I, Record Jr MT. *E. coli* RNA polymerase determinants of open complex lifetime and structure. *J Mol Biol* 2015;427:2435–50.
- [118] Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. rRNA promoter regulation by nonoptimal binding of  $\sigma$  region 1.2: an additional recognition element for RNA polymerase. *Cell* 2006;125:1069–82.
- [119] Winkelman JT, Chandrangu P, Ross W, Gourse RL. Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. *Proc Natl Acad Sci U. S. A* 2016;113:E1787–95.
- [120] Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* 2015;4.
- [121] Browning DF, Busby SJ. Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol* 2016;14(10):638–50.
- [122] Hawley DK, McClure WR. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* 1983;11:2237–55.
- [123] Harley CB, Reynolds RP. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* 1987;15:2343–61.
- [124] Aiyar SE, Gourse RL, Ross W. Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase  $\alpha$  subunit. *Proc Natl Acad Sci U. S. A* 1998;95:14652–7.
- [125] Sivaraman K, Seshasayee AS, Swaminathan K, Muthukumar G, Pennathur G. Promoter addresses: revelations from oligonucleotide profiling applied to the *Escherichia coli* genome. *Theor Biol Med Model* 2005;2:20.
- [126] Feklistov A, Bae B, Hauver J, Lass-Napiorkowska A, Kalesse M, Glaus F, et al. RNA polymerase motions during promoter melting. *Science* 2017;356:863–6.
- [127] Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, et al. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One* 2009;4:e7526.
- [128] Heyduk E, Kuznedelov K, Severinov K, Heyduk T. A consensus adenine at position -11 of the nontemplate strand of bacterial promoter is important for nucleation of promoter melting. *J Biol Chem* 2006;281:12362–9.
- [129] Revyakina A, Liu C, Ebricht RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 2006;314:1139–43.
- [130] Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebricht RH. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 2006;314:1144–7.
- [131] De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol* 2007;7:34.
- [132] Kawano M, Storz G, Rao BS, Rosner JL, Martin RG. Detection of low-level promoter activity within open reading frame sequences of *Escherichia coli*. *Nucleic Acids Res* 2005;33:6268–76.
- [133] Bron PA, Hoffer SM, Van Il S, De Vos WM, Kleerebezem M. Selection and characterization of conditionally active promoters in *Lactobacillus plantarum*, using alanine racemase as a promoter probe. *Appl Environ Microbiol* 2004;70:310–7.
- [134] Stout V. Identification of the promoter region for the colanic acid polysaccharide biosynthetic genes in *Escherichia coli* K-12. *J Bacteriol* 1996;178:4273–80.
- [135] Slauch JM, Camilli A. IVET and RIVET: use of gene fusions to identify bacterial virulence factors specifically induced in host tissues. *Methods Enzymol* 2000;326:73–96.
- [136] Mahan MJ, Slauch JM, Mekalanos JJ. Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* 1993;259:686–8.
- [137] Valdivia RH, Falkow S. Fluorescence-based isolation of bacterial genes expressed within host cells. *Science* 1997;277:2007–11.
- [138] Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, et al. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J Bacteriol* 2015;197:18–28.
- [139] Yan B, Núñez C, Ueki T, Esteve-Núñez A, Puljic M, Adkins RM, et al. Computational prediction of RpoS and RpoD regulatory sites in *Geobacter sulfurreducens* using sequence and gene expression information. *Gene* 2006;384:73–95.
- [140] Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R. Genome-wide analysis of the general stress response network in *Escherichia coli*:  $\sigma^S$ -dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* 2005;187:1591–603.
- [141] Kröger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK, Hokamp K, et al. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U. S. A* 2012;109:E1277–86.
- [142] Ray-Soni A, Bellecourt MJ, Landick R. Mechanisms of bacterial transcription termination: all good things must end. *Annu Rev Biochem* 2016;85:319–47.
- [143] Banerjee S, Chalissery J, Bandey I, Sen R. Rho-dependent transcription termination: more questions than answers. *J Microbiol* 2006;44:11–22.
- [144] Peters JM, Vangeloff AD, Landick R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol* 2011;412:793–813.
- [145] Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, et al. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* 2014;5:e01442-14.
- [146] Epshtein V, Toulmé F, Rahmouni AR, Borukhov S, Nudler E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J* 2003;22:4719–27.
- [147] Lee F, Yanofsky C. Transcription termination at the *trp* operon attenuators of *Escherichia coli* and *Salmonella typhimurium*: RNA secondary structure and regulation of termination. *Proc Natl Acad Sci U. S. A* 1977;74:4365–9.
- [148] Gottesman ME, Adhya S, Das A. Transcription antitermination by bacteriophage lambda N gene product. *J Mol Biol* 1980;140:57–75.
- [149] Yang XJ, Hart CM, Grayhack EJ, Roberts JW. Transcription antitermination by phage  $\lambda$  gene Q protein requires a DNA segment spanning the RNA start site. *Genes Dev* 1987;1:217–26.
- [150] Olson ER, Flamm EL, Friedman DI. Analysis of *nutR*: a region of phage lambda required for antitermination of transcription. *Cell* 1982;31:61–70.
- [151] Grayhack EJ, Roberts JW. The phage  $\lambda$  Q gene product: activity of a transcription antiterminator *in vitro*. *Cell* 1982;30:637–48.
- [152] Friedman DI, Court DL. Transcription antitermination: the  $\lambda$  paradigm updated. *Mol Microbiol* 1995;18:191–200.
- [153] Turnbough Jr CL, Switzer RL. Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* 2008;72:266–300.
- [154] Richardson JP, Grimley C, Lowery C. Transcription termination factor rho activity is altered in *Escherichia coli* with *suA* gene mutations. *Proc Natl Acad Sci U. S. A* 1975;72:1725–8.
- [155] Platt T. Transcription termination and the regulation of gene expression. *Annu Rev Biochem* 1986;55:339–72.
- [156] Oppenheim DS, Yanofsky C. Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* 1980;95:785–95.
- [157] Harinarayanan R, Gowrishankar J. Host factor titration by chromosomal R-loops as a mechanism for runaway plasmid replication in transcription termination-defective mutants of *Escherichia coli*. *J Mol Biol* 2003;332:31–46.
- [158] Leela JK, Syeda AH, Anupama K, Gowrishankar J. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc Natl Acad Sci U. S. A* 2013;110:258–63.
- [159] Aguilera A, García-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 2012;46:115–24.
- [160] Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U. S. A* 1987;84:7024–7.
- [161] Stueber D, Bujard H. Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO J* 1982;1:1399–404.
- [162] Kotlajich MV, Hron DR, Boudreau BA, Sun Z, Lyubchenko YL, Landick R.

- Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife* 2015;4.
- [163] Landick R. Transcriptional pausing without backtracking. *Proc Natl Acad Sci U. S. A* 2009;106:8797–8.
- [164] Dutta D, Chalissery J, Sen R. Transcription termination factor rho prefers catalytically active elongation complexes for releasing RNA. *J Biol Chem* 2008;283:20243–51.
- [165] Hein PP, Kolb KE, Windgassen T, Bellecourt MJ, Darst SA, Mooney RA, et al. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* 2014;21:794–802.
- [166] Orosz A, Boros I, Venetianer P. Analysis of the complex transcription termination region of the *Escherichia coli rrmB* gene. *Eur J Biochem* 1991;201:653–9.
- [167] Uptain SM, Chamberlin MJ. *Escherichia coli* RNA polymerase terminates transcription efficiently at rho-independent terminators on single-stranded DNA templates. *Proc Natl Acad Sci U. S. A* 1997;94:13548–53.
- [168] Telesnitsky AP, Chamberlin MJ. Sequences linked to prokaryotic promoters can affect the efficiency of downstream termination sites. *J Mol Biol* 1989;205:315–30.
- [169] Scholtissek S, Grosse F. A cloning cartridge of  $\lambda$  t(o) terminator. *Nucleic Acids Res* 1987;15:3185.
- [170] McDowell JC, Roberts JW, Jin DJ, Gross C. Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. *Science* 1994;266:822–5.
- [171] Hajimorad M, Gray PR, Keasling JD. A framework and model system to investigate linear system behavior in *Escherichia coli*. *J Biol Eng* 2011;5:3.
- [172] Eisen H, Brachet P, Pereira da Silva L, Jacob F. Regulation of repressor expression in  $\lambda$ . *Proc Natl Acad Sci U. S. A* 1970;66:855–62.
- [173] Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. Switches in bacteriophage lambda development. *Annu Rev Genet* 2005;39:409–29.
- [174] Schubert RA, Dodd IB, Egan JB, Shearwin KE. Cro's role in the CI Cro bistable switch is critical for  $\lambda$ 's transition from lysogeny to lytic development. *Genes Dev* 2007;21:2461–72.
- [175] Johnson A, Meyer BJ, Ptashne M. Mechanism of action of the cro protein of bacteriophage  $\lambda$ . *Proc Natl Acad Sci U. S. A* 1978;75:1783–7.
- [176] Herskowitz I, Hagen D. The lysis-lysogeny decision of phage  $\lambda$ : explicit programming and responsiveness. *Annu Rev Genet* 1980;14:399–445.
- [177] Benson N, Youderian P. Phage  $\lambda$  Cro protein and cI repressor use two different patterns of specific protein-DNA interactions to achieve sequence specificity *in vivo*. *Genetics* 1989;121:5–12.
- [178] Humayun Z, Jeffrey A, Ptashne M. Completed DNA sequences and organization of repressor-binding sites in the operators of phage lambda. *J Mol Biol* 1977;112:265–77.
- [179] Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* 2005;15:136–44.
- [180] Lieb M. Studies of heat-inducible lambda bacteriophage. I. Order of genetic sites and properties of mutant prophages. *J Mol Biol* 1966;16:149–63.
- [181] Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebricht RH. Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol* 2004;14:10–20.
- [182] Ebricht RH, Ebricht YW, Gunasekera A. Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E. coli lac* DNA site. *Nucleic Acids Res* 1989;17:10295–305.
- [183] Zheng D, Constantinidou C, Hobman JL, Minchin SD. Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling. *Nucleic Acids Res* 2004;32:5874–93.
- [184] Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 2008;6:613–24.
- [185] Igarashi K, Ishihama A. Bipartite functional map of the *E. coli* RNA polymerase  $\alpha$  subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. *Cell* 1991;65:1015–22.
- [186] Niu W, Kim Y, Tau G, Heyduk T, Ebricht RH. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell* 1996;87:1123–34.
- [187] Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter. *J Bacteriol* 1995;177:4121–30.
- [188] Blommel PG, Becker KJ, Duvnjak P, Fox BG. Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol Prog* 2007;23:585–98.
- [189] Lewis M. The lac repressor. *C R Biol* 2005;328:521–48.
- [190] Gilbert W, Maxam A. The nucleotide sequence of the lac operator. *Proc Natl Acad Sci U. S. A* 1973;70:3581–4.
- [191] Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 1996;271:1247–54.
- [192] Dickson RC, Abelson J, Barnes WM, Reznikoff WS. Genetic regulation: the Lac control region. *Science* 1975;187:27–35.
- [193] Daber R, Staybrook S, Rosenberg A, Lewis M. Structural analysis of Lac repressor bound to allosteric effectors. *J Mol Biol* 2007;370:609–19.
- [194] Jobe A, Bourgeois S. Lac Repressor-operator interaction. VI. The natural inducer of the lac operon. *J Mol Biol* 1972;69:397–408.
- [195] Oehler S, Eismann ER, Krämer H, Müller-Hill B. The three operators of the lac operon cooperate in repression. *EMBO J* 1990;9:973–9.
- [196] Swigon D, Coleman BD, Olson WK. Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. *Proc Natl Acad Sci U. S. A* 2006;103:9879–84.
- [197] Schlax PJ, Capp MW, Record Jr MT. Inhibition of transcription initiation by lac repressor. *J Mol Biol* 1995;245:331–50.
- [198] Becker NA, Kahn JD, Maher 3rd LJ. Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol* 2005;349:716–30.
- [199] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science* 2002;297:1183–6.
- [200] Jensen PR, Westerhoff HV, Michelsen O. The use of lac-type promoters in control analysis. *Eur J Biochem* 1993;211:181–91.
- [201] Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci U. S. A* 1957;43:553–66.
- [202] Maloney PC, Rotman B. Distribution of suboptimally induced  $\beta$ -D-galactosidase in *Escherichia coli*: the enzyme content of individual cells. *J Mol Biol* 1973;73:77–91.
- [203] Lee N, Francklyn C, Hamilton EP. Arabinose-induced binding of AraC protein to *araI* activates the *araBAD* operon promoter. *Proc Natl Acad Sci U. S. A* 1987;84:8814–8.
- [204] Schleif R. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet* 2000;16:559–65.
- [205] Khelebnikov A, Risa Ø, Skaug T, Carrier TA, Keasling JD. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol* 2000;182:7029–34.
- [206] Siegle DA, Hu JC. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci U. S. A* 1997;94:8168–72.
- [207] Ramos JL, Martínez-Bueno M, Molina-Henares AJ, Terán W, Watanabe K, Zhang X, et al. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 2005;69:326–56.
- [208] Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Mol Biol* 2000;7:215–9.
- [209] Hillen W, Schollmeier K, Gatz C. Control of expression of the Tn10-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the tet operon regulatory region. *J Mol Biol* 1984;172:185–201.
- [210] Hinrichs W, Kisker C, Düvel M, Müller A, Tovar K, Hillen W, et al. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science* 1994;264:418–20.
- [211] Miller MB, Bassler BL. Quorum sensing in bacteria. *Annu Rev Microbiol* 2001;55:165–99.
- [212] Fuqua WC, Winans SC, Greenberg EP. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* 1994;176:269–75.
- [213] Fuqua C, Greenberg EP. Listening in on bacteria: acyl-homoserine lactone signalling. *Nat Rev Mol Cell Biol* 2002;3:685–95.
- [214] Engebrecht J, Nealon K, Silverman M. Bacterial bioluminescence: isolation and genetic analysis of functions from *Vibrio fischeri*. *Cell* 1983;32:773–81.
- [215] Urbanowski ML, Lostroh CP, Greenberg EP. Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR protein. *J Bacteriol* 2004;186:631–7.
- [216] Finney AH, Blick RJ, Murakami K, Ishihama A, Stevens AM. Role of the C-terminal domain of the alpha subunit of RNA polymerase in LuxR-dependent transcriptional activation of the lux operon during quorum sensing. *J Bacteriol* 2002;184:4520–8.
- [217] Devine JH, Shadel GS, Baldwin TO. Identification of the operator of the lux regulon from the *Vibrio fischeri* strain ATCC7744. *Proc Natl Acad Sci U. S. A* 1989;86:5688–92.
- [218] Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U. S. A* 2004;101:6355–60.
- [219] Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. A synthetic multicellular system for programmed pattern formation. *Nature* 2005;434:1130–4.
- [220] Nodadello S, Swennen EF. The new pLAI (*lux* regulon based auto-inducible) expression system for recombinant protein production in *Escherichia coli*. *Microb Cell Fact* 2012;11:3.
- [221] Pai A, Tanouchi Y, You L. Optimality and robustness in quorum sensing (QS)-mediated regulation of a costly public good enzyme. *Proc Natl Acad Sci U. S. A* 2012;109:19810–5.
- [222] Federle MJ, Bassler BL. Interspecies communication in bacteria. *J Clin Invest* 2003;112:1291–9.
- [223] Song H, Ding MZ, Jia XQ, Ma Q, Yuan YJ. Synthetic microbial consortia: from systematic analysis to construction and applications. *Chem Soc Rev* 2014;43:6954–81.
- [224] You L, Cox III RS, Weiss R, Arnold FH. Programmed population control by cell-cell communication and regulated killing. *Nature* 2004;428:868–71.
- [225] Saedi N, Wong CK, Lo TM, Nguyen HX, Ling H, Leong SS, et al. Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol Syst Biol* 2011;7:521.
- [226] Balagaddé FK, You L, Hansen CL, Arnold FH, Quake SR. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* 2005;309:137–40.
- [227] Copeland MF, Politz MC, Pfeleger BF. Application of TALEs, CRISPR/Cas and

- sRNAs as *trans*-acting regulators in prokaryotes. *Curr Opin Biotechnol* 2014;29:46–54.
- [228] Bogdanove AJ, Schornack S, Lahaye T. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol* 2010;13:394–401.
- [229] Bonas U, Stall RE, Staskawicz B. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol Gen Genet* 1989;218:127–36.
- [230] Römer P, Hahn S, Jordan T, Strauss T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science* 2007;318:645–8.
- [231] Mak AN, Bradley P, Bogdanove AJ, Stoddard BL. TAL effectors: function, structure, engineering and applications. *Curr Opin Struct Biol* 2013;23:93–9.
- [232] Boch J, Bonas U. *Xanthomonas AvrBs3* family-type III effectors: discovery and function. *Annu Rev Phytopathol* 2010;48:419–36.
- [233] Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 2012;335:716–9.
- [234] Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, et al. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 2012;335:720–3.
- [235] Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009;326:1509–12.
- [236] Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* 2009;326:1501.
- [237] Streubel J, Blücher C, Landgraf A, Boch J. TAL effector RVD specificities and efficiencies. *Nat Biotechnol* 2012;30(7):593–5.
- [238] Rinaldi FC, Doyle LA, Stoddard BL, Bogdanove AJ. The effect of increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids Res* 2017;45:6960–70.
- [239] Mahfouz MM, Li L, Shammuzzaman M, Wibowo A, Fang X, Zhu JK. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U. S. A* 2011;108:2623–8.
- [240] Polstein LR, Perez-Pinera P, Kocak DD, Vockley CM, Bledsoe P, Song L, et al. Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res* 2015;25:1158–69.
- [241] Copeland MF, Politz MC, Johnson CB, Markley AL, Pfeleger BF. A transcription activator-like effector (TALE) induction system mediated by proteolysis. *Nat Chem Biol* 2016;12:254–60.
- [242] Politz MC, Copeland MF, Pfeleger BF. Artificial repressors for controlling gene expression in bacteria. *Chem Commun (Camb)* 2013;49:4325–7.
- [243] Lienert F, Torella JP, Chen JH, Norsworthy M, Richardson RR, Silver PA. Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res* 2013;41:9967–75.
- [244] Cong L, Zhou R, Kuo YC, Cunniff M, Zhang F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* 2012;3:968.
- [245] Mercer AC, Gaj T, Sirk SJ, Lamb BM, Barbas III CF. Regulation of endogenous human gene expression by ligand-inducible TALE transcription factors. *ACS Synth Biol* 2014;3:723–30.
- [246] Cuculis L, Abil Z, Zhao H, Schroeder CM. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat Commun* 2015;6:7277.
- [247] Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 2013;14:49–55.
- [248] Gao H, Wu X, Chai J, Han Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* 2012;22:1716–20.
- [249] Lucast LJ, Batey RT, Doudna JA. Large-scale purification of a stable form of recombinant tobacco etch virus protease. *Biotechniques* 2001;30:544–6.
- [250] Kapust RB, Waugh DS. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr Purif* 2000;19:312–8.
- [251] Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 2013;82:237–66.
- [252] Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 2010;327:167–70.
- [253] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007;315:1709–12.
- [254] Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 2008;322:1843–5.
- [255] Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 2012;482:331–8.
- [256] Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirezada ZA, et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 2011;471:602–7.
- [257] Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011;9:467–77.
- [258] Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014;156:935–49.
- [259] Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 2016;351:867–71.
- [260] Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014;513:569–73.
- [261] Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. *Science* 2013;339:823–6.
- [262] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–21.
- [263] Mekler V, Minakhin L, Severinov K. Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc Natl Acad Sci U. S. A* 2017;114:5443–8.
- [264] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819–23.
- [265] Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;346:1258096.
- [266] Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. *Elife* 2013;2:e00471.
- [267] Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 2010;11:181–90.
- [268] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013;152:1173–83.
- [269] Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 2013;41:7429–37.
- [270] Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-mediated control of gene repression and activation. *Cell* 2014;159:647–61.
- [271] Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013;154:442–51.
- [272] Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, et al. Comparison of Cas9 activators in multiple species. *Nat Methods* 2016;13:563–7.
- [273] Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 2015;160:339–50.
- [274] De Mey M, De Maeseneire S, Soetaert W, Vandamme E. Minimizing acetate formation in *E. coli* fermentations. *J Ind Microbiol Biotechnol* 2007;34:689–700.
- [275] Hein S, Tran H, Steinbüchel A. *Synechocystis* sp. PCC6803 possesses a two-component polyhydroxyalkanoic acid synthase similar to that of anoxygenic purple sulfur bacteria. *Arch Microbiol* 1998;170:162–70.
- [276] Gründel M, Scheunemann R, Lockau W, Zilliges Y. Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 2012;158:3032–43.
- [277] Rodriguez GM, Atsumi S. Toward aldehyde and alkane production by removing aldehyde reductase activity in *Escherichia coli*. *Metab Eng* 2014;25:227–37.
- [278] Muro-Pastor MI, Reyes JC, Florencio FJ. Ammonium assimilation in cyanobacteria. *Photosynth Res* 2005;83:15–50.
- [279] Osanai T, Imamura S, Asayama M, Shirai M, Suzuki I, Murata N, et al. Nitrogen induction of sugar catabolic gene expression in *Synechocystis* sp. PCC 6803. *DNA Res* 2006;13:185–95.
- [280] Herrero A, Muro-Pastor AM, Flores E. Nitrogen control in cyanobacteria. *J Bacteriol* 2001;183:411–25.
- [281] Zhang S, Bryant DA. The tricarboxylic acid cycle in cyanobacteria. *Science* 2011;334:1551–3.
- [282] Steinhauser D, Fernie AR, Araújo WL. Unusual cyanobacterial TCA cycles: not broken just different. *Trends Plant Sci* 2012;17:503–9.
- [283] Laub MT, Goulian M. Specificity in two-component signal transduction pathways. *Annu Rev Genet* 2007;41:121–45.
- [284] Rowland MA, Deeds EJ, Crosstalk and the evolution of specificity in two-component signaling. *Proc Natl Acad Sci U. S. A* 2014;111:5550–5.
- [285] Capra EJ, Perchuk BS, Skerker JM, Laub MT. Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell* 2012;150:222–32.
- [286] Schell MA. Transcriptional control of the *nah* and *sal* hydrocarbon-degradation operons by the *nahR* gene product. *Gene* 1985;36:301–9.
- [287] Isberg RR, Voorhis DL, Falkow S. Identification of invasins: a protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell* 1987;50:769–78.
- [288] Hughes J, Lamparter T, Mittmann F, Hartmann E, Gärtner W, Wilde A, et al. A prokaryotic phytochrome. *Nature* 1997;386:663.
- [289] Lamparter T, Mittmann F, Gärtner W, Börner T, Hartmann E, Hughes J. Characterization of recombinant phytochrome from the cyanobacterium *Synechocystis*. *Proc Natl Acad Sci U. S. A* 1997;94:11792–7.
- [290] Yeh KC, Wu SH, Murphy JT, Lagarias JC. A cyanobacterial phytochrome two-component light sensory system. *Science* 1997;277:1505–8.
- [291] Forst S, Delgado J, Inouye M. Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the *ompF* and *ompC* genes in *Escherichia coli*. *Proc Natl Acad Sci U. S. A* 1989;86:6052–6.
- [292] Levskaya A, Weiner OD, Lim WA, Voigt CA. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* 2009;461:997–1001.

- [293] Mascher T, Helmann JD, Uden G. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* 2006;70:910–38.
- [294] Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, et al. A synthetic genetic edge detection program. *Cell* 2009;137:1272–81.
- [295] Hirose Y, Shimada T, Narikawa R, Katayama M, Ikeuchi M. Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein. *Proc Natl Acad Sci U. S. A* 2008;105:9528–33.
- [296] Tabor JJ, Levskaya A, Voigt CA. Multichromatic control of gene expression in *Escherichia coli*. *J Mol Biol* 2011;405:315–24.
- [297] Möglich A, Ayers RA, Moffat K. Design and signaling mechanism of light-regulated histidine kinases. *J Mol Biol* 2009;385:1433–44.