Transcription control engineering and applications in synthetic biology

Michael D. Engstrom a, b, 1, Brian F. Pfleger b, c, * 

* Genetics-Biotechnology Center, University of Wisconsin-Madison School of Medicine and Public Health, USA  
b Department of Chemical and Biological Engineering, University of Wisconsin-Madison College of Engineering, USA  
c Microbiology Doctoral Training Program, University of Wisconsin-Madison, USA

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. Cl 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 (B.F. Pfleger).  
Peer review under responsibility of KeAi Communications Co., Ltd.

1 Present address: 2730 Engineering Hall, 1415 Engineering Dr. Madison, WI 53706, USA.

https://doi.org/10.1016/j.synbio.2017.09.003  
2405-805X/© 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/).
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...
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 and the evolution time required to optimize natural systems. These well-characterized examples of transcriptional control in model organisms are beyond its scope. Although we will focus primarily on control and applications for engineered systems, but many control transcriptional adjustments 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, σ70, 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 σ70 (σ70, E. coli), and common synthetic biology systems such as 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 methods are 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.

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
disrupted in the AT-rich –10 sequence, the base at position –11 (usually an A) is flipped into $\sigma^{35}$ 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 β-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_Lac-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 106-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 phospho-enolpyruvate carboxylase) and ddx (encoding deoxy-xylulose-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 Ryttter et al. [33] took consensus sequences for the C. glutamicum –10 region (TGTNNTATTGT), the E. coli –35 region with and without an AT-rich region upstream, and primarily changed the spacing 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 Synochoccus elongatus strain PCC 7002 promoter P_Cpce 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_Cpce –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 mechanisms (i.e., translational coupling) [91,144,150,153–156], potentially arrestingdownstream 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 be 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 backtrackling (albeit, nucleoid structure may serve a function perturbing RNAP elongation dynamics along with promoting RNAP backtrackling) [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-trace downstream of this stem loop is dehybridized from the associated DNA poly A-trace [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-trace is sometimes present upstream of some terminator stem-loops [35,142,144], but the precise function for these sequences remains unclear (e.g., this sequence may be the poly U-trace for transcription directed from the opposite strand of DNA, promote extension of the terminator stem into the poly U-trace, etc). Well known Rho-independent transcription terminators include the T_RnB1 and T_RnB2 terminators [166–168], the T_R T7 phage terminator [167,168], and the λ phage T0 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
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, PBAD (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.cellod.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.

3.3. 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_{BS} and represses cl and CI binds P_{k8} and represses cro [172–176]. This toggle switch controls the fate of a cell infected with bacteriophage λ. If the Cro binding state is predominant a lytic infection with this phage may enter into the lysogenic cycle [172–176]. The Cro and CI binding sites have the same consensus sequence (TATCACCGGCGGTGATA), 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 CIB57 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_{k8} and P_{BS} 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 (AATGTGATCTAGTACACCATT), which are often upstream of operons related to carbohydrate metabolism [181–184]. Indeed, the αCTD 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, adenyllyl cyclase activity and cAMP synthesis is modulated by the availability of glucose in the extracellular milieu [184]. When glucose levels are high, adenyllyl cyclase activity is low, and when glucose levels are low, adenyllyl 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. Lac

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 (TGGCGATAACAAATT) in the absence of the small molecule allo-lactose (or isopropyl-β-D-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_{lac} (a strong synthetic promoter generated by fusing P_{lac} 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 heterogonous 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 cell expression levels similar to a fully induced system [202,201]. 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 PActivator (green arrow). Activator (bound to inducer) also binds target sites on a high copy number plasmid activating reporter2 (green suns) again at PActivator, 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 PInducer1 and PInducer2 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 PRepress. Inducer 3 is not secreted (red arrow with a cross), and Regulator 3 (without inducer 3 bound) binds PInducer3 repressing repressor. Subsequently, Repressor does not repress reporter, at PRepress 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 PRepress. Repressor subsequently represses reporter, blocking Reporter production (0 and a red box in the truth table).
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 α-arabinose-mediated isomerizations of this regulator [203,204]. Indeed, when α-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 RNA polymerase promoter binding [115,212,204]. AraC recognizes an AT-rich sequence (usually 6-bp long) flanked by the sequence TRTGGA on the upstream side and GCTA on the downstream side [203]. In the presence of α-arabinose, AraC binds this metabolite and isomerizes to a collapsed configuration [204]; AraC and CAP (Pαara is a weak promoter without CAP activation) occupy binding sites proximal to Pαara and Pαrac and induce RNA polymerase holoenzyme promoter binding and open complex formation [204]. This may be achieved by locally bending DNA such that the αCTDs of the RNA polymerase 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 saturating levels of α-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 Pαara, 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 Ptet is another model of tight E. coli gene repression. In its natural context, TetR binds as a dimer to the operator sequence (TCTATATAGTAGG), which in turn obstructs RNA polymerase-promoter interactions [115,207–209]. In the presence of tetracycline (or the analogue αTc) 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 αTc 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, Siutti et al. [19] used TetR- and αTc-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 Ptet [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 αTc-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, LuxR 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 (ACTGTAGGATGTACGGT) upstream of the extended – 10 sequence of Plux [211–217]. In the case of Plux, LuxR dimers make contact with the αCTD subunits of the RNA polymerase holoenzyme, which in turn facilitates promoter recognition (of the weak Plux) [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 (αTc-mediated derepression of P(tet-O1)) to produce AHL (sender cells) that in turn binds LuxR in receiver cells leading to a pulse of gfp expression (under control of Plux). 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 Bogdanove [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–Y, N/D–C, N/K–G, H–Y/T, H/C–H, N/G–N, N/A–G, and I/C–T]. Thus, shuffling different repeat regions impacts 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]. However, Politz et al. [242] demonstrated that a TALE recognizing the Lacl 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 Ptek (aTc inducible). Upon TetR repression 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 tfp 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 ω 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-Rodriguez 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]), gatC (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 θ-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 θ-ketoglutarate levels, activate NtcA (a catabolic gene regulator activated by reduced ammonium assimilation and the modest increase in θ-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 gatC, sdoA, and sdoB (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 (in eukaryotic 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 PhoQ system (PhoQ kinase is inactive under high Mg\textsuperscript{2+} concentrations [64,65]), the cognate P\textsubscript{PhoR}, P\textsubscript{Lux}, and P\textsubscript{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\textsubscript{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 CaCR response regulator on or off, respectively [285,286]); the YFI receptor (switched off by blue light, which switches off response regulator FixJ [297]); Cl and PifH 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] Bai N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 2000;289: 905–20.
[4] Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity: 2000. Science 2001;289:2900–2.
[5] Muth CW, Ortoleva-Dennelly 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. Microbiol Cell Fact 2016;15:196.
[8] Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemam MR, Pfleger 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 lyocycne production in Escherichia coli by engineering metabolic control. Nat Biotechnol 2000;18:337–7.
[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, Neovig E, Stephanopoulos G. Tuning genetic control with precise in vivo DNA writing in living cell populations. Science 2015;4:107–11.
[13] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature 2000;403:392–9.
[14] Purcell O, Lu TK. Synthetic analog and digital circuits for cellular computation and memory. Curr Opin Biotechnol 2014;29:146–55.
[15] Tamsir A, Tabor JJ, Voigt CA. Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. Nature 2011;469:212–5.
[16] Siuci P, Yabezk J, Lu TK. Synthetic circuits integrating logic and memory in living cells. Nat Biotechnol 2013;31:448–52.
[17] 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.
[18] Anderson JC, Voigt CA, Arkin AP. Construction of a genetic toggle switch in Escherichia coli. Nat Chem Biol 2017;13:706–8.
[19] Alper H, Fischer C, Neovig E, Stephanopoulos G. Tuning genetic control with precise in vivo DNA writing in living cell populations. Nat Biotechnol 2013;10:203–8.
[20] Anderson JC, Voigt CA, Arkin AP, Environmental signal integration by a modular AND gate. Mol Syst Biol 2007;3:133.
[21] Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. Science 2016;352. aac7341.
[22] Kotula JW, Kerns SJ, Shaker 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.
[23] Daniel R, Rubens JR, Sarpeshkar R, Lu TK. Synthetic analog computation in living cells. Nature 2013;497:619–23.
[24] Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. Synthetic gene regulation. Curr Opin Biotechnol 2012;30:354–60.
[25] Farzadfard F, Lu TK. Synthetic biology. Genomically encoded NOR gates and chemical ‘wires’. Nature 2011;469:212–5.
[26] 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.
promoter libraries for Corynebacterium glutamicum. Appl Microbiol Bio-technol 2014;98:2617–23.

[34] Mackey AL, Bengtson MB, Clarke RE, Gordon GC, Pfeger 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 5′- and 3′-UTR-deleted and synthetic 60S and 50S Bic and quantification of their design constraints.Nat Methods 2013;10:659–64.

[36] Zaslawer A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A comprehensive library of fluorescent transcripational reporters for Escherichia coli. Nat Methods 2008;5:623–8.

[37] Jensen PR, Hammer K. The sequence of spacers between the consensus sequence of riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[38] Winkler WC, Breaker RR. Regulation of bacterial gene expression by riboswitch elements. Annu Rev Microbiol 2005;59:487–517.

[39] Belfort M, Pedersen-Lane J, West D, Ehrenman K, Maley G, Chu F, et al. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012:4.

[40] Stragier P, Losick R. Molecular genetics of sporulation in Bacillus subtilis. Annu Rev Genet 1996;30:297–308.

[41] LaBell TL, Trempy JE, Haldenwang WG. Sporulation-speciﬁc promoters and DNA-logging in a cyanobacterium. J Biol Eng 2014;8:4.

[42] Kanehara K, Ito K, Akiyama Y. YaeL (EcfE) activates the yeaT-yoeB toxin-antitoxin operon. J Bacteriol 2001;183:1835–40.

[43] Camsund D, Heidorn T, Lindblad P. Design and analysis of LacI-repressed promoters for cyanobacteria. J Biol Eng 2013;7:10.

[44] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome protection circuits. J Am Chem Soc 1998;120:4452–3.

[45] Sleytr UB. Heterologous reattachment of regular arrays of glycoproteins on insect cell membranes. J Mol Biol 1976;108:1–18.

[46] Markley AL, Begemann MB, Clarke RE, Gordon GC, Pabo CO, Engstrom M.D. Tol2 transposon-mediated transformation of Drosophila melanogaster embryos. Nucleic Acids Res 1997;25:11613–21.

[47] Perriman R, Ares Jr M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. RNA 1998;4:1047–54.

[48] Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell 2000;100:391–8.

[49] Kruguer K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autocleavage and autocyclization of the ribosomal RNA inter-

[50] Perriman R, Ares Jr M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. RNA 1998;4:1047–54.

[51] Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell 2000;100:391–8.

[52] Stragier P, Losick R. Molecular genetics of sporulation in Bacillus subtilis. Annu Rev Genet 1996;30:297–341.

[53] Alba BM, 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.

[54] Salis HM, Mistry EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009;27:946–50.

[55] 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.

[56] Kruguer K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autocleavage and autocyclization of the ribosomal RNA inter-

[57] Spero RW. Involvement of the yefM-yoeB toxin-antitoxin operon in Escherichia coli. Cell 1982;31:147–158.

[58] LaBell TL, Trempy JE, Haldenwang WG. Sporulation-speciﬁc promoters and DNA-logging in a cyanobacterium. J Biol Eng 2014;8:4.

[59] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[60] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[61] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[62] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[63] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[64] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[65] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[66] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[67] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[68] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[69] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[70] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[71] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[72] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[73] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[74] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.

[75] Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 2012;4.
three domains of life. Nat Rev Microbiol 2011;9:85–98.

[100] Werner F. Structural evolution of multisubunit RNA polymerases. Trends Microbiol 2008;16:213–20.

[101] Deuschle U, Kammerer W, Ganz H, Bujard H. Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. EMBO J 1986;5:5978–94.

[102] Mahan MJ, Slauch JM, Mekalanos JJ. Selection of bacterial virulence genes naturally induced in host tissues. Methods Enzymol 1983;77:1–28.

[103] Kuznedelov K, Severinov K, Heyduk T. A consensus adenine at Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science 2006;314:1144–7.

[104] Kuznedelov K, Severinov K, Heyduk T. A consensus adenine at Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science 2006;314:1144–7.

[105] 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 genetic metabolic engineering. BMC Biotechnol 2007;7:34.

[106] Kuznedelov K, Severinov K, Heyduk T. A consensus adenine at Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science 2006;314:1144–7.

[107] 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 genetic metabolic engineering. BMC Biotechnol 2007;7:34.

[108] Kuznedelov K, Severinov K, Heyduk T. A consensus adenine at Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science 2006;314:1144–7.

[109] 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 genetic metabolic engineering. BMC Biotechnol 2007;7:34.

[110] 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;33:6288–76.

[111] 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;33:6288–76.

[112] 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;33:6288–76.

[113] 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;33:6288–76.

[114] 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;33:6288–76.

[115] 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;33:6288–76.

[116] 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;33:6288–76.

[117] 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;33:6288–76.

[118] 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;33:6288–76.

[119] 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;33:6288–76.

[120] 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;33:6288–76.

[121] 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;33:6288–76.

[122] 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;33:6288–76.

[123] 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;33:6288–76.

[124] 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;33:6288–76.

[125] 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;33:6288–76.

[126] 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;33:6288–76.

[127] 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;33:6288–76.

[128] 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;33:6288–76.

[129] 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;33:6288–76.
Bridged filaments of histone-like nucleoducting protein cause RNA polymerase and aid inactivation 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 2013;288:20242–51.

[165] Hein PP, Kolb KE, Windgassen T, Bellecourt MJ, Darst SA, Mooney RA, et al. DNA looping may affect the efficiency of downstream transcription in E. coli. Nucleic Acids Res 2013;41:1145–53.

[166] Oehler S, Eismann ER, Kröger B, Stülke J. Carbon catabolite repression in bacteria: many ways to achieve consistent repression. Nat Rev Microbiol 2010;8:589–99.

[167] Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. 51–56.

[168] Copeland MF, Politz MC, Palleroni NJ. Listening in on bacteria: acyl-homoserine lactone signaling. Vibrio cholerae. Proc Natl Acad Sci U. S. A 2007;104:3479–84.

[169] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science 2002;297:1183–6.

[170] Jansen PR, Westerhoff HV, Michelsen O. The use of loc-type promoters in control analysis. J Biochem 1993;211:181–91.

[171] Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. Proc Natl Acad Sci U. S. A 1975:43:53–66.

[172] Maloney PC, Bottema R. Distribution of suboptimally induces β-galactosidase in Escherichia coli: the enzyme content of individual cells. J Mol Biol 1973;71:77–91.

[173] Lee N, Franklyn C, Hamilton EP. Arabinose-induced binding of AraC protein to araZ activates the araBAD operon promoter. Proc Natl Acad Sci U. S. A 1987;84:8148–8.

[174] Schleif R. Regulation of the l-arabinose operon of Escherichia coli. Trends Genet 2000;16:559–63.

[175] Klahebnikov A, Risa O, Skagl 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.

[176] Siegel DA, Hu JC. Gene expression from plasmids containing the araBAD promoter at suboptimal induction. Microbiol Mol Biol Rev 2006;70:98–130.

[177] Smith BD, White RS. Listening in on bacteria: acyl-homoserine lactone binding to purified regulators. Microbiol Mol Biol Rev 2006;70:358–74.

[178] Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. A synthetic multicellular system for programmed pattern formation. Nature 2005;434:1130–4.

[179] Nocadello S, Swennes EN. The new pF1 (lux regulon based auto-inducible) expression system for recombinant protein production in Escherichia coli. Microcell Fact 2012:113.

[180] Moser SV, Ashkenazi A. Stochasticity 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.

[181] Federle MJ, Basler BL. Intercellular communication in bacteria. J Clin Invest 2003;112:1291–9.

[182] Song H, Ding MZ, Jia QX, Ma Q, Yuan YJ. Synthetic microbial consortia: from systematic analysis to construction and applications. Chem Soc Rev 2014;43:6954–81.

[183] You L, Cox III RS, Weiss R, Arnold FH. Programmed population control by cell-cell communication and regulated killing. Nature 2004;428:868–71.

[184] Saeidi 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;7:721.

[185] Balagaddé FK, You L, Hansen CL, Arnold FH, Quake SR. Long-term monitoring of bacteria undergoing programmed population control in a micro-chromatob: Syst Biol 2005;50:137–40.

[186] Copeland MF, Politz MC, Pfleger BF. Application of TALEs, CRISPR/Cas and

[187] Zheng D, Constantinidou C, Hobman JL, Minchin SD. Identification of the operator of the araBAD operon based auto-inducible system for programmed pattern formation. Mol Syst Biol 2011;7:521.
Mascher T, Helmann JD, Unden C. Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev 2006;70:910–38.

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.

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.

Tabor JJ, Levskaya A, Voigt CA. Multichromatic control of gene expression in Escherichia coli. J Mol Biol 2011;405:315–24.

Möglich A, Ayers RA, Moffat K. Design and signaling mechanism of light-regulated histidine kinases. J Mol Biol 2009;385:1433–44.