CRISPR-Cas activation for regulating multi-gene expression in bacteria

Jason Fontana

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2020

Reading Committee:
James M. Carothers, Chair
Jesse G. Zalatan, Chair
Maitreya Dunham
Eric Klavins

Program Authorized to Offer Degree
Molecular Engineering

University of Washington

Abstract

CRISPR-Cas activation for regulating multi-gene expression in bacteria

Jason Fontana

Chairs of the Supervisory Committee:
James M. Carothers
Department of Chemical Engineering

Jesse G. Zalatan
Department of Chemistry

Bacterial metabolism is comprised of large and complex gene networks that can produce valuable chemical products. In principle, these networks can be reengineered using synthetic multi-gene transcriptional programs to optimize production of high-value compounds. However, our limited ability to regulate multi-gene expression with precision makes this goal difficult to achieve. While programmable gene repression in bacteria is well-established, there is a lack of robust tools for programmable gene activation. This work describes our efforts to develop new tools for bacteria to independently target and predictably manipulate the expression levels of multiple genes. We developed a programmable CRISPR-Cas transcriptional activation (CRISPRa) system for *E. coli* that uses modified guide RNAs to recruit a transcriptional activator. When used in concert with CRISPRi, our tools can be used to simultaneously activate and repress multiple genes. However, to expand our ability to regulate synthetic and endogenous pathways, we need predictive rules for selecting effective target sites for CRISPRa. To uncover these rules, we systematically investigated the requirements for CRISPRa in bacteria. In contrast to comparable systems in eukaryotes, we show that bacterial CRISPRa is sensitive to the geometry of the interaction between the CRISPRa complex and the synthetic promoter. Last, we discuss engineering dynamically regulated CRISPR-Cas systems to control the timing of the expression of multiple genes. These next-generation tools will enable us to engineer sophisticated multi-gene transcriptional networks that more closely resemble natural gene networks. Together, our work provides a foundation toward rapidly discovering and implementing new multi-gene programs that improve production of high-value chemicals.

In memory of Egidio Fontana.

Table of Contents

Reading Guide

Chapter 1 is comprised of our manuscript published in Nature Communication, “Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria”, which describes our efforts in developing new tools for programmable transcriptional activation using the CRISPR-Cas system (CRISPRa). When used in concert with CRISPRi gene repression, our CRISPRa tools can be used to activate and repress multiple genes at the same time.

Chapter 2 is our manuscript published in Nature Communications, “Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements”. In this paper, we systematically investigated the rules underlying CRISPRa targeting in bacteria. These findings will help us design highly effective promoters for regulating heterologous pathways in *E. coli* and establishes a foundation for further studies focusing on the rules for activating endogenous genes.

Chapter 3 contains our manuscript published in the Journal of Industrial and Microbial Biotechnology, “Prospects for engineering dynamic CRISPR–Cas transcriptional circuits to improve bioproduction”. This review serves as a brief introduction to the topics of CRISPR-Cas transcriptional regulation and dynamic control of gene expression in the context of metabolic engineering.

Chapter 4 is our manuscript published in the Biotechnology Journal, “Regulated expression of sgRNAs tunes CRISPRi in *E. coli*”, which describes how currently available genetic regulators could be useful to dynamically regulate CRISPR-Cas transcriptional control programs.

Chapter 5 is our manuscript published in Current Opinion in Biotechnology, “Challenges and opportunities with CRISPR activation in bacteria for data-driven metabolic engineering”. Here, we first summarize the recent developments in bacterial CRISPRa. Second, we provide our perspective on the future directions of CRISPRa engineering.

Chapter 1:
Synthetic CRISPR-Cas Gene Activators for Transcriptional Reprogramming in Bacteria

Chen Dong[1], Jason Fontana[2], Anika Patel[1], James M. Carothers[2,3,4], & Jesse G. Zalatan[1,2,4]

[1]Department of Chemistry
[2]Molecular Engineering & Sciences Institute
[3]Department of Chemical Engineering
[4]Center for Synthetic Biology
University of Washington, Seattle, WA 98195, USA

Published as a research article in *Nature Communications* on June 27th, 2018.
DOI: 10.1038/s41467-018-04901-6

Abstract

Methods to regulate gene expression programs in bacterial cells are limited by the absence of effective gene activators. To address this challenge, we have developed synthetic bacterial transcriptional activators in *E. coli* by linking activation domains to programmable CRISPR-Cas DNA binding domains. Effective gene activation requires target sites situated in a narrow region just upstream of the transcription start site, in sharp contrast to the relatively flexible target site requirements for gene activation in eukaryotic cells. Together with existing tools for CRISPRi gene repression, these bacterial activators enable programmable control over multiple genes with simultaneous activation and repression. Further, the entire gene expression program can be switched on by inducing expression of the CRISPR-Cas system. This work will provide a foundation for engineering synthetic bacterial cellular devices with applications including diagnostics, therapeutics, and industrial biosynthesis.

Introduction

Bacteria are attractive targets for a wide variety of engineering applications. Bacterial strains with the ability to utilize carbon sources like CO_2, CO, methane, or lignocellulose, and alternative energy sources such as light or H_2 could provide the foundation for cost-effective and environmentally-friendly industrial biosynthesis[1,2]. Microbial communities, such as those that reside in the human gut, play an important role in human health and disease, and tools to engineer these bacteria have great potential as both diagnostics and therapeutics[3-5]. To harness, regulate, and modify the behavior of these and other bacteria, there is a compelling need to develop genetic tools to control gene expression and implement complex, multi-gene regulatory programs. Ideally, we want to build circuits that can regulate many genes at once, dynamically respond to external inputs or the internal state of the cell, and be easily reprogrammed to explore different functional architectures. While capabilities to edit and modify genomes are rapidly expanding, our ability to encode a precisely-defined and dynamically-responsive gene expression program with cis-regulatory sequences at the DNA level remains difficult. Thus, we sought to develop synthetic transcription factors in bacteria, which could be coupled to programmable DNA binding domains and controlled by inducible promoters to engineer complex, dynamically-responsive multi-gene expression programs.

Synthetic control of gene expression has recently become much more straightforward with the emergence of programmable transcription factors using the CRISPR-Cas system (Fig. 1). A catalytically-inactive Cas9 (dCas9) protein can be used to target specific DNA sequences with guide RNAs (gRNAs) that recognize their targets based on predictable Watson-Crick base pairing. This approach can be used to repress genes by physically blocking RNA polymerase (CRISPR interference or CRISPRi)[6,7]. To activate genes (CRISPR activation or CRISPRa), the CRISPR complex can be linked to a transcriptional activator by direct fusion to dCas9 or via recruitment domains on the gRNA[7-11]. In bacteria, however, there are very few transcriptional activation domains that have been reported to be effective when fused to modular DNA binding domains. Bacterial two-hybrid systems have been constructed with pairs of candidate interacting proteins separately fused to RNA polymerase subunits and DNA binding proteins. It is also possible to fuse

RNA polymerase subunits directly to DNA binding domains to activate transcription[12-14]. One of the RNA polymerase subunits, RpoZ, has been coupled to the CRISPR system to activate gene expression[7,15-18]. For comparison, in eukaryotic systems there are many effective activators and CRISPRa has been extensively used in a variety of applications[19]. The paucity of reports of CRISPRa in bacteria suggests that RpoZ may not be effective as a general activator of transcription, or that we lack a complete understanding of the design rules to predictably activate gene expression in bacteria.

To develop an improved toolkit for gene activation in bacteria, we screened a broad set of candidate proteins for transcriptional activity in *E. coli*. We identified several proteins that can effectively activate gene expression when recruited via the CRISPR-Cas system. Using the most effective activator, SoxS, we can activate one target gene while simultaneously repressing a different target gene with CRISPRi, and we can control the entire multi-gene expression program with inducible promoters driving CRISPR-Cas system components. We find that gene activation in *E. coli* is highly sensitive to the location of the gRNA target site, consistent with prior results[7], and suggesting a possible explanation for why it has been difficult to develop synthetic activators in bacteria. Finally, we show that bacterial CRISPRa can be used to increase the output of a heterologous ethanol biosynthesis pathway. These results provide a framework for implementing CRISPRa in bacteria with a wide variety of potential applications. Further, because SoxS interacts with a highly conserved site on RNA polymerase, our bacterial CRISPRa toolkit may be portable to a broad range of bacterial species.

Results

Identifying Transcriptional Activation Domains for Bacteria

To recruit transcriptional activators to the CRISPR-Cas system, we used gRNAs that are extended with hairpin sequences to recruit RNA binding proteins (RBPs), which are in turn fused to candidate activators (Fig. 1A)[10]. These modified gRNAs, termed scaffold RNAs (scRNAs), encode both the target sequence and the regulatory action to execute at that target. By using scRNAs to recruit activators (CRISPRa) to some genes and gRNAs to physically block RNA polymerase (CRISPRi) at other genes, we can encode complex expression programs where some genes are activated and others are repressed (Fig. 1B)[10].

To screen for potential activators, we first constructed an *E. coli* strain with a genomically-integrated, weakly-expressed GFP reporter gene. The upstream region of the reporter gene includes a number of potential gRNA target sites and is identical to that used previously to evaluate the dCas9-RpoZ fusion protein (see Supplementary Methods)[7]. We targeted the CRISPR-Cas complex to the W108 gRNA site located 91 bases upstream of the transcriptional start site (TSS), as this site previously demonstrated the strongest activation with dCas9-RpoZ[7]. For candidate activators we chose endogenous transcriptional regulators SoxS, MarA, Rob, and CAP[20,21]; hijackers (i.e. bacteriophage or transposon effectors) TetD, λcII, GP33, $N4_{SSB}$, and AsiA[22-26]; and RNA polymerase subunits RpoZ (ω), RpoD (σ^{70}), and the N-terminal domain of RpoA (αNTD)[12,13,27]. Of these candidates, αNTD[12], RpoZ[7,13], and AsiA[26] have been previously

reported to activate transcription when fused to heterologous DNA binding domains, and other candidates were selected based on literature reports suggesting that they could recruit RNA polymerase to activate transcription. Each of these candidate proteins was fused to the MS2 coat protein (MCP), which binds to an MS2 hairpin on the scRNA (Fig. 2A)[10].

Several candidate activators produced significant GFP reporter expression in late stationary phase cultures, with the largest effects from SoxS and TetD, and smaller but still detectable GFP expression from λcII, αNTD, and RpoZ (Fig. 2A). Activation with MCP-RpoZ is significantly increased in a Δ*rpoZ* host strain, consistent with that observed previously for other RpoZ fusion proteins, including dCas9-RpoZ (Supplementary Fig. 1A)[7,13]. We also obtained significant GFP expression with the T4 bacteriophage activator AsiA (Supplementary Fig. 1B) by co-transforming with a σ^{70} F563Y mutant, which prevents toxicity that occurs when AsiA is expressed alone and inhibits the endogenous σ^{70} subunit[26]. The most effective activator, SoxS, is a member of the AraC family of transcription factors. SoxS is expressed during oxidative stress and activates expression of a number of genes by binding to RNA polymerase in a pre-recruitment complex and scanning the genome to find DNA targets located in promoter regions[20,28]. This mechanism of action could explain the effectiveness of SoxS as a candidate activator for CRISPRa.

To confirm that the observed GFP expression arises from recruitment of the candidate activator upstream of the GFP reporter gene, we performed several negative controls with the SoxS activator. First, we demonstrated that activation requires the presence of the activator protein. When dCas9 and the 1x MS2 scRNA are expressed without MCP-SoxS, the CRISPR-Cas complex can bind upstream of the GFP reporter, but there is no significant GFP expression (Fig. 2B). Similarly, when the gRNA lacks the 1x MS2 recruitment hairpin, there is no significant GFP activation. Finally, we expressed dCas9, MCP-SoxS, and an off-target 1x MS2 scRNA and observed weak but detectable GFP expression (Fig. 2B). This result suggests that a small fraction of the GFP expression observed with MCP-SoxS arises from non-specific gene activation, which is plausible given that SoxS is an endogenous regulator of transcription in *E. coli* with many gene targets and a relatively degenerate binding site[29,30].

Optimizing Activity of the SoxS Activator

To optimize the activity of CRISPRa with SoxS and minimize off-target effects at endogenous SoxS gene targets, we systematically varied several design parameters of our system. First, we modified the scRNA structure to optimize it for bacterial expression. In our original design, the MS2 hairpin is appended at the 3' end of the gRNA sequence[10], just downstream of an endogenous tracr terminator hairpin[31]. Removing this terminator hairpin while retaining the 3' MS2 hairpin leads to a 1.5 to 2-fold increase in GFP expression, likely due to increased steady-state levels of the full length scRNA (Fig. 3A). For comparison, in eukaryotic cells the same modified scRNA lacking the terminator reduces CRISPRa-mediated reporter gene expression ~2-fold (Supplementary Fig. 2A), possibly because the terminator hairpin interacts with Cas9[32]. Presumably, the same detrimental effect is present in bacteria, but it is outweighed by the benefit from ensuring that the entire scRNA construct is expressed. Using this optimized scRNA design

(scRNA.b1), we then varied the length of the amino acid linker connecting MCP and SoxS. 5 and 10 amino acid linkers increased GFP expression by ~2 to 3-fold compared to our original 2 amino acid linker (Fig. 3B). To increase activity further, we tested an scRNA design with two MS2 hairpins (2x MS2), but observed no significant increase in GFP expression compared to the 1x MS2 scRNA (Supplementary Fig. 2B). We also tested an alternative gRNA design in which MS2 hairpins are embedded within existing hairpins in the gRNA (sgRNA 2.0)[11]. We observed no significant GFP expression with sgRNA 2.0 (Supplementary Fig. 2C), which was surprising given that this design is highly effective in eukaryotic cells.

To minimize potential off-target effects at endogenous SoxS gene targets, we attempted to decouple its endogenous DNA binding activity from its transcriptional activation function. We identified several candidate residues at the SoxS DNA binding interface that are known to disrupt activity at endogenous sites when mutated[33,34]. We expected that if transcriptional activation could be decoupled from endogenous DNA binding, we would retain activity with our CRISPRa system even with SoxS mutants that are defective for activity at endogenous promoters. For two SoxS mutants, R93A and S101A, we observed no loss in activity in a CRISPRa assay; rather, we observed a 2-fold increase in GFP expression compared to wild type (wt) SoxS (Fig. 3C). The increased activity could arise if less SoxS is sequestered at endogenous DNA sites in these mutants. In contrast, the mutations R40A and F88A result in a substantial loss of GFP fluorescence in a CRISPRa assay, which could result from perturbations to protein structure or stability. To assess whether we effectively decoupled transcriptional activation from endogenous DNA binding, we plotted CRISPRa versus endogenous promoter activity for the series of SoxS mutants (Fig. 3D). To directly measure endogenous activity at SoxS gene targets when MCP-SoxS fusion proteins are expressed, we used LacZ fusion constructs for two different SoxS target genes, *zwf* and *fumC* (Supplementary Fig. 3), which are representative of class I and class II SoxS target genes, respectively[34]. The mutants that retain CRISPRa function, R93A and S101A, are approximately 2-fold weaker in endogenous activity than wt SoxS at *zwf* and >4-fold weaker at *fumC*. Thus, with SoxS R93A or S101A, we can partially decouple transcriptional activation from endogenous DNA binding. It may be possible to further decouple transcriptional activation from DNA binding with additional SoxS mutations.

We proceeded to test a fully optimized system with the 1xMS2 scRNA.b1 design, a longer 5 amino acid linker between MCP and SoxS, and the SoxS R93A mutant to reduce non-specific activity at endogenous SoxS targets. We observed a substantial increase in GFP fluorescence using an on-target guide RNA and a significant decrease in background activity with an off-target guide RNA (Supplementary Fig. 4) compared to the original, unoptimized system (Fig. 2B). Using RT-qPCR, we observed a 50-fold increase in GFP mRNA levels relative to negative control strains (Supplementary Fig. 5). Importantly, we observed no significant growth burdens associated with expression of the CRISPRa system (Supplementary Fig. 6A & B). Further, while we typically measured GFP levels in late stationary phase, we found similar trends in gene activation with smaller overall effects when GFP levels were measured in exponential phase (Supplementary Fig. 6C). We therefore proceeded with MCP-(5aa)-SoxS$_{R93A}$ and 1x MS2 scRNA.b1 in future experiments. In some cases, we also used the functionally equivalent 1x MS2

scRNA.b2 design (Fig. 3A), which differs by one base (see Supplementary Methods). In direct comparisons, this fully optimized SoxS-based CRISPRa system performs better than the previously reported dCas9-RpoZ system bacterial CRISPRa system[7]. Activation with SoxS is effective in an MG1655 strain where dCas9-RpoZ activity is undetectable, and in a Δ*rpoZ* MG1655 strain the SoxS-based system outperforms dCas9-RpoZ by >2-fold (Supplementary Fig. 7).

Inducible and simultaneous control of multiple genes

With CRISPR-Cas transcriptional programs, it is straightforward to target multiple genes for activation or repression[6,8,10,35], although simultaneous activation and repression has not yet been demonstrated in bacteria. Using our optimized bacterial CRISPRa system, we tested whether we could simultaneously activate one gene while repressing another. We constructed an *E. coli* strain with two genomically-integrated fluorescent reporter genes, a weakly expressed GFP and a strongly expressed RFP. We expressed dCas9, MCP-(5aa)-SoxS$_{R93A}$, a 1x MS2 scRNA.b1 for GFP activation, and a gRNA targeting RFP for CRISPRi repression (Fig. 4A). When both the scRNA and the gRNA are expressed, we observe a simultaneous increase in GFP expression and decrease in RFP expression, and the magnitudes of the effects are indistinguishable from those observed when each gRNA is expressed alone (Fig. 4A). Thus, when all components are expressed constitutively, we can effectively control multiple genes simultaneously and have different effects at individual gene targets.

To develop the capability to dynamically regulate multi-gene programs, we first attempted to control the CRISPRa system with inducible promoters. In previous work, using Tet-inducible promoters to control CRISPRi in bacteria required pTet controlling both dCas9 and the gRNA[6]. We tested whether a similar strategy could effectively control CRISPRa. When dCas9 and the 1x MS2 scRNA are controlled by pTet and MCP-(5aa)-SoxS$_{R93A}$ is expressed constitutively, however, we observed leaky GFP expression in the absence of the anhydrotetracycline (aTc) inducer (Fig. 4B), likely due to leaky expression from pTet[36]. In contrast, when either MCP-(5aa)-SoxS$_{R93A}$ alone or all three components of the CRISPRa system are controlled by pTet, we observed no leaky GFP expression and pTet-inducible GFP levels comparable to that observed with constitutive expression (Fig. 4B). GFP levels for pTet-inductions were measured in late stationary phase cultures; in early stationary phase cultures we observed modestly weaker overall GFP induction compared to that observed with constitutive promoters.

We also tested arabinose-inducible pBAD promoters as an alternative to pTet. With pBAD controlling all three components of the CRISPRa system, we observed inducible GFP expression, but at a level 2-fold weaker than that observed with constitutive expression. Alternatively, with pBAD controlling both dCas9 and the 1x MS2 scRNA, and constitutive expression of MCP-(5aa)-SoxS$_{R93A}$, we observed strong inducible GFP expression with no significant leaky expression in the absence of arabinose (Fig. 4B). These results were obtained with cultures in early stationary phase, unlike the results with pTet which were obtained in late stationary phase. In preliminary experiments with arabinose-inducible CRISPRi, we observed leaky repression in late stationary phase cultures, possibly because glucose depletion can relieve

catabolite repression of the pBAD promoter[37]. We therefore performed the arabinose-inducible experiments in early stationary phase cultures, where no leaky CRISPRi repression was observed. Taken together, these results indicate that effective inducible control of CRISPRa can be achieved with pTet controlling MCP-(5aa)-SoxS$_{R93A}$, or with pBAD controlling dCas9 and the 1x MS2 scRNA.

With inducible control over CRISPRa and CRISPRi, we can implement a two-gene switch from ON/OFF to OFF/ON. We used separate pTet promoters to control all components of the system: dCas9, MCP-(5aa)-SoxS$_{R93A}$, the 1x MS2 scRNA for GFP, and the gRNA for RFP. Upon addition of aTc, GFP expression increases while RFP expression decreases (Fig. 4C). Alternatively, with pBAD promoters controlling dCas9, the 1x MS2 scRNA for GFP, and the gRNA for RFP, and constitutively expressed MCP-(5aa)-SoxS$_{R93A}$, we could also inducibly switch GFP on and RFP off (Fig. 4C).

Bacterial CRISPRa is highly sensitive to gRNA target site

In prior reports of bacterial CRISPRa with dCas9-RpoZ, reporter gene expression depended on the distance of the gRNA target from the TSS, with peak expression occurring at a site located ~90 bases upstream[7]. To determine if a similar dependence applies to different activators recruited via 1x MS2 scRNA.b2 to the CRISPR-Cas complex, we designed a synthetic promoter (J1) that has gRNA target sites with the necessary PAM sequences every 10 bases on both strands upstream of a weak BBa_J23117 promoter driving an RFP reporter gene (Fig. 5A). With the MCP-(5aa)-SoxS$_{R93A}$ activator, we observed a sharp dependence of RFP expression on target site position, with significant RFP expression occurring for sites at 80-90 bases upstream of the TSS on the non-template strand and at 60-80 bases upstream on the template strand (Fig. 5B). To determine if activators with weak activity in our initial assays (Fig. 2A) might be more effective at different target sites, we tested MCP fusion proteins of TetD, λcII, and αNTD with this promoter. We observed peak activities for these activators at target site positions similar to the most effective sites for SoxS (Fig. 5B), and in no case did we observe dramatic increases in activity relative to our initial assays (Fig. 2A). Similarly, we tested whether an alternative gRNA design (sgRNA 2.0)[11] with MS2 sites embedded within internal hairpins in the gRNA and the MCP-(5aa)-SoxS$_{R93A}$ activator would be more effective at different target sites; we observed no detectable CRISPRa activity with this gRNA design at any target site. Finally, we tested an alternative RNA hairpin (1x PP7 scRNA.b1), that recruits the PP7 coat protein (PCP) fused to SoxS$_{R93A}$. MCP and PCP are structurally homologous proteins, but the MS2 and PP7 RNA hairpins differ significantly and interact in distinct orientations with their cognate binding proteins[38]. Thus, we do not expect activators recruited by MS2 or PP7 to be positioned in similar orientations when recruited to the CRISPR-Cas complex. Perhaps because of this structural difference, PCP-(5aa)-SoxS$_{R93A}$ behaved very differently than MCP-(5aa)-SoxS$_{R93A}$. The PCP construct was relatively ineffective at all sites on the non-template strand and moderately effective on the template strand (Fig. 5B). These results suggest that bacterial CRISPRa is highly sensitive to target position upstream of the TSS, and that it may be possible to find different combinations of activators, recruitment domains, and target sites that perform better than the most effective SoxS constructs tested here. Because all of the activators tested

here interact directly with RNA polymerase, albeit at different protein interfaces, we suggest that strong activation requires a CRISPR target site that allows RNA polymerase to bind to the minimal promoter (i.e. the -35 and -10 sites) and the activator simultaneously, which puts a significant constraint on the upstream target sites that can be used for CRISPRa.

To extend the range of effective target sites, we hypothesized that a longer, flexible linker between MCP and $SoxS_{R93A}$ might allow activators to recruit RNA polymerase to the promoter from a broader range of target sites. We therefore tested linkers with 2, 5, 10, or 20 amino acids. Extending the original 5 amino acid linker to 20 amino acids should lengthen the linker by ~57 Å (assuming a worm-like chain model for a flexible peptide linker with 3.8 Å/residue)[39]. Because target sites on the J1 promoter are spaced ~33 Å apart (10 bp spacing with 3.3 Å/bp for B-DNA), we expect that the extended linker should broaden the range of target sites by at least 10-20 bp in either direction from the optimal range observed with the 5 amino acid linker. Surprisingly, we observed no significant broadening of the effective target site range with longer linkers (Fig. 5C). We observed maximal activity with 5 and 10 amino acid linkers, while both 2 and 20 amino acid linkers show reduced activity but no significant difference in the position of effective target sites. While it remains to be seen if it is possible to design a bacterial CRISPRa complex that functions over a broader target range, simply extending linker lengths between MCP and the $SoxS_{R93A}$ activator was not effective.

Activating a metabolic gene cluster for ethanol biosynthesis

An important practical application of bacterial CRISPRa will be to enable complex and dynamic multi-gene circuits to control biosynthetic pathways. The capability to express pathway enzymes while repressing competing enzymes may lead to significant improvements in product yields[40]. As a proof of concept, we tested whether we could upregulate a heterologous *pdc adhB* gene cassette from *Zymomonas mobilis*, which converts pyruvate to ethanol and has previously been used for ethanol production in *E. coli* (Fig. 6)[41,42]. When we targeted the heterologous gene cassette with our CRISPRa system, we observed a 3-fold increase in ethanol production relative to cells without CRISPRa (Fig. 6). This result suggests that CRISPRa-based transcriptional control can be used to activate biosynthetic pathways. We expect that combining CRISPRa-based control of heterologous biosynthetic genes with CRISPRi on competing endogenous genes will enable rapid explorations of a large space of genetic circuit architectures to improve biosynthesis yields, and further improvements may be realized by targeting additional metabolic genes for activation or repression, and by coupling the system to dynamically-regulated or inducible promoters.

Discussion

In this work, we have identified multiple synthetic transcriptional activators compatible with CRISPRa in *E. coli*, including SoxS, TetD, and AsiA. The most effective activator, SoxS, is substantially stronger than the previously reported RpoZ-based synthetic activator and does not require additional host-strain modifications for function. SoxS may be an effective activator in part because it interacts with the C-terminal domain of RpoA, which is connected to RNA polymerase by a relatively flexible tether[12,43]. Using

SoxS and other activators, we find a surprisingly sharp dependence on the CRISPR target site distance from the TSS for effective gene activation, with the optimal gRNA sites positioned within a narrow window roughly 60-90 bases upstream of the transcription start site of a heterologous reporter gene. This range is strikingly smaller than the broad target site range observed for effective CRISPRa in eukaryotic cells, where many target sites in the 1-500 base range upstream of the TSS are effective[44]. One notable difference between bacteria and eukaryotes is that the effective bacterial activators identified in this and other work all appear to directly interact with RNA polymerase, while in eukaryotic cells transcription factors may act over longer distances via indirect chromatin modifications. Nevertheless, even given these constraints the ability to perform effective CRISPRa in bacteria will open significant avenues for engineering bacterial systems. Most importantly, we expect that controlling branch points in metabolic networks with simultaneous gene activation and repression will improve biosynthetic yields beyond that obtained simply by constitutively overexpressing heterologous pathways. Further, we can use inducible promoters to control the timing of gene expression, and we can build more sophisticated dynamic gene expression programs using protein or RNA biosensors to sense and respond to cellular metabolic states[45,46].

Two future challenges remain for broad applicability of CRISPRa in bacteria. First, it will be necessary to identify predictive rules for targeting and activating gene expression at endogenous sites. In preliminary experiments, we have observed only modest increases in gene expression at endogenous sites, and the position of effective target sites does not necessarily correspond to the optimal target sites that we observed for heterologous promoter activation. It is possible that our observed distance dependence is not generalizable to different promoters, that endogenous regulatory factors interfere with CRISPRa, or that different types of promoters require distinct transcriptional activation domains. In practice, if predictive rules for targeting arbitrary endogenous genes remain elusive, we envision using gene editing to introduce heterologous promoters at target sites of interest, which will enable us to regulate endogenous genes as part of a multi-gene CRISPRi/a control program.

A second outstanding challenge for bacterial CRISPRa is to develop a system that is portable across bacterial species of high commercial and industrial value, such as strains that have the ability to utilize carbon sources like CO_2, CO, methane, or lignocellulose, and alternative energy sources such as H_2 or light[1,2]. For gene repression, CRISPRi has been successfully used in a broad range of bacterial species[16,47-50], suggesting that the programmable DNA targeting component of the system is portable. For CRISPRa, many of our candidate gene activators interact with motifs on bacterial RNA polymerase that are highly conserved across a broad range of bacteria[51], including several that are relevant for industrial biosynthesis or microbiome engineering. In particular, SoxS interacts with a surface on the RNA polymerase α subunit that is well-conserved in gammaproteobacteria, alphaproteobacteria, bacteroides, gram-positive bacteria, and even to a lesser extent in cyanobacteria (Supplementary Fig. 8). This conserved interface may allow CRISPRa systems developed in *E. coli* to be ported to non-model bacteria with a wide range of useful biological functions.

Methods

Bacterial Strain Construction and Manipulation

Plasmid constructs were cloned and *E. coli* cells were cultured using standard molecular biology methods. Gene knockouts and integrations of fluorescent reporters were performed by recombineering in a strain with a genomically-integrated lambda red system under the control of a temperature-sensitive promoter (NM700)[52-54]. Modified genome fragments were then transferred to MG1655 by P1 transduction[55]. The sfGFP reporter was integrated at the *nfsA* locus. The mRFP reporter (for CRISPRi experiments) was integrated at *rbsAR*[56]. The ΔrpoZ strain was constructed by recombineering a pKD13-derived KanR linear PCR cassette with 36 base homology overhangs to flanking sites at the rpoZ locus. The ΔrpoZ::KanR knockout genomic fragment was transferred to MG1655 by P1 transduction, and the FRT-flanked KanR cassette was eliminated by transformation with the pCP20 helper plasmid to express FLP recombinase, followed by incubation at 42 °C to cure the plasmid. CRISPR-Cas system components were delivered on plasmids as described in Supplementary Table 3. *S. pyogenes* dCas9 was expressed from its endogenous *S. pyogenes* Cas9 promoter (cloned from pWJ66,[7] addgene #46570). Candidate effector proteins fused to RNA binding proteins (i.e. MCP and PCP) were expressed with the medium-strength BBa_J23107 promoter. Guide RNAs were expressed from the strong BBa_J23119 promoter[6]. BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org). zwf-LacZ and fumC-LacZ reporter genes were constructed following previously described designs (see Supplementary Methods)[57,58]. Complete annotated sequences of the reporter genes and CRISPR-Cas system components are included in the Supplementary Methods, along with a list the engineered *E. coli* strains (Supplementary Table 1). Guide RNA target sites are listed in Supplementary Table 2.

Flow Cytometry

Cells were inoculated in EZ-RDM (Teknova) supplemented with appropriate antibiotics and grown in 96-deep-well plates at 37 °C, 220 RPM overnight. Late stationary phase cultures were then diluted 1:40 in PBS and analyzed on a MACSQuant VYB flow cytometer (Miltenyi Biotec). To enrich for single cells, a side scatter threshold trigger (SSC-H) was applied. To gate for single bacterial cells, we first selected events along the diagonal of the SSC-H vs. SSC-A plot[59]. We then excluded events that appeared on the edges of the SSC-A vs. FSC-A plot, and events that appeared on the edge of the fluorescence histogram (Supplementary Fig. 9).

For inducible CRISPR system construction with pTet or pBAD promoters, strains were inoculated in 3 mL LB medium supplemented with antibiotics and grown overnight at 37 °C 220 RPM. On the next day, they were diluted 1:100 in 500 µL EZ-RDM supplemented with antibiotics and induced with 1 µM anhydrotetracycline (aTc) for pTet or 100 mM L-arabinose for pBAD. Non-induced controls were prepared for each strain. For pTet, cells were grown at 37 °C 220 RPM overnight. For pBAD, cells were grown at 37 °C 220 RPM for 6 hours. In both cases, fluorescence was assayed via flow cytometry as described above.

LacZ Reporter Assays

LacZ reporter assays for SoxS activation of *zwf* and *fumC* promoters were performed following a previously reported protocol with minor modifications[60,61]. Cultures were grown for 18 hrs in EZ-RDM supplemented with antibiotics. Absorbance at 600 nm (OD_{600}) was measured from 150 µL samples in a Biotek Synergy HTX plate reader. 20 µL aliquot from each culture were added to 80 µL permeabilization solution [100 mM Na_2HPO_4, 20 mM KCl, 2 mM $MgSO_4$, 0.8 mg/mL hexadecyltrimethylammonium bromide, 0.4 mg/mL sodium deoxycholate, 5.4 µL/mL β-mercaptoethanol]. Samples were incubated at 30 °C for 20 minutes. To initiate the reaction, 600 µL of substrate solution [60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/mL *o*-nitrophenyl-β-D-galactopyranoside, 2.7 µL/mL β-mercaptoethanol] was added. Samples were incubated at 30 °C until visible color developed, at which point 700 µL of stop solution [1 M Na_2CO_3] was added and the reaction time was recorded. Samples were centrifuged for 10 minutes to pellet cell debris, and the supernatants were removed. Absorbance at 420 nm (A_{420}) was measured from 150 µL samples in a plate reader. LacZ activity was calculated according to the formula: Miller Units = $(1000 \times A_{420})/(OD_{600} \times 0.02\ mL \times time)$.

J1 Reporter Sequence Design

Custom Python scripts were used to generate a tiling array containing an NGG PAM site every 10 nucleotides on each strand according to the following base unit: 5'-NNNCCNNNGG-3'. 10000 sequences of length 500 nt were generated by randomly sampling N nucleotides, adjusted for endogenous GC content in *E. coli* (BNID '100528 [http://bionumbers.hms.harvard.edu/bionumber.aspx?id=100528]')[62]. We discarded any sequences with four consecutive identical nucleotides. Sequences containing known transcription factor binding sites (using the experimental TF binding site dataset from RegulonDB)[63] were also discarded. Of the remaining sequences, we arbitrarily chose one, labelled as J1, and confirmed that it had no detectable homology to the *E. coli* genome. We placed a 170 bp fragment of this sequence upstream of the weak BBa_J23117 constitutive promoter (http://parts.igem.org).

Plate Reader Experiments

For fluorescent reporter experiments with the J1 reporter, cells were inoculated in 2 mL EZ-RDM supplemented with appropriate antibiotics and grown at 37 °C 220 RPM overnight. OD_{600} and observed fluorescence values were measured in a Biotek Synergy HTX plate reader using 150 µL of the overnight culture in flat, clear-bottomed 96-well plates (Corning). For mRFP detection, the excitation wavelength was 540 nm and the emission wavelength was 600 nm.

Ethanol Fermentations

E. coli MG1655 cells were transformed with the pCD355 plasmid containing the *pdc adhB* gene cassette from *Z. mobilis* under the control of a weak promoter (Supplementary Table 3) and with or without

components of the CRISPRa system. Transformed cells were inoculated in 5 mL LB (2% glucose) supplemented with appropriate antibiotics and grown overnight at 37 °C, 220 RPM. OD_{600} measurements were taken from the overnight cultures, and 2.5 OD·mL of the culture was diluted into 50 mL LB supplemented with antibiotics. Cultures were grown aerobically for 18 hours at 37 °C, 220 RPM. OD_{600} measurements were taken, and 150 OD·mL was pelleted and resuspended into 25 mL M9 media with 2% dextrose and antibiotics in a 125 mL conical flask. Flasks were sealed with rubber septa with a needle connected to a balloon to relieve pressure from CO_2 production. Cultures were incubated at 37 °C without shaking for 4 days. Ethanol concentrations in supernatants were measured with an Ethanol assay kit (R-Biopharm). We counted viable cells by plating on selective media before and after fermentation and observed no decrease in viable cells after the fermentation. To evaluate plasmid stability over the course of the fermentation, we performed minipreps on cells after the fermentation. Analytical restriction digests indicated no large-scale recombination events, and no point mutations were detected by sequencing.

Quantitative RT-PCR

Strains were inoculated in 5 mL LB supplemented with appropriate antibiotics and grown overnight at 37 °C, 220 RPM. Cultures were then diluted 1:100 in 5 mL EZ-RDM supplemented with antibiotics and grown to OD_{600} 0.5 (using 150 µL samples in flat clear bottomed 96-well plates in a Biotek Synergy HTX plate reader). Cultures were pelleted, flash-frozen in liquid nitrogen, and stored at -80 °C. Total RNA was extracted using an Aurum Total RNA Mini Kit (Bio-Rad). 1 µg of RNA was converted to cDNA using iScript reverse transcriptase (Bio-Rad) in 20 µL reactions. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) with a 58 °C annealing temperature and 10 µL reaction volumes. qPCR reactions were performed in triplicate on a CFX Connect (Bio-Rad) using 0.5 ng of cDNA, 400 nM primer concentration, and 15 s extension time. 16S rRNA was used for normalization. Control samples without a template and without reverse transcriptase were analyzed to confirm gene-specific amplification and absence of gDNA contamination. Primer sequences are listed in Supplementary Table 4. Expression levels for each gene were calculated relative to 16S using the $\Delta\Delta C_T$ method[64].

Code Availability

Custom Python code to generate tiling arrays with PAM sequences (*J1 Reporter Sequence Design*) is available upon request.

Data Availability

All data from this study are available upon request.

Acknowledgements

The authors thank Maureen Thomason, Mary Lidstrom, Frances Chu, Willy Voje, Jason Stevens, Chuhern Hwang, and members of the Zalatan and Carothers groups for technical assistance, advice, and helpful discussions. The NM700 strain was a gift from Nadim Majdalani and Susan Gottesman. This work was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (J.G.Z.), an NSF Award MCB 1517052 (J.M.C.), and a University of Washington Presidential Innovation Award (J.M.C.)

Author Contributions

J.G.Z. conceived the project. C.D., J.F., J.M.C., and J.G.Z. designed experiments, analyzed data and wrote the manuscript. C.D., J.F., and A.P. performed experiments.

Competing interests

The authors declare no competing interests.

References

1. Haynes, C. A. & Gonzalez, R. Rethinking biological activation of methane and conversion to liquid fuels. *Nat. Chem. Biol.* **10,** 331–339 (2014).
2. Lan, E. I. & Liao, J. C. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. *Bioresour. Technol.* **135,** 339–349 (2013).
3. Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. *Cell Systems* **1,** 62–71 (2015).
4. Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable expression tools enable single-cell strain distinction in the gut microbiome. *Cell* **169,** 538–546.e12 (2017).
5. Riglar, D. T. *et al.* Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat. Biotechnol.* **35,** 653–658 (2017).
6. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).
7. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41,** 7429–7437 (2013).
8. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).
9. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10,** 957–963 (2013).
10. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160,** 339–350 (2015).
11. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9

complex. *Nature* **517,** 583–588 (2015).

12. Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386,** 627–630 (1997).
13. Dove, S. L. & Hochschild, A. Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12,** 745–754 (1998).
14. Dove, S. L. & Hochschild, A. A bacterial two-hybrid system based on transcription activation. *Methods Mol. Biol.* **261,** 231–246 (2004).
15. Otoupal, P. B., Erickson, K. E., Escalas-Bordoy, A. & Chatterjee, A. CRISPR perturbation of gene expression alters bacterial fitness under stress and reveals underlying epistatic constraints. *ACS Synth. Biol.* **6,** 94–107 (2017).
16. Peters, J. M. *et al.* Bacterial CRISPR: accomplishments and prospects. *Curr. Opin. Microbiol.* **27,** 121–126 (2015).
17. Choi, K. R. & Lee, S. Y. CRISPR technologies for bacterial systems: Current achievements and future directions. *Biotechnol. Adv.* **34,** 1180–1209 (2016).
18. Peng, R. *et al.* CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in *Myxococcus xanthus*. *Microb. Cell Fact.* **17,** 15 (2018).
19. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17,** 5–15 (2016).
20. Martin, R. G., Gillette, W. K., Martin, N. I. & Rosner, J. L. Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in *Escherichia coli*. *Mol. Microbiol.* **43,** 355–370 (2002).
21. Busby, S. & Ebright, R. H. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293,** 199–213 (1999).
22. Griffith, K. L., Becker, S. M. & Wolf, R. E. Characterization of TetD as a transcriptional activator of a subset of genes of the *Escherichia coli* SoxS/MarA/Rob regulon. *Mol. Microbiol.* **56,** 1103–1117 (2005).
23. Jain, D. *et al.* Crystal structure of bacteriophage lambda cII and its DNA complex. *Molecular Cell* **19,** 259–269 (2005).
24. Twist, K.-A. F. *et al.* Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the β-subunit flap domain of *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA* **108,** 19961–19966 (2011).
25. Miller, A., Wood, D., Ebright, R. H. & Rothman-Denes, L. B. RNA polymerase β' subunit: a target of DNA binding-independent activation. *Science* **275,** 1655–1657 (1997).
26. Gregory, B. D., Deighan, P. & Hochschild, A. An artificial activator that contacts a normally occluded surface of the RNA polymerase holoenzyme. *J. Mol. Biol.* **353,** 497–506 (2005).
27. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57,** 441–466 (2003).

28. Griffith, K. L. & Wolf, R. E. Genetic evidence for pre-recruitment as the mechanism of transcription activation by SoxS of *Escherichia coli*: the dominance of DNA binding mutations of SoxS. *J. Mol. Biol.* **344,** 1–10 (2004).
29. Pomposiello, P. J., Bennik, M. H. & Demple, B. Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. *J. Bacteriol.* **183,** 3890–3902 (2001).
30. Griffith, K. L. & Wolf, R. E. Systematic mutagenesis of the DNA binding sites for SoxS in the *Escherichia coli zwf* and *fpr* promoters: identifying nucleotides required for DNA binding and transcription activation. *Mol. Microbiol.* **40,** 1141–1154 (2001).
31. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
32. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).
33. Rhee, S., Martin, R. G., Rosner, J. L. & Davies, D. R. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc. Natl. Acad. Sci. USA* **95,** 10413–10418 (1998).
34. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: identifying amino acids important for DNA binding and transcription activation. *J. Mol. Biol.* **322,** 237–257 (2002).
35. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–1171 (2013).
36. Fontana, J., Dong, C., Ham, J. Y., Zalatan, J. G. & Carothers, J. M. Regulated Expression of sgRNAs Tunes CRISPRi in *E. coli*. *Biotechnol. J.* (2018). doi:10.1002/biot.201800069
37. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177,** 4121–4130 (1995).
38. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* **15,** 103–105 (2008).
39. Zhou, H.-X. Polymer models of protein stability, folding, and interactions. *Biochemistry* **43,** 2141–2154 (2004).
40. Tan, S. Z. & Prather, K. L. Dynamic pathway regulation: recent advances and methods of construction. *Curr. Opin. Chem. Biol.* **41,** 28–35 (2017).
41. Ingram, L. O., Conway, T., Clark, D. P., Sewell, G. W. & Preston, J. F. Genetic engineering of ethanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* **53,** 2420–2425 (1987).
42. Alterthum, F. & Ingram, L. O. Efficient ethanol production from glucose, lactose, and xylose by recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* **55,** 1943–1948 (1989).
43. Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. & Ebright, R. H. Domain organization of RNA

polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78,** 889–896 (1994).

44. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159,** 647–661 (2014).
45. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30,** 354–359 (2012).
46. Skjoedt, M. L. *et al.* Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat. Chem. Biol.* **12,** 951–958 (2016).
47. Peters, J. M. *et al.* A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* **165,** 1493–1506 (2016).
48. Yao, L., Cengic, I., Anfelt, J. & Hudson, E. P. Multiple gene repression in cyanobacteria using CRISPRi. *ACS Synth. Biol.* **5,** 207–212 (2016).
49. Cleto, S., Jensen, J. V., Wendisch, V. F. & Lu, T. K. *Corynebacterium glutamicum* metabolic engineering with CRISPR interference (CRISPRi). *ACS Synth. Biol.* **5,** 375–385 (2016).
50. Tan, S. Z., Reisch, C. R. & Prather, K. L. J. A robust CRISPR interference gene repression system in *Pseudomonas*. *J. Bacteriol.* **200,** (2018).
51. Murakami, K. S. Structural biology of bacterial RNA polymerase. *Biomolecules* **5,** 848–864 (2015).
52. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97,** 6640–6645 (2000).
53. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97,** 5978–5983 (2000).
54. Court, D. L. *et al.* Mini-lambda: a tractable system for chromosome and BAC engineering. *Gene* **315,** 63–69 (2003).
55. Thomason, L. C., Costantino, N. & Court, D. L. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **Chapter 1,** Unit 1.17 (2007).
56. Sabri, S., Steen, J. A., Bongers, M., Nielsen, L. K. & Vickers, C. E. Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci. *Microb. Cell Fact.* **12,** 60 (2013).
57. Fawcett, W. P. & Wolf, R. E. Genetic definition of the *Escherichia coli zwf* 'soxbox,' the DNA binding site for SoxS-mediated induction of glucose 6-phosphate dehydrogenase in response to superoxide. *J. Bacteriol.* **177,** 1742–1750 (1995).
58. Fawcett, W. P. & Wolf, R. E. Purification of a MalE-SoxS fusion protein and identification of the control sites of *Escherichia coli* superoxide-inducible genes. *Mol. Microbiol.* **14,** 669–679 (1994).
59. Shapiro, H. M. Multiparameter flow cytometry of bacteria: implications for diagnostics and therapeutics. *Cytometry* **43,** 223–226 (2001).
60. Zhang, X. & Bremer, H. Control of the *Escherichia coli* rrnB P1 promoter strength by ppGpp. *J. Biol. Chem.* **270,** 11181–11189 (1995).

61. Beta-Galactosidase Assay (A better Miller). *OpenWetWare* Available at: https://openwetware.org/mediawiki/index.php?title=Beta-Galactosidase_Assay_(A_better_Miller)&oldid=620416. (Accessed: 17 May 2018)
62. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers - the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38,** D750–3 (2010).
63. Gama-Castro, S. *et al.* RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* **44,** D133–43 (2016).
64. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25,** 402–408 (2001).

Figures

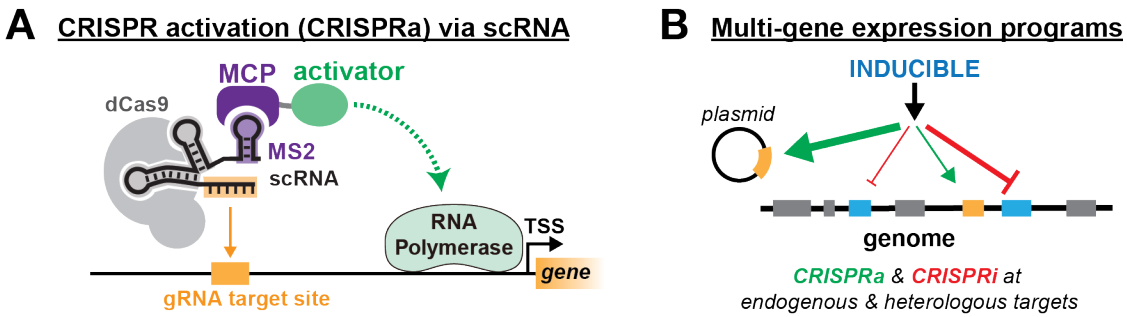


Figure 1. CRISPR activation in bacteria enables complex, multi-gene expression programs.

A) To activate gene expression, we target a CRISPR-Cas complex upstream of a target gene. dCas9 binds a scaffold RNA (scRNA), which is a modified gRNA that encodes both the target sequence and an RNA hairpin to recruit effector proteins that interact with RNA polymerase. The schematic depicts a 1x MS2 scRNA containing an MS2 RNA hairpin, which binds the MS2 coat protein (MCP) that is fused to candidate activator proteins[10].

B) Combining CRISPRi with CRISPRa enables multi-gene expression programs for simultaneous activation and repression. scRNAs that recruit activators can target genes for activation, while gRNAs targeted within a gene result in CRISPRi-based repression. If the CRISPR-Cas system components are controlled by inducible promoters, the entire gene expression program can be dynamically regulated.

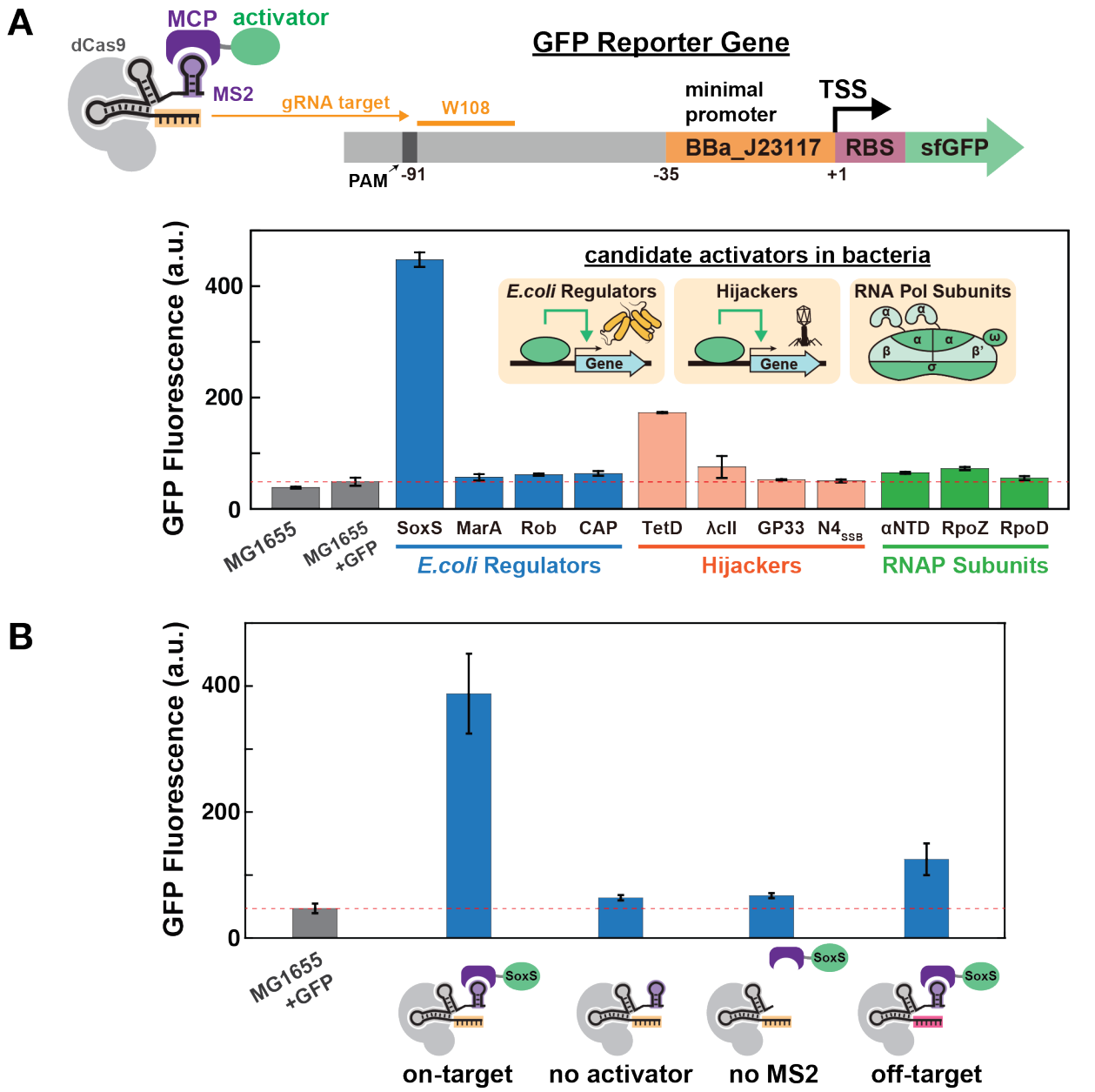


Figure 2. Effector proteins can activate reporter gene expression in *E. coli*.

A) The CRISPRa complex targets a GFP reporter gene (sfGFP, superfolder GFP) driven by a weak BBa_J23117 promoter. The gRNA target is the W108 sequence, located 91 bases upstream of the transcription start site (TSS). Several candidate activator proteins fused to MCP result in significant increases in GFP expression, including SoxS and TetD. The dotted red line indicates the background fluorescence level observed in the parental MG1655 *E. coli* strain containing the GFP reporter gene. GFP levels were measured in late stationary phase. Similar trends with smaller overall effects were observed in exponential phase.

B) All components of the CRISPRa complex must be present for gene activation. Significant GFP expression is observed when dCas9, the 1x MS2 scRNA, and MCP-SoxS are all expressed. When MCP-

SoxS is omitted (no activator) or the MS2 hairpin is removed (no MS2), there is no significant GFP expression. Modest GFP expression is observed when an off-target scRNA is used that has no target site in this strain (RR2, Supplementary Table 2), suggesting that overexpression of SoxS may have some off-target gene activation effects.

Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three biological replicates (specific values are indicated by black dots).

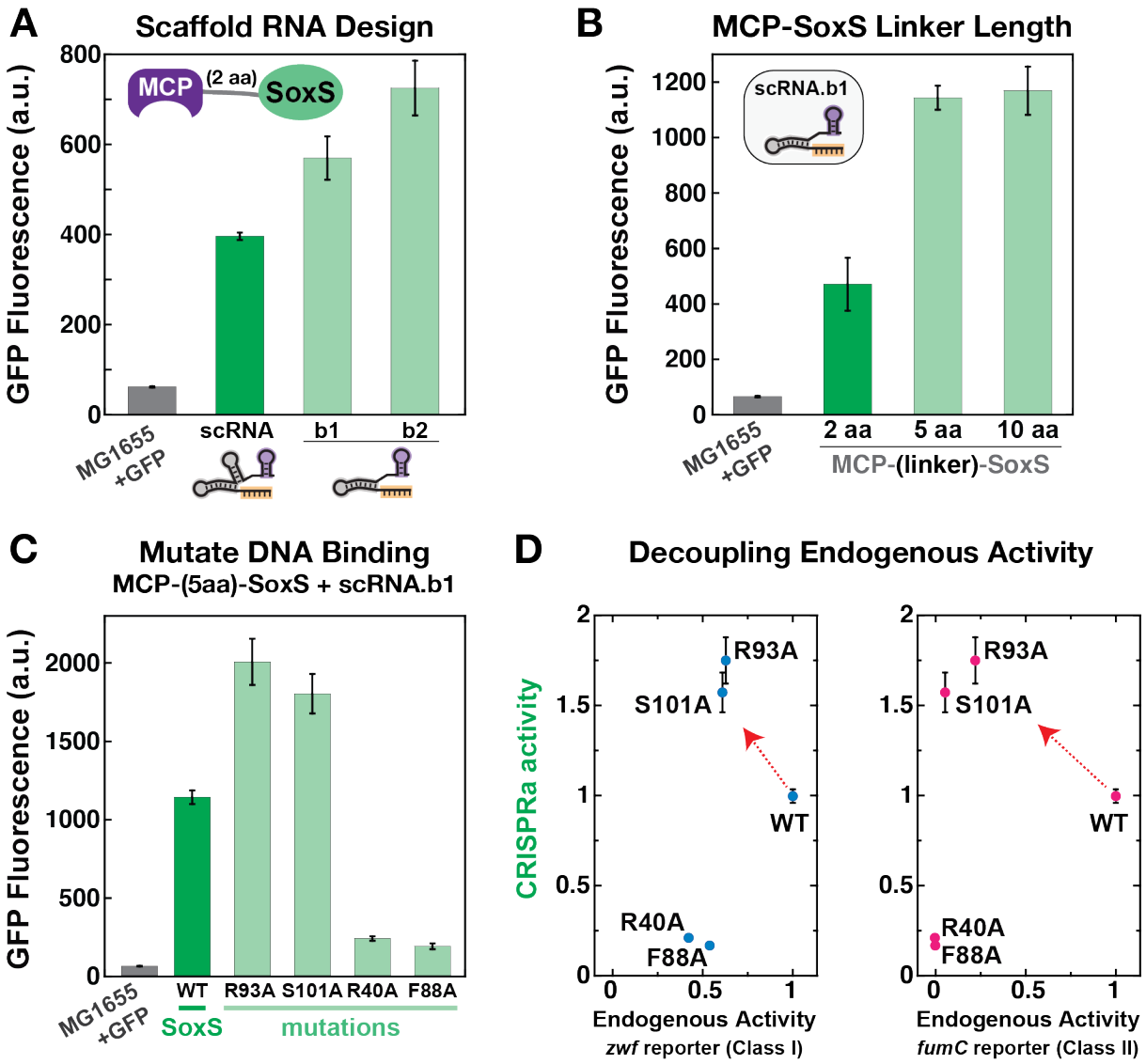


Figure 3. Optimization of gene activation.

A) Modifying the scRNA design to remove the tracr terminator hairpin results in a 1.5 to 2-fold increase in gene expression. Complete sequences of the original scRNA, scRNA.b1, and scRNA.b2 designs are included in the Supplementary Methods. scRNA.b1 and scRNA.b2 differ by one base from the 5' end of the terminator hairpin.

B) Increasing the linker length between MCP and SoxS from 2 to 5 or 10 amino acids increases GFP expression by 2 to 3-fold.

C) Mutations in SoxS at the binding interface to endogenous DNA targets have variable effects on activity in a CRISPRa assay. Point mutations at SoxS R93A or S101A result in a 2-fold increase in GFP expression, while SoxS R40A and F88A lead to substantial decreases in GFP expression. Based on the structure of the SoxS homolog MarA[33], residues R93, S101, and R40 are surface exposed while the F88 side chain points into the hydrophobic core.

D) Plots of CRISPRa activity vs endogenous activity for wild type (wt) and mutant SoxS proteins indicate that transcriptional activation can be decoupled from binding to endogenous targets. CRISPRa activity values are GFP fluorescence levels (Fig. 3C) normalized to the value obtained for wt SoxS. Endogenous activity values are from LacZ reporter assays with endogenous SoxS promoters for the *zwf* and *fumC* genes (Supplementary Fig. 3A & B), corrected for background reporter activity in the absence of SoxS and normalized to the value obtained for wt SoxS. *zwf* is representative of class I SoxS target genes in which the SoxS site is located upstream of the -35 site, and *fumC* is representative of class II SoxS target genes in which the SoxS site overlaps the -35 site (Supplementary Fig. 3A). The data shown for CRISPRa and endogenous activity were obtained in separate reporter strains. Similar results were obtained when CRISPRa values were measured in strains with both the GFP and LacZ reporters (Supplementary Fig. 3C). Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three biological replicates (specific values are indicated by black dots).

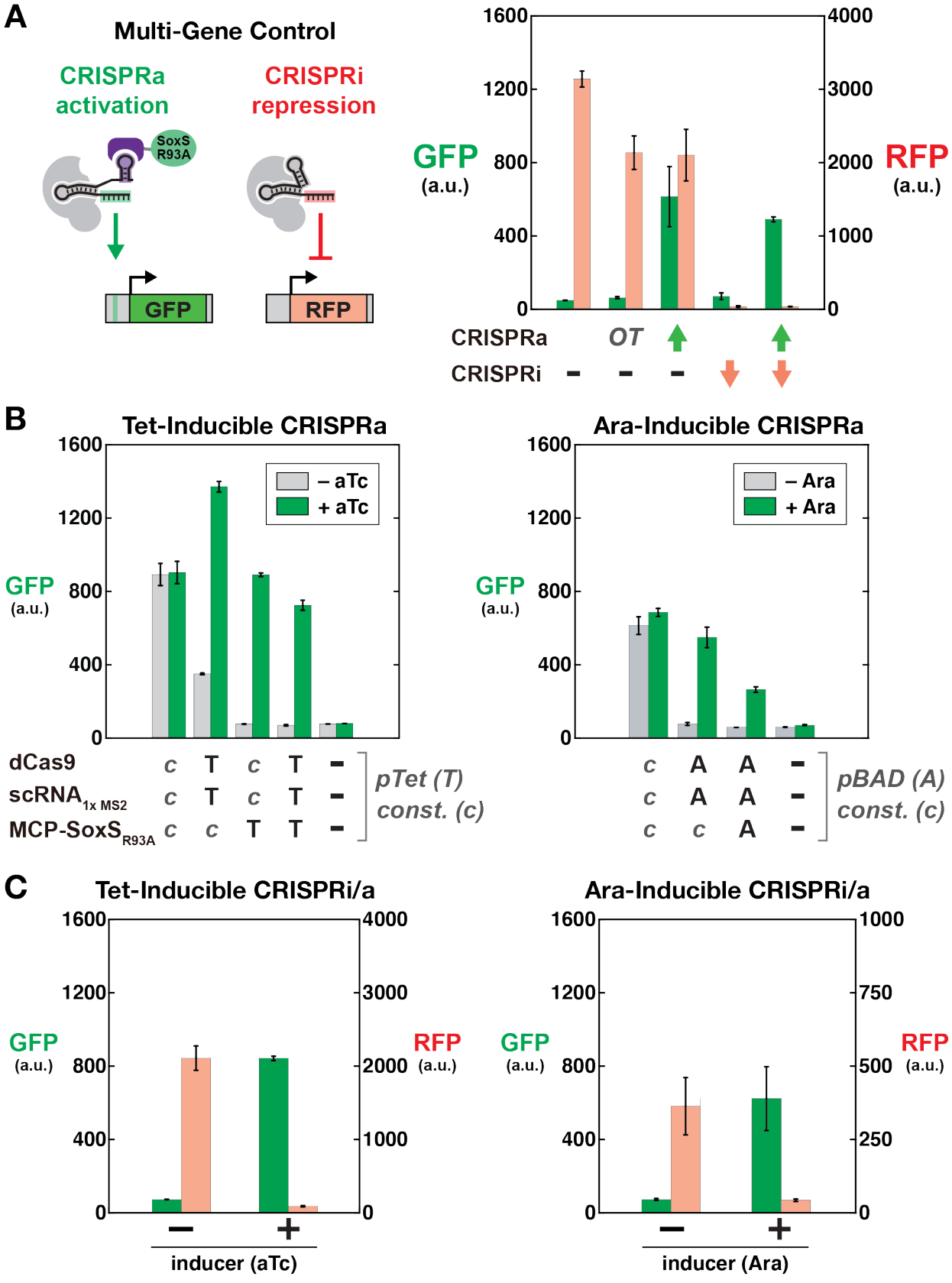

A
Multi-Gene Control
CRISPRa activation
CRISPRi repression
SoxS R93A
GFP
RFP
GFP (a.u.)
RFP (a.u.)
1600
1200
800
400
0
4000
3000
2000
1000
0
CRISPRa
OT
CRISPRi
B
Tet-Inducible CRISPRa
– aTc
+ aTc
GFP (a.u.)
dCas9
scRNA1x MS2
MCP-SoxSR93A
c T c T –
c T c T –
c c T T –
pTet (T)
const. (c)
Ara-Inducible CRISPRa
– Ara
+ Ara
c A A –
c A A –
c c A –
pBAD (A)
const. (c)
C
Tet-Inducible CRISPRi/a
Ara-Inducible CRISPRi/a
1000
750
500
250
– +
inducer (aTc)
inducer (Ara)

Figure 4. Multi-gene, inducible control.

A) Expressing multiple gRNAs allows simultaneous regulation of multiple genes. Using an *E. coli* strain with integrated GFP and RFP reporters, a 1x MS2 scRNA targets GFP for activation by recruiting MCP-(5aa)-$SoxS_{R93A}$, while an unmodified gRNA targets RFP for repression. When both gRNAs are expressed, GFP expression increases while RFP expression decreases, and the observed effects are similar to those observed when each gRNA is expressed alone. OT indicates an off-target control for the J106 target site (Supplementary Table 2), which is not present in this strain. Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for three biological replicates (specific values are indicated by black dots).

B) CRISPRa can be inducibly controlled with pTet or pBAD (Ara) promoters. In each case, different components of the CRISPRa system (i.e. dCas9, the 1x MS2 scRNA, or MCP-(5aa)-$SoxS_{R93A}$) are controlled by constitutive (c) or inducible (pTet, T or pBAD, A) promoters. pTet is induced with aTc and pBAD is induced with arabinose. For pTet inductions, cultures were harvested after overnight growth (late stationary phase), while for arabinose inductions cultures were harvested after 6 hours (early stationary phase). Values reported are GFP or RFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for three biological replicates (specific values are indicated by black dots).

C) CRISPRa and CRISPRi can be simultaneously induced with pTet or pBAD (Ara) promoters. For pTet-induction, all components of the CRISPR system are controlled by pTet. For arabinose induction the pBAD promoter controls dCas9 and the guide RNAs, while MCP-(5aa)-$SoxS_{R93A}$ is constitutively expressed. The RFP axis in the Ara panel is smaller than all other RFP axes in this figure because we observed consistently lower fluorescent protein expression in early stationary phase cultures versus late stationary phase cultures, even in the absence of CRISPRi.

Values reported are GFP or RFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for three biological replicates, except for the +aTc tet inductions, for which two biological replicates were obtained (specific values are indicated by black dots).

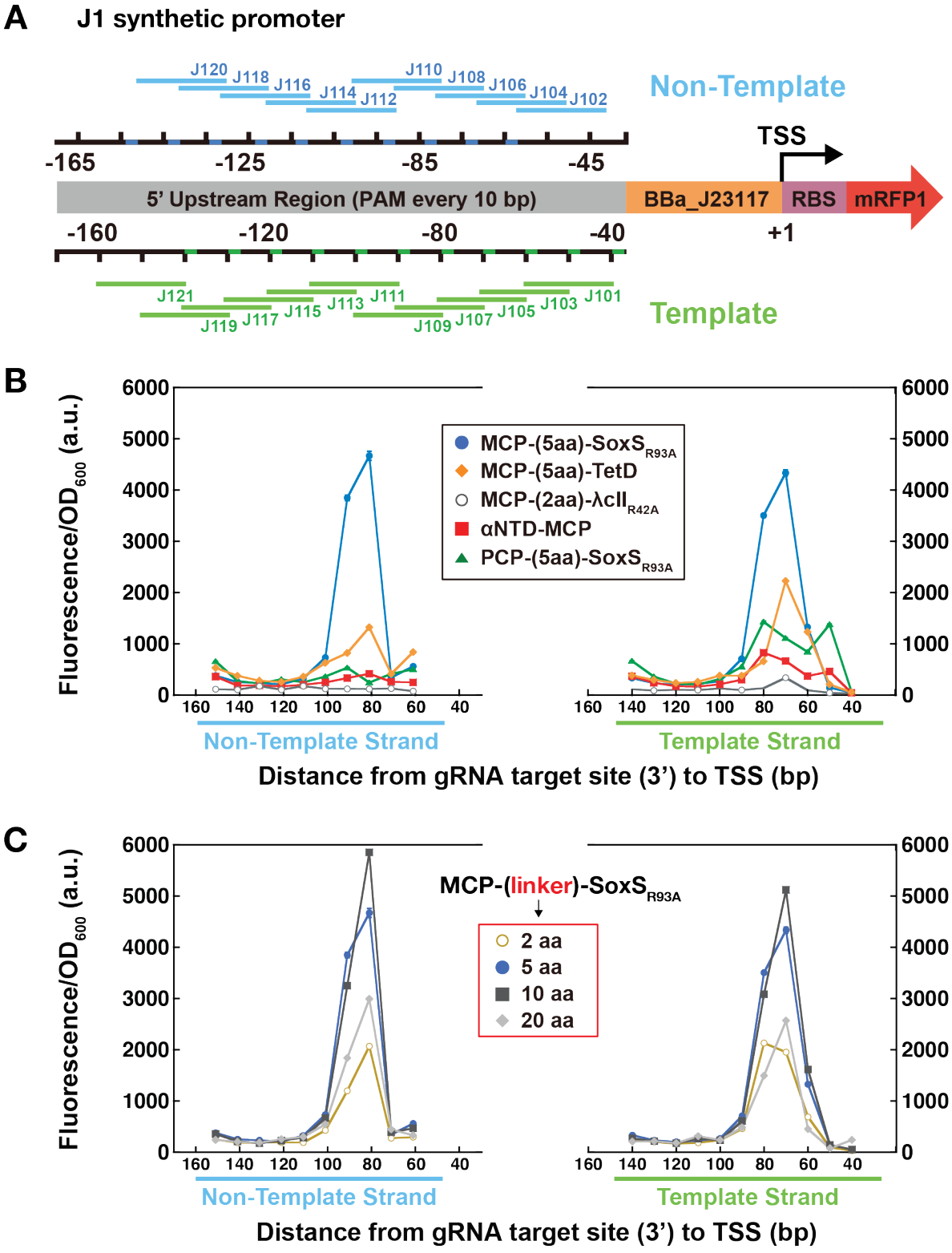


Figure 5. Gene activation is highly sensitive to CRISPR target site position.

A) The J1 synthetic promoter has potential gRNA target sites with an appropriately positioned PAM sequence every 10 bases on both strands upstream of a weak BBa_J23117 minimal promoter driving an

RFP reporter gene.

B) A plot of RFP expression level versus target site position indicates a narrow region for effective CRISPRa from -80 to -90 upstream of the TSS on the non-template strand and from -50 to -80 on the template strand. The plot shows CRISPRa with several different activators recruited by a 1x MS2 scRNA.b2, including MCP-(5aa)-SoxS$_{R93A}$, MCP-(5aa)-TetD (see Supplementary Fig. 2D), MCP-(2aa)-λcII$_{R42A}$ (see Supplementary Fig. 2E), and αNTD-MCP. We also used a 1x PP7 scRNA.b1 to recruit PCP-(5aa)-SoxS$_{R93A}$.

C) Increasing the linker length between MCP and SoxS$_{R93A}$ does not extend the range of gRNA target sites that are effective for CRISPRa. Maximal activity is observed with 5 and 10 amino acid linkers, while both 2 and 20 amino acid linkers show reduced activity but no significant difference in the position of effective target sites.

Values reported are RFP fluorescence levels normalized by the cell density (fluorescence/OD$_{600}$) measured in a plate reader. Values are mean ± s.d. for at least three biological replicates.

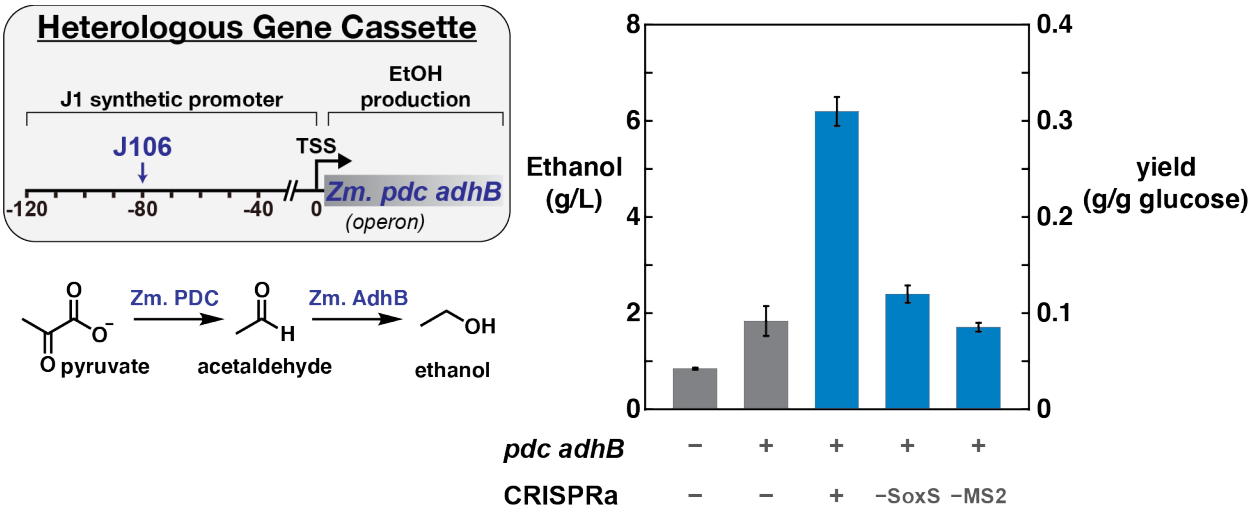


Figure 6. CRISPRa-mediated ethanol production.

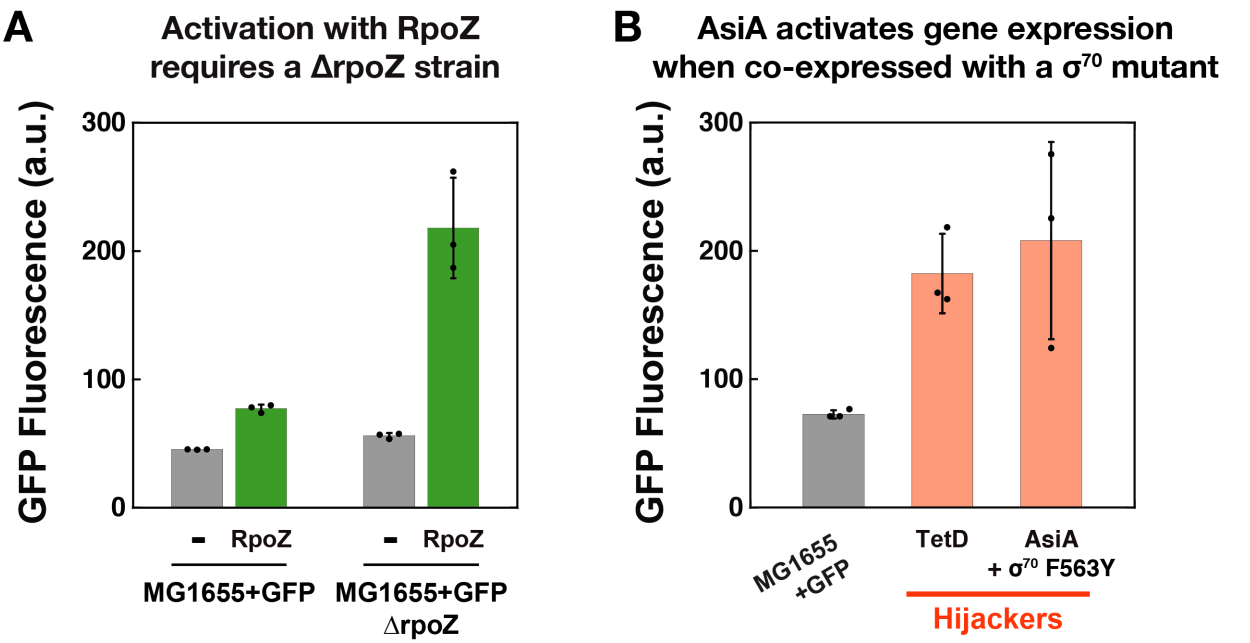
A heterologous gene cassette for pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adhB*) from *Z. mobilis* converts pyruvate to ethanol in *E. coli*. The gene cassette is controlled by a weak promoter (see Supplementary Methods). When a CRISPRa complex (dCas9, MCP-(5aa)-SoxS$_{R93A}$, and 1x MS2 scRNA.b1) is targeted to the J106 site upstream of the promoter, ethanol production increases approximately 3-fold. In the absence of MCP-(5aa)-SoxS$_{R93A}$ (-SoxS) or with a gRNA lacking the MS2 hairpin (-MS2), ethanol levels are indistinguishable from background strain containing just the *pdc adhB* gene cassette. The left y-axis is ethanol titer (g/L) and the right y-axis is the same data represented as yield (g ethanol/g glucose). Yield is calculated relative to the initial 2% glucose (20 g/L). Values are the mean for two biological replicates (specific values are indicated by black dots).

Chapter 1: Supplementary Information

Chen Dong[1], Jason Fontana[2], Anika Patel[1], James M. Carothers[2,3,4], & Jesse G. Zalatan[1,2,4]

[1]Department of Chemistry
[2]Molecular Engineering & Sciences Institute
[3]Department of Chemical Engineering
[4]Center for Synthetic Biology
University of Washington, Seattle, WA 98195, USA

Supplementary Figures

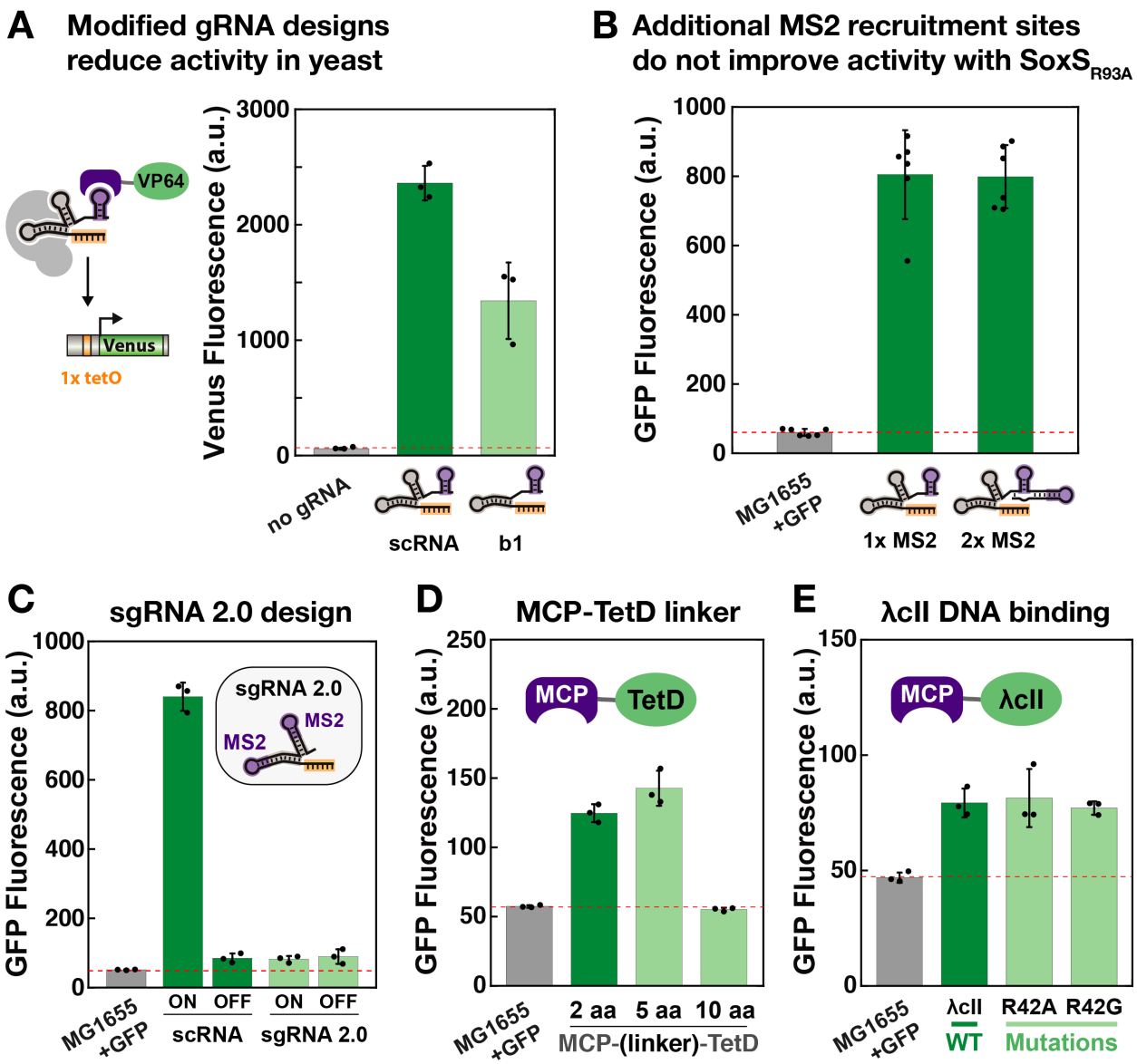


Supplementary Figure 1. Some candidate transcriptional activation domains require additional strain modifications for effective reporter gene activation.

A) CRISPRa with the MCP-RpoZ activator recruited via an scRNA is significantly increased in a ΔrpoZ host strain, consistent with that observed previously for other RpoZ fusion proteins including dCas9-RpoZ[1,2]. GFP reporter strains were transformed with dCas9 and a 1x MS2 scRNA, and either with or without the MCP-RpoZ fusion protein.

B) CRISPRa with an MCP-AsiA transcriptional activator, co-transformed with a σ^{70} F563Y mutant, produces GFP expression levels comparable to that obtained with MCP-TetD. The σ^{70} F563Y mutant prevents toxicity that occurs when AsiA is expressed alone and inhibits the activity of the endogenous σ^{70} subunit[3]. Attempts to transform MCP-AsiA into *E. coli* without co-transforming σ^{70} F563Y were unsuccessful unless a substantially weaker promoter (BBa_J23112) was used for MCP-AsiA, and no detectable GFP expression was observed with this construct (data not shown).

Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three biological replicates (specific values are indicated by black dots).



Supplementary Figure 2. Optimization of gene activation.

A) Removing the 3' terminator hairpin from a 1x MS2 scRNA decreases CRISPRa in a yeast reporter system. See Supplementary Methods for complete sequences of 1x MS2 scRNA and 1x MS2 scRNA.b1. Experiments were performed in yeast as previously described using a single TetO target site to activate a Venus fluorescent reporter gene by recruiting MCP-VP64, an activator of eukaryotic transcription[4].

B) A 2x MS2 scRNA, which increases CRISPRa in yeast and human cells relative to a 1x MS2 scRNA[4], does not improve GFP expression with CRISPRa in *E. coli*. The activator in this experiment is MCP-(5aa)-$SoxS_{R93A}$ (Fig. 3).

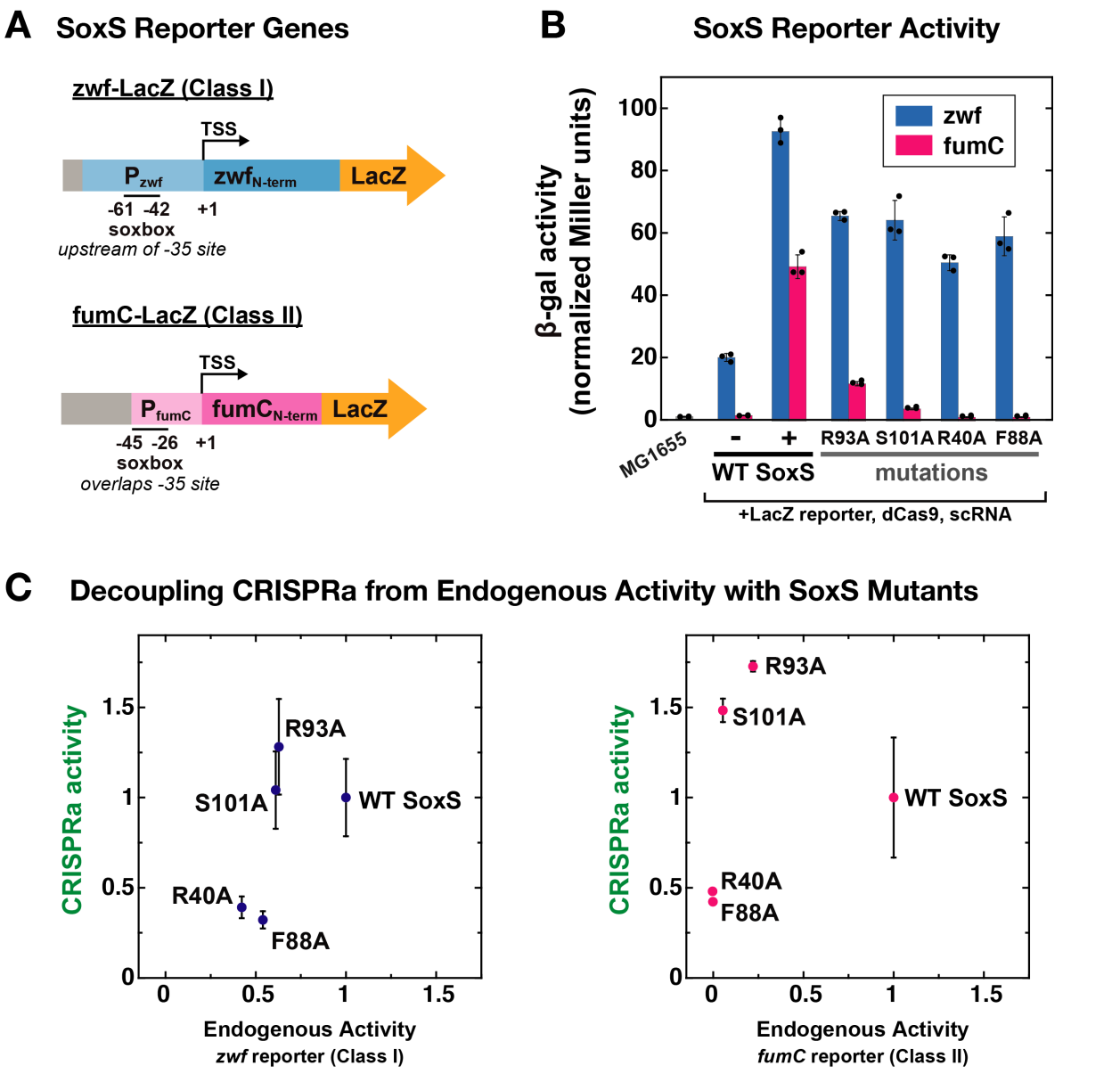
C) sgRNA 2.0 does not activate GFP expression in *E. coli*. The sgRNA 2.0 design has two MS2 hairpins embedded wthin internal hairpins of the sgRNA and is very effective for CRISPRa in eukaryotic cells[5]. We

expressed either 1x MS2 scRNA or sgRNA 2.0 with a W108 target site (ON) or an off-target RR2 sequence (OFF). The activator in this experiment is MCP-(5aa)-SoxS$_{R93A}$ (Fig. 3).

D) Increasing the linker length between MCP and TetD from 2 to 5 amino acids modestly increases GFP expression, while increasing the linker further to 10 amino acids decreases GFP expression to background levels. The optimized 1x MS2.b1 scRNA was used in this experiment.

E) Mutations in the MCP-λcII activator at R42, which is essential for DNA binding, retain activity when recruited via CRISPRa in *E. coli*, although the absolute level of GFP expression remains low compared to CRISPRa with MCP-SoxS. The unmodified 1x MS2 scRNA was used in this experiment. λcII R42G was previously described[6].

Values reported are Venus or GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three measurements (specific values are indicated by black dots).



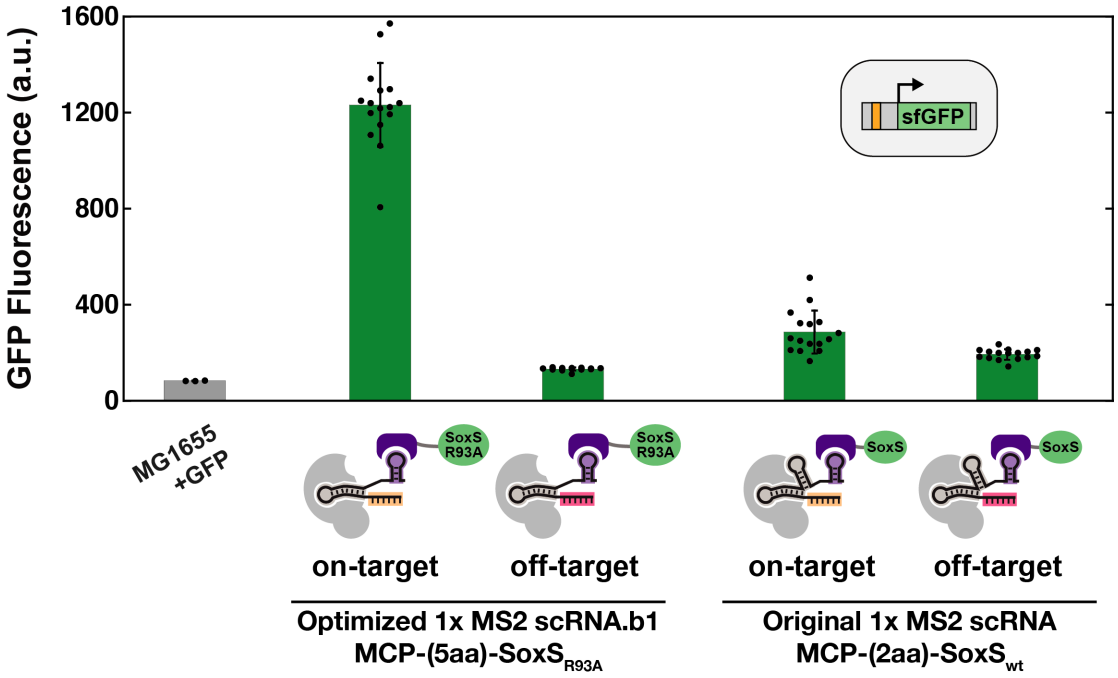
Supplementary Figure 3. Mutations in SoxS reduce activity at endogenous SoxS promoters.

A) Plasmids containing zwf-LacZ and fumC-LacZ fusion constructs were used as reporters of endogenous SoxS transcriptional activity[7]. The *zwf* promoter is representative of SoxS class I promoters in which the SoxS target site (soxbox) is upstream of the -35 site. The *fumC* promoter is representative of SoxS class II promoters in which the soxbox overlaps the -35 site. Complete sequences of the reporter constructs are included in Supplementary Methods.

B) LacZ (β-Gal) activity was measured in *E. coli* strain CD06 (Supplementary Table 1), an MG1655 strain modified to include a GFP reporter for CRISPRa activity. This strain was further transformed with a plasmid containing either the zwf-LacZ or fumC-LacZ reporter, along with a plasmid with the CRISPRa system components dCas9, a 1x MS2 scRNA.b1, and an MCP-SoxS fusion protein. Point mutants of SoxS that

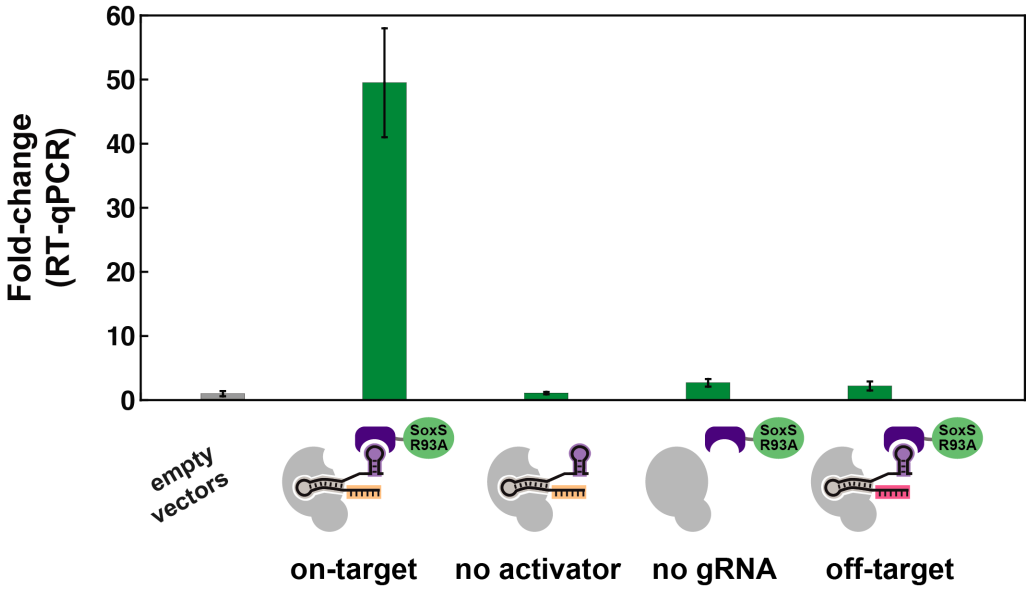
disrupt the DNA binding interface lead to reduced LacZ activity, and the results are consistent with previously reported activity trends for these mutants[7]. LacZ activity was measured as previously described[8], and values are reported in Miller units normalized to the value obtained in the MG1655 parent strain with no reporter plasmids. Values are median ± s.d. for three measurements (specific values are indicated by black dots).

C) Plots of CRISPRa activity vs endogenous activity for wild type (wt) and mutant SoxS proteins indicate that transcriptional activation can be decoupled from binding to endogenous targets. In the main text Fig. 3D, CRISPRa activity values were measured in strains that did not contain LacZ reporters (Fig. 3C). The CRISPRa activity values plotted here were obtained in the same strains used to measure LacZ activity. Similar trends were observed in both cases. CRISPRa activity values were normalized to the value obtained for wt SoxS. Endogenous activity values are LacZ activity (Supplementary Fig. 3B), corrected for background reporter activity in the absence of SoxS and normalized to the value obtained for wt SoxS. Values are median ± s.d. for at least three measurements.



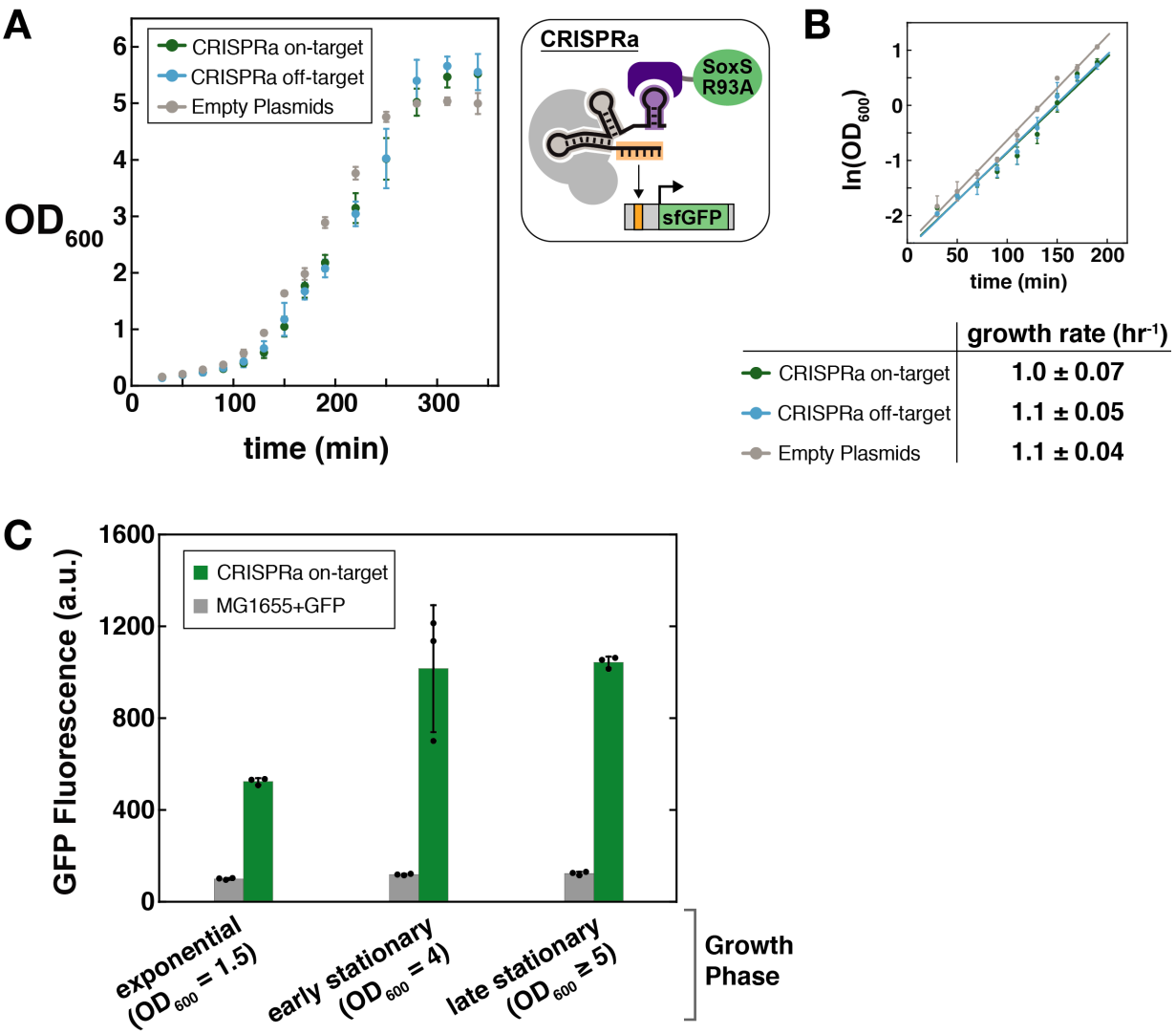
Supplementary Figure 4. An optimized CRISPRa system improves activity.

The optimized CRISPRa system with 1x MS2 scRNA.b1 and MCP-(5aa)-SoxS$_{R93A}$ produces significantly more on-target GFP expression and less off-target GFP expression than the original system with 1x MS2 scRNA with MCP-(2aa)-SoxS. The on-target site is W108 and the off-target site is RR2 (Supplementary Table 2). Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three biological replicates (specific values are indicated by black dots).



Supplementary Figure 5. CRISPRa significantly increases GFP mRNA levels.

GFP expression with a fully optimized scRNA.b1, MCP-(5aa)-SoxS$_{R93A}$ CRISPRa system increases 50-fold relative to an empty vector control, as measured by RT-qPCR. Negative controls without MCP-(5aa)-SoxS$_{R93A}$ (no activator) or scRNA.b1 produce 1.1-fold and 2.7-fold changes in GFP mRNA levels, respectively, relative to the empty vector control. When an off-target scRNA is expressed (RR2, Supplementary Table 2), there is a 2.2-fold change in GFP mRNA relative to the empty vector control. For comparison to the 50-fold increase in GFP mRNA levels with CRISPRa, we observe a 30-fold increase in GFP fluorescence in the same strain (Fig. 3E). The fluorescence change may be an underestimate of the true change in protein levels, as there is a significant autofluorescence background (see for example the observed GFP levels in the MG1655 parental strain versus MG1655+GFP in Fig. 2A). Expression levels from RT-qPCR were calculated from at least three replicates using the $\Delta\Delta C_T$ method[9] with three independent measurements each of the target gene (GFP) and reference gene (16S rRNA). Error bars represent s.e.m. Because the mean and error of fold changes in the $\Delta\Delta C_T$ method are calculated from the average and error of the individual C_T values, individual data points are not plotted.

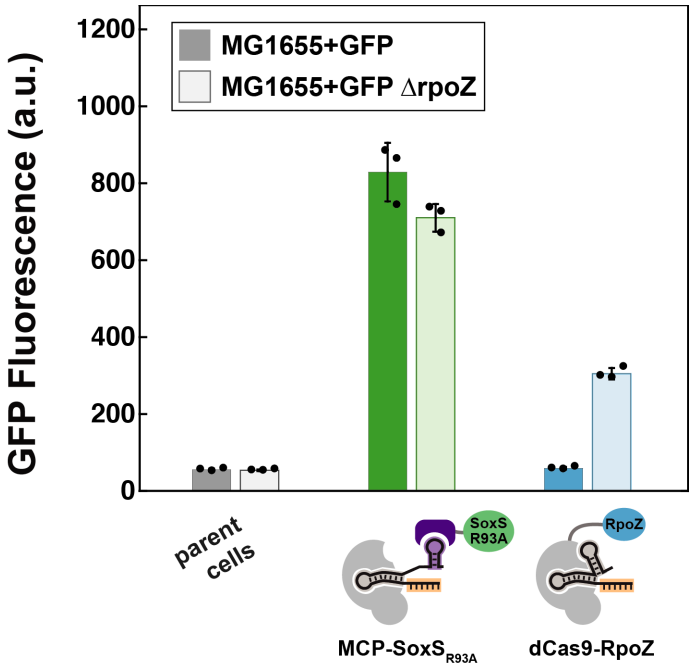


Supplementary Figure 6. Relationship between cell growth and CRISPRa.

A) The CRISPR-Cas system does not introduce a significant growth burden. Cell growth (measured by OD_{600}) versus time for *E. coli* cells transformed with empty plasmids or with CRISPRa system components, using 1x MS2 scRNAs with either on or off-target sequences. The growth curves are similar in all cases. Experiments were performed in biological triplicate, and error bars represent the standard deviation from at least three measurements.

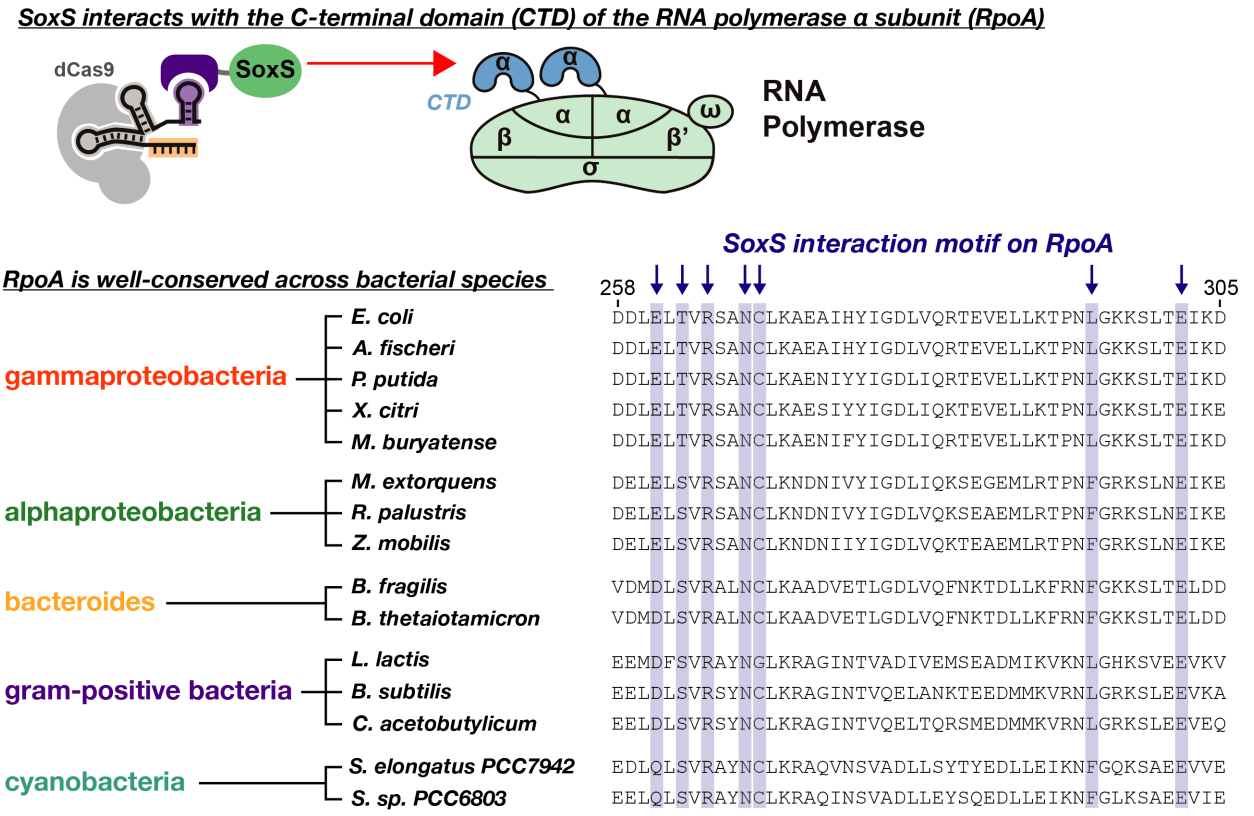
B) Quantification of growth rates using exponential phase data only (up to 200 minutes). The slope of the plot of $\ln(OD_{600})$ versus time gives the growth rate. The observed growth rates are identical within error.

C) CRISPRa-mediated GFP expression increases as cells enter stationary phase, although gene expression is still detectable in exponential phase. Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three biological replicates (specific values are indicated by black dots).



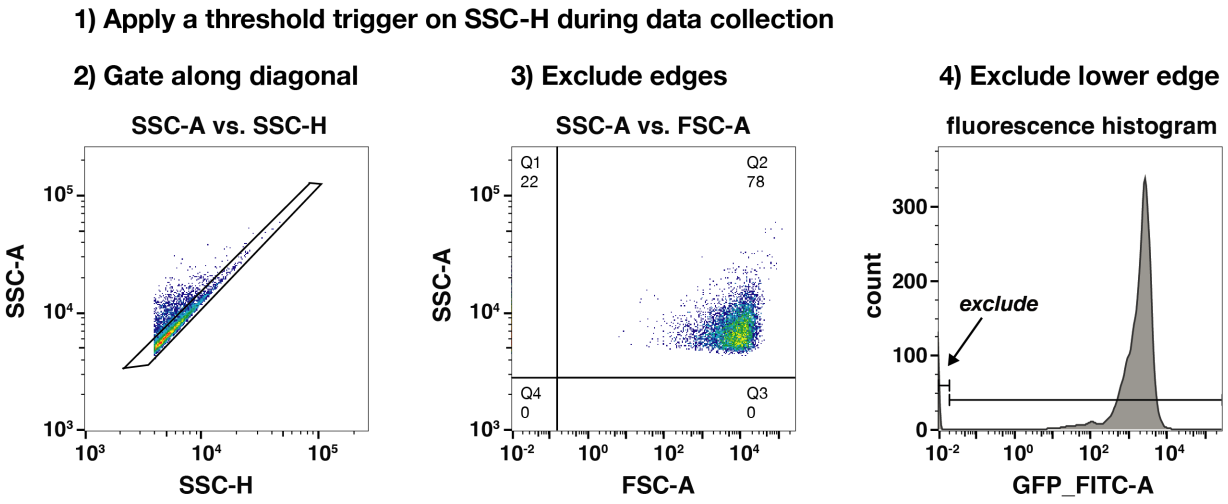
Supplementary Figure 7. CRISPRa with fully optimized MCP-SoxS$_{R93A}$ is more effective than dCas9-RpoZ.

MCP-SoxS$_{R93A}$ activates gene expression in an MG1655 strain, where dCas9-RpoZ [1] is ineffective. In a Δ*rpoZ* strain, where dCas9-RpoZ is effective, the SoxS-based system outperforms dCas9-RpoZ by >2-fold. Values reported are GFP fluorescence levels measured by flow cytometry. Values are median ± s.d. for at least three measurements (specific values are indicated by black dots).



Supplementary Figure 8. The SoxS recruitment site on RNA polymerase is highly conserved.

E. coli SoxS interacts with the C-terminal domain of RpoA, the polymerase α-subunit (αCTD), and specific amino acids on RpoA important for interacting with SoxS have been identified[10]. These amino acids in RpoA are largely conserved across a broad range of bacterial species. The observed conservation suggests that *E. coli* SoxS may be able to recruit RNA polymerase in other bacteria in addition to *E. coli*. Complete RpoA sequences were aligned using Clustal Omega[11], and the sequences corresponding *E. coli* residues 258-305 are shown. At least seven residues have been reported to affect the *E. coli* SoxS-αCTD interaction upon alanine substitution[10]. Alanine substitutions at E261, T263, and R265 have significant detrimental effects on the SoxS-αCTD interaction. R265 is absolutely conserved among the species examined here, and E261 and T263 are largely conserved with a few homologous substitutions (i.e. E->D/Q or T->S). Alanine substitutions at L295 and E302 have modest detrimental effects. L295 is largely conserved with some homologous substitutions (L->F), and E302 is absolutely conserved. Alanine substitutions at N268 and C269 have modest beneficial effects on the SoxS-αCTD interaction; both residues are absolutely conserved except for a C->G substitution in *L. lactis*.



Supplementary Figure 9. Flow cytometry gating strategy.

To select for single bacterial cells, we (1) applied a side scatter threshold trigger (SSC-H) during data collection. We then gated for single cells by (2) selecting events along the diagonal of the SSC-H vs. SSC-A plot[12], (3) excluding events that appeared on the edges of the SSC-A vs. FSC-A plot, and (4) excluding events that appeared on the edges of the fluorescence histogram.

Supplementary Tables

Supplementary Table 1. *E. coli* Strains

Strain	**Description**	**Genotype**
MG1655	parent *E. coli* strain	F- λ- ilvG- rfb-50 rph-1
CD03	MG1655/ΔrpoZ	MG1655 ΔrpoZ
CD06	MG1655/sfGFP (weak promoter)	MG1655 *W1-BBa_J23117-sfGFP KanR::nfsA*
CD08	MG1655/sfGFP (weak promoter)/ΔrpoZ	MG1655 *W1-BBa_J23117-sfGFP KanR::nfsA ΔrpoZ*
CD10	MG1655/sfGFP (weak promoter)/ mRFP1 (strong promoter	MG1655 *W1-BBa_J23117-sfGFP::nfsA, BBa_J23119-mRFP1 KanR::rbsAR*

Supplementary Table 2. gRNA Target Sites

sgRNA target	DNA Sequence	Target Strand[c]	Distance to TSS[d]
W108[a]	GAAGATCCGGCCTGCAGCCA	NT	91
RFP R2[b]	TGGAACCGTACTGGAACTGC	NT	-215
J101	TGGGTTCCACCGGATACCTC	T	40
J103	AGGCGTCCTTTGGGTTCCAC	T	50
J105	CGGTTACCAAAGGCGTCCTT	T	60
J107	CGGTGTCCTGCGGTTACCAA	T	70
J109	AGGTATCCTGCGGTGTCCTG	T	80
J111	GGGCGACCTCAGGTATCCTG	T	90
J113	GGGCCACCACGGGCGACCTC	T	100
J115	TGGTGACCATGGGCCACCAC	T	110
J117	GGGTGACCTATGGTGACCAT	T	120
J119	TGGTTGCCAAGGGTGACCTA	T	130
J121	AGGACACCTTTGGTTGCCAA	T	140
J102	AGGTATCCGGTGGAACCCAA	NT	61
J104	TGGAACCCAAAGGACGCCTT	NT	71
J106	AGGACGCCTTTGGTAACCGC	NT	81
J108	TGGTAACCGCAGGACACCGC	NT	91
J110	AGGACACCGCAGGATACCTG	NT	101
J112	AGGATACCTGAGGTCGCCCG	NT	111
J114	AGGTCGCCCGTGGTGGCCCA	NT	121
J116	TGGTGGCCCATGGTCACCAT	NT	131
J118	TGGTCACCATAGGTCACCCT	NT	141
J120	AGGTCACCCTTGGCAACCAA	NT	151

[a] The W108 site for GFP activation was described previously using a 30 base gRNA targeting site (i.e. spacer sequence)[1]. We use a 20 base gRNA targeting site for the same site (i.e. the same PAM site).

[b] The RFP R2 site for RFP repression by CRISPRi was described previously[13].

[c] Template strand (T) or non-template strand (NT).

[d] Distance to TSS is the distance from the 3' end (PAM proximal) of the guide target site to the transcription start site. For synthetic promoters driven by BBa_J23117 or BBa_J23119 (http://parts.igem.org), the TSS is immediately downstream of the BBa sequence (see complete maps below).

Supplementary Table 3. *E. coli* Expression Plasmids[a]

Plasmid	Marker	origin	Promoter	Gene	Terminator
pCD067	*1) AmpR* *2) KanR*	*R6K*	*BBa_J23117*	sfGFP	*BBa_B0015*
pJF023	*1) AmpR* *2) KanR*	*R6K*	*BBa_J23119*	mRFP1	*BBa_B0015*
p_gRNA_bact_RR2[b]	*AmpR*	*ColE1*	*BBa_J23119*	sgRNA (RR2)	*TrrnB*
pCD005	*AmpR*	*ColE1*	*BBa_J23119*	1x MS2 scRNA (W108 target)	*TrrnB*
pCD006	*AmpR*	*ColE1*	*BBa_J23119*	2x MS2 scRNA (W108 target)	*TrrnB*
pCD015	*AmpR*	*ColE1*	*BBa_J23119*	sgRNA (W108 target)	*TrrnB*
pCD061	*AmpR*	*ColE1*	*BBa_J23119*	1x MS2 scRNA.b1 (W108 target)	*TrrnB*
pJF67-2	*AmpR*	*ColE1*	*BBa_J23119*	1x MS2 scRNA.b2 (W108 target)	*TrrnB*
pCD034	*AmpR*	*ColE1*	*BBa_J23119*	1x MS2 scRNA (RR2)	*TrrnB*
pCD315	*AmpR*	*ColE1*	*pBAD*	1x MS2 scRNA.b1 (W108 target)	*TrrnB*
pCD325	*AmpR*	*ColE1*	*pBAD*	sgRNA (RR2)	*TrrnB*
pCD326	*AmpR*	*ColE1*	*1) pBAD* *2) pBAD*	1) 1x MS2 scRNA.b1 (W108 target) 2) sgRNA (RR2)	*1) TrrnB* *2) TrrnB*
pCD372	*AmpR*	*ColE1*	*1)* *BBa_J23119* *2)* *BBa_J23119*	1) 1x MS2 scRNA.b1 (W108 target) 2) sgRNA (RR2)	*1) TrrnB* *2) TrrnB*
pCD403	*AmpR*	*ColE1*	*BBa_J23119*	1x MS2 scRNA.b1 (RR2)	*TrrnB*
pCD017	*CmR*	*p15A*	*Sp.pCas9*	dCas9	*BBa_B0015*
pCD018	*CmR*	*p15A*	*1) Sp.pCas9* *2)*	1) dCas9 2) MCP-rpoZ	*1)* *BBa_B0015*

			BBa_J23107		*2) BBa_B1002*
pCD037	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) αNTD-MCP	*1) BBa_B0015 2) BBa_B1002*
pCD064	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-rpoD	*1) BBa_B0015 2) BBa_B1002*
pCD087	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-N4SSB (Y75A)	*1) BBa_B0015 2) BBa_B1002*
pCD123	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(2aa)-SoxS (wt)	*1) BBa_B0015 2) BBa_B1002*
pCD134	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(2aa)-MarA	*1) BBa_B0015 2) BBa_B1002*
pCD141	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(2aa)-Rob	*1) BBa_B0015 2) BBa_B1002*
pCD146	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(2aa)-TetD	*1) BBa_B0015 2) BBa_B1002*
pCD151	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(2aa)-λcII	*1) BBa_B0015 2) BBa_B1002*
pCD175	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-AsiA 3) rpoD (F563Y)	*1) BBa_B0015 2)*

			3) BBa_J23107		*BBa_B1002 3) BBa_B1002*
pCD226	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23107*	1) dCas9 2) MCP-GP33	*1) BBa_B0015 2) BBa_B1002*
pCD351	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-CAP	*1) BBa_B0015 2) BBa_B1002*
pCD156	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (wt)	*1) BBa_B0015 2) BBa_B1002*
pCD157	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(10aa)-SoxS (wt)	*1) BBa_B0015 2) BBa_B1002*
pCD185	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(10aa)-SoxS (R93A)	*1) BBa_B0015 2) BBa_B1002*
pCD186	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (R93A)	*1) BBa_B0015 2) BBa_B1002*
pCD404	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (F88A)	*1) BBa_B0015 2) BBa_B1002*
pCD406	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (R40A)	*1) BBa_B0015 2) BBa_B1002*
pCD408	*CmR*	*p15A*	*1) Sp.pCas9*	1) dCas9	*1)*

			2) BBa_J23107	2) MCP-(5aa)-SoxS (S101A)	*BBa_B0015 2) BBa_B1002*
pJF093	*CmR*	*p15A*	*1) Sp.pCas9 2) TetR-pTet*	1) dCas9 2) MCP-(5aa)-SoxS (R93A)	*1) BBa_B0015 2) BBa_B1002*
pJF094	*CmR*	*p15A*	*1) TetR-pTet 2) Bba_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (R93A)	*1) BBa_B0015 2) BBa_B1002*
pJF104B	*CmR*	*p15A*	*1) TetR-pTet 2) TetR-pTet*	1) dCas9 2) MCP-(5aa)-SoxS (R93A)	*1) BBa_B0015 2) BBa_B1002*
pJF121	*AmpR*	*ColE1*	*TetR-pTet*	1x MS2 scRNA (W108 target)	*TrrnB*
pCD227	*CmR*	*p15A*	*1) pBAD 2) BBa_J23107 3) araC*	1) dCas9 2) MCP-(5aa)-SoxS (R93A) 3) AraC	*1) BBa_B0015 2) BBa_B1002 3) N/A*
pJF077.1~21[c]	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A) 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD294.1~21	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) PCP-(5aa)-SoxS (R93A) 3) 1x PP7 scRNA.b1 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD296.1~21	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-TetD 3) 1x MS2 scRNA.b2	*1) BBa_B0015 2)*

			3) BBa_J23119	(J101~121 targets)	*BBa_B1002 3) TrrnB*
pCD297.1~21	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) αNTD-MCP 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD298.1~21	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(2aa)-λcII (R42A) 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD416.1~21	CmR	p15A	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(2aa)-SoxS (R93A) 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD417.1~21	CmR	p15A	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(10aa)-SoxS (R93A) 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD441.1~21	CmR	p15A	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(20aa)-SoxS (R93A) 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD290	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A) 3) 1x MS2 scRNA.b2 (J106 target)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD477	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (wt) 3) 1x MS2 scRNA.b1 (W108)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*

pCD478	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A) 3) 1x MS2 scRNA.b1 (W108)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD479	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (S101A) 3) 1x MS2 scRNA.b1 (W108)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD481	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (F88A) 3) 1x MS2 scRNA.b1 (W108)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pCD482	*CmR*	*p15A*	*1) Sp.pCas9 2) BBa_J23107 3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R40A) 3) 1x MS2 scRNA.b1 (W108)	*1) BBa_B0015 2) BBa_B1002 3) TrrnB*
pJF076	*AmpR*	*pSC101*	*J1_BBa_J23117*	mRFP1	*BBa_B0015*
pCD469	*AmpR*	*pSC101*	*zwf*	$zwf_{N\text{-}term}$-LacZ	*BBa_B0015*
pCD470	*AmpR*	*pSC101*	*fumC*	$fumC_{N\text{-}term}$-LacZ	*BBa_B0015*
pCD355	*AmpR*	*pSC101*	*J1_BBa_J23117*	*Z. mobilis* pdc/adhB (single operon)	*BBa_B0015*
pJF009[d]	*AmpR*	*ColE1*	*n/a*	*n/a*	*n/a*
pJF043 [d]	*CmR*	*p15A*	*n/a*	*n/a*	*n/a*

[a] BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org). dCas9 is the catalytically inactive form of *S. pyogenes* Cas9[13]. Sp.pCas9 is the endogenous Cas9 promoter from *S. pyogenes*. pBAD is an arabinose-inducible promoter, described previously[14]. Complete sequences of the GFP, RFP, and LacZ reporter constructs are included in the Supplementary Experimental Procedures.

[b] The plasmid with the RFP R2 gRNA for RFP repression was described previously[13].

[c] Plasmid designations such as pJF077.1~21 indicate a set of plasmids (pJF077.1, pJF077.2…) where the

final number corresponds to guide RNA target sites (J101, J102…, Supplementary Table 2) used for the J1 mRFP1 reporter (Fig. 4).

[d] pJF009 and pJF043 are empty vector control plasmids with expression cassettes removed from p_gRNA_bact_RR2 and pCD017, respectively. These control plasmids were used in RT-qPCR experiments (Supplementary Fig. 5) and growth burden experiments (Supplementary Fig. 6).

Supplementary Table 4. Primer Sequences for RT-qPCR

Primer	**Sequence**
qCD010_sfGFP_f	GAGGGTGAAGGTGATGCTACAA
qCD011_sfGFP_r	GGTCAGAGTAGTGACAAGTGTTGG
qCD012_16S_f	AAAGTTAATACCTTTGCTCATTGACGTT
qCD013_16S_r	GACTACCAGGGTATCTAATCCTGTTT

Supplementary Methods

Scaffold RNA (scRNA) Sequence Designs

gRNA and 1x MS2 scRNA sequences with RNA recruitment hairpins were initially constructed following previous designs[4,13]. 1x MS2 scRNA.b1 removes the final tracr RNA terminator hairpin, based on secondary structure predictions of the hairpin structure[15]. 1x MS2 scRNA.b2 removes the final tracr RNA terminator hairpin, based on the x-ray crystal structure of gRNA in complex with Cas9 and a DNA target[16]. Compared to scRNA.b1, the scRNA.b2 design includes one additional G from the tracr sequence and removes the two base GC linker to the MS2 hairpin.

Parent sgRNA
GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

1x MS2 scRNA
GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGCGCACATGAGGATCACCCATGT
GCTTTTTTT

2x MS2 scRNA
GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGGGAGCACATGAGGATCACCCAT
GTGCCACGAGCGACATGAGGATCACCCATGTCGCTCGTGTTCCCTTTTTTT

1x MS2 scRNA.b1
GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGCGCACATGAGGATCACCCATGTGCTTTTTTT

1x MS2 scRNA.b2
GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGCACATGAGGATCACCCATGTGCTTTTTTT

1x PP7 scRNA.b1
GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTAACATAAGGAGTTTATATGGAAACCCTTATGTTTTTTT

Annotations:
20 base target site (W108), 1x MS2, tracr RNA terminator, 1x PP7

Candidate activator sequences

Sequences shown below are those candidate activators that gave detectable GFP expression in initial experiments. RNA binding proteins MCP (MS2 coat protein) or PCP (PP7 coat protein) were fused to the N-termini of candidate activator proteins, except for the RpoA αNTD. RpoA has a native flexible linker connecting its N and C terminal domains[17], so we cloned the αNTD with its flexible linker and fused this sequence to the N-terminus of MCP. The MCP and PCP proteins are modified versions ($MCP_{\Delta FG,\ V29I}$ and $PCP_{\Delta FG}$) to prevent oligomerization and improve RNA binding activity[4].

> MCP-(2aa)-SoxS (wtSoxS with 2aa linker used in Fig. 2)
MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK
GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSMSHQKIIQDLIAWIDEHIDQPL
NIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFSRVFR
RQFDRTPSDYRHRL

> MCP-(5aa)-$SoxS_{R93A}$ (fully optimized version, Fig. 3 and subsequent)
MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK
GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMSHQKIIQDLIAWIDEHI
DQPLNIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFS
RVFARQFDRTPSDYRHRL

> MCP-(5aa)-TetD (optimized linker for Fig. 5; 2aa linker used in Fig. 2)
MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK
GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMYIEQHSRYQNKANNIQ
LRYDDKQFHTTVIKDVLLWIEHNLDQSLLLDDVANKAGYTKWYFQRLFKKVTGVTLASYIRARRLTKAAVEL
RLTKKTILEIALKYQFDSQQSFTRRFKYIFKVTPSYYRRNKLWELEAMH

> MCP-λcII
MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK
GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSMVRANKRNEALRIESALLNKI
AMLGTEKTAEAVGVDKSQISRWKRDWIPKFSMLLAVLEWGVVDDDMARLARQVAAILTNKKRPAATERSE
QIQMEF

> MCP-RpoZ
MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK
GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSARVTVQDAVEKIGNRFDLVLV

AARRARQMQVGGKDPLVPEENDKTTVIALREIEEGLINNQILDVRERQEQQEQEAAELQAVTAIAEGRR

> αNTD-MCP

MQGSVTEFLKPRLVDIEQVSSTHAKVTLEPLERGFGHTLGNALRRILLSSMPGCAVTEVEIDGVLHEYSTK
EGVQEDILEILLNLKGLAVRVQGKDEVILTLNKSGIGPVTAADITHDGDVEIVKPQHVICHLTDENASISMRIK
VQRGRGYVPASTRIHSEEDERPIGRLLVDACYSPVERIAYNVEAARVEQRTDLDKLVIEMETNGTIDPEEAI
RRAATILAEQLEAFVDLRDVRQPEVKEEKPEASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRS
QAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAA
NSGIY

> MCP-AsiA

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK
GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSMNKNIDTVREIITVASILIKFSR
EDIVENRANFIAFLNEIGVTHEGRKLNQNSFRKIVSELTQEDKKTLIDEFNEGFEGVYRYLEMYTNK

> PCP-(5aa)-SoxS$_{R93A}$

MGPSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRLRLTASLRQNGAKTAYRVNLKLDQADVVDSGL
PKVRYTQVWSHDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLVPLGRGGGGSMSHQKIIQDLIAWID
EHIDQPLNIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQ
TFSRVFARQFDRTPSDYRHRL

Fluorescent Protein Reporter Sequences

GFP Reporter (Fig. 2, 3, & 5)

This construct reports on gene activation using a GFP reporter driven by a weak promoter. The promoter was constructed following the design by Bikard et al., 2013[1], driving expression of superfolder GFP (sfGFP)[18]. BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org).
W108 target site, BBa_J23117 promoter, Bujard RBS, sfGFP, BBa_B0015 terminator

GCATGCCCGATCAACGTCTCATTTTCGCCAGATATCAAGCAGAGGAGCAAAAGCTCATTTCTGAAGAG
GACTTGTTGCGGAAACGACGAGAACAGTTGAAACACAAACTTGAACAGCTACGGAACTCTTGTGCGT
AAGGAAAAGTAAGGAAAACGATTCCTTCTAACAGAAATGTCCTGAGCAATCACCTATGAACTGTCGAC
TCGAGCCTCTATGGATTATCACCTTGGCTGCAGGCCGGATCTTCCACAACACGCACGGTGTTACATTA
GGCATACCGGTCttgacagctagctcagtcctagggattgtgctagcGAATTCATTAAAGAGGAGAAAGGTACCATGAG
CAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCA
CAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTT
GCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGC

TTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTA
CAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACA
CAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAA
AGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAA
TACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTC
GAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACAC
ATGGCATGGATGAGCTCTACAAAtaaggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaa
gactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata

RFP Reporter (Fig. 5)

This construct reports on gene silencing using an RFP reporter driven by a strong promoter. The reporter was constructed following the design from Qi et al., 2013[1,13]. The RFP R2 gRNA target site is underlined.
BBa_J23119 promoter, Bujard RBS, mRFP1, BBa_B0015 terminator

GCCCTCTAGAGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGG
TGGTGAATttgacagctagctcagtcctaggtataatagatctGAATTCATTAAAGAGGAGAAAGGTACCATGGCGAGTA
GCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAG
TTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTA
CCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAG
CTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGG
GAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACG
GTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAA
AAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAA
TCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCT
AAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACG
AAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAggatcca
aactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctac
tagagtcacactggctcaccttcgggtgggcctttctgcgtttata

J1 RFP reporter (Fig. 4)

This construct reports on gene activation using an RFP reporter driven by a weak promoter. The J1 upstream region includes PAM sites on both strands every 10 bases to systematically map out the relationship between gRNA target site position and gene activation.
J1 upstream region, BBa_J23117 promoter, Bujard RBS, mRFP1, BBa_B0015 terminator

actcttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgc

gcacatttccccgaaaagtgccacctgtggcaattccgacgtcGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCG
GCCGCCATAGGACACCTTTGGTTGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACC
TCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGAC
ttgacagctagctcagtcctagggattgtgctagcGAATTCATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGA
CGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAA
TCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGG
TGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTT
AAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTG
TTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTT
CATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCA
TGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAAT
GCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAA
CCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACT
ACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAggatccaaactcgagt
aaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcac
actggctcaccttcgggtgggcctttctgcgtttata

LacZ reporter genes for activity assays at endogenous SoxS target genes

zwf-LacZ reporter

This construct reports on activation of the endogenous Class I SoxS target gene zwf using a zwf-LacZ operon fusion constructed following a previously described design[19]. The construct includes 140 bases upstream of the TSS, the 5' UTR, and coding sequence corresponding to zwf amino acids 1-163 fused in frame to LacZ. The TSS and the soxbox were experimentally identified previously[19-21]. BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org).

zwf promoter, $zwf_{1\text{-}161}$, LacZ, BBa_B0015 terminator. The soxbox is underlined.

GATCAGTGTCAGATTTTTACCCAATGGAAAACGATGATTTTTTTATCAGTTTTGCCGCACTTTGCGCGCTTTTCCCG
TAATCGCACGGGTGGATAAGCGTTTACAGTTTTCGCAAGCTCGTAAAAGCAGTACAGTGCACCGTAAGAAAATTAC
AAGTATACCCTGGCTTAAGTACCGGGTTAGTTAACTTAAGGAGAATGACATGGCGGTAACGCAAACAGCCCAGGC
CTGTGACCTGGTCATTTTCGGCGCGAAAGGCGACCTTGCGCGTCGTAAATTGCTGCCTTCCCTGTATCAACTGGA
AAAAGCCGGTCAGCTCAACCCGGACACCCGGATTATCGGCGTAGGGCGTGCTGACTGGGATAAAGCGGCATATA
CCAAAGTTGTCCGCGAGGCGCTCGAAACTTTCATGAAAGAAACCATTGATGAAGGTTTATGGGACACCCTGAGTG
CACGTCTGGATTTTTGTAATCTCGATGTCAATGACACTGCTGCATTCAGCCGTCTCGGCGCGATGCTGGATCAAAA
AAATCGTATCACCATTAACTACTTTGCCATGCCGCCCAGCACTTTTGGCGCAATTTGCAAAGGGCTTGGCGAGGC
AAAACTGAATGCTAAACCGGCACGCGTAGTCATGGAGAAACCGCTGGGGACGTCGCTGGCGACCTCGCAGGAA
ATCAATGACATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTA
CCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGC
CCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGG
AAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTAC

GATGCGCCCATCTACACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACG
GGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCG
TTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAA
TTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCA
GTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTA
CACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACTGGAGGCTGAAGTTCA
GATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCG
GCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAAC
GTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGA
CGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGC
TGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATG
AGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGA
ACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACG
GCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAATG
GTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCA
CGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACA
CCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAA
TGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATG
GGTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTC
TGGGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTT
GGCGATACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCT
GACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATA
CCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTG
AAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCC
GGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCG
CCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCT
GACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTT
TCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGA
TAACGACATTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGC
CATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGC
TCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAAT
GGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGG
CGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCC
TGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGC
TGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAG
TCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACG
GTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGT
CGCTACCATTACCAGTTGGTCTGGTGTCAAAAATAAggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtc
gaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata

fumC-LacZ reporter

This construct reports on activation of the endogenous Class II SoxS target gene fumC using a fumC-LacZ operon fusion. The construct includes 46 bases upstream of the TSS, the 5' UTR, and coding sequence corresponding to fumC amino acids 1-56 fused in frame to LacZ. This construct is similar but not identical to a previously described fumC-LacZ reporter gene[22]. The TSS and the soxbox were experimentally identified previously[21,23]. BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org).

fumC promoter, $fumC_{1-56}$, LacZ, BBa_B0015 terminator. The soxbox is underlined.

ATGGCACGAAAGACCAAACATTTGTTATCAAATGGTAAATAATAAGTGAGCTAAAAGTTGCTTAACGAAAGCAAAAC
AGAAAGAAAAAATTAATCAGGTGAGGAGCAGGTCATGAATACAGTACGCAGCGAAAAAGATTCGATGGGGGCGAT
TGATGTCCCGGCAGATAAGCTGTGGGGCGCACAAACTCAACGCTCGCTGGAGCATTTCCGCATTTCGACGGAGA
AAATGCCCACCTCACTGATTCATGCGCTGGCGCTAACCAAGCGTGCAGCGGCCATGACCATGATTACGGATTCAC
TGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCC
CTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGC
GAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGG
CCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATC
CCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGA
AAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGG
GCGCTGGGTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAG
AAAACCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATG
AGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCT
TTAATGATGATTTCAGCCGCGCTGTACTGGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGG
TAACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGAT
GAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAAT
CCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATG
TCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTT
AACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATG
AAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGC
TACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGAT
CCGCGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGA
TCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTCG
ATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTAC
GCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGG
AGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTG
GCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAAATA
TGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTTCTGTAT
GAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTC
CAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTG
CACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTA
AACAGTTGATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTG

CAACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAAC
CTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGA
GCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAACAA
CTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCC
GCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCA
GTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCT
TATTTATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAG
CGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTC
GGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCA
GACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCC
ACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATC
GCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGAC
TCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTCTGGTGTCA
AAAATAAggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgct
ctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata

Ethanol Production Pathway Cassette (pCD355)

This construct includes the *Z. mobilis pdc-adhB* gene cluster driven by a weak promoter. In Fig. 6, the CRISPRa complex targets the J106 site upstream of the promoter. J106 is the site with maximal activity in reporter assays (Fig. 4B).

J1 upstream region (J106 site underlined), BBa_J23117 promoter, Bujard RBS, Zm.pdc, Zm.adhB, BBa_B0015 terminator

gcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttag
aaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgACGTCGCGGCCGCCTACGGTATCCACCGG
AGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTGACCTATGGTGACCAT
GGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCGTCCTTTGGGTT
CCACCGGATACCTCCGGACttgacagctagctcagtcctagggattgtgctagcGAATTCATTAAAGAGGAGAAAGGTA
CCatgagttatactgtcggtacctatttagcggagcggcttgtccagattggtctcaagcatcacttcgcagtcgcgggcgactacaacctcgtccttct
tgacaacctgcttttgaacaaaaacatggagcaggtttattgctgtaacgaactgaactgcggtttcagtgcagaaggttatgctcgtgccaaaggcg
cagcagcagccgtcgttacctacagcgtcggtgcgctttccgcatttgatgctatcggtggcgcctatgcagaaaaccttccggttatcctgatctccgg
tgctccgaacaacaatgaccacgctgctggtcacgtgttgcatcacgctcttggcaaaaccgactatcactatcagttggaaatggccaagaacatc
acggccgccgctgaagcgatttataccccggaagaagctccggctaaaatcgatcacgtgattaaaactgctcttcgtgagaagaagccggtttatc
tcgaaatcgcttgcaacattgcttccatgccctgcgccgctcctggaccggcaagcgcattgttcaatgacgaagccagcgacgaagcttctttgaat

gcagcggttgaagaaaccctgaaattcatcgccnaccgcgacaaagttgccgtcctcgtcggcagcaagctgcgcgcagctggtgctgaagaag

ctgctgtcaaatttgctgatgctcttggtggcgcagttgctaccatggctgctgcaaaaagcttcttcccagaagaaaacccgcattacatcggtacctc

atggggtgaagtcagctatccgggcgttgaaaagacgatgaaagaagccgatgcggttatcgctctggctcctgtctttaacgactactccaccactg

gttggacggatattcctgatcctaagaaactggttctcgctgaaccgcgttctgtcgtcgttaacggcattcgcttccccagcgtccatctgaaagactat

ctgacccgtttggctcagaaagtttccaagaaaaccggtgctttggacttcttcaaatccctcaatgcaggtgaactgaagaaagccgctccggctga

tccgagtgctccgttggtcaacgcagaaatcgcccgtcaggtcgaagctcttctgaccccgaacacgacggttattgctgaaaccggtgactcttggtt

caatgctcagcgcatgaagctcccgaacggtgctcgcgttgaatatgaaatgcagtggggtcacattggttggtccgttcctgccgccttcggttatgc

cgtcggtgctccggaacgtcgcaacatcctcatggttggtgatggttccttccagctgacggctcaggaagtcgctcagatggttcgcctgaaactgcc

ggttatcatcttcttgatcaataactatggttacaccatcgaagttatgatccatgatggtccgtacaacaacatcaagaactgggattatgccggtctga

tggaagtgttcaacggtaacggtggttatgacagcggtgctggtaaaggcctgaaggctaaaaccggtggcgaactggcagaagctatcaaggtt

gctctggcaaacaccgacggcccaaccctgatcgaatgcttcatcggtcgtgaagactgcactgaagaattggtcaaatggggtaagcgcgttgct

gccgccaacagccgtaagcctgttaacaagctcctctaacaattcaaaaGATCTAAAGAGGAGAAATCTAGAatggcttcttcaacttttt

tatattcctttcgtcaacgaaatgggcgaaggttcgcttgaaaaagcaatcaaggatcttaacggcagcggctttaaaaatgcgctgatcgtttctgatg

ctttcatgaacaaatccggtgttgtgaagcaggttgctgacctgttgaaagcacagggtattaattctgctgtttatgatggcgttatgccgaacccgact

gttaccgcagttctggaaggccttaagatcctgaaggataacaattcagacttcgtcatctccctcggtggtggttctccccatgactgcgccaaagcc

atcgctctggtcgcaaccaatggtggtgaagtcaaagactacgaaggtatcgacaaatctaagaaacctgccctgcctttgatgtcaatcaacacga

cggctggtacggcttctgaaatgacgcgtttctgcatcatcactgatgaagtccgtcacgttaagatggccattgttgaccgtcacgttaccccgatggtt

tccgtcaacgatcctctgttgatggttggtatgccaaaaggcctgaccgccgccaccggtatggatgctctgacccacgcatttgaagcttattcttcaa

cggcagctactccgatcaccgatgcttgcgctttgaaagcagcttccatgatcgctaagaatctgaagaccgcttgcgacaacggtaaggatatgcc

ggctcgtgaagctatggcttatgcccaattcctcgctggtatggccttcaacaacgcttcgcttggttatgtccatgctatggctcaccagttgggcggtta

ctacaacctgccgcatggtgtctgcaacgctgttctgcttccgcatgttctggcttataacgcctctgtcgttgctggtcgtctgaaagacgttggtgttgct

atgggtctcgatatcgccaatctcggtgataaagaaggcgcagaagccaccattcaggctgttcgcgatctggctgcttccattggtattccagcaaa

cctgaccgagctgggtgctaagaaagaagatgtgccgcttcttgctgaccacgctctgaaagatgcttgtgctctgaccaacccgcgtcagggtgat

cagaaagaagttgaagaactcttcctgagcgctttctaaggatccaaactcgagtaaggatctCCAGGCATCAAATAAAACGAAAGG

CTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCAC

ACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA

Supplementary References

1. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41,** 7429–7437 (2013).
2. Dove, S. L. & Hochschild, A. Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12,** 745–754 (1998).
3. Gregory, B. D., Deighan, P. & Hochschild, A. An artificial activator that contacts a normally occluded surface of the RNA polymerase holoenzyme. *J. Mol. Biol.* **353,** 497–506 (2005).
4. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160,** 339–350 (2015).
5. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).
6. Ho, Y. S., Mahoney, M. E., Wulff, D. L. & Rosenberg, M. Identification of the DNA binding domain of the phage lambda cII transcriptional activator and the direct correlation of cII protein stability with its oligomeric forms. *Genes Dev.* **2,** 184–195 (1988).
7. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: identifying amino acids important for DNA binding and transcription activation. *J. Mol. Biol.* **322,** 237–257 (2002).
8. Zhang, X. & Bremer, H. Control of the *Escherichia coli* rrnB P1 promoter strength by ppGpp. *J. Biol. Chem.* **270,** 11181–11189 (1995).
9. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25,** 402–408 (2001).
10. Shah, I. M. & Wolf, R. E. Novel protein-protein interaction between Escherichia coli SoxS and the DNA binding determinant of the RNA polymerase alpha subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *J. Mol. Biol.* **343,** 513–532 (2004).
11. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
12. Shapiro, H. M. Multiparameter flow cytometry of bacteria: implications for diagnostics and therapeutics. *Cytometry* **43,** 223–226 (2001).
13. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).
14. Lee, T. S. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* **5,** 12 (2011).
15. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
16. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

17. Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. & Ebright, R. H. Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78,** 889–896 (1994).
18. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).
19. Fawcett, W. P. & Wolf, R. E. Genetic definition of the *Escherichia coli zwf* 'soxbox,' the DNA binding site for SoxS-mediated induction of glucose 6-phosphate dehydrogenase in response to superoxide. *J. Bacteriol.* **177,** 1742–1750 (1995).
20. Rowley, D. L. & Wolf, R. E. Molecular characterization of the Escherichia coli K-12 zwf gene encoding glucose 6-phosphate dehydrogenase. *J. Bacteriol.* **173,** 968–977 (1991).
21. Martin, R. G., Gillette, W. K., Rhee, S. & Rosner, J. L. Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in Escherichia coli: sequence, orientation and spatial relationship to the core promoter. *Mol. Microbiol.* **34,** 431–441 (1999).
22. Fawcett, W. P. & Wolf, R. E. Purification of a MalE-SoxS fusion protein and identification of the control sites of *Escherichia coli* superoxide-inducible genes. *Mol. Microbiol.* **14,** 669–679 (1994).
23. Taliaferro, L. P., Keen, E. F., Sanchez-Alberola, N. & Wolf, R. E. Transcription activation by Escherichia coli Rob at class II promoters: protein-protein interactions between Rob's N-terminal domain and the σ(70) subunit of RNA polymerase. *J. Mol. Biol.* **419,** 139–157 (2012).

Chapter 2:

Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements

Jason Fontana*,[1], Chen Dong*,[2], Cholpisit Kiattisewee[1], Venkata P. Chavali[3], Benjamin I. Tickman[1], James M. Carothers[1,3,4], Jesse G. Zalatan[2,3,4]

* these authors contributed equally

[1]Molecular Engineering & Sciences Institute
[2]Department of Chemistry
[3]Department of Chemical Engineering
[4]Center for Synthetic Biology
University of Washington, Seattle, WA 98195, USA

Published as a research article in *Nature Communications* on April 1st, 2020.
DOI: 10.1038/s41467-020-15454-y

Abstract

In bacterial systems, CRISPR-Cas transcriptional activation (CRISPRa) has the potential to dramatically expand our ability to regulate gene expression, but we lack predictive rules for designing effective gRNA target sites. Here, we identify multiple features of bacterial promoters that impose stringent requirements on CRISPRa target sites. Notably, we observe narrow, 2-4 base windows of effective sites with a periodicity corresponding to one helical turn of DNA, spanning ~40 bases and centered ~80 bases upstream of the TSS. However, we also identify two features suggesting the potential for broad scope: CRISPRa is effective at a broad range of σ^{70}-family promoters, and an expanded PAM dCas9 allows the activation of promoters that cannot be activated by *S. pyogenes* dCas9. These results provide a roadmap for future engineering efforts to further expand and generalize the scope of bacterial CRISPRa.

Introduction

Developing tools to activate the expression of arbitrary genes has been transformative for biotechnology and biological research[1]. In metabolic engineering, regulating the timing and levels of the expression of complex multi-gene pathways is critical for reducing cellular burden and improving production of valuable metabolites[2]. To enable these goals, we recently developed a CRISPR-Cas transcriptional activation (CRISPRa) system that is effective in *E. coli*. Our system can be combined with CRISPRi gene repression to programmably target multiple genes for simultaneous activation and repression[3]. While our CRISPRa system can be used with heterologous genes, an outstanding challenge is to understand the rules that define effective target sites at arbitrary promoters in the genome.

To programmably downregulate target genes, we use nuclease defective Cas9 (dCas9) with a guide RNA (gRNA) that specifies a target site on the DNA. Targeting this complex to a promoter or an open reading frame (ORF) results in gene repression (CRISPRi)[4]. To enable simultaneous activation, we use modified guide RNAs, termed scaffold RNAs (scRNAs), that include a 3' MS2 hairpin to recruit a transcriptional activator fused to the MS2 coat protein (MCP)[3]. We can express multiple gRNAs and scRNAs to inhibit and activate genes simultaneously; gRNAs targeted to a promoter or ORF result in CRISPRi and scRNAs targeted to an appropriate site upstream of a minimal promoter result in CRISPRa.

We demonstrate here that the rules for targeting CRISPRa to effective sites in *E. coli* are surprisingly stringent. In prior work, we found that CRISPRa *in E. coli* was effective at target sites located in a narrow 40 base window between 60 and 100 bases upstream of the transcriptional start site (TSS)[3]. Here, we show that multiple factors combine to make the requirements for effective sites even stricter. We demonstrate that the basal promoter strength of the target gene and the sequence composition between the target site and the minimal promoter can have dramatic effects on gene activation. Further, by scanning the 40 base window at single base resolution, we find sharp peaks of activity and broad regions of inactivity that occur in a periodic 10-11 base pattern, corresponding to one helical turn along the DNA target. The observation that only a few precisely-positioned target sites upstream of the TSS are effective for CRISPRa poses a significant challenge, as many genes will likely lack an NGG PAM sequence at exactly the right

position necessary for *S. pyogenes* dCas9. These stringent requirements may explain why CRISPRa and other tools for gene activation in bacteria have lagged far behind comparable tools in eukaryotic systems, where such strict target site requirements are absent[5].

Although the requirements for bacterial CRISPRa target sites pose challenges, our data also demonstrate CRISPRa has the potential to be effective at a broad range of target genes. In addition to σ^{70}–dependent genes, CRISPRa can activate expression from genes that use the σ^{70} family members σ^{38}, σ^{32}, and σ^{24}. We further demonstrate that the strict requirement for a precisely positioned PAM site can be partially overcome using a re-engineered dCas9 protein that targets an expanded set of PAM sequences[6]. Recently, some of the rules that we describe here were independently reported for an alternative bacterial CRISPRa system that can target genes regulated by σ^{54} promoters[7]. Our results demonstrate that this behavior applies to a much broader range of σ^{70}–family promoters, which cover the majority of the E. coli genome[8]. The availability of these complementary systems should further extend the scope of bacterial CRISPRa. More broadly, by systematically defining the rules for effective CRISPRa sites, we identify strategies for improving and generalizing synthetic gene regulation in bacteria.

Results

A SoxS mutant reduces off-target activation

Ideally, a synthetic transcriptional activator should only activate its programmed target genes. The activation domain for our CRISPRa system is SoxS, a native *E. coli* transcription factor that directly binds DNA and activates endogenous gene targets as part of a stress response program[3]. We previously demonstrated that point mutations in the SoxS DNA binding site can reduce activation of endogenous SoxS targets while maintaining CRISPRa activity at a heterologous reporter gene. However, the most effective single point mutants, R93A and S101A, did not completely abolish activity at endogenous targets. To further minimize off-target SoxS activity, we tested a double mutant SoxS(R93A/S101A). This double mutant SoxS retained full CRISPRa activity and showed a reduction in endogenous SoxS-dependent gene expression to levels indistinguishable from background (Figure 1). Thus, SoxS(R93A/S101A) is an effective modular transcriptional effector that can activate gene expression only when recruited to a target gene via the CRISPR-Cas complex.

A distance metric for target sites is not effective

To determine if we could predictably activate endogenous genes with CRISPRa, we selected three candidate genes with appropriately positioned PAM sites upstream of the TSS. Previously, we demonstrated that CRISPRa can activate heterologous promoters up to 50-fold with target sites positioned within a 40 base window between 60 and 100 bases upstream of the transcriptional start site (TSS)[3]. We therefore targeted the CRISPR-Cas complex to the same window upstream of the candidate target genes. First, we targeted the *aroK-aroB* operon, which expresses enzymes involved in aromatic amino acid biosynthesis, whose programmed overexpression could be useful for bioproduction[9]. Targeting the

CRISPR-Cas complex to two sites within the optimal 40 base window resulted in no statistically significant increases in gene expression. Further, sites inside and outside of the 40 base window gave similar effects (Figure 2A). Next, we targeted *cysK*, an enzyme involved in cysteine biosynthesis[10]. Similar to what we observed with *aroK-aroB*, targeting three sites within the 40 base window resulted in no statistically significant increases in gene expression (Figure 2B). Finally, we targeted *ldhA*, an enzyme involved in mixed acid fermentation[11]. We selected 8 sites and observed no apparent relationship between the position of the target site and *ldhA* expression (Supplementary Figure 1). Together, these results suggest that endogenous genes cannot be activated simply by targeting the CRISPR-Cas complex to sites positioned between 60 and 100 bases upstream of the TSS.

There are several possible explanations for our inability to activate endogenous bacterial genes with CRISPRa. First, we originally demonstrated CRISPRa using a relatively weak synthetic promoter. The basal levels of expression of endogenous genes vary significantly[12], and it may be difficult to increase the transcription of genes that are already strongly expressed[13]. In addition, some endogenous target genes might require an alternative sigma factor. Our original reporter gene is controlled by the σ^{70} housekeeping sigma factor, and we do not know if our CRISPRa system is effective at gene targets that use alternative sigma factors. Another possibility is that native transcriptional regulator binding sites near endogenous gene promoters could disrupt CRISPRa. Finally, the optimal distance window metric that we previously identified may have been oversimplified. We initially identified the optimal window from an experiment with target sites spaced 10 bases apart, which may not be sufficient to generalize to any site within the 40 base window. To systematically explore these possibilities, we proceeded to test the efficacy of CRISPRa with a new set of synthetic promoters engineered with variable basal expression levels, alternative sigma factors, variable regulator binding sites, and variable scRNA target site positions.

CRISPRa is sensitive to promoter strength

To evaluate whether the intrinsic strength of the promoter affects CRISPRa, we tested activation on a set of fluorescent reporter genes with minimal promoters spanning a 200-fold range in basal expression level [http://parts.igem.org] (Figure 3A). We observed the most effective gene activation with a moderately weak J23117 promoter. With the weakest promoters, we could not detect any activation, even though their basal expression levels were only 2-fold weaker than the J23117 promoter. With stronger promoters, we observed progressively smaller CRISPRa-mediated activation of gene expression; the basal expression level increased, while the maximal, CRISPRa-induced expression remained roughly constant. These results indicate that the bacterial CRISPRa activity varies considerably with promoter strength, similar to effects observed in eukaryotic systems[14,15]. Thus, when targeting arbitrary endogenous genes, the level of activation that can be achieved may depend on the basal level of expression of its promoter.

CRISPRa is effective with alternative sigma factors

Bacterial transcription is initiated by a sigma factor binding to the minimal promoter and the RNA

polymerase holoenzyme[16].The SoxS activator binds directly to the α subunit of RNA polymerase[17], which suggests that our CRISPRa system could be compatible with genes that are controlled by non-housekeeping sigma factors. To investigate this possibility, we built synthetic promoters regulated by σ^{38} (RpoS), σ^{32} (RpoH), σ^{24} (RpoE), and σ^{54} (RpoN) to compare with our original housekeeping σ^{70} (RpoD) promoter (Figure 3B)[18–21]. CRISPRa was able to activate reporter gene expression when we targeted σ^{38}, σ^{32}, and σ^{24}-dependent promoters; these σ factors are all members of the σ^{70} family. CRISPRa was not active on the σ^{54} promoter, possibly because σ^{54} initiates gene expression using a distinct mechanism that requires additional *cis*-regulatory elements[16]. These results suggest that CRISPRa can activate promoters regulated by non-housekeeping sigma factors such as σ^{38}, σ^{32}, and σ^{24}, and likely other members of the homologous σ^{70} family.

A recent paper described an alternative CRISPRa system that is capable of activating σ^{54}-dependent genes[7], which comprise a small fraction of the genome[8] (Supplementary Figure 2). The availability of multiple, complementary CRISPRa systems should further extend the scope of bacterial CRISPRa. Both systems effectively activate expression from synthetic and heterologous promoters, and each system has the potential to target a different, non-overlapping set of endogenous genes.

CRISPRa is sensitive to intervening sequence composition

To determine if the sequence composition between the target site and the -35 site affects CRISPRa, we constructed a promoter library with randomized sequences in this intervening region. We analyzed single colonies from this library and observed gene activation with a broad distribution over a 27-fold range (Figure 3C). Although most variant sequences can still be activated (>2-fold) with CRISPRa, the large variation in activity was unexpected because each reporter gene was driven by the same minimal promoter and contained the same scRNA target site. One possible interpretation of this result is that these randomized intervening sequences contain binding sites for endogenous transcriptional regulators; there is evidence that binding sites can emerge with relatively high frequency from random sequences[22]. These sites could potentially affect CRISPRa by directly blocking access to a scRNA target site, by blocking RNA polymerase binding, or by interfering with the ability of a CRISPRa effector protein to engage with RNA polymerase.

To directly test the hypothesis that a bound transcriptional effector can disrupt CRISPRa, we introduced a binding site for the transcriptional repressor TetR upstream of the -35 region[23]. The presence of a bound TetR significantly disrupted CRISPRa-mediated gene activation. Further, adding anhydrotetracycline (aTc), which releases TetR from the DNA, restored CRISPRa activity to the levels observed when TetR was not present (Figure 3D). Because endogenous genes contain binding sites for a variety of transcriptional activators and repressors upstream of the minimal promoter[24,25], this effect could be contributing to the inconsistent and variable effects we observed when targeting endogenous genes for CRISPRa (Figure 2).

To determine if transcription factor binding sites appear in the library of randomized intervening sequences, we sequenced 29 variants spanning the full range of observed activation levels (Supplementary

Table 6). Only five intervening sequences contained exact matches to a known consensus transcription factor binding motif. However, all sequences contained at least one match within a single base of a known motif, and it is well established that DNA binding proteins can recognize sites that deviate from the consesus[26]. There was no significant correlation between gene activation by CRISPRa and the number of these motifs (Spearman rank order correlation r_s = 0.29, p = 0.11, Supplementary Figure 3A), but we note that it is not known which of these motifs actually bind endogenous transcription factors. We did find that intervening sequences that give more effective CRISPRa tend to be more GC-rich, though we do not yet understand the basis for this trend (r_s = 0.42, p = 0.02, Supplementary Figure 3B & C). Nonetheless, these experiments indicate that the composition of the intervening sequence between the CRISPR-Cas complex and the minimal promoter is an important factor determining the level of CRISPRa.

CRISPRa is sharply dependent on single base shifts

Our original hypothesis that optimal target sites are located -60 to -100 bases upstream of the TSS was based on an experiment with scRNA sites spaced every 10 bases[3]. To further test this hypothesis, we targeted the CRISPRa complex to a window from -61 to -113 at single base resolution. We used a reporter gene with 5 scRNA sites located at -61, -71, -81, -91, and -101 relative to the TSS, and we inserted 1-12 bases upstream of the -35 site to generate a set of reporter genes that allowed the CRISPRa complex to target every possible distance in the optimal targeting window. Using this reporter gene set, we found that shifting the target site by 1-3 bases caused significant decreases in activation (Figure 4A). Shifting the target site further by 4-9 bases decreased expression to levels nearly indistinguishable from background. At 10-11 base shifts, corresponding to one full turn of a DNA helix, gene expression increased again. This periodic positional dependence of CRISPRa extended over the entire -60 to -100 window, with the strongest peaks centered at -81 and -91 and smaller peaks centered at -102 and -70. There is no recovery of activity when the site at -101 is shifted to -111, outside of the -60 to -100 window. This sharp periodic relationship suggests that the criteria for effective target sites are quite stringent, and that both distance and relative periodicity to the TSS are critical factors.

Notably, the distance to the TSS is not the sole determining factor for CRISPRa-mediated expression level. Sites that overlap at the same distance, such as the original -81 site and the -71 site shifted by 10, do not give the same gene expression output (Figure 4A). These discrepancies could arise from intrinsic differences in the activity of the 20 base scRNA target sequence (Supplementary Figure 4) or from the effect of different intervening sequence composition between the scRNA target site and the minimal promoter (Figure 3).

Because we demonstrated that sequence composition can have unexpected effects on CRISPRa (Figure 3), we tested whether the periodicity of CRISPRa was similar in different sequence contexts. We obtained comparable periodic phase dependence when different nucleotide sequences were used to shift the scRNA target site, and when the bases were inserted at a different location in the promoter (Supplementary Figure 5A). Similar results were also obtained when we performed the base shift

experiment with a reporter that had a different 5' upstream sequence (Supplementary Figure 5B) or where the minimal BBa_J23117 promoter was replaced by endogenous *aroK* promoter (Supplementary Figure 5C). Further, the sharp positioning dependence was observed when targeting the template or non-template strand of the reporter (Supplementary Figure 5D). Finally, one possible confounding effect could arise if the basal expression level of the reporter gene changes when bases are inserted, which can affect the efficacy of CRISPRa (Figure 3A). However, we observed that basal expression from the original reporter and the +5 base shifted reporter were indistinguishable (Supplementary Figure 5E). Together, these experiments confirm that bacterial CRISPRa is sensitive to periodicity in multiple different sequence contexts.

In the experiments described above, comparisons between single base shifted scRNA sites were performed with different reporter gene constructs, each with a differing number of inserted bases. To test the positional dependence of CRISPRa at single base resolution in a single reporter construct, we designed an alternative reporter gene with 6 adjacent scRNA target sites between -81 and -86. We again observed sharp drops in gene expression when targeting sites one or more bases away from the optimal site at -81 (Supplementary Figure 5F).

The finding that CRISPRa displays the same ~10 base periodicity as the DNA helix suggests that the angular phase of the CRISPRa complex relative to the minimal promoter is critical for effective activation. Our bacterial CRISPRa system requires a direct interaction between the SoxS activation domain and RNA polymerase[3], and this interaction appears to be highly sensitive to both the distance and relative phase of the target site to the minimal promoter. The sharp phase dependence of CRISPRa may be a general feature of transcriptional regulation in *E. coli*. The native SoxS protein and other transcription factors such as CAP and LacI have restrictive positioning requirements that correspond to DNA periodicity[27–34]; we confirmed this result with an endogenous SoxS reporter (Supplementary Figure 6). In practice, this periodic behavior means that effective target sites must be located at one of the narrow peaks of activation within the optimal distance range. These stringent requirements suggest that targeting endogenous genes will be extremely challenging. There is ~1 PAM site every 10 bases in the regions upstream of endogenous promoters in *E. coli* (Supplementary Figure 7 A & B), and the likelihood that a PAM site will be located at the appropriate phase within a 10 base window is low (Supplementary Figure 7C).

Tuning structure to expand target site range is ineffective

If rotating the CRISPRa complex out of phase along the DNA prevents SoxS from interacting with RNA polymerase, then a longer amino acid linker to SoxS might allow effective CRISPRa at more scRNA sites. To test this possibility, we extended the linker between MCP and SoxS from 5 amino acids (aa) to 10 or 20 aa, but even with these longer linkers we observed the same sharp dependence on the target site position as with the original 5 aa linker (Figure 4B). We obtained similar results using a linker with a different amino acid composition (Supplementary Figure 8A).

Another potential approach to expand the range of effective CRISPRa sites would be to change the spatial position of the MCP-SoxS protein by altering the position of the MS2 hairpin that binds MCP. We

therefore tested multiple alternative scRNA designs that present the MS2 hairpin at different locations. Extending the MS2 stem by 2, 5, 10, and 20 bp resulted in progressively lower CRISPRa activity, but no change in the position of the target sites that were most effective (Supplementary Figure 8B). Similarly, no changes were observed with alternative scRNA designs with one or two MS2 hairpins presented from different locations within the scRNA structure (Supplementary Figure 8C).

Finally, we assessed whether any alternative activation domains could produce a different phase dependent behavior. Previously, these constructs all produced weaker activation than SoxS[3], perhaps because they have each distinct optimal target site positions. We tested MCP fused to TetD, αNTD, lambda cII, and RpoZ[3], and dCas9 fused to RpoZ[35]; however, none of these constructs produced gene activation at any site that was not already effective with SoxS (Supplementary Figure 9).

Although endogenous bacterial transcription factors exhibit a sharp periodic dependence on distance[27–34], it remains surprising that no structural modifications of the CRISPRa complex produced any changes in the phase dependence. If SoxS is simply tethered to the CRISPRa complex by a flexible linker, we would have expected the peak of effective CRISPRa sites to broaden with longer linkers. The failure of this prediction suggests that our understanding of the CRISPR-Cas complex and its interactions with bacterial transcriptional machinery is fundamentally incomplete, or that the linker tethering SoxS to the CRISPRa complex is not truly flexible. Practically, it means that we still lack a way to expand the range of effective CRISPRa target sites.

A dCas9 variant expands the range of targetable sites

Because there is a limited number of genes with an appropriate NGG PAM site at precisely the optimal position upstream of the promoter (Supplementary Figure 7C), we attempted to expand the scope of targetable PAM sites for CRISPRa. We used a recently characterized dCas9 variant, dxCas9(3.7), that has improved activity at a variety of non-NGG PAM sites including NGN, GAA, GAT, and CAA[6]. We generated reporter plasmids by replacing AGG PAM sites with alternative PAM sequences and delivered a CRISPRa system with dxCas9(3.7) to target these reporters. dxCas9(3.7) maintained the ability to target the AGG PAM and showed significantly increased levels of activation at alternative PAM sites compared to dCas9 (Figure 5A). Activation levels varied with different PAM sites and correlated well with dxCas9(3.7) activity previously reported in human cells (Supplementary Figure 10A)[6]. dxCas9(3.7) showed similar distance and phase dependent target site preferences as dCas9 (Supplementary Figure 10B & C), but its expanded PAM scope makes it more likely that an arbitrary gene will have a targetable PAM site at an effective position. Bioinformatic analysis of the sequences between transcriptional units in *E. coli* revealed that there are on average 6.4 times more dxCas9(3.7)-compatible PAM sites than NGG PAM sites (Supplementary Figure 10D). Accounting for the fact that dCas9 has some activity at non-NGG sites[6] (Figure 5A), there are still on average ~2.2-fold more dxCas9(3.7)-compatible PAM sites than dCas9-compatible PAM sites (Supplementary Figure 10D).

To demonstrate the utility of dxCas9(3.7) for CRISPRa at sites inaccessible to dCas9, we

constructed a reporter plasmid that contains an AGG PAM site at the original position with maximum CRISPRa activity and an AGT PAM 5 bases downstream. Using this reporter, we observe that both dCas9 and dxCas9(3.7) are effective for CRISPRa at the optimally-positioned NGG PAM site, but neither is capable of activating the AGT PAM site, which is 5 bases out of phase from the optimal site (Figure 5B). We then inserted 5 bases into the reporter to shift the AGT PAM site into the peak activation range. With this reporter, neither dCas9 nor dxCas9(3.7) can activate the NGG PAM site, which is now out of phase. dxCas9(3.7) was now able to effectively activate the AGT PAM site, and dCas9 was ineffective at this site (Figure 5B). This result confirms that dxCas9(3.7) is able to activate optimally-positioned target sites that are inaccessible to dCas9. We expect that this behavior will be effective at many σ^{70}-family promoters (Figure 3B), and a recent report demonstrated a similar behavior of dxCas9(3.7) at σ^{54}-dependent promoters[7].

Defined rules enable endogenous gene activation

Our systematic characterization of the requirements for effective CRISPRa in *E. coli* demonstrates that candidate genes must have a targetable PAM site located at one of the sharp peaks of activity upstream of the TSS. In hindsight, the scRNA sites at endogenous genes that we initially targeted in Figure 2 did note meet this criterion. To determine if the revised rules would enable activation of endogenous *E. coli* genes, we surveyed the genome for candidate genes with appropriately positioned, dxCas9(3.7)-compatible PAM sites (Supplementary Methods) (Supplementary Figure 7C). We selected candidates with multiple potentially effective PAM sites and further narrowed the pool based on two additional criteria: (1) genes should not be too highly expressed (Figure 3A) and (2) genes should be regulated by σ^{70}, which is the sigma factor that regulates most genes[8] (Figure 3B). Ideally, we would also exclude genes with tightly bound transcriptional regulators in the promoter region (Figure 3D), but this information is not readily available. We chose six genes that could be tested using reporter strains from the *E. coli* promoter collection[36] and targeted two PAM sites for each gene.

We first examined the *yajG* gene, which had two plausible target sites, one of which was only compatible with dxCas9(3.7). We also included an additional site predicted to be out of phase and ineffective for CRISPRa. We observed significant, ~4-6-fold gene activation for the two sites located at the predicted peak of activity at -80/-81, and no activation at the out of phase site at -87 (Figure 6A). The site at -81 is inaccessible to dCas9, and we only observed activation with dxCas9(3.7). We proceeded to test an additional five genes with partial success. We observed significant activation at *poxB* (~10-fold) and *uxuR* (~2-fold) (Figure 6B). We validated these results by performing RT-qPCR on the endogenous *yajG* and *poxB* loci. Targeting CRISPRa to these genes resulted in increases in RNA levels (Supplementary Figure 11). Targeting CRISPRa to *araE* produced a statistically significant difference in expression, but the activation measured was modest (1.13-fold). For the remaining two candidate genes, *ansB* was modestly repressed at one of the target sites and we did not observe a statistically significant difference in expression at *ppiD*. Similarly, one of the *ldhA* sites that we targeted in initial experiments

(Supplementary Figure 1) was at a predicted optimal site at -91 and failed to give substantial activation. Thus, of seven endogenous genes tested with target sites that we predict should be effective (the six genes from Figure 6B and *ldhA* from Supplementary Figure 1), we were able to activate three genes with >2-fold increases in gene expression.

Although any success at endogenous gene activation is encouraging, significant challenges remain for predictable CRISPRa in bacteria. Our results suggest that even with a precise distance metric for effective target sites, some genes will not be predictably activated. There are several possible explanations: (1) tightly bound negative regulators could interfere with CRISPRa (Figure 3D), and (2) small errors in transcription start site annotation could lead to inaccurate predictions for effective sites, given that 1-2 base shifts can have dramatic effects on CRISPRa (Figure 4), and (3) intrinsic differences in the activity of the 20 base scRNA target sequence (Supplementary Figure 4).

Discussion

Bacterial CRISPRa is sensitive to a number of factors, including (i) the strength of the target promoter, (ii) the sigma factor regulating the promoter, (iii) the sequence composition immediately upstream of the minimal promoter, (iv) the composition of the scRNA target sequence, (v) the position of the scRNA target site with respect to the TSS at single base resolution. Some of these factors, such as promoter strength and scRNA target sequence composition, are also relevant in eukaryotic systems[13,15,37,38]. Other factors are plausible given our understanding of bacterial transcription. Sigma factor levels are regulated to control gene expression in response to cell state and external signals[16], so it is reasonable that we observed variable levels of activation from promoters with alternative sigma factors. Many bacterial genes are controlled by negative regulators[39], and different sequences upstream of the minimal promoter could be recruiting repressors.

The most unexpected property that we observed with bacterial CRISPRa was its sharp, periodic dependence on target site position. This behavior is quite distinct from CRISPRa in eukaryotes, where a broad range of sites upstream of the TSS are effective[40], possibly because eukaryotic activators typically recruit transcription factors and chromatin modifying machinery rather than directly recruiting RNA polymerase. There is precedent for bacterial transcriptional activators that are sensitive to target site periodicity[27–34], but the dramatic changes in activity with only single base shifts is surprising. Moreover, it is puzzling that we were unable to predictably alter or broaden the range of sites that are effective. Our models for how activators interact with bacterial transcription machinery may be incomplete. It will likely be productive to continue screening for activity at out-of-phase target sites using additional systematic modifications to the CRISPRa complex structure, alternative CRISPR-Cas systems, and additional candidate transcriptional activation domains.

Despite the challenges described above for identifying effective CRISPRa sites in *E. coli*, our systematic characterization provides a framework for immediate practical applications and a path for future improvements. We now have a clear understanding of the criteria needed to design synthetic promoters

that can be regulated by CRISPRa, which will enable the construction of complex, tunable synthetic multi-gene circuits. To extend the scope of CRISPRa to endogenous target genes, expanded PAM variants like dxCas9(3.7)[6], or orthologous dCas9 proteins with alternate PAM specificities[41,42] will open more DNA sites for targeting, increasing the likelihood of finding a targetable site at an optimal position relative to the TSS. These strategies lay the groundwork for more widespread use of bacterial CRISPRa in basic research and practical applications including functional genomics screens, metabolic engineering, and synthetic microbial communities.

Methods

Bacterial strain construction and manipulation

Plasmids were cloned using standard molecular biology protocols. Bacterial strains with sfGFP or mRFP1 reporter strains are described in Supplementary Table 1. The CRISPRa system used for each figure panel is described in Supplementary Table 2. Guide RNA target sequences are described in Supplementary Table 3. Plasmid containing the reporter genes and the CRISPR components are described in Supplementary Table 4. *S. pyogenes* dCas9 (*Sp*-dCas9) or dxCas9(3.7) were expressed from the endogenous *Sp.pCas9* promoter in a p15A vector. MCP-SoxS containing wild-type and mutant SoxS were expressed using the BBa_J23107 promoter [http://parts.igem.org] in the same plasmid with dCas9. The scRNAs were expressed using the BBa_J23119 promoter, either in the same plasmid with the dCas9 protein and the activation domain or in a separate ColE1 plasmid. The scRNA.b1 or scRNA.b2 designs, where the endogenous tracr terminator hairpin upstream of MS2 was removed[3], were used in all experiments except otherwise noted. The *zwfp-lacZ* and *fumCp-lacZ* reporter plasmids were generated in a previous study[3]. mRFP1 and sfGFP reporters were expressed from the weak BBa_J23117 minimal promoter [http://parts.igem.org] in a low-copy pSC101** vector. Variant versions of reporter genes are described in the Supplementary Methods. Plasmid libraries containing N26 sequences between the scRNA target site and BBa_J23117 minimal promoter were constructed by PCR amplification using mixed bases oligos (IDT). The dxCas9(3.7)-VPR plasmid was a gift from David Liu (Addgene #108383)[6].

Flow cytometry

Single colonies from LB plates were inoculated in 500 µL EZ-RDM (Teknova) supplemented with appropriate antibiotics and grown in 96-deep-well plates at 37 °C and shaking. Cultures were grown overnight at 37 °C and shaking and then diluted in 1:50 in DPBS and analyzed on a MACSQuant VYB flow cytometer with the MACSQuantify 2.8 software (Miltenyi Biotec). A side scatter threshold trigger (SSC-H) was applied to enrich for single cells until 10000 events were collected. The FlowJo 10.0.7 software was used to apply a narrow gate along the diagonal line on the SSC-H vs SSC-A plot was selected to exclude the events where multiple cells were grouped together. Within the selected population, events that appeared on the edges of the FSC-A vs. SSC-A plot and the fluorescence histogram were excluded.

Plate reader experiments

Single colonies from LB plates were inoculated in 500 µL EZ-RDM (Teknova) supplemented with appropriate antibiotics and grown in 96-deep-well plates at 37 °C and shaking overnight. For experiments with the *E. coli* promoter collection[36] the activation domain was placed under the control of a tet-inducible promoter. Attempts to use constitutive CRISPRa were unsuccessful due to plasmid instability, possibly because of toxicity arising from increased expression of the target genes. Single colonies from LB plates were inoculated in 500 µL EZ-RDM supplemented with appropriate antibiotics and 400 nM anhydrotetracycline (aTc) and grown in 96-deep-well plates at 37 °C and shaking overnight. 150 µL of the overnight culture were transferred into a flat, clear-bottomed black 96-well plate and the OD_{600} and fluorescence were measured in a Biotek Synergy HTX plate reader and analyzed using the BioTek Gen5 2.07.17 software. For mRFP1 detection, the excitation wavelength was 540 nm and emission wavelength was 600 nm. For sfGFP detection, the excitation wavelength was 485 nm and emission wavelength was 528 nm.

Quantitative RT-PCR

Single colonies from LB plates were inoculated in 5 mL LB containing appropriate antibiotics and grown overnight at 37 °C and shaking. Overnight cultures were diluted 1:100 into 5 mL EZ-RDM supplemented with appropriate antibiotics and grown at 37 °C and shaking until an OD_{600} of 0.5 (using 150 µL of culture in a 96 well plate) was reached. For the experiments targeting *yajG* and *poxB*, the activation domain was placed under the control of a tet-inducible promoter and cultures in EZ-RDM were supplemented with 400 nM aTc. Cultures were pelleted and total RNA was extracted using the Aurum Total RNA Mini Kit (Bio-rad). Reverse transcription reactions were performed from 1 µg RNA in 20 µL reactions using iScript reverse transcriptase (Bio-Rad). qPCR reactions were prepared in triplicate in a final volume of 10 µL using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), 0.5-5 ng of cDNA and 400 nM primers. The reaction was performed in a CFX Connect (Bio-Rad) with a 58 °C annealing temperature and 30 s extension time. A list of the qPCR primer sequences is provided in Supplementary Table 5. Expression levels for each gene were calculated in the Bio-Rad CFX Maestro 4.0.23225.0418 software by normalizing to the 16S rRNA gene and relative to a negative control carrying an off target-scRNA using the ΔΔCT method[43].

Statistics and reproducibility

Statistical significance was calculated using two-tailed unpaired Welch's t-tests. To ensure reproducibility, experiments were performed using n = 3 biologically independent samples, unless otherwise noted.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The datasets generated and analyzed during the current study are available from the corresponding author upon request.

The source data underlying Figures 1B, 2A-B, 3A-D, 4A-B, 5A-B, 6A-B and Supplementary Figures 1, 2, 3A-B, 4 , 5A-F, 6, 7A-C, 8A-C, 9, 10A-D, 11 are provided as a Source Data file.

Code availability

Custom Python code to generate the DNA sequences between transcriptional units in *E. coli* and analyze the density of PAM sites in these sequences (detailed in the Supplementary Methods) is available on GitHub [https://github.com/carothersresearch/Fontana-Dong_2020_NatComm].

Acknowledgements

The dxCas9(3.7)-VPR plasmid was a gift from David Liu (Addgene #108383). The authors thank Mary Lidstrom, Joanne Wong, Semira Beraki and members of the Zalatan and Carothers groups for technical assistance, advice, and helpful discussions. This work was supported by NSF Award 1817623 (J.M.C, J.G.Z.) and NSF Award 1844152 (J.M.C.).

Author Contributions

J.F., C.D., C.K., B.I.T., J.M.C., and J.G.Z. designed experiments and analyzed data. J.F., C.D., C.K., V.P.C., and B.I.T. performed experiments. J.F., C.D., J.M.C. and J.G.Z. wrote the manuscript.

Competing Interests

The authors declare no competing interests.

References

1. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11**, 508–520 (2014).
2. Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
3. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat Commun* **9**, 2489 (2018).
4. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
5. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome gditing and beyond. *Annual Review of Biochemistry* **85**, 227–264 (2016).
6. Johnny H. Hu, D. R. L., Shannon M. Miller, Maarten H. Geurts, Weixin Tang, Liwei Chen, Ning Sun, Christina M. Zeina, Xue Gao, Holly A. Rees, Zhi Lin. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
7. Yang Liu, B. W., Xinyi Wan. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nature Communications* **10**, 3693 (2019).

8. Keseler, I. M. *et al.* The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, D543–D550 (2017).
9. Rodriguez, A. *et al.* Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microb Cell Fact* **13**, 126 (2014).
10. C R Byrne, N. M. K., R. S. Monroe, K. A. Ward. DNA sequences of the *cysK* regions of *Salmonella typhimurium* and *Escherichia coli* and linkage of the *cysK* regions to *ptsH*. *Journal of Bacteriology* **170**, 3150–3157 (1988).
11. Gene Ruijun Jiang, D. P. C., Sonia Nikolova. Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology* **147**, 2437–2446 (2001).
12. Olin K. Silander, M. A., Nela Nikolic, Alon Zaslaver, Anat Bren, Ilya Kikoin, Uri Alon. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genetics* **8**, 5 (2012).
13. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12**, 326–328 (2015).
14. Alejandro Chavez, G. C., Marcelle Tuttle, Benjamin W. Pruitt, Ben Ewen-Campen, Raj Chari, Dmitry Ter-Ovanesyan, Sabina J. Haque, Ryan J. Cecchi, Emma J. K. Kowal, Joanna Buchthal, Benjamin E. Housden, Norbert Perrimon, James J. Collins. Comparison of Cas9 activators in multiple species. *Nature Methods* **13**, 563–567 (2016).
15. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
16. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).
17. Ishita M. Shah, R. E. W. Novel protein–protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase α subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *Journal of Molecular Biology* **343**, 513–532 (2004).
18. Amy Strohmeier Gort, J. A. I., Daniel M. Ferber. The regulation and role of the periplasmic copper, zinc superoxide dismutase of *Escherichia coli*. *Molecular Microbiology* **32**, 179–191 (1999).
19. Gen Nonaka, V. A. R., Matthew Blankschien, Christophe Herman, Carol A. Gross. Regulon and promoter analysis of the *E. coli* heat-shock factor, σ^{32}, reveals a multifaceted cellular response to heat stress. *Genes & Development* **20**, 1776–1789 (2006).
20. Virgil A Rhodius, C. A. G., Won Chul Suh, Gen Nonaka, Joyce West. Conserved and variable functions of the σ^{E} stress response in related genomes. *PLoS Biology* **4**, e2 (2006).
21. Zhe-Xian Tian, Y. W., Quan-Sheng Li, Martin Buck, Annie Kolb. The CRP–cAMP complex and downregulation of the *glnAp2* promoter provides a novel regulatory linkage between carbon metabolism and nitrogen assimilation in *Escherichia coli*. *Molecular Microbiology* **41**, 911–924 (2001).
22. Avihu H. Yona, J. G., Eric J. Alm. Random sequences rapidly evolve into de novo promoters. *Nature Communications* **9**, 1530 (2018).

23. Lee, T. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* **5**, 12 (2011).
24. Alfredo Mendoza-Vargas, E. M., Leticia Olvera, Maricela Olvera, Ricardo Grande, Leticia Vega-Alvarado, Blanca Taboada, Verónica Jimenez-Jacinto, Heladia Salgado, Katy Juárez, Bruno Contreras-Moreira, Araceli M. Huerta, Julio Collado-Vides. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* **4**, e7526 (2009).
25. Georgina Lloyd, S. B., Paolo Landini. Activation and repression of transcription initiation in bacteria. *Essays In Biochemistry* **37**, 17–31 (2001).
26. Schneider, T. D. 70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence. *Nucleic Acids Res.* **38**, 5995–6006 (2010).
27. Timothy I. Wood, R. E. W., Kevin L. Griffith, William P. Fawcett, Kam-Wing Jair, Thomas D. Schneider. Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Molecular Microbiology* **34**, 414–430 (1999).
28. Yi Zhou, Y.-P. W., Annie Kolb, Stephen J. W. Busby. Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Research* **42**, 9209–9216 (2014).
29. Johannes Müller, B. M.-H., Andrew Barker, Stefan Oehler. Dimeric *lac* repressors exhibit phase-dependent co-operativity. *Journal of Molecular Biology* **284**, 851–857 (1998).
30. Straney, D. C., Straney, S. B. & Crothers, D. M. Synergy between *Escherichia coli* CAP protein and RNA polymerase in the *lac* promoter open complex. *J. Mol. Biol.* **206**, 41–57 (1989).
31. Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. Stringent spacing requirements for transcription activation by CRP. *Cell* **62**, 733–743 (1990).
32. Ushida, C. & Aiba, H. Helical phase dependent action of CRP: effect of the distance between the CRP site and the −35 region on promoter activity. *Nucleic Acids Res* **18**, 6325–6330 (1990).
33. Müller, J., Oehler, S. & Müller-Hill, B. Repression of *lac* Promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *Journal of Molecular Biology* **257**, 21–29 (1996).
34. Martin, R. G., Gillette, W. K., Rhee, S. & Rosner, J. L. Structural requirements for marbox function in transcriptional activation of *mar*/*sox*/*rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. *Mol. Microbiol.* **34**, 431–441 (1999).
35. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429–7437 (2013).
36. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3**, 623–628 (2006).
37. Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat Methods* **11**, 723–726 (2014).

38. Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M. & Klavins, E. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat Commun* **8**, 15459 (2017).
39. Rojo, F. Repression of transcription initiation in bacteria. *Journal of bacteriology* **181**, 2987–91 (1999).
40. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
41. Ryan T. Leenay, C. L. B. Deciphering, communicating, and engineering the CRISPR PAM. *Journal of Molecular Biology* **429**, 177–191 (2017).
42. Sergey Shmakov, E. V. K., Aaron Smargon, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, Jonathan S. Gootenberg, Kira S. Makarova, Yuri I. Wolf, Konstantin Severinov, Feng Zhang. Diversity and evolution of class 2 CRISPR-Cas systems. *Nature Reviews Microbiology* **15**, 169–182 (2017).
43. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
44. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: Identifying amino acids important for DNA binding and transcription activation. *Journal of Molecular Biology* **322**, 237–257 (2002).

Figures

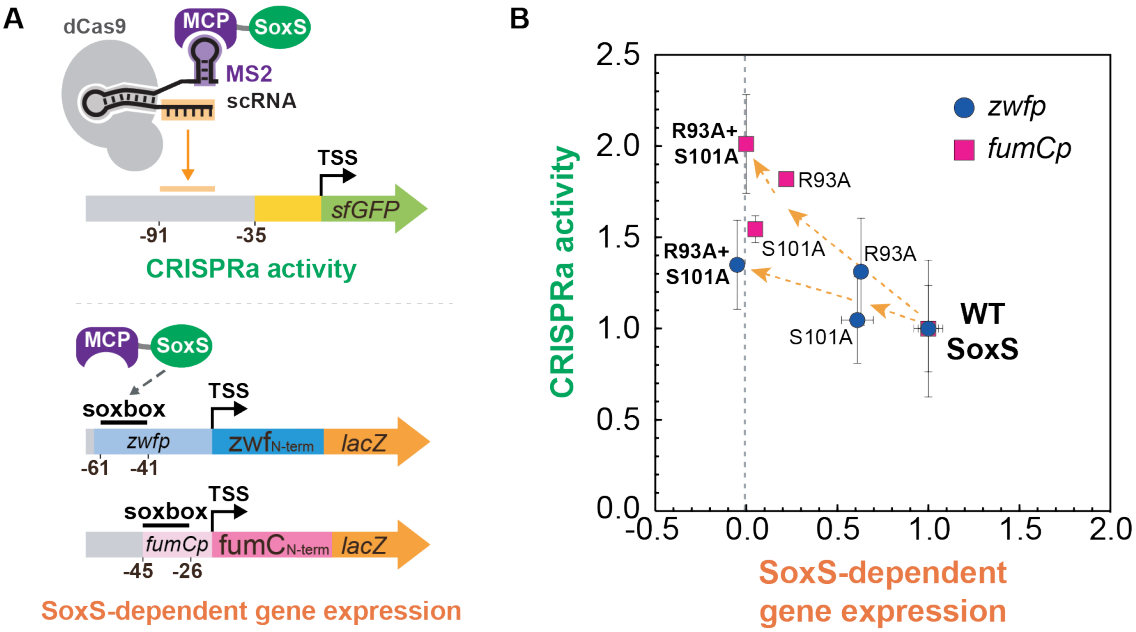


Figure 1. A SoxS double mutant maintains CRISPRa activity and does not activate endogenous SoxS targets.

A) Reporter system for measuring the CRISPRa activity and endogenous SoxS-dependent gene expression of wild-type or mutant SoxS constructs. CRISPRa activity was determined in a strain harboring a genomically-integrated sfGFP reporter (CD06, Supplementary Table 1). The endogenous SoxS-dependent gene expression was determined by monitoring *lacZ* expression from reporter plasmids where *lacZ* was driven by SoxS-regulated promoters *zwfp* and *fumCp*[44]. GFP fluorescence was measured by flow cytometry and *lacZ* activity was measured using a β-galactosidase assay. **B)** SoxS(R93A/S101A) maintains CRISPRa activity and does not activate expression from the endogenous expression from the *zwfp* and *fumCp* reporters. Fluorescence and *lacZ* activity values were baseline-subtracted using a strain that does not express a scRNA. Both GFP levels and *lacZ* activities were normalized to the values observed in the strain with wild-type SoxS. Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panel B are provided as a Source Data file.

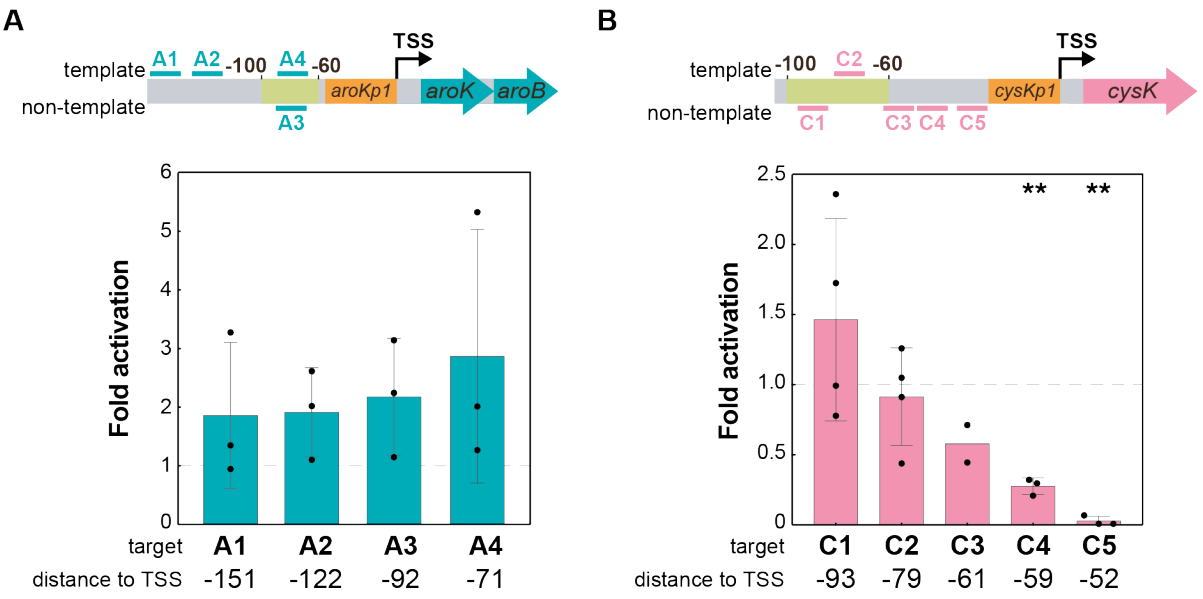


Figure 2. A simple distance metric does not predict CRISPRa activity.

A) CRISPRa on the *aroK-aroB* operon. Two scRNA target sites within the 40 base window where CRISPRa is effective (-100 to -60) in heterologous reporter genes (A3-A4) and two sites further upstream (A1-A2) were chosen for the *aroKp1* promoter. **B)** CRISPRa on the *cysK* gene. Three scRNA target sites within the 40 base window where CRISPRa is effective in heterologous reporter genes (C1-C3) and two sites further downstream (C4-C5) were chosen for the *cysKp2* promoter. The C4 and C5 sites resulted in repression; targeting these sites close to the core promoter may interfere with RNA polymerase binding. Gene expression was measured using RT-qPCR. Fold activation represents expression levels relative to a strain expressing an off-target scRNA (hAAVS1). In panels A and B, bars represent the average +/- standard deviation calculated from n = 4 (C1, C2), n = 3 (A1-A4, C4,C5) or n = 2 (C3) biologically independent samples. Some individual replicates in samples A1-A4 and C1-C2 appear to show activation, but a two-tailed unpaired Welch's t-test indicates that the average differences relative to the off-target control are not statistically significant (*p*-value > 0.05). Targeting CRISPRa to C4-C5 resulted in repression, likely due the short distance between the sites and the promoter. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's t-test (**: *p*-value < 0.01). Exact *p*-values: A1: 0.18, A2: 0.27, A3: 0.36, A4: 0.17, C1: 0.29, C2: 0.66, C3: 0.098, C4: 0.0022, C5: 0.00043. Source data for panels A and B are provided as a Source Data file.

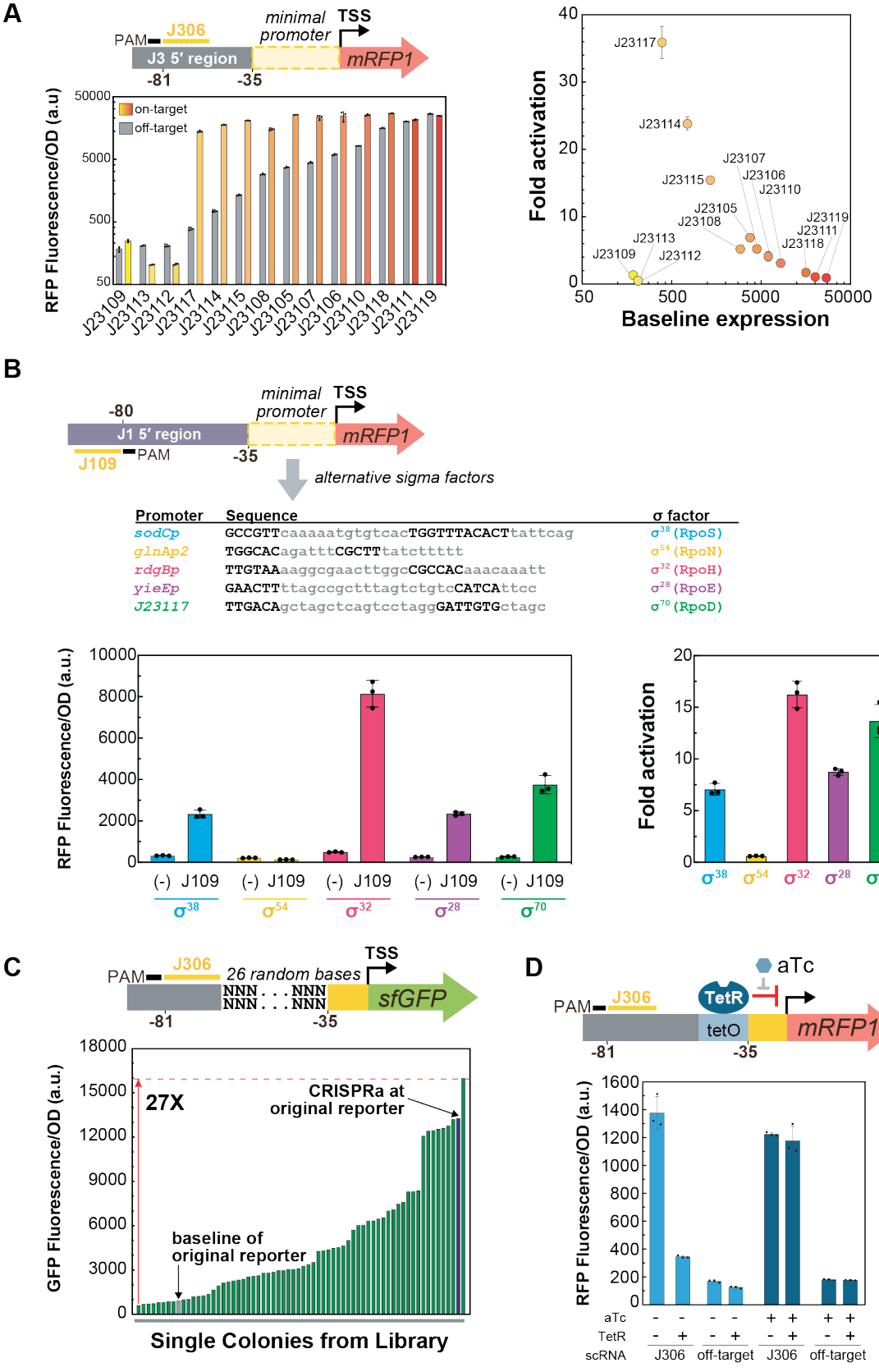

A
PAM
J306
minimal promoter
TSS
J3 5′ region
mRFP1
-81
-35
RFP Fluorescence/OD (a.u)
on-target
off-target
J23109 J23113 J23112 J23117 J23114 J23115 J23108 J23105 J23107 J23106 J23110 J23118 J23111 J23119
Fold activation
Baseline expression
B
-80
J1 5′ region
J109
PAM
alternative sigma factors
Promoter Sequence σ factor
sodCp GCCGTTcaaaaatgtgtcacTGGTTTACACTtattcag σ38 (RpoS)
glnAp2 TGGCACagatttCGCTTtatcttttt σ54 (RpoN)
rdgBp TTGTAAaaggcgaacttggcCGCCACaaacaaatt σ32 (RpoH)
yieEp GAACTTttagccgctttagtctgtcCATCAttcc σ28 (RpoE)
J23117 TTGACAgctagctcagtcctaggGATTGTGctagc σ70 (RpoD)
(-) J109
σ38 σ54 σ32 σ28 σ70
C
26 random bases
NNN . . . NNN
sfGFP
GFP Fluorescence/OD (a.u.)
27X
CRISPRa at original reporter
baseline of original reporter
Single Colonies from Library
D
aTc
TetR
tetO
aTc
TetR
scRNA
J306 off-target J306 off-target

Figure 3. CRISPRa is sensitive to promoter identity and local sequence.

A) CRISPRa is sensitive to promoter strength. Promoters contain a scRNA target site at -81 from the TSS of the indicated J231NN minimal promoter, on the non-template strand[3]. The panel on the left shows the Fluorescence/OD_{600} of strains expressing an on-target or off-target scRNA. The panel on the right shows the fold activation measured at each promoter relative to their baseline expression with an off-target scRNA (J206). **B)** CRISPRa can activate promoters regulated by σ^{38} (RpoS), σ^{32} (RpoH), and σ^{24} (RpoE) sigma factors. The minimal promoter from the reporter plasmid was replaced with *sodCp*, *glnAp2*, *rdgBp*, or *yieEp*. The -35 and -10 regions are highlighted in bold. The plot on the left shows the Fluorescence/OD_{600} when CRISPRa targeted each promoter at the J109 target site (-80 from the TSS on the template strand) or with an off-target scRNA (hAAVS1, labeled (–)). The plot on the right shows the fold activation measured at each promoter relative to an off-target scRNA (J206). **C)** CRISPRa activity differs significantly among promoters with varying sequence composition between the scRNA target and the -35 region. Green bars represent the Fluorescence/OD_{600} of overnight cultures from individual colonies. The blue bar represents the Fluorescence/OD_{600} of a strain expressing the J3-J23117-sfGFP reporter, activated by CRISPRa with the J306 scRNA. The grey bar represents a negative control expressing the J3-J23117-sfGFP reporter plasmid with CRISPRa targeting an off-target site (J206). **D)** CRISPRa was inhibited binding of the TetR transcriptional repressor binding to a tet operator (tetO) site placed upstream of the -35 region. Cultures where CRISPRa was targeted to the J306 site or to an off-target site (J206) were grown overnight in media +/- 1 μM aTc. In panels A, B and D, values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. In panel C, bars represent the value of n = 1 biologically independent samples. Source data for panels A, B, C and D are provided as a Source Data file.

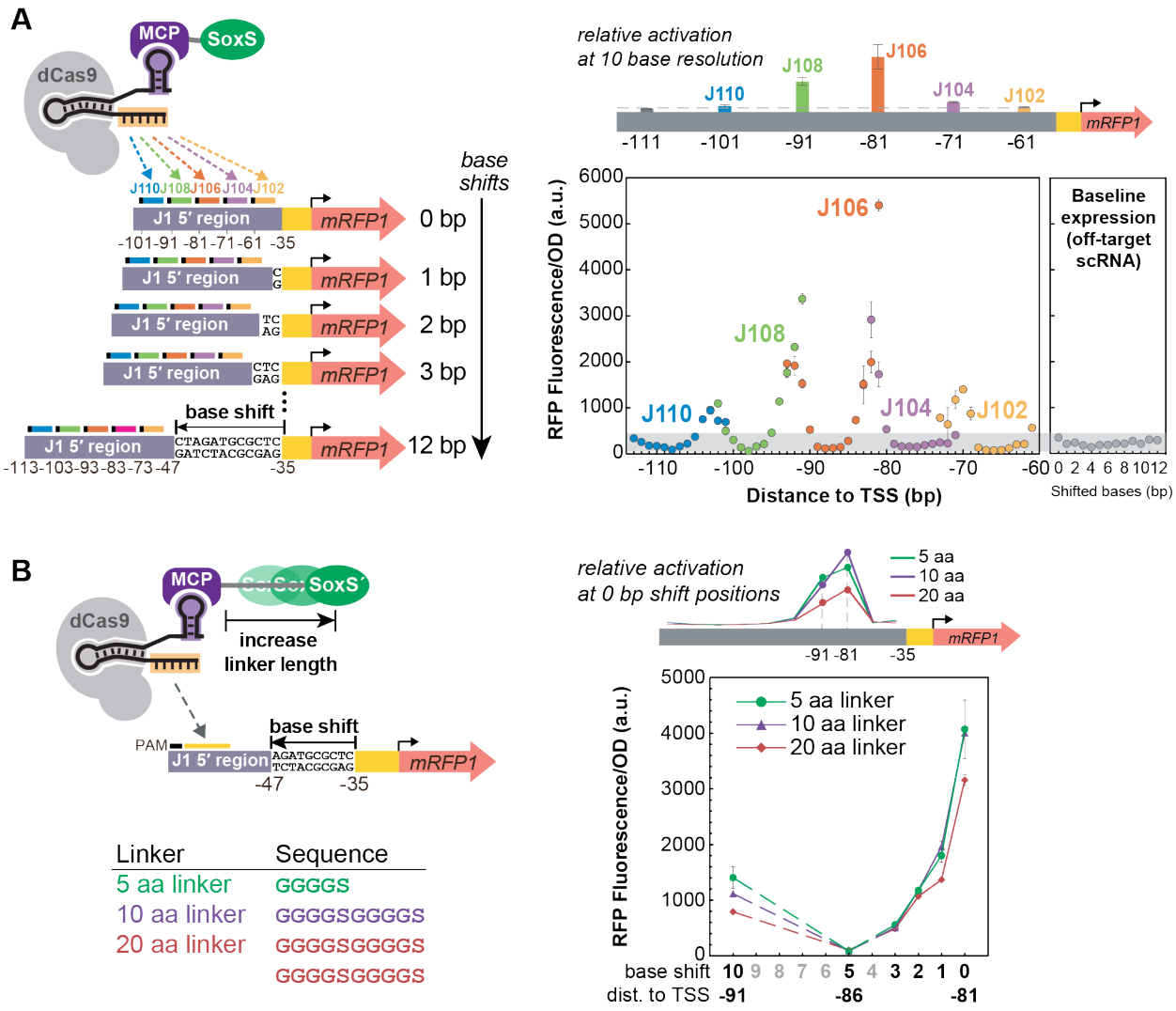


Figure 4. CRISPRa is sensitive to the precise position of the scRNA target.

A) CRISPRa displays periodic positioning dependence with peak activities every 10-11 bases between -60 to -100 from the TSS. Reporter genes were constructed by inserting 0-12 bases upstream of the -35 region of the J1-J23117-mRFP1 reporter. Five scRNA sites (J102, J104, J106, J108, J110) with positions -61, -71, -81, -91, -101 from the TSS on the non-template strand of the original promoter were targeted. In this way, the complete -61 to -113 region can be covered at single base resolution. The color coding indicates data for the same target site shifted across a 12 base window. The panel on the right shows the baseline expression of reporters with shifted bases when an off-target scRNA was used (J206). The grey area represents the range of the baselines among the reporter series. For comparison, previous CRISPRa data for the J102, J104, J106, J108, J110 target positions at 10 base resolution are shown on the schematic above the plot[3]. **B)** Extending the linker length between MCP and SoxS does not change the position dependence of CRISPRa. The J1-J23117-mRFP1 reporter plasmid series with base shifts were delivered together with CRISPRa components for targeting J106. The MCP-SoxS(R93A) effector contained 5aa,

10aa, and 20aa linkers. For comparison, previous CRISPRa data with 5aa, 10aa, or 20aa linker between MCP and SoxS targeting at the -81 and -91 positions are shown on the schematic above the plot[3]. Values in panels A and B represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A and B are provided as a Source Data file.

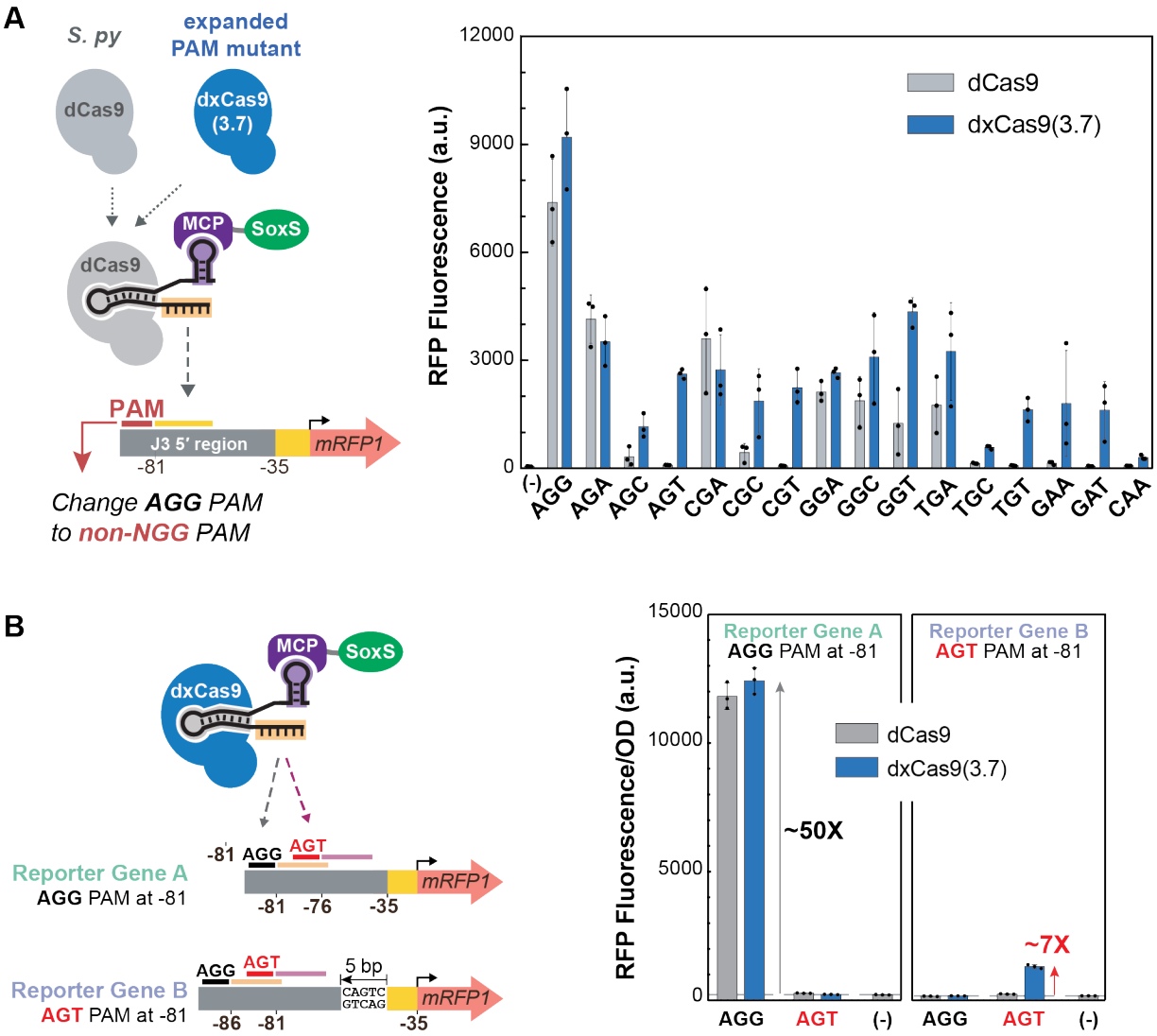


Figure 5. dxCas9(3.7) expands the range of targetable scRNA target sites by recognizing alternative PAMs.

A) CRISPRa with dxCas9(3.7) displayed activity on non-NGG PAM sites with AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA sequences. CRISPRa activity with dxCas9(3.7) on non-NGG PAM sites was generally lower (6-fold to 89-fold activation relative to a control without a scRNA) compared to the AGG PAM site (188-fold activation). *Sp*-dCas9 also displayed moderate CRISPRa activity at non-NGG PAM sites with AGA, CGA, GGA, GGC, GGT, TGA sequences, consistent with published reports[6]. Reporter plasmids were constructed by replacing the AGG PAM site for the J306 target in the J3-J23117-mRFP1 reporter with alternative PAM sequences that have been previously reported to be recognized by dxCas9(3.7) in human cells[6]. The (-) sign indicates a control expressing the original reporter with the AGG PAM and the CRISPRa components with *Sp*-dCas9, the activation domain and no scRNA. **B)** dxCas9(3.7) can activate promoters that cannot be activated by *Sp*-dCas9. When the scRNA target at the optimal position (-81 to the TSS) has an AGG PAM site, both *Sp*-dCas9 and dxCas9(3.7)

increased gene expression by 50-fold. When the scRNA target at the optimal position has an AGT PAM site, only dxCas9(3.7) displayed a 7-fold increase in gene expression while *Sp*-dCas9 was inactive. The reporter gene has a target with an AGG PAM (M1) and a target with an AGT PAM (M2) upstream of a BBa_J23117 minimal promoter. In reporter gene A, the AGG target was located -81 to the TSS on the non-template strand and the AGT target was located -76 to the TSS on the non-template strand. In reporter gene B, 5 bases were inserted upstream of the -35 region, shifting the locations of the AGG target and AGT target to -86 and -81, respectively. The (-) sign indicates a negative control strain that contains the reporter plasmid and a plasmid expressing Sp-dCas9, the activation domain and an off-target scRNA (J206). Bars in panels A and B represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A and B are provided as a Source Data file.

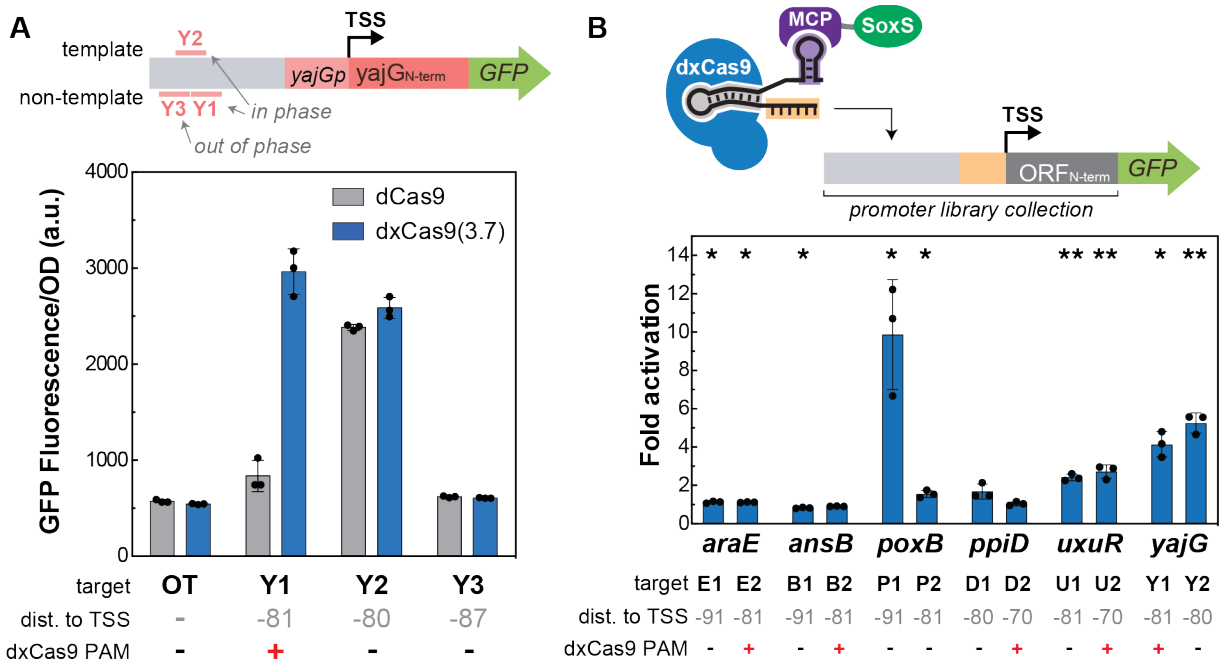


Figure 6. Predictive rules enable endogenous activation.

A) CRISPRa using dCas9 and dxCas9(3.7) was targeted to a *yajG* reporter plasmid from the *E. coli* promoter collection[36]. Three scRNA target sites were selected; two sites were located at the positions where CRISPRa was most effective (Y1-Y2), and one was located out of phase (Y3). A negative control (OT) expressing an off-target scRNA (J306) was included. **B)** CRISPRa was targeted to *yajG* and five additional promoters from the *E. coli* promoter collection (Supplementary Methods). Two scRNA sites located at the positions where CRISPRa was most effective were targeted for each gene using dxCas9(3.7). Samples are arranged by baseline expression of the target genes, in ascending order left to right. Fold activation indicates the median fluorescence of strains relative to an off-target control (J306). Values in panels A and B represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's t-test (*: *p*-value < 0.05, **: *p*-value < 0.01). Exact *p*-values: E1: 0.036, E2: 0.024, B1: 0.031, B2: 0.141, P1: 0.033, P2: 0.021, D1: 0.088, D2: 0.585, U1: 0.0008, U2: 0.001, Y1: 0.013, Y2: 0.003. Source data for panels A and B are provided as a Source Data file.

Chapter 2: Supplementary Information

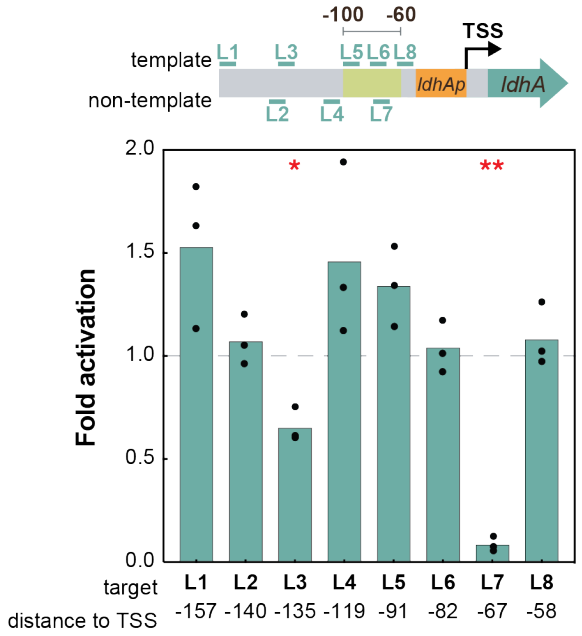
Jason Fontana*,[1], Chen Dong*,[2], Cholpisit Kiattisewee[1], Venkata P. Chavali[3], Benjamin I. Tickman[1], James M. Carothers[1,3,4], Jesse G. Zalatan[2,3,4]

* these authors contributed equally

[1]Molecular Engineering & Sciences Institute
[2]Department of Chemistry
[3]Department of Chemical Engineering
[4]Center for Synthetic Biology
University of Washington, Seattle, WA 98195, USA

Supplementary figures

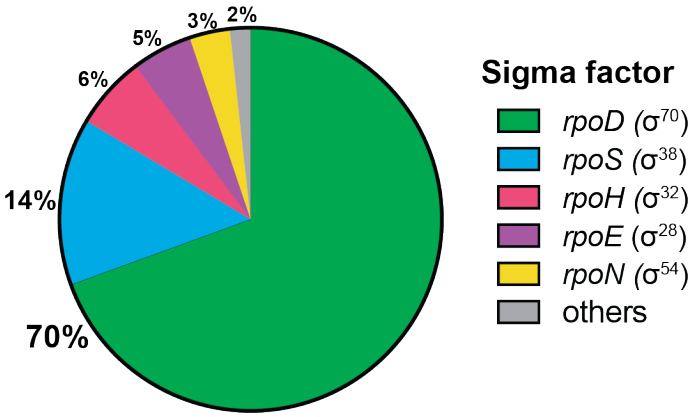
Figure S1



Supplementary Figure 1: CRISPRa at the endogenous gene target *ldhA* does not follow predicted trends.

Eight scRNA target sites (L1-L8) upstream of the *ldhA* promoter were selected. Three of the target sites (L5-L7) were within the 40 bp window where CRISPRa is effective (-100 to -60). While L1, L4, and L5 resulted in weak increases in gene expression, there was no apparent relationship between the position of the sites and *ldhA* expression levels. Gene expression was measured using RT-qPCR. Fold activation represents expression levels relative to an off-target control (hAAVS1). Values represent the average calculated from n = 3 technical replicates. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's t-test (*: *p-value* < 0.05, **: *p-value* < 0.01). Exact *p-values*: L1: 0.12, L2: 0.43, L3: 0.02, L4: 0.20, L5: 0.09, L6: 0.67, L7: 0.0005, L8: 0.45. Source data are provided as a Source Data file.

Figure S2

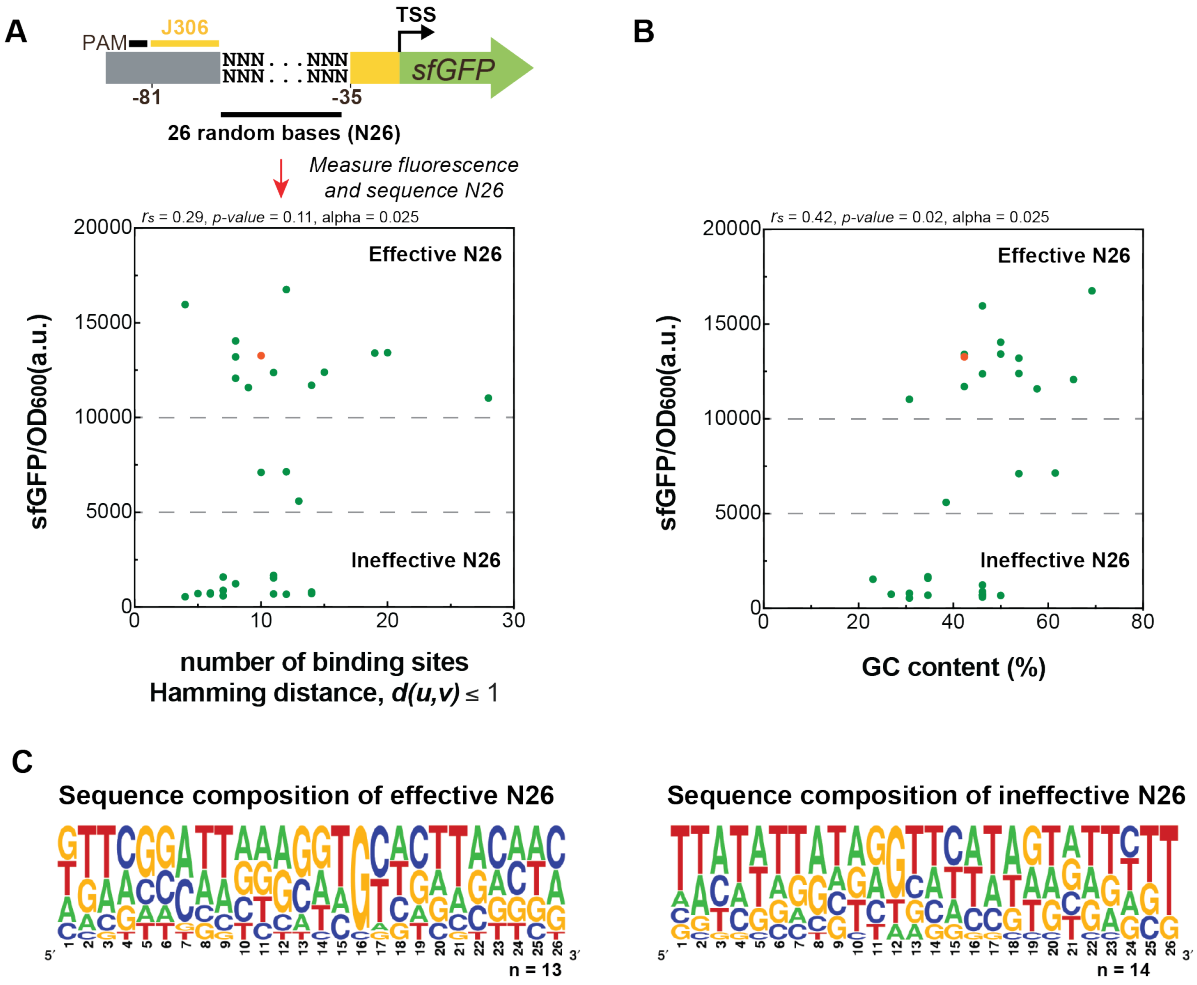


Total transcriptional units = 1609

Supplementary Figure 2: Distribution of transcriptional units regulated by sigma factors.

The number of *E. coli* transcriptional units regulated by sigma factors was obtained from Ecocyc[1] on the "Regulon" tab of the page relative to each sigma factor. The total number of transcriptional units represents the sum of the transcriptional units regulated by each sigma factor. Source data are provided as a Source Data file.

Figure S3

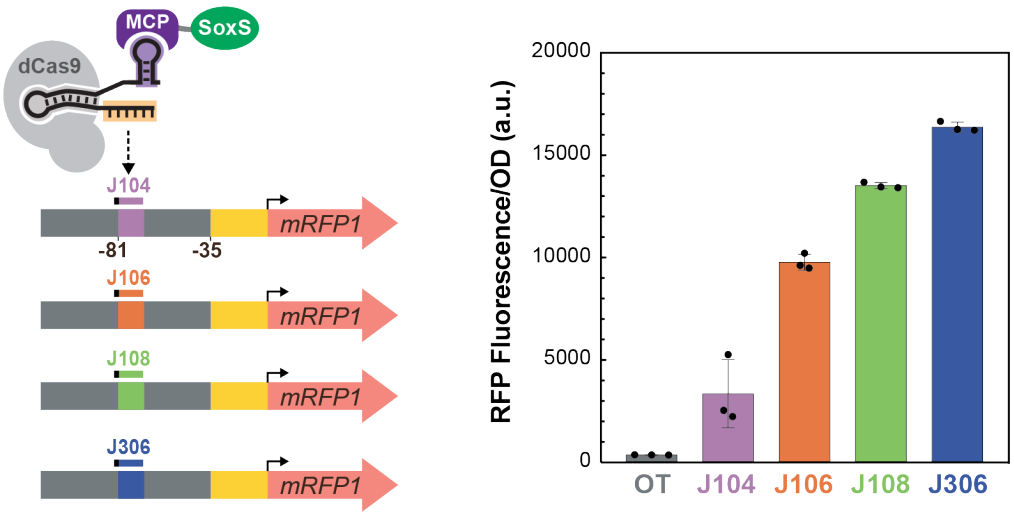


Supplementary Figure 3: Intervening sequences between the scRNA target site and minimal promoter that interfere with CRISPRa tend to be more AT-rich.

A) Plot of gene activation vs. transcription factor binding sites present in the intervening sequence between the scRNA target site and the minimal promoter. CRISPRa was targeted to a reporter library with 26 randomized bases between the scRNA target site and the -35 region on the J3-J23117-sfGFP reporter (N26), and we observed a 27-fold variation in gene activation (Figure 3C). We sequenced 29 variants and identified a number of motifs within one base of known consensus transcription factor binding (Supplementary Table 6). In parallel, we measured gene activation (sfGFP/OD_{600}) for each strain. The plot shows the sfGFP/OD_{600} of each strain versus the number of sequences within a Hamming distance $d(u,v) \leq 1$ from the consensus sequences of transcription factor binding sites, obtained from RegulonDB[2]. r_s indicates the Spearman rank order correlation coefficient between sfGFP/OD_{600} and number of binding sites, and its associated two-tailed *p-value* is relative to the null hypothesis of no correlation between sfGFP/OD_{600} and number of binding sites, with a Bonferroni-corrected $\alpha = 0.025$. Green dots indicate the sfGFP/OD_{600} values and the number of binding sites calculated from individual colonies, and the orange

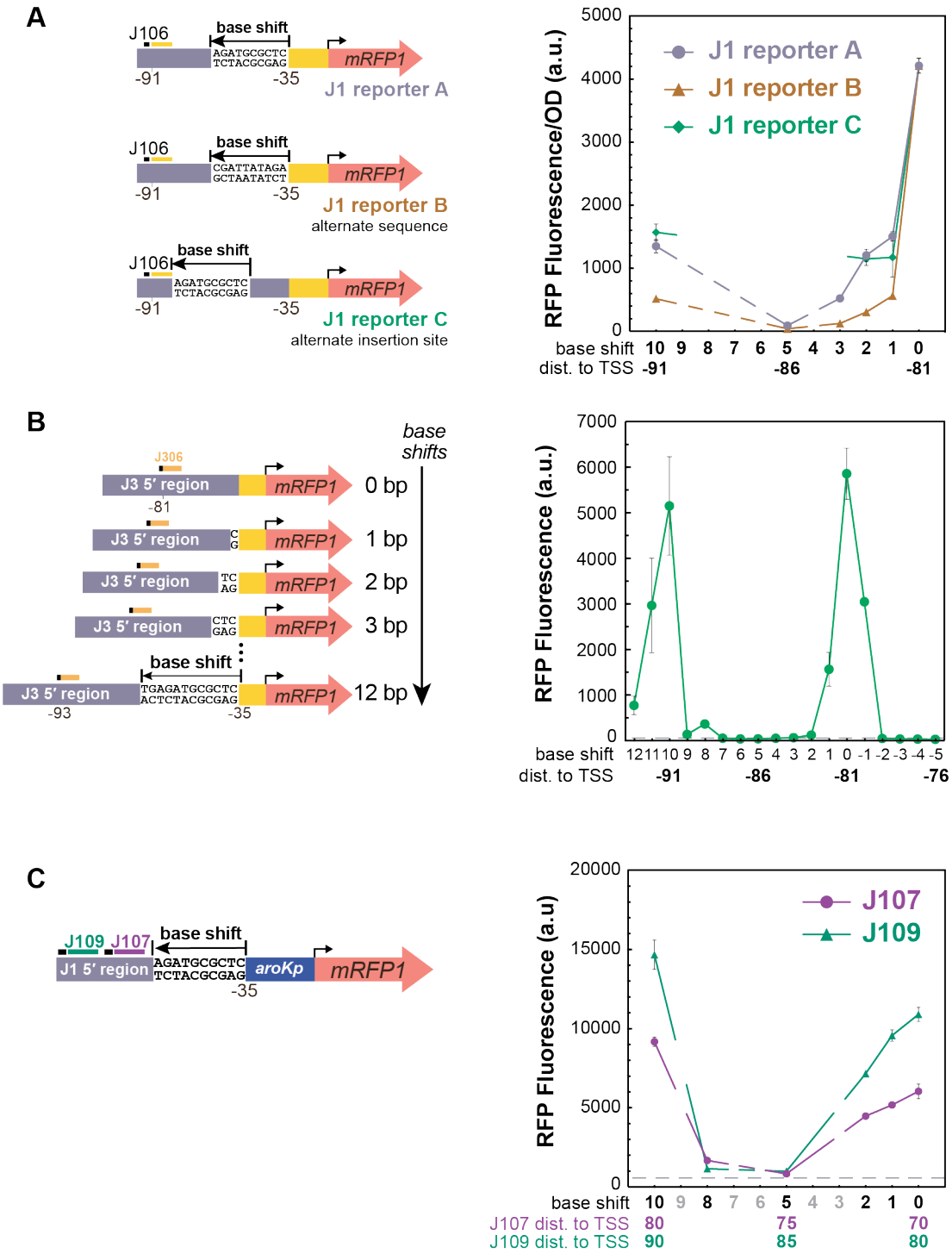
dot indicates a strain where CRISPRa is targeting the original J3-J23117-sfGFP promoter. **B)** Plot of gene activation vs. GC content in the intervening sequence between the scRNA target site and the minimal promoter. r_s indicates the Spearman rank order correlation coefficient between sfGFP/OD_{600} and GC content, and its associated two-tailed *p-value* is relative to the null hypothesis of no correlation between sfGFP/OD_{600} and GC content, with a Bonferroni-corrected $\alpha = 0.025$. In panels A and B, green dots indicate the sfGFP/OD_{600} values and GC content calculated from n = 1 biologically independent samples, and the orange dot indicates a strain where CRISPRa targets the original J3-J23117-sfGFP promoter. **C)** Logo plots (https://weblogo.berkeley.edu/) showing the base composition of N26 sequences effective for CRISPRa (sfGFP/OD_{600} > 10000 a.u., 13 sequences) and N26 sequences ineffective for CRISPRa (sfGFP/OD_{600} < 5000 a.u., 14 sequences). Source data for panels A and B are provided as a Source Data file.

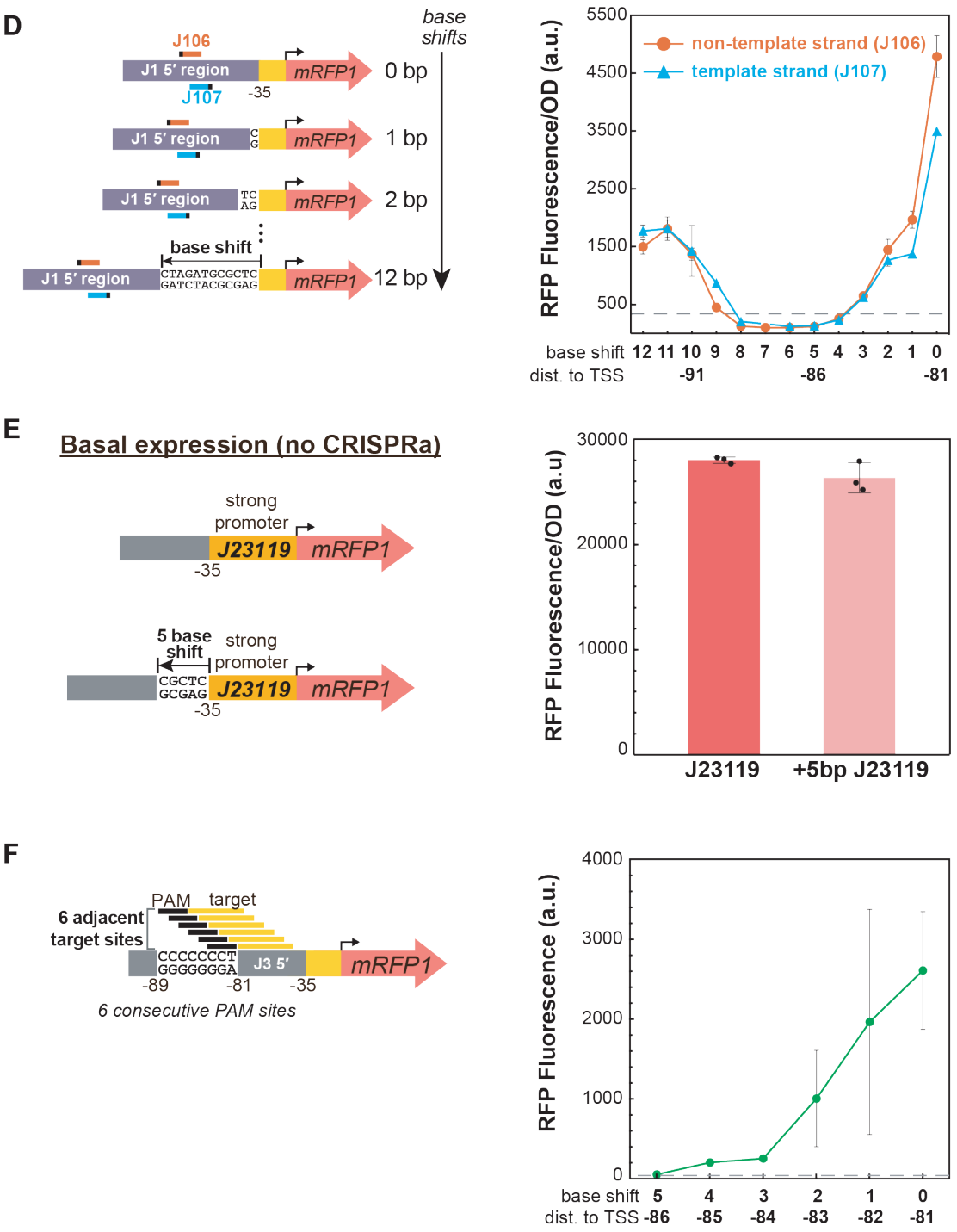
Figure S4



Supplementary Figure 4: CRISPRa activity depends on the target sequence on the scRNA. Reporter cassettes that differ only by the sequence of the 20 base scRNA target site give a broad range of gene expression levels, demonstrating that the sequence of the scRNA target site can have a substantial effect on CRISPRa. Three new reporter plasmids were constructed where the J306 target site, located at -81 from the TSS, on the J3-J23117-mRFP1 reporter was replaced by the J104, J106, and J108 sequence. Activation at each promoter was tested when CRISPRa was targeted to their cognate scRNA site. The off-target negative control (OT) represents a strain expressing the original reporter with the J306 site and the CRISPRa components to target an off-target site (J206). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

Figure S5


A
J106
base shift
AGATGCGCTC
TCTACGCGAG
mRFP1
-91
-35
J1 reporter A
J106
base shift
CGATTATAGA
GCTAATATCT
mRFP1
-91
-35
J1 reporter B
alternate sequence
J106
base shift
AGATGCGCTC
TCTACGCGAG
mRFP1
-91
-35
J1 reporter C
alternate insertion site
RFP Fluorescence/OD (a.u.)
J1 reporter A
J1 reporter B
J1 reporter C
base shift 10 9 8 7 6 5 4 3 2 1 0
dist. to TSS -91 -86 -81
B
base shifts
J306
J3 5′ region
mRFP1
-81
0 bp
1 bp
2 bp
3 bp
base shift
TGAGATGCGCTC
ACTCTACGCGAG
12 bp
-93
-35
RFP Fluorescence (a.u.)
base shift 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5
dist. to TSS -91 -86 -81 -76
C
J109 J107
base shift
J1 5′ region
AGATGCGCTC
TCTACGCGAG
aroKp
mRFP1
-35
RFP Fluorescence (a.u)
J107
J109
base shift 10 9 8 7 6 5 4 3 2 1 0
J107 dist. to TSS 80 75 70
J109 dist. to TSS 90 85 80



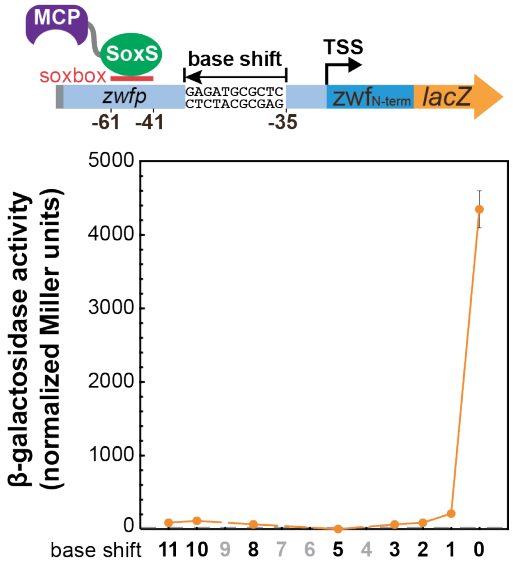
Supplementary Figure 5: The sharp positioning dependence of CRISPRa is observed across multiple promoters.

A) The sharp positioning requirements of CRISPRa are not significantly affected by the location or composition of the inserted sequence. Reporters were based on the J1-J23117-mRFP1 (Figure 4) with base shifts introduced in different ways. In the J1 reporter A, bases were inserted upstream of the -35 region. In the J1 reporter B, a different sequence was inserted at the same site. In the J1 reporter C, bases

were inserted downstream of the J106 target site. There were modest differences between reporter A and reporter B that could indicate a contribution from sequence composition. **B)** The sharp positioning requirements of CRISPRa were observed with a different heterologous promoter. CRISPRa was targeted at the -81 site on the J3-J23117 promoter that has a different upstream sequence and a different scRNA target site (J306) compared to J1-J23117. Peaks in gene expression were observed at -81 and -91 to the TSS, displaying a 10 bp periodicity. The peaks of gene expression on the J3-J23117 promoter were sharper than in the J1-J23117 promoter (Figure 4A); gene expression decreased to baseline levels after shifting only 2 bp from the peak position. After a complete 10 bp-shift period, CRISPRa activity was fully restored to the original peak expression. Reporter gene sets were constructed by inserting 0-12 bp or deleting 1-5 bp upstream of the -35 of the J3-J23117-mRFP1 reporter. The grey line represents the baseline activity of the J3-J23117-mRFP1 reporter strain containing an empty vector instead of the CRISPRa component plasmid. For comparison, previous CRISPRa data at -81 and -91 are shown on the schematic above the plot[3]. **C)** The sharp positioning requirements of CRISPRa are observed when targeting a different minimal promoter. The J1-*aroKp2* promoter displayed positioning requirements similar to the J1-J23117 promoter. Decreases in CRISPRa activity after 3 bp shifts and recovery at a 10 bp shift were observed on both the J107 and the J109 target sites. The J1-*aroKp2* promoter was constructed by replacing the BBa_J23117 minimal promoter from the J1-J23117 promoter to the aroKp2 minimal promoter. A J1-*aroKp2*-mRFP1 reporter series was constructed by adding 0 bp, 1 bp, 2 bp, 5 bp, 8 bp, and 10 bp upstream of the -35 region. The grey line represents the baseline activity of the J1-*aroKp2*-mRFP1 reporter reporter strain expressing a CRISPRa component plasmid with an off-target scRNA. **D)** The sharp positioning requirements of CRISPRa are observed on both the template and the non-template strand. For both the J106 target site on the non-template strand and the J107 target on the template strand, the CRISPRa activity decreased to baseline levels after shifting 3 bp from its original position and recovered after shifting 10 bp. The reporter plasmid was the same as in Figure 4A. The dotted grey line indicates the negative control where an off-target scRNA (J206) was co-transformed with the original reporter with no inserted bases. **E)** Adding 5 bases adjacent to the -35 region does not dramatically alter the expression of the promoter. The 5 bases added were the same as those used to shift the J1 and J3 promoters (Figure 4A and Supplementary Figure 5B). This experiment was performed using a strong minimal promoter (BBa_J23119), so that any detrimental effects would be detectable. **F)** The sharp positioning requirements of CRISPRa were observed when tested in a single reporter with multiple consecutive PAM sites. 6 consecutive PAM sites on the non-template strand were introduced by placing a CCCCCCCT sequence between -89 and -81 bp to the TSS on the J3-J23117-mRFP1 reporter. Maximum gene expression was observed at the original -81 site, after which expression gradually decreased to one third of the maximum activity after moving 2 bp away (-83). Gene expression decreased further when the scRNA target was moved 3 bp and 4 bp away from the TSS (-84 and -85), and reached the baseline when moved 5 bp (-86). The grey line represents the baseline activity of the reporter strain containing an empty vector instead of the CRISPRa component plasmid. Values in panels A-F represent the average +/- standard deviation

calculated from n = 3 biologically independent samples. Source data for panels A, B, C, D, E and F are provided as a Source Data file.

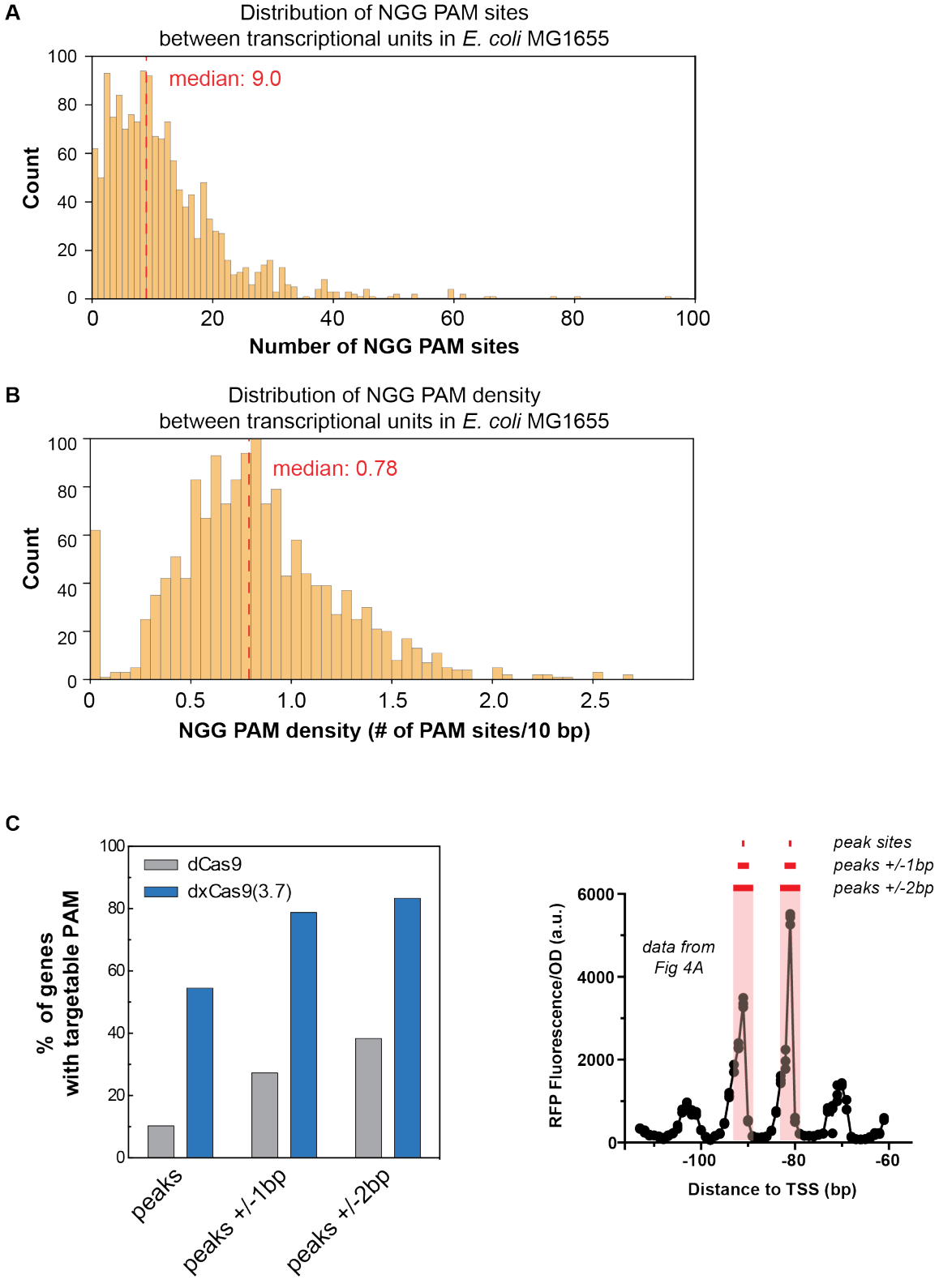
Figure S6



Supplementary Figure 6: Wild type SoxS displays a sharp positioning requirement when targeting a SoxS-dependent promoter.

When the wild type SoxS binding site (soxbox) on the endogenous *zwfp* promoter is shifted by 1 bp, gene expression decreases significantly, consistent with previous reports[4]. Shifting the soxbox further upstream causes gene expression to completely reduce to the baseline. Reporter plasmids were constructed by adding 1, 2, 3, 5, 8, 10, and 11 bp upstream of the -35 on the *zwfp-lacZ* reporter (Figure 1). Values represent the average β-galactosidase activity in normalized Miller +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

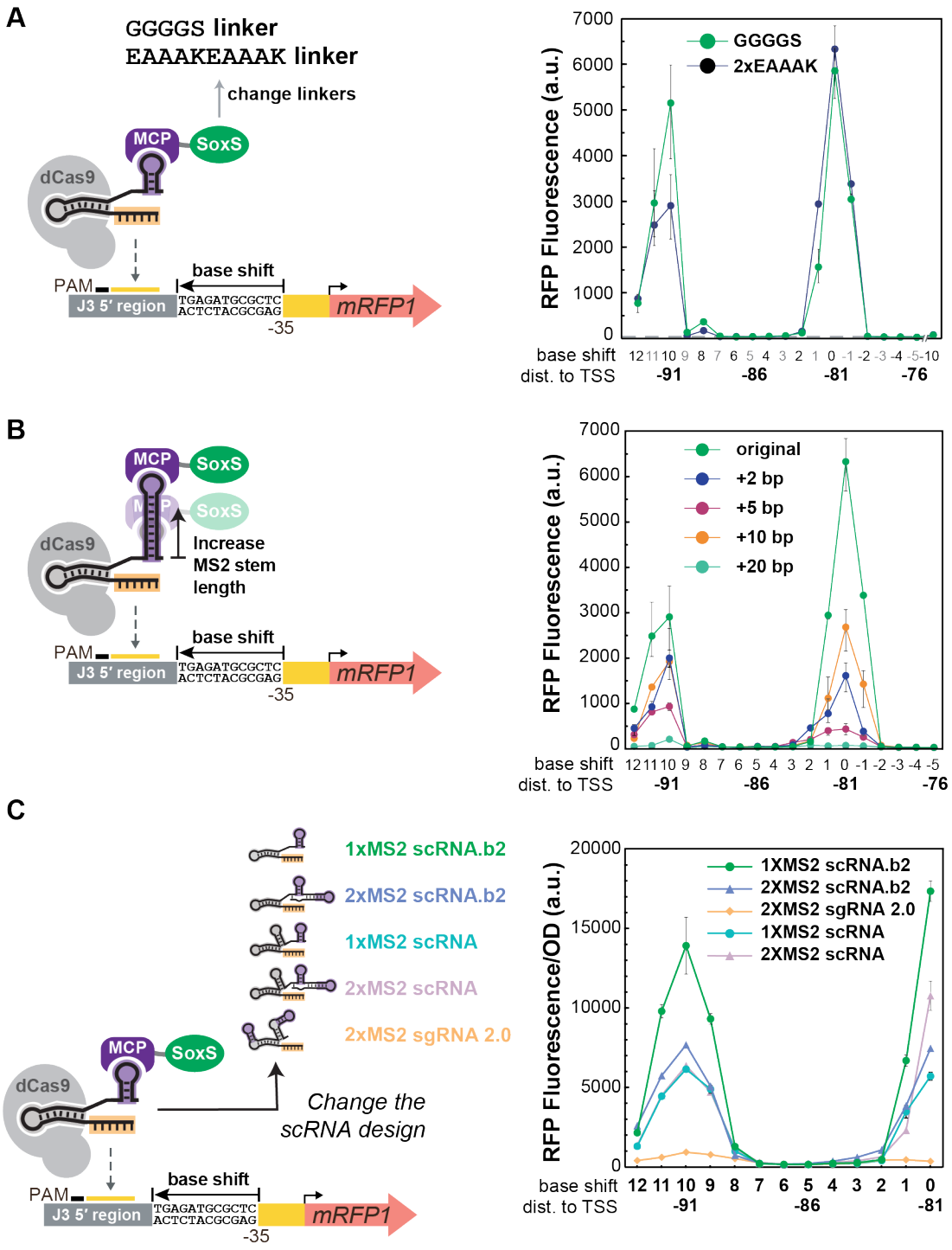
Figure S7



Supplementary Figure 7: Availability of PAM sites between transcriptional units in *E. coli* MG1655.

A) Distribution of the number of NGG PAM sites between transcriptional units in *E. coli* MG1655. The median number of NGG PAM sites between transcriptional units in *E. coli* MG1655 is 9.0. Methods for extracting the sequence of the DNA regions between transcriptional units in *E. coli* MG1655 and identifying available PAM sites are described in the Supplementary Methods. **B)** Distribution of the density of NGG PAM sites between transcriptional units in *E. coli* MG1655. The density of PAM sites is reported over a 10 bp window. This window size was chosen because it corresponds to a full turn of the DNA helix. The density of PAM sites in a 10 bp window was calculated for each sequence between transcriptional units as the total number of PAM sites in the sequence divided by the length of the intergenic sequence and then multiplied by 10. The median density of NGG PAM sites per 10 bp between transcriptional units in *E. coli* MG1655 is 0.78. **C)** The number of genes in the *E.coli* genome that can be targeted by dCas9 is small, and the expanded PAM variant dxCas9(3.7) significantly increases the number of genes with predicted effective target sites for CRISPRa. The plot to the left shows percentage of genes with at least one PAM site targetable by dCas9 or dxCas9(3.7) at: the positions where CRISPRa displays a peak in activity. The plot to the right illustrates the range of positions on the non-template strand chosen for the analysis. Corresponding peaks on the template strand were also included. Analyses were performed using data generated when selecting candidate endogenous genes for activation (Supplementary Methods). Source data for panels A, B and C are provided as a Source Data file.

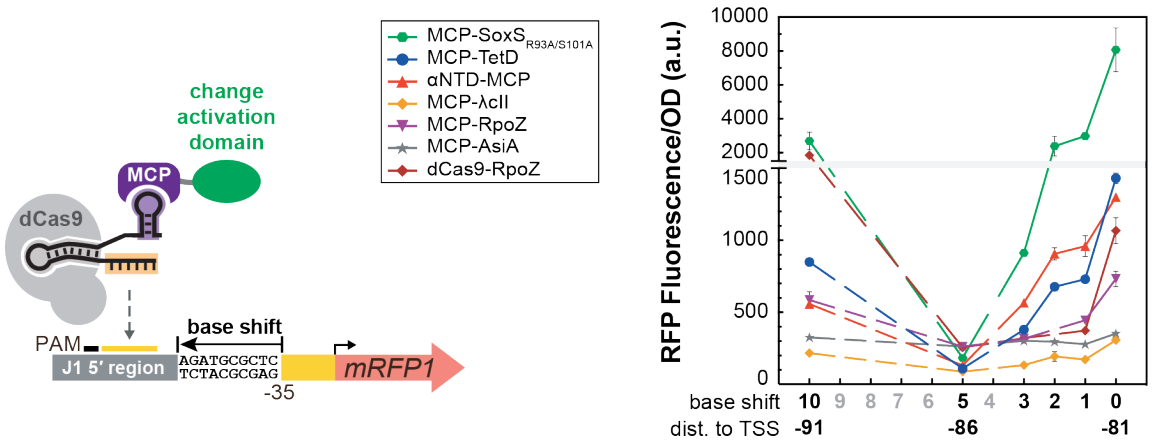
Figure S8


A
GGGGS linker
EAAAKEAAAK linker
change linkers
MCP
SoxS
dCas9
PAM
base shift
J3 5′ region
TGAGATGCGCTC
ACTCTACGCGAG
mRFP1
-35
RFP Fluorescence (a.u.)
GGGGS
2xEAAAK
base shift
dist. to TSS
-91
-86
-81
-76
B
Increase
MS2 stem
length
original
+2 bp
+5 bp
+10 bp
+20 bp
C
1xMS2 scRNA.b2
2xMS2 scRNA.b2
1xMS2 scRNA
2xMS2 scRNA
2xMS2 sgRNA 2.0
Change the
scRNA design
RFP Fluorescence/OD (a.u.)

Supplementary Figure 8: Modifying the CRISPRa complex structure does relax the sharp positioning requirements of CRISPRa.

A) Changing the linker between MCP and SoxS does not change the positioning dependence of CRISPRa. A CRISPRa complex with the MCP-(EAAAKEAAAK)-SoxS(R93A/S101A) activation domain displayed 10 bp periodicity of peak expression similar to that with the MCP-(GGGGS)-SoxS(R93A/S101A) activation domain. The EAAAKEAAAK linker is predicted to be more rigid than the GGGGS linker[5]. The grey line represents the baseline activity of the J3-J23117-mRFP1 reporter strain containing an empty vector instead of the CRISPRa component plasmid. **B)** Extending the length of the MS2 stem does not change the positioning dependence of CRISPRa. CRISPRa systems with scRNAs that have +2, +5, and +10 RNA base pairs added to the bottom of the MS2 stem displayed the same 10 bp periodicity but lower peak activity compared to the original 1xMS2 scRNA.b2. The strain having the scRNA with +20 bp extended MS2 stem did not show any CRISPRa activity. **C)** No improvements in the effective target range were observed for CRISPRa systems with alternative scRNA designs. The scRNA designs tested are: 1xMS2 scRNA.b2 and 2XMS2 scRNA.b2 with the tracrRNA hairpin removed[3], the original 1xMS2 and 2XMS2 scRNA design having the tracrRNA hairpin[6], and sgRNA 2.0 where the MS2 hairpins are extended from the RNA stems on the sgRNA[7]. CRISPRa systems expressing a 2XMS2 scRNA.b2, 1xMS2 scRNA, and 2XMS2 scRNA displayed the same 10 bp periodicity but lower peak activity than 1xMS2 scRNA.b2. Expressing a sgRNA 2.0 resulted in no CRISPRa activity. Values in panels A-C represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A, B and C are provided as a Source Data file.

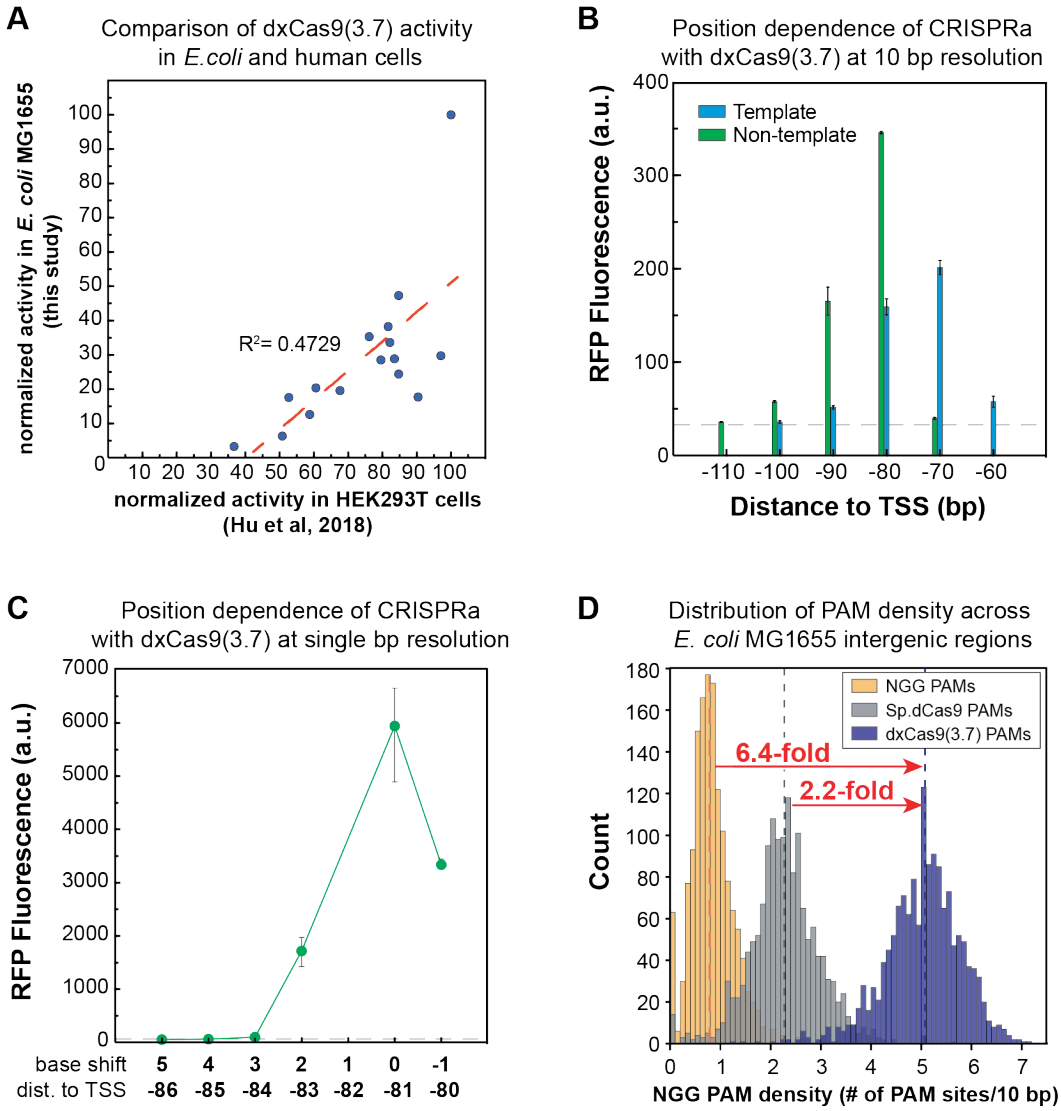
Figure S9



Supplementary Figure 9: Performing CRISPRa with alternative activation domains does not expand the range of targetable positions.

CRISPRa with MCP-TetD and αNTD-MCP activation domains[3] displayed similar positioning dependence as MCP-SoxS(R93A/S101A), where gene expression decreases as the position shifts from 0 to 3 bp, reaches the baseline at a 5 bp shift, and increases again at a 10 bp shift. CRISPRa with MCP-RpoZ[3] and dCas9-RpoZ[8] activation domains displayed more stringent positioning dependence, where gene expression approaches the baseline at 1 bp shift. All alternative activation domains gave weaker peak gene expression at 0 bp shift compared to MCP-SoxS(R93A/S101A). MCP-λcII and MCP-AsiA[3] activation domains did not show any significant CRISPRa activity. All activation domains were cloned into the CRISPRa component plasmid containing *Sp*-dCas9 and 1xMS2 scRNA.b2 targeting J106. To reduce the toxicity of AsiA, the MCP-AsiA plasmid also contains a co-expressed *rpoD(F563Y)* gene[3]. The MCP-SoxS(R93A/S101A), MCP-TetD, αNTD-MCP, MCP-λcII and MCP-AsiA plasmids were tested in *E. coli* MG1655. The MCP-*rpoZ* and dCas9-*rpoZ* plasmids were tested in the *E. coli* strain CD03 (MG1655/Δ*rpoZ*) (Supplementary Table 1)[3]. Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

Figure S10

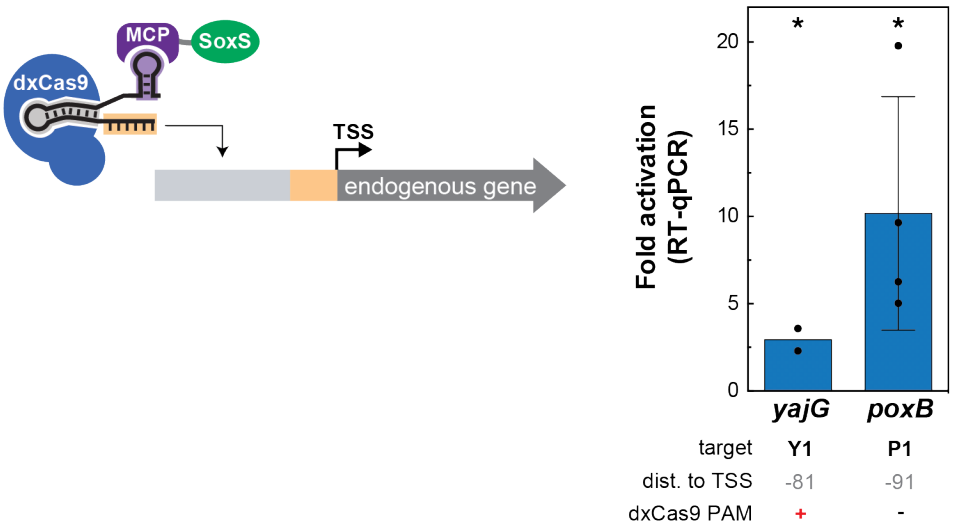


Supplementary Figure 10: dxCas9(3.7) can target an expanded range of PAM sites and is sensitive to target site position for CRISPRa.

A) Relative CRISPRa activities of dxCas9(3.7) on different PAM targets in *E. coli* correlates well with the corresponding data obtained by Hu et al.[9] in human cells. The CRISPRa activity on the NGG PAM site in both studies were normalized to 100 and the CRISPRa activity on all the other PAM sites were normalized to the value of NGG PAM site. Normalized data in this study (y axis) were plotted against the normalized data obtained by Hu et al. (x axis). **B)** CRISPRa with dxCas9(3.7) displayed similar sensitivity to target site position as *Sp*-dCas9. Peaks of activation were observed at -81 and -91 on the non-template strand and -70 and -80 on the template strand. CRISPRa was targeted to a J1-J23117-mRFP1 reporter integrated into the genome (*E. coli* CD13, Supplementary Table 1). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. The grey dotted line represents the baseline fluorescence of a strain containing the dxCas9(3.7) and MCP-SoxS(R93A/S101A) and an empty vector

with no scRNAs. **C)** CRISPRa with dxCas9(3.7) displayed similar positioning dependence with single base shifts. Gene expression was significantly reduced when the scRNA target site was shifted 1-2 bp in either directions from the optimal position -81 bp to the TSS on the non-template strand. Shifting the scRNA target site 3-5 bp causes gene expression to fall to the baseline level. Reporter gene sets were constructed with 1 bp deleted or 2-5 bp inserted upstream of the -35 of the J3-J23117-mRFP1 reporter. The grey line represents the baseline activity of a strain containing the J3-J23117-mRFP1 reporter plasmid and a CRISPRa component plasmid with an off-target scRNA (hAAVS1). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. **D)** dxCas9(3.7) increases the probability of finding targetable PAM sites for CRISPRa. Histograms showing the distribution of the PAM density (defined in Supplementary Figure 7B) between transcriptional units in *E. coli* MG1655 suggest that the average likelihood of finding a dxCas9(3.7)-compatible PAM (blue) is ~6.4-times higher than finding an NGG PAM (yellow), and ~2.2-times higher than finding a *Sp*-dCas9-compatible PAM (grey). Vertical dotted lines indicate the median PAM density for each group. The median PAM density is 5.08 for dxCas9(3.7)-compatible PAMs (NGG, AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA), 2.33 for *Sp*-dCas9-compatible PAMs (NGG, AGA, CGA, GGA, GGC, GGT, TGA). and 0.78 for NGG PAMs. Methods for extracting the sequence of *E. coli* MG1655 intergenic regions, counting targetable PAM sites and calculating the PAM density are described in the Supplementary Methods. Source data for panels A, B, C and D are provided as a Source Data file.

Figure S11



Supplementary Figure 11. Predictive rules for CRISPRa enable activation of endogenous genes *yajG* and *poxB*.

The scRNA sites displaying the highest activity on the *yajG* and *poxB* reporters from the *E. coli* promoter collection[10] (Figure 6B) were tested for activity with dxCas9(3.7) at the endogenous *yajG* and *poxB* genes using RT-qPCR. Fold activation represents expression levels relative to a control expressing an off-target scRNA (J306). For Y1, the value represents the average calculated from n = 2 biologically independent samples. For P1, the value represents the average +/- standard deviation calculated from n = 3 biologically independent samples. The (*) symbol indicates a statistically significant difference from the off-target control (*p-value* < 0.05 using a two-tailed unpaired Welch's t-test). Exact *p-values*: Y1: 0.02, P1: 0.04. Source data are provided as a Source Data file.

Supplementary tables

Supplementary Table 1. *E. coli* Strains

Strain	Description	Genotype	Reference
MG1655	parent *E. coli* strain	F- λ- ilvG- rfb-50 rph-1	
CD03	MG1655 with *rpoZ* knocked out	MG1655 Δ*rpoZ*	[3]
CD06	MG1655/sfGFP (weak promoter)	MG1655 *W1-BBa_J23117-sfGFP KanR::nfsA*	[3]
CD13	MG1655/mRFP1 (weak promoter)	MG1655 *J1-BBa_J23117-mRFP1::nfsA*	This study, Suppl. Fig. 8

Supplementary Table 2. Description of the CRISPRa systems used in each figure

Figure	Cas protein	scRNA design	scRNA target	Activation domain
1B	dCas9	1xMS2 scRNA.b1	W108	MCP-(5aa)-SoxS(wild-type or mutant)
2A	dCas9	1xMS2 scRNA.b1	A1-A4, hAAVS1	MCP-(5aa)-SoxS(R93A/S101A)
2B	dCas9	1xMS2 scRNA.b2	C1-C5, hAAVS1	MCP-(5aa)-SoxS(R93A)
3A	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)
3B	dCas9	1xMS2 scRNA.b2	J101-J120, hAAVS1	MCP-(5aa)-SoxS(R93A/S101A)
3C	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)
3D	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A/S101A)
4A	dCas9	1xMS2 scRNA.b2	J102, J104, J106, J108, J110, J206	MCP-(5aa)-SoxS(R93A)
4B	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A), MCP-(10aa)-SoxS(R93A), MCP-(20aa)-SoxS(R93A)
5A	dCas9 / dxCas9(3.7)	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A)
5B	dCas9 / dxCas9(3.7)	1xMS2 scRNA.b2	M1, M2, J206	MCP-(5aa)-SoxS(R93A/S101A)
6A	dCas9 / dxCas9(3.7)	1xMS2 scRNA.b2	Y1-Y3, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible
6B	dxCas9(3.7)	1xMS2 scRNA.b2	Y1, Y2, P1, P2, U1, U2, D1, D2, B1, B2, E1, E2, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible
Supplementary Figure 1	dCas9	1xMS2 scRNA.b2	L1-L8, hAAVS1	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 3A	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)

Supplementary Figure 4	dCas9	1xMS2 scRNA.b2	J104, J106, J108, J306, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5A	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5B	dCas9	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 5C	dCas9	1xMS2 scRNA.b2	J107, J109	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 5D	dCas9	1xMS2 scRNA.b2	J106, J107, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5F	dCas9	1xMS2 scRNA.b2	J306, J306+1-5	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 8A	dCas9	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A), MCP-(2xEAAAK)-SoxS(R93A/S101A)
Supplementary Figure 8B	dCas9	1xMS2 scRNA.b2, 1xMS2 scRNA.b2 + 2/5/10 bp MS2 stem extension	J306	MCP-(2xEAAAK)-SoxS(R93A/S101A)
Supplementary Figure 8C	dCas9	1xMS2 scRNA.b2, 2xMS2 scRNA.b2, 2xMS2 sgRNA2.0, 1xMS2 scRNA, 2xMS2 scRNA	J306	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 9	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A/S101A), Alternative activators
Supplementary Figure 10B	dxCas9(3.7)	1xMS2 scRNA.b2	J104-J113	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 10C	dxCas9(3.7)	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 11	dxCas9(3.7)	1xMS2 scRNA.b2	Y1, P1, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible

Supplementary Table 3. gRNA Target Sites

sgRNA target	DNA Sequence	Target Strand[a]	Distance to TSS[b]
W108	GAAGATCCGGCCTGCAGCCA	NT	91
J101[c]	TGGGTTCCACCGGATACCTC	T	40
J103[c]	AGGCGTCCTTTGGGTTCCAC	T	50
J105[c]	CGGTTACCAAAGGCGTCCTT	T	60
J107[c]	CGGTGTCCTGCGGTTACCAA	T	70
J109[c]	AGGTATCCTGCGGTGTCCTG	T	80
J111[c]	GGGCGACCTCAGGTATCCTG	T	90
J113[c]	GGGCCACCACGGGCGACCTC	T	100
J115[c]	TGGTGACCATGGGCCACCAC	T	110
J117[c]	GGGTGACCTATGGTGACCAT	T	120
J119[c]	TGGTTGCCAAGGGTGACCTA	T	130
J121[c]	AGGACACCTTTGGTTGCCAA	T	140
J102[c]	AGGTATCCGGTGGAACCCAA	NT	61
J104[c]	TGGAACCCAAAGGACGCCTT	NT	71
J106[c]	AGGACGCCTTTGGTAACCGC	NT	81
J108[c]	TGGTAACCGCAGGACACCGC	NT	91
J110[c]	AGGACACCGCAGGATACCTG	NT	101
J112[c]	AGGATACCTGAGGTCGCCCG	NT	111
J114[c]	AGGTCGCCCGTGGTGGCCCA	NT	121
J116[c]	TGGTGGCCCATGGTCACCAT	NT	131
J118[c]	TGGTCACCATAGGTCACCCT	NT	141
J120[c]	AGGTCACCCTTGGCAACCAA	NT	151
hAAVS1[c]	GGGGCCACTAGGGACAGGAT	off-target	n/a
J206	TAGTAGCCGAACACGTCCTC	off-target	n/a
J306	TTGTGTCCAGAACGCTCCGT	NT	81
J306+1	TGTGTCCAGAACGCTCCGTA	NT	82
J306+2	GTGTCCAGAACGCTCCGTAG	NT	83
J306+3	TGTCCAGAACGCTCCGTAGG	NT	84
J306+4	GTCCAGAACGCTCCGTAGGG	NT	85
J306+5	TCCAGAACGCTCCGTAGGGG	NT	86
M1	AGCAGAAGTGTCAGCAGTGT	NT	81 (reporter A), 86 (reporter B)
M2	CGACGAGCAGAAGTGTCAGC	NT	76 (reporter A), 81 (reporter B)

aroKB_A1	GGGCAATTATTTCGTCATGA	T	151
aroKB_A2	AGATGAACGACGCGAGTTAG	T	122
aroKB_A3	TTTTACGGCTGTTTACTCAC	NT	92
aroKB_A4	TGAGTAAACAGCCGTAAAAG	T	71
cysK_C1	CCACCCCTGTTTCACACAAA	NT	93
cysK_C2	AAACCGTTTGTGTGAAACAG	T	79
cysK_C3	GACATGCAAGATGGAATAAG	NT	61
cysK_C4	ATGACATGCAAGATGGAATA	NT	59
cysK_C5	GGAAATAATGACATGCAAGA	NT	52
ldhA_L1	CAGTAATAACAGCGCGAGAA	T	157
ldhA_L2	GGATATTAACTACCCATGCT	NT	140
ldhA_L3	GCTTTATATTTACCCAGCAT	T	135
ldhA_L4	GCTTAATTTTTCGCTAAATC	NT	119
ldhA_L5	GAAAAATTAAGCATTCAATA	T	91
ldhA_L6	AGCATTCAATACGGGTATTG	T	82
ldhA_L7	AGGCGCAACCTTCAACTGAA	NT	67
ldhA_L8	ATGTTTAACCGTTCAGTTGA	T	58
yajG_Y1	TTGACGAAATAATCGCCCCT	NT	81
yajG_Y2	CATCAGTGTTTCTTTTACCA	T	80
yajG_Y3	AAATAATCGCCCCTGGTAAA	NT	87
poxB_P1	CCCGATGAAAGGAATATCAT	NT	91
poxB_P2	GGTTAAATAGCCCGATGAAA	NT	81
uxuR_U1	TGATTGACCAGTAAGTCTGT	NT	81
uxuR_U2	GATTACCCTACAGACTTACT	T	70
ppiD_D1	ACTAAGCGTTGTCCCCAGTG	T	80
ppiD_D2	GTCCCCAGTGGGGATGTGAC	T	70
ansB_B1	AGATCTACAAAGTTAGAGGC	NT	91
ansB_B2	TATATTTTGGAGATCTACAA	NT	81
araE_E1	TGCGACATGTCGTTATGTGA	NT	91
araE_E2	ATTAAATTGCTGCGACATGT	NT	81

[a] Template strand (T) or non-template strand (NT).

[b] Distance to TSS is the distance from the 3' end (PAM proximal) of the guide target site to the transcription start site. For synthetic promoters driven by BBa_J23117 or BBa_J23119 (http://parts.igem.org), the TSS is immediately downstream of the BBa sequence (see complete maps below).

[c] The J101-J121 sites were the same target sites used to test the positioning dependence of CRISPRa at a 10 bp resolution on the J1-J23117 promoter[3].

Supplementary Table 4. Select *E. coli* Expression Plasmids[a]

Plasmid	Marker	origin	Promoter	Gene	Terminator
pCD442	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A)	*1) BBa_B0015* *2) BBa_B1002*
pCK005.1-21[b]	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J101-121 targets)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pCD564	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107*	1) dxCas9(3.7) 2) MCP-(5aa)-SoxS (R93A/S101A)	*1) BBa_B0015* *2) BBa_B1002*
pCD565	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dxCas9(3.7) 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J306 target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pCD580.-1~-5	*AmpR*	*pSC101***	*J1_BBa_J23117 (with 1-5 bp deleted from the J1 region)*	mRFP1	*BBa_B0015*
pCD581	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J306 target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pJF215.x	*CmR*	*p15A*	*1) Sp.pCas9* *2) TetR-pTet* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (variable target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pJF076Sa[c]	*AmpR*	*pSC101***	*J3_BBa_J23117*	mRFP1	*BBa_B0015*
pJF143-J3[d]	*AmpR*	*pSC101***	*J3_BBa_J23117*	mRFP1	*BBa_B0015*
pJF155.1-12[c]	*AmpR*	*pSC101***	*J1_BBa_J23117 (with 1-12 bp*	mRFP1	*BBa_B0015*

			inserted upstream of -35)		
pJF161.1-12[d]	*AmpR*	*pSC101***	*J3_BBa_J23117 (with 1-12 bp inserted upstream of -35)*	mRFP1	*BBa_B0015*

[a] BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org). dCas9 is the catalytically inactive form of *S. pyogenes* Cas9. Sp.pCas9 is the endogenous Cas9 promoter from *S. pyogenes*.

[b] pCK005.1-21 indicates a set of plasmids (pCK005.1, pCK005.2…) where the final number corresponds to guide RNA target sites (J101, J102…, Supplementary Table 3) used for the J1-117-mRFP1 reporter.

[c] Originally described in previous work[3]. Modified versions of this plasmid are available where BBa_J23117 is replaced with minimal promoters regulated by alternative sigma factors (Figure 3B).

[d] Modified version of this plasmid are available where: BBa_J23117 is replaced with Anderson promoters of different strength (Figure 3A) and different PAM sites at the -81 J306 site (Figure 5A).

Supplementary Table 5. Primer Sequences for RT-qPCR

Primer	Sequence	Reference
16S_f	AAAGTTAATACCTTTGCTCATTGACGTT	[3]
16S_r	GACTACCAGGGTATCTAATCCTGTTT	[3]
aroK_f	TCTGGTTGGGCCTATGGGTG	This study
aroK_r	TACGAATGGTCACGTCGGCA	This study
cysK_f	TGCTGAAACCAGGCGTTGAA	This study
cysK_r	TCCCAACGCCAGCAATAAATACA	This study
ldhA_f	TGGCTGCGAAGCGGTATGTA	This study
ldhA_r	GAACGCCAGCAGACGCATAC	This study
yajG_fw	AAGTCACCCGCGATAAT	This study
yajG_rev	CTTTGGTCGCGATGTTG	This study
poxB_fw	GGTCTTAGTGACAGTCTTAATC	This study
poxB_r	GGAATATGAGCGGCAATC	This study

Supplementary Table 6. Intervening sequences between scRNA target site and -35 region with respective CRISPRa activity and number of transcription factor (TF) binding sites.

Index	sfGFP/OD_{600}[a]	Sequence[b]	TF binding sites, consensus[c]	TF binding sites, $d(u,v) \leq 1$[c]
1	549.87	TATCATAGTGATGACCAGTAAACATT	0	4
2	602.70	TGCAGAAAAGGGCTCTAGTGACTTGT	0	7
3	682.05	ATCTATGCGACGTCAAACGTGATGGG	0	12
4	702.47	GCCTTTCAGTGATCCCTTGTGATCTT	0	11
5	704.82	CAATTCTAGCAAGAAATTGACTTCGT	0	6
6	710.25	TTTACGGATTGGTGCGCTATTGGGTT	0	14
7	716.04	TAATGTTGATCGCCTCTATGCGTCCT	0	5
8	746.08	TTATTATAATAGTTTTGAACGTGCCT	0	6
9	794.37	CAATATGACGTGTTGTTAATTTGGTT	1	14
10	883.74	GGTCTGTGTACTCGAACACGAGATTT	0	7
11	1230.90	TTGCTCCGCGGTTTCTCTGTAAATGT	0	8
12	1532.80	ATAAATTGCATGATTAAAGACTATTG	0	11
13	1586.70	TTCGGATTTAAGGAGATATGTTTACG	0	7
14	1679.31	TAATAGGCTAGGAATTTGAAGGGATT	0	11
15	5604.77	GCATCTACATAGTGGTACAATTGAAG	0	13
16	7110.46	GCGTCCATATATGCTGATGGTGTGGG	0	10
17	7143.40	GCGGGCTATGCCTTGGAGGGTTGTAG	0	12
18	11027.55	ATTACACACGAAGGTGTATTTACTAT	2	28
19	11589.46	TTATCTAATAGGGGCGCCCCCGCTTC	0	9
20	11714.54	TTAAACAAACGAAAAGTTCGCGCCTG	1	14
21	12073.23	GATCGGGTAGGGCGCCGACAAAGGGA	0	8
22	12374.02	TGAGGGATATAGAGAGATGTTCAGCC	0	11
23	12399.33	CATCGCCGTTGCGATGCATTTTGACG	0	15
24	13207.42	GTTGTCCTTCTAGTCGCCCATGACTC	0	8
J3	**13269.13**	**CGTCGTCTTGAAGTTGCGATTATAGA**	0	10
25	13413.88	GTCGTAAATAAGTAAGTCACTCCCAC	2	19
26	13434.32	TTGAGGCCCATGCTTGTGGAAATGAC	1	20
27	14053.61	GGCAAGATGCCTCGTGCAGTAGAATA	0	8
28	15973.97	AGTCGCTGAGTAGATGTTCGTAGAGA	0	4

29	16760.50	ACACCGACTACCCCTGCTGGGCCCAG	0	12

[a] sfGFP/OD_{600} values represent the CRISPRa activity of strains where CRISPRa is targeted to individual elements of a reporter library where the 26 bases between the scRNA target site and the -35 region on the J3-J23117-sfGFP reporter were replaced with random bases. The sfGFP/OD_{600} of a strain expressing the J3-J23117-sfGFP promoter are included for comparison and are in bold.

[b] The sequence corresponds to the 26 bases between the scRNA target site and the -35 region of the promoter depicted in Figure 3C.

[c] Number of exact matches to consensus transcription factor binding sites[2] or sequences within a Hamming distance $d(u,v) \leq 1$ from the consensus transcription factor binding sites.

Supplementary methods

Computational analysis of PAM site availability

The intergenic sequences from *E. coli* were obtained from the RegulonDB database[2]. To obtain the intergenic sequences upstream of promoters, the intergenic sequences between convergent genes and between intra-operon coding sequences were removed. 5' untranslated regions (UTRs) were removed from sequences using transcription start site (TSS) information from Kim et al.[11] and RegulonDB. Intergenic sequences upstream of genes with no known TSS or where no intergenic sequence remained after removing the UTRs were discarded, yielding 1504 transcriptional units for further analysis. The number of PAM sites was calculated by counting the number of PAM sites on both strands (Supplementary Figure 10D). The PAM density over a 10 bp window was calculated by dividing the number of PAMs found at each intergenic sequence by the length of the intergenic sequence and multiplying by 10. Analyses were performed using Python 2.7.

Computational selection of candidate endogenous genes for CRISPRa

PAM sites found in *E. coli* intergenic sequences upstream of promoters were assigned a "sequence score" and a "distance score". The sequence score indicates the relative activity of a given PAM sequence compared to an AGG PAM, and was calculated using data from Figure 5A with dxCas9(3.7) (e.g. AGG = 1, CGT = 0.24). All NGG sequences were assigned a "sequence score" of 1. The distance score indicated the relative activity of a given PAM site position compared to a PAM site found at -81 on the non-template strand or -70 on the template from the TSS. Distance scores for the non-template strand were calculated using data from Figure 4A (e.g. -81 = 1, -91 = 0.62). For the template strand, the same scores were used, but each score was shifted upstream by 11 bases to match the position of the site of maximum activation (e.g. -70 = 1, -80 = 0.62). Each PAM site was assigned a final calculated score as "sequence score" x "position score". Promoters were then ranked by sorting for higher PAM scores at the peaks of activation (sum of the scores for the PAMs at -70, -80, -81, -91). This analysis was performed using Python 2.7. Candidates were then manually selected from among the top scoring promoters by the following criteria: (1) two or more candidate PAM sites, (2) regulation by σ^{70}, and (3) relatively weak basal expression level (<10% of the level of the maximally expressed *E. coli* gene according to data from the *E. coli* promoter collection[10]. Using these criteria, we chose the following six candidates for further characterization: *yajG*, *uxuR*, *ansB*, *poxB*, *araE*, and *ppiD*. Each of these genes is represented in the *E. coli* promoter collection (Dharmacon), a commercially available library of promoter-GFPmut2 fusions[10].

Sequence of the double mutant activation domain: MCP-(5aa)-SoxS(R93A/S101A)

> MCP-(5aa)-SoxS(R93A/S101A) (optimized for minimal endogenous activity, Figure 1 and subsequent)
$MCP_{\Delta FG,\ V29I}$., 5 aa linker, SoxS(R93A/S101A) (underlined are the alanine point mutations)

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPK

GAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMSHQKIIQDLIAWIDEHI
DQPLNIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFS
RVFARQFDRTPADYRHRL

Reporter genes

The integrated sfGFP reporter, the *zwfp-lacZ* and *fumCp-lacZ* reporters used for testing the CRISPRa and endogenous activities of mutant SoxS (Figure 1) and the J1-J23117-mRFP1 reporter (Figure 3 & Supplementary Figure 5 and subsequent) for testing the distance dependence property were described in previous study[3].

J3-J23117_mRFP1 reporter (Figure 3 & Supplementary Figure 4 and subsequent)

The J3 upstream sequence contains a PAM site allowing guide RNAs to target at -81 bp to the TSS on the non-template strand, which is the same distance where maximum CRISPRa activity was observed on the J1 upstream region described previously[3].

J3 upstream region, BBa_J23117 promoter, Bujard RBS, mRFP1, BBa_B0015 terminator. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGATTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGC
GAATTCATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATGCGTTT
CAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGT
CCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGG
GACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGG
ACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGG
TGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTG
GTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGA
ACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGT
GGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTT
ACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGAA
CGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAGGATCCAAACTCGAGTAAGGATCTCCAGGCATCA
AATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTC
TCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA

J3(N26)-J23117-sfGFP reporter library (Figure 3)

J3 upstream region (without 26 bases at 3'), N26, BBa_J23117 promoter, Bujard RBS, sfGFP, BBa_B0015 terminator. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAANNNNNNNNNNNNNNNNNNNNNNNNNNNTTGACAGCTAGCTCAGTCCTAGGGATTGTGCTA
GCGAATTCATTAAAGAGGAGAAAGGTACCATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCC
AATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTG
ATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAA
CACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATG
ACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGG
ACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGG
TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTA
TACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGA
TGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACC
AGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGG
TCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAGGATCCA
AACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTT
TTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTC
TGCGTTTATA

Modified synthetic promoters

Promoters with different basal expression levels (Figure 3)

Annotations:

J3 upstream region, variable promoter (BBa_J231xx), Bujard RBS, Start codon of mRFP. The J106 target site is underlined

BBa_J23105

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtactatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23106

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtatagtgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23107

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagccctaggtattatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23108

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGActgacagctagctcagtcctaggtataatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23109

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacagctagctcagtcctagggactgtgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23110

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtacaatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23111

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtatagtgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23112

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGActgatagctagctcagtcctagggattatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23113

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGActgatggctagctcagtcctagggattatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23114
AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttatggctagctcagtcctaggtacaatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23115
AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAtttatagctagctcagcccttggtacaatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23118
AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtattgtgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

BBa_J23119
AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctaggtataatgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

Promoters regulated by alternative sigma factors (Figure 3)
Annotations:
J1 upstream region, variable promoter (color coded), Bujard RBS, Start codon of mRFP. The J106 target site is underlined.

J1-*sodCp* promoter
GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAG
GGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTAC

CAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGCCGTTcaaaaatgtgtcacTGGTTTACACTtattcagG
GAATTCATTAAAGAGGAGAAAGGTACCATG

In the J1-*sodCp* promoter, we deleted 3 bp between the J1 sequence and the *sodCp* minimal promoter to maintain the same -80 bp spacing between the target site and the J109 target site as in the rpoD promoter. Without this deletion, activation from the *sodCp* promoter is substantially weaker (not shown), likely due to the sensitive positioning requirements described in Figure 4.

J1-*glnAp2* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAG
GGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTAC
CAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACTGGCACagatttCGCTTtatctttttTacggcgacGAA
TTCATTAAAGAGGAGAAAGGTACCATG

J1-*rdgBp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAG
GGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTAC
CAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACTTGTAAaaggcgaacttggcCGCCACaaacaaatt
GAATTCATTAAAGAGGAGAAAGGTACCATG

J1-*yieEp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAG
GGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTAC
CAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACGAACTTTtagccgctttagtctgtcCATCAttccAGAA
TTCATTAAAGAGGAGAAAGGTACCATG

J3(tetO)-J23117_mRFP1 reporter (Figure 3)

J3 upstream region (without 19 bases at 3'), tetO, BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCACTCTATCGTTGATAGAGTtgacagctagctcagtcctagggattgtgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

J1-J23117 promoter with shifted bases (Figure 4, Supplementary Figure 5 and subsequent)

For promoters shifted by fewer than 12bp, bases were removed from the 5' end of the inserted sequence (starting with C).

J1 upstream region, inserted sequence (12 bases), BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J106 target site is underlined.

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAG
GGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTAC
CAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACCTAGATGCGCTCttgacagctagctcagtcctaggga
ttgtgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

J3-J23117 promoter with shifted bases (Supplementary Figure 5 and subsequent)
For promoters shifted by fewer than 12bp, bases were removed from the 5' end of the inserted sequence (starting with T). For promoters with negative shifts, no sequence was inserted between the J3 upstream region and BBa_J23117, and bases were removed from the 3' end of the J3 upstream region (starting with A).
J3 upstream region, inserted sequence (12 bases), BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTG
GACACAACGTCGTCTTGAAGTTGCGATTATAGATGAGATGCGCTCttgacagctagctcagtcctagggattgtgctag
cGAATTCATTAAAGAGGAGAAAGGTACCATG

Modified J3-J23117 promoter with 6 adjacent PAMs around -81 bp to TSS (Supplementary Figure 5)
J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP. Region with additional PAM sites inserted. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACCCCCCCCTACGGAGCGTTCT
GGACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATTCATTAA
AGAGGAGAAAGGTACCATG

Modified J3-J23117 promoter with non-NGG PAMs around -81 bp to TSS (Figure 5A)
J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP, region with modified PAMs. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTNNNACGGAGCGTTCT
GGACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATTCATTAA

AGAGGAGAAAGGTACCATG

Promoter to demonstrate dxCas9(3.7) versatility (Figure 5B)

J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP. The modified target site (M) is underlined. AGT PAM site.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACACTGCTGACAC
TTCTGCTCGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATTCATTAAA
GAGGAGAAAGGTACCATG

Supplementary references

1. Keseler, I. M. *et al.* The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, D543–D550 (2017).
2. Gama-Castro, S. *et al.* RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res* **44**, D133–D143 (2016).
3. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat Commun* **9**, 2489 (2018).
4. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: Identifying amino acids important for DNA binding and transcription activation. *Journal of Molecular Biology* **322**, 237–257 (2002).
5. Ryoichi Arai, T. N., Hiroshi Ueda, Atsushi Kitayama, Noriho Kamiya. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering, Design and Selection* **14**, 529–532 (2001).
6. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
7. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
8. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429–7437 (2013).
9. Johnny H. Hu, D. R. L., Shannon M. Miller, Maarten H. Geurts, Weixin Tang, Liwei Chen, Ning Sun, Christina M. Zeina, Xue Gao, Holly A. Rees, Zhi Lin. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
10. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3**, 623–628 (2006).
11. Donghyuk Kim, B. Ø. P., Jay Sung-Joong Hong, Yu Qiu, Harish Nagarajan, Joo-Hyun Seo, Byung-Kwan Cho, Shih-Feng Tsai. Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genetics* **8**, e1002867 (2012).

Chapter 3:
Prospects for engineering dynamic CRISPR-Cas transcriptional circuits to improve bioproduction

Jason Fontana[1], William E. Voje[1, 2], Jesse G. Zalatan[1,3] and James M. Carothers[1,2]
University of Washington, Seattle, WA 98195

[1] Molecular Engineering & Sciences Institute and Center for Synthetic Biology
[2] Department of Chemical Engineering
[3] Department of Chemistry

Published as a mini-review in the *Journal of Industrial Microbiology & Biotechnology (JIMB) on May 8th, 2018.*
DOI: 10.1007/s10295-018-2039-z

Abstract

Dynamic control of gene expression is emerging as an important strategy for controlling flux in metabolic pathways and improving bioproduction of valuable compounds. Integrating dynamic genetic control tools with CRISPR-Cas transcriptional regulation could significantly improve our ability to fine-tune the expression of multiple endogenous and heterologous genes according to the state of the cell. In this mini-review, we combine an analysis of recent literature with examples from our own work to discuss the prospects and challenges of developing dynamically-regulated CRISPR-Cas transcriptional control systems for applications in synthetic biology and metabolic engineering.

Introduction

Microbes have the potential to produce a vast array of valuable chemical compounds starting from a variety of carbon sources [48]. The development and optimization of microorganisms for bioproduction typically requires lengthy and extensive study and manipulation, oftentimes because the underlying gene and regulatory networks are large and complex [48]. In principle, synthetic multi-gene programs could be constructed to perturb these networks and potentially improve the yield of engineered microbes. In seminal examples of metabolic engineering, the production of valuable biosynthetic products such as isoprenoids [45], opioid precursors [14, 22, 24] and artemisinin [50] was improved by tuning the expression of heterologous enzymes and regulating competing pathways. Creating new methods for rapidly engineering complex multi-gene programs in industrially-promising microorganisms could be transformative for both fundamental research and for accelerating strain design.

To regulate complex, multi-gene programs, we need methods to easily target and manipulate individual genes. The development of the CRISPR-Cas system as an engineering platform has resulted in widely-used tools for programmable gene regulation [17] (Figure 1). A nuclease-deficient Cas9 protein (dCas9) can be used to target arbitrary DNA sequences using dCas9-binding single guide RNAs (sgRNAs). Transcriptional repression can be performed by targeting dCas9 at or near the promoter of genes to act as a physical block for the RNA polymerase (CRISPRi). Alternatively, fusing or recruiting a transcriptional activator to dCas9 and targeting the complex upstream of the gene of interest results in transcriptional activation (CRISPRa). This approach can be used to target multiple genes in parallel by expressing multiple sgRNAs corresponding to each gene of interest. Improvements in bioproduction obtained through multi-gene CRISPR-Cas transcriptional control have already been demonstrated in several different industrially-relevant microbial species including *E. coli* [37, 44, 72, 73], *S. cerevisiae* [77], *C. glutamicum* [11], a clostridial consortium [69], and cyanobacteria [29, 75].

Unlike their natural counterparts which can dynamically up- or down-regulate metabolic genes in response to changing conditions [9, 32, 36], engineered strains often encode static patterns of gene expression. As a result, these engineered systems can experience significant burdens associated with pathway gene overexpression and the complete removal of endogenous biochemical functions, leading to sub-optimal bioproduction yields [30]. Engineering dynamic control of gene expression has been

demonstrated to be a promising approach to improve bioproduction. For instance, overexpressing a polyhydroxybutyrate pathway only after late-exponential phase improved production 2-fold [35], possibly because cells were able to reach higher densities before experiencing burdens due to heterologous gene expression. Similarly, a transcriptional regulator responsive to a toxic intermediate in a fatty acid ethyl ester pathway was used to minimize the buildup of the toxic intermediate, improving production yield 3-fold [78]. Other strategies dynamically modulate the expression of endogenous genes like efflux pumps to increase clearance of toxic compounds and product export [18].

Several strategies to implement dynamic control of gene expression have been recently reviewed in this journal [15], including ways to reduce the accumulation of toxic intermediates during production or to initiate production based on the growth status of the cell. These approaches for dynamic control could be integrated with CRISPR-Cas transcriptional regulation tools to engineer sophisticated regulatory control systems with capabilities more closely resembling their natural counterparts. Combining metabolite-responsive or growth-phase responsive sensors with tools for CRISPRi and CRISPRa would enable us to dynamically activate and repress multiple genes simultaneously (Figure 2). Compared to existing approaches for implementing dynamic gene regulation, dynamically-controlled CRISPR-Cas transcriptional regulation tools could enable regulating the expression of multiple genes in a more straightforward manner that does not involve directly engineering the cis-regulatory sequences of target genes [23]. Practically speaking, because CRISPRi and CRISPRa can be directed by plasmid-borne sgRNAs, variations in the timing and expression levels of combinations of multiple genes can be rapidly probed without needing to engineer the genome at multiple sites for every new condition to be tested [52, 67]. Developing dynamic CRISPR-Cas tools that can control the expression of multiple genes would introduce a versatile technology for rapidly regulating metabolic flux in engineered microbes (Figure 3) and identifying dynamic regulatory architectures that improve biosynthesis titers, rates, and yields.

In this manuscript, we will examine the prospects for such an undertaking, using recent results from the literature and ongoing work in our own laboratories as illustrative examples. This mini-review provides (i) a brief introduction to CRISPR-Cas transcriptional control in microbes, (ii) considerations for the development of dynamically-controlled CRISPR-Cas transcriptional networks, and (iii) practical challenges in building control systems through the integration of metabolite-responsive sensing and CRISPR-Cas transcriptional regulation. Finally, we discuss possible avenues for completing the development of dynamic CRISPR-Cas transcriptional programs and extending these technologies to non-model organisms of high industrial interest.

CRISPR-Cas transcriptional control in microbes

Several CRISPR-Cas-based methods have recently been reported to enable complex transcriptional regulation [17]. In a flexible approach developed by Zalatan *et al.* [77], the sgRNA sequence is extended to include RNA sequences that recruit RNA binding proteins (RBPs) fused to transcriptional activators or repressors. These modified sgRNAs, referred to as scaffold RNAs (scRNAs) encode both the regulatory

action (activation or repression) and the genetic locus that will be the target of that action. Sets of scRNAs and sgRNAs constructs can then be used to generate synthetic multi-gene transcriptional programs where some genes are activated and others are repressed simultaneously. This approach was used to redirect flux through a complex, branched metabolic pathway in *S. cerevisiae* by inducing expression of the dCas9 protein, which acts as a single master regulatory control point.

In bacteria, CRISPRi transcriptional repression is very effective in the absence of any additional repressor proteins fused to dCas9 or recruited to the CRISPR-Cas complex [52]. In contrast to yeast, where a variety of transcriptional activation domains are available and have been used for CRISPRa [38], few transcriptional activators have been described in bacteria. Bacterial CRISPRa has been achieved by fusing the ω subunit of RNAP to dCas9 and directing the CRISPR-Cas complex upstream of the target promoter [4]. However, this strategy requires knocking out the genomic ω subunit of the RNAP (*rpoZ*), which may result in undesired changes in existing regulatory networks [74]. In work from our laboratories, we have utilized the scRNA design to recruit a mutant SoxS transcriptional activator to the CRISPR-Cas complex, generating even higher levels of CRISPRa in wild-type *E. coli*. Using this strategy, CRISPRa and CRISPRi can be performed concurrently and independently in bacteria in the same way as in yeast (Dong *et al.,* submitted).

Toward developing dynamic CRISPR-Cas transcriptional regulation

Many dynamic genetic controllers regulate gene expression in response to variations in the concentration of cognate metabolites [24, 48]. By regulating the CRISPR-Cas system with dynamic genetic controllers, it may be possible to engineer complex responses to external inputs or changes in cellular state that improve production phenotypes. Naturally-occurring genetic controllers operating through growth phase-responsive promoters [46], metabolite-responsive transcriptional regulators [32] or riboswitches [19] could be utilized as biosensors for many different metabolites of interest [56] (Figure 4). However, many intermediates and products that are the targets of metabolic engineering have no cognate regulators that can be repurposed as biosensors, or are responsive to concentrations that are not relevant for high-titer bioproduction. In this case, generating ligand-binding RNA aptamers for target molecules through *in vitro* selection and then assembling the aptamers into dynamic RNA-based genetic regulators like aptazymes [43] or riboswitches [19] may be possible.

Engineering dynamically responsive CRISPRi/a tools utilizing existing dynamic genetic controllers could be successful if CRISPRi/a actuation can be tuned by changing the expression of its components. Varying the expression levels of either dCas9 and sgRNA, or of genomically integrated dCas9, are sufficient for changing the level of transcriptional repression generated by CRISPRi [40, 42, 51, 52]. We have shown in our recent work that CRISPRa can be tuned by titrating the expression of dCas9, the sgRNA and the recruitable transcriptional activator (Dong *et al*., submitted). However, to initiate different regulatory effects at different genes independently and in response to different metabolic inputs, the ability to efficiently tune CRISPRi/a transcriptional control through the regulated expression of the sgRNAs is required. There are now several demonstrations where CRISPRi is dynamically regulated by controlling the availability and

function of sgRNAs [20, 65]. We recently evaluated the performance of dynamic genetic controllers in the context of CRISPR-Cas transcriptional regulation circuits where only the expression of sgRNA is regulated [21]. Our analyses reveal that in *E. coli*, regulating the expression of the sgRNA by as little as 2.5-fold can lead to meaningful 16-fold changes in CRISPRi repression. These results are significant because they reveal the required transcriptional regulation characteristics for dynamic genetic controllers to effectively tune CRISPRi. Importantly, many of the dynamic regulators mentioned above are able to generate the required changes in transcription, suggesting they could already be useful for constructing dynamically-regulated CRISPR-Cas circuits.

In eukaryotic cells, alternative methods for sgRNA expression provide additional opportunities for dynamic regulation. Expression of sgRNA in yeast hosts can be mediated by Pol III or Pol II promoters [16, 26]. Though Pol III promoters can be used to express non-coding RNAs that are readily retained in the nucleus, the use of engineered Pol II promoters allows for greater sequence diversity, inducible transcription, and tunable levels of expression [25]. The post-transcriptional removal of the 5' cap and 3' poly-A tail nuclear export signals from Pol II-expressed sgRNAs is thought to be essential for obtaining high levels of sgRNA-directed actuation [26]. Using Ribozyme-sgRNA-Ribozyme (RGR) architectures that remove the nuclear export signals from sgRNA transcripts through 5' and 3' *cis* ribozyme cleavage, CRISPRi circuits comprised of as many as seven sgRNAs have been successfully constructed in yeast [25]. These results support the notion that complex systems operating through regulated sgRNA expression can operate in eukaryotic microbes as well as bacteria. We have recently replaced the 3' ribozyme in the RGR architecture with an aptazyme and shown that ligand-dependent cleavage can lead to up to 10-fold increases in sgRNA-directed CRISPRi activity, providing yet another route for engineering metabolite-responsive CRISPR-Cas regulation (Voje *et al.,* unpublished).

Challenges in implementing dynamic CRISPR-Cas transcriptional programs

In this section, we describe two challenges specific to the task of coupling sensing functions to CRISPR-Cas transcriptional control actuators in order to create dynamic circuits in microbes engineered for bioproduction.

Matching the time-scale of the CRISPR-Cas response to the control problem

In our recent work described above, we identified the dynamic range in the expression of the CRISPRi components required for tuning gene expression as an important constraint for dynamic control, yet other hurdles may exist in creating useful systems. The timescales required to dynamically turn on and off the transcriptional program are likely critical for improving bioproduction. In this context, the time that dynamic CRISPRi/a requires for modulating actuation in response to changes in metabolite concentrations will impact the performance of the engineered system. For instance, if the accumulation of a toxic intermediate takes place on timescales much shorter compared to the response and initiation time of dynamic CRISPRi/a, cells could experience toxicity before a programmed response can be delivered through

feedback regulation. In experiments regulating the expression of sgRNA and dCas9 using an anhydrotetracycline responsive promoter in rapidly dividing cells, maximal repression was reached 4 h after induction [52], which may be too long to be relevant for solving many control problems in metabolic engineering [61].

Just as important, CRISPRi/a transcriptional control needs to be reversible on relatively short timescales to permit fully-dynamic feedback regulation. The CRISPRi/a program would need to be reversed when the concentration of the intermediate varies due to changing external conditions, or if the CRISPRi/a program under- or overshoots the set-point for optimized flux through the pathway. *In vitro* experiments suggest that dCas9:sgRNA complexes bound to their DNA targets take as long as ~6 h to unbind [55]. Experiments investigating the binding of dCas9 to sgRNAs also reveal that dCas9:sgRNA complexes may be long-lived [54]. Consistent with these reports, we observed persistence of the CRISPRi repression phenotype for up to 12 h after induction of the expression of a competing sgRNA [21]. The factors responsible for the persistence of CRISPRi repression are likely the lifetimes of the dCas9:sgRNA bound to DNA and of the dCas9:sgRNA complexes already in the cell. Increasing the turnover of dCas9 proteins using a degradation tag [7] could possibly permit to overcome the persistence of CRISPRi.

While there are many examples of metabolic pathways where flux is controlled at the level of enzyme transcription [41, 76], some pathways are controlled post-transcriptionally [12]. Because CRISPRi and CRISPRa control gene expression by acting on transcription, dynamically regulating flux in these pathways with the current tools could be challenging. Strategies to directly regulate translation, perhaps by using new CRISPR-Cas systems that target RNA [2, 62], could provide additional means for dynamic control.

Repercussions of bacterial genome organization for CRISPRa

Executing complex programs of gene expression requires the ability to simultaneously repress and activate gene expression. scRNA-encoded programs of CRISPRi gene repression and CRISPRa gene activation have been achieved in eukaryotic systems using RNA-mediated recruitment of activators or repressors to the dCas9 complex [77]. While the rules for CRISPRa in eukaryotes are fairly permissive, and targeting a relatively broad region upstream of the transcription start site (TSS) is sufficient for activation [27], our recent work and other efforts activating heterologous genes show that bacterial CRISPRa is characterized by much more strict targeting requirements [4]. We observed high levels of transcriptional activation only when targeting CRISPRa between 60 and 90 bases upstream of the TSS (Dong *et al.*, submitted). Even with these constraints, CRISPRa can be performed fairly trivially on heterologous genes by inserting appropriately positioned sgRNA target sites (Protospacer Adjacent Motifs, or PAMs) upstream of their promoters. However, when targeting endogenous genes in *E. coli*, the availability of PAM sites is often limited. Based on our preliminary bioinformatics analysis, the average intergenic region upstream of a gene is ~140 bases and contains nine PAM sites. Additionally, native transcriptional regulators that bind upstream of target genes may interfere with dCas9 binding, and whether dCas9 is able to displace endogenous

bacterial regulators is not known. Clearly, much more work will be needed to determine generalizable rules for performing CRISPRa in bacteria, and whether bacterial CRISPRa can be as functionally versatile as eukaryotic CRISPRa.

Conclusions and future directions

Despite the numerous demonstrations of the utility of CRISPR-Cas transcriptional regulation, engineering complex dynamic control systems using these tools is still challenging. The creation of dynamic control systems requires inputs and outputs of the interacting biological parts to be well matched [8]. There have been impressive, though time-consuming, efforts which use characterized libraries of parts to predict and then validate functions of assembled systems [47, 68]. In our recent work (Voje *et al*., unpublished), we found that experimental outcomes generated from a large, 7-sgRNA CRISPRi logical evaluation circuit can be predicted using parameter ranges derived from characterization data collected for the individual components. Elsewhere, Cello has been developed as a computational tool that uses characterization data of individual genetic regulators to automate the design of genetic circuits that display user-defined logic behaviors [47]. However, when put into the context of larger systems, parts characterized in isolation have been shown to indirectly interact with one another resulting in unintuitive failures, an effect known as “retroactivity” [53]. For instance, we observed that decreasing the levels of dCas9 expression can change the response of multi-sgRNA CRISPRi cascades, an effect likely caused by sgRNA competition that limited our ability to predict the behavior of CRISPRi cascades (Voje *et al*., unpublished). Integrating further understanding of the mechanisms responsible for retroactivity in CRISPR-Cas transcriptional regulation networks with computational design-driven tools such as Cello could reduce the amount of trial and error experimentation needed to generate the multi-sgRNA circuits useful for engineering dynamic metabolic control circuits.

Ultimately, the hope is that dynamic control strategies can be applied in industrial bioprocessing settings to improve titers, rates and yields of bioproduction strains. However, industrial-scale bioproduction is often a very complex process that involves multiple changing parameters, such as oxygen and nutrients levels, pH, and temperatures, that differ significantly from laboratory conditions. Engineering production strains with new expression programs could significantly improve metabolic production in process conditions. In fact, perturbations to gene regulatory networks are known to be key drivers of species and strain adaptation [66, 71]. A body of results obtained in *S. cerevisiae* through gene knockouts, heterologous overexpression, screening and directed evolution suggest that new expression programs could have positive impacts on bioproduction [13, 28, 33, 63, 64]. CRISPR-Cas transcriptional regulation tools that enable targeting of multiple arbitrary genes could be immediately useful for systematically exploring the effects of multiple, simultaneous variations in the timing and expression levels of genes on product yields, tolerance to process conditions and biomass accumulation. Given that our current understanding of network dynamics is not sufficient for purely-rational engineering approaches [48], screening strategies may provide a unique practical route for rapidly improving bioproduction phenotypes. Several approaches to generate

barcoded, multi-sgRNA libraries have been described [58, 59, 70]. Recovery and sequencing of the barcoded cassettes could be used to identify multi-sgRNA programs conferring improved bioproduction phenotypes. Alternatively, new single-cell sequencing technologies could be employed to associate the individual sgRNAs comprising successful multi-sgRNA programs essentially without any specific barcoding scheme [57]. Looking ahead, it will be extremely interesting to perform screens of dynamically-regulated programs alongside constitutive programs comprised of the same sets of sgRNAs to determine the extent to which dynamic control accesses different parts of the underlying design space and generates improved production phenotypes [15, 61].

A possible challenge for implementing dynamic control programs and performing screens using CRISPR-Cas tools is the issue of genetic stability, particularly when stresses are generated by bioproduction. In fact, it will rarely be in the interest of cells to express high levels of heterologous genes. Even when dynamic feedback control can help mitigate these stresses, there is some expectation that the metabolic burden associated with expressing the control system may lead to selective pressures that break its functionality [60]. For instance, targeting multiple genes simultaneously requires the delivery of multiple sgRNAs. Expression vectors for transiently delivering multi-gRNA cassettes are often constructed by concatenating the RGR architecture [25] or by using the Csy4 nuclease to cleave multiple gRNA units from single transcripts [49]. However, both of these strategies fail to eliminate repeating sequences that could easily recombine. Understanding how to design low-overhead, genetically-stable circuit architectures will be a crucial aspect of determining whether dynamic CRISPR-Cas control systems are practically useful for industrial biotechnology. Addressing these challenges would permit production screens to identify expression programs that improve tolerance to process conditions and increase production. For instance, chromatography (e.g. LC-MS or HPLC), cell-based biosensors [48], or cell-free biosensors [1, 5, 6] could be employed to screen strains and quantify secreted metabolites. Genetically-encoded intracellular biosensors [10, 56] that generate fluorescent outputs in response to increases in intracellular metabolite concentrations could be combined with cell sorting to isolate cells with high intracellular metabolite levels (Figure 5).

Much of the work described in this review utilizes genetically-tractable, well-established model organisms such as *E. coli* or *S. cerevisiae*, which are common hosts for metabolic engineering [48]. There are many other microbial species with unique metabolic capabilities that may allow for more cost-effective industrial bioproduction. Among these are non-model bacteria that can utilize alternative carbon sources like CO_2, CO, or methane, and alternative energy sources such as light or H_2 and non-model thermotolerant yeasts that can efficiently utilize the full spectrum of sugars from lignocellulosic feedstocks [31, 39]. Moving these unique capabilities into standard model strains has proven difficult and many genetic parts are not portable between organisms [3, 34]. In contrast, CRISPR-Cas genetic control systems, which rely on fundamentally modular aspects of protein and RNA structure, may provide a new suite of tools that can be easily implemented in non-model organisms to take advantage of their unique capabilities. In addition, a variety of molecular biology tools, including CRISPR-based genome editing have been demonstrated in

non-model, but industrially-relevant organisms, suggesting that CRISPR-Cas transcriptional control programs could be developed for those organisms [11, 29, 37, 44, 69, 72, 73, 75]. Despite the challenges, given the rate of progress, dynamic multi-gene CRISPR-Cas transcriptional control programs operating on endogenous and heterologous genes to improve bioproduction in a wide range of microbial hosts may soon be within reach.

Funding

Related work in the authors' laboratories was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (J.G.Z.), a NSF Award MCB 1517052 and a University of Washington Presidential Innovation Award (J.M.C.).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Abatemarco J, Sarhan MF, Wagner JM, Lin J-L, Liu L, Hassouneh W, Yuan S-F, Alper HS, Abate AR (2017) RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. Nat Commun 8:332. doi: 10.1038/s41467-017-00425-7

2. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, Verdine V, Cox DBT, Kellner MJ, Regev A, Lander ES, Voytas DF, Ting AY, Zhang F (2017) RNA targeting with CRISPR–Cas13. Nature 550:280–284. doi: 10.1038/nature24049

3. Antonovsky N, Gleizer S, Noor E, Zohar Y, Herz E, Barenholz U, Zelcbuch L, Amram S, Wides A, Tepper N, Davidi D, Bar-On Y, Bareia T, Wernick DG, Shani I, Malitsky S, Jona G, Bar-Even A, Milo R (2016) Sugar Synthesis from CO2 in Escherichia coli. Cell 166:115–125. doi: 10.1016/j.cell.2016.05.064

4. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:gkt520–7437. doi: 10.1093/nar/gkt520

5. Burke CR, Carothers JM RNA-Based Molecular Sensors for Biosynthetic Pathway Design, Evolution, and Optimization. Biotechnology for Biofuel Production and Optimization 117–138.

6. Burke CR, Sparkman-Yager D, Carothers JM (2017) Multi-state design of kinetically-controlled RNA aptamer ribosensors. bioRxiv 213538. doi: 10.1101/213538

7. Cameron DE, Collins JJ (2014) Tunable protein degradation in bacteria. Nature Biotechnology

32:1276–1281. doi: 10.1038/nbt.3053

8. Cardinale S, Arkin AP (2012) Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. Biotechnol J 7:856–866. doi: 10.1002/biot.201200085

9. Carothers JM (2013) Design-driven, multi-use research agendas to enable applied synthetic biology for global health. Syst Synth Biol 7:79–86. doi: 10.1007/s11693-013-9118-2

10. Carothers JM, Goler JA, Juminaga D, Keasling JD (2011) Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression. Science 334:1716–1719. doi: 10.1126/science.1212209

11. Cleto S, Jensen JV, Wendisch VF, Lu TK (2016) Corynebacterium glutamicumMetabolic Engineering with CRISPR Interference (CRISPRi). ACS Synth Biol 5:375–385. doi: 10.1021/acssynbio.5b00216

12. Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MAH, de Groot MJL, Slijper M, Heck AJR, Daran J-M, de Winde JH, Westerhoff HV, Pronk JT, Bakker BM (2007) The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. Proceedings of the National Academy of Sciences 104:15753–15758. doi: 10.1073/pnas.0707476104

13. Deaner M, Mejia J, Alper HS (2017) Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in Saccharomyces cerevisiae. ACS Synth Biol 6:1931–1943. doi: 10.1021/acssynbio.7b00163

14. DeLoache WC, Russ ZN, Narcross L, Gonzales AM, Martin VJJ, Dueber JE (2015) An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol 11:465–471. doi: 10.1038/nchembio.1816

15. Di Liu, Mannan AA, Han Y, Oyarzún DA, Zhang F (2018) Dynamic metabolic control: towards precision engineering of metabolism. J Ind Microbiol Biotechnol 8:1–9. doi: 10.1007/s10295-018-2013-9

16. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343. doi: 10.1093/nar/gkt135

17. Dominguez AA, Lim WA, Qi LS (2016) Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. Nature Reviews Molecular Cell Biology 17:5–15. doi:

10.1038/nrm.2015.2

18. Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, Keasling JD, Hadi MZ, Mukhopadhyay A (2011) Engineering microbial biofuel tolerance and export using efflux pumps. Mol Syst Biol 7:487–487. doi: 10.1038/msb.2011.21

19. Espah Borujeni A, Mishler DM, Wang J, Huso W, Salis HM (2016) Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. Nucleic Acids Res 44:1–13. doi: 10.1093/nar/gkv1289

20. Ferry QRV, Lyutova R, Fulga TA (2017) Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. Nat Commun 8:14633. doi: 10.1038/ncomms14633

21. Fontana J, Dong C, Ham JY, Zalatan JG, Carothers JM (2018) Regulated expression of sgRNAs tunes CRISPRi in E. coli. Biotechnol J 1800069. doi: 10.1002/biot.201800069

22. Fossati E, Narcross L, Ekins A, Falgueyret J-P, Martin VJJ (2015) Synthesis of Morphinan Alkaloids in Saccharomyces cerevisiae. PLoS ONE 10:e0124459. doi: 10.1371/journal.pone.0124459

23. Freed EF, Winkler JD, Weiss SJ, Garst AD, Mutalik VK, Arkin AP, Knight R, Gill RT (2015) Genome-Wide Tuning of Protein Expression Levels to Rapidly Engineer Microbial Traits. ACS Synth Biol 4:1244–1253. doi: 10.1021/acssynbio.5b00133

24. Galanie S, Thodey K, Trenchard IJ, Interrante MF, Smolke CD (2015) Complete biosynthesis of opioids in yeast. Science 349:1095–1100. doi: 10.1126/science.aac9373

25. Gander MW, Vrana JD, Voje WE, Carothers JM, Klavins E (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. Nat Commun 8:15459. doi: 10.1038/ncomms15459

26. Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. - PubMed - NCBI. J Integr Plant Biol 56:343–349. doi: 10.1111/jipb.12152

27. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS (2014) Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 159:647–661.

28. Gold ND, Gowen CM, Lussier F-X, Cautha SC, Mahadevan R, Martin VJJ (2015) Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics. Microbial Cell Factories 2015 14:1 14:73. doi: 10.1186/s12934-015-0252-2

29. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, Pfleger BF (2016) CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002. Metab Eng 38:170–179. doi: 10.1016/j.ymben.2016.07.007

30. Gupta A, Reizman IMB, Reisch CR, Prather KLJ (2017) Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. Nature Biotechnology 35:273–279. doi: 10.1038/nbt.3796

31. Haynes CA, Gonzalez R (2014) Rethinking biological activation of methane and conversion to liquid fuels. Nat Chem Biol 10:331–339. doi: 10.1038/nchembio.1509

32. Holtz WJ, Keasling JD (2010) Engineering Static and Dynamic Control of Synthetic Pathways. Cell 140:19–23. doi: 10.1016/j.cell.2009.12.029

33. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, Hernday AD, Mahatdejkul-Meadows T, Szeto W, Chandran SS, Newman JD (2015) Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Systems 1:88–96. doi: 10.1016/j.cels.2015.02.001

34. Kalyuzhnaya MG, Puri AW, Lidstrom ME (2015) Metabolic engineering in methanotrophic bacteria. Metab Eng 29:142–152. doi: 10.1016/j.ymben.2015.03.010

35. Kang Z, Wang Q, Zhang H, Qi Q (2008) Construction of a stress-induced system in Escherichia coli for efficient polyhydroxyalkanoates production. Appl Microbiol Biotechnol 79:203–208. doi: 10.1007/s00253-008-1428-z

36. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330:1355–1358. doi: 10.1126/science.1193990

37. Kim SK, Han GH, Seong W, Kim H, Kim S-W, Lee D-H, Lee S-G (2016) CRISPR interference-guided balancing of a biosynthetic mevalonate pathway increases terpenoid production. Metab Eng 38:228–240. doi: 10.1016/j.ymben.2016.08.006

38. La Russa MF, Qi LS (2015) The New State of the Art: Cas9 for Gene Activation and Repression. Mol Cell Biol 35:3800–3809. doi: 10.1128/MCB.00512-15

39. Lan EI, Liao JC (2013) Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. Bioresource Technology 135:339–349.

40. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS (2013) CRISPR interference

(CRISPRi) for sequence-specific control of gene expression. Nature Protocols 8:2180–2196. doi: 10.1038/nprot.2013.132

41. Latimer LN, Lee ME, Medina-Cleghorn D, Kohnz RA, Nomura DK, Dueber JE (2014) Employing a combinatorial expression approach to characterize xylose utilization in Saccharomyces cerevisiae. Metab Eng 25:20–29. doi: 10.1016/j.ymben.2014.06.002

42. Li X-T, Jun Y, Erickstad MJ, Brown SD, Parks A, Court DL, Jun S (2016) tCRISPRi: tunable and reversible, one-step control of gene expression. Scientific Reports 6:39076. doi: 10.1038/srep39076

43. Lupták A (2016) In vitro selection and evolution. Methods 106:1–2. doi: 10.1016/j.ymeth.2016.07.022

44. Lv L, Ren Y-L, Chen J-C, Wu Q, Chen G-Q (2015) Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metab Eng 29:160–168. doi: 10.1016/j.ymben.2015.03.013

45. Meadows AL, Hawkins KM, Tsegaye Y, Antipov E, Kim Y, Raetz L, Dahl RH, Tai A, Mahatdejkul-Meadows T, Xu L, Zhao L, Dasika MS, Murarka A, Lenihan J, Eng D, Leng JS, Liu C-L, Wenger JW, Jiang H, Chao L, Westfall P, Lai J, Ganesan S, Jackson P, Mans R, Platt D, Reeves CD, Saija PR, Wichmann G, Holmes VF, Benjamin K, Hill PW, Gardner TS, Tsong AE (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. Nature 537:694–697. doi: 10.1038/nature19769

46. Miksch G, Bettenworth F, Friehs K, Flaschel E, Saalbach A, Twellmann T, Nattkemper TW (2005) Libraries of synthetic stationary-phase and stress promoters as a tool for fine-tuning of expression of recombinant proteins in Escherichia coli. Journal of Biotechnology 120:25–37. doi: 10.1016/j.jbiotec.2005.04.027

47. Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA (2016) Genetic circuit design automation. Science 352:aac7341–aac7341. doi: 10.1126/science.aac7341

48. Nielsen J, Keasling JD (2016) Engineering Cellular Metabolism. Cell 164:1185–1197. doi: 10.1016/j.cell.2016.02.004

49. Nissim L, Perli SD, Fridkin A, Perez-Pinera P, Lu TK (2014) Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. Molecular Cell 54:698–710. doi: 10.1016/j.molcel.2014.04.022

50. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Jiang H, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievense J, Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496:528–532. doi: 10.1038/nature12051

51. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo B-M, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross CA (2016) A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. Cell 165:1493–1506. doi: 10.1016/j.cell.2016.05.003

52. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183. doi: 10.1016/j.cell.2013.02.022

53. Qian Y, Huang H-H, Jiménez JI, Del Vecchio D (2017) Resource Competition Shapes the Response of Genetic Circuits. ACS Synth Biol 6:1263–1272. doi: 10.1021/acssynbio.6b00361

54. Raper AT, Stephenson AA, Suo Z (2018) Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. J Am Chem Soc 140:2971–2984. doi: 10.1021/jacs.7b13047

55. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nature Biotechnology 34:339–344. doi: 10.1038/nbt.3481

56. Rogers JK, Taylor ND, Church GM (2016) Biosensor-based engineering of biosynthetic pathways. Curr Opin Biotechnol 42:84–91. doi: 10.1016/j.copbio.2016.03.005

57. Rosenberg AB, Roco C, Muscat RA, Kuchina A, Mukherjee S, Chen W, Peeler DJ, Yao Z, Tasic B, Sellers DL, Pun SH, Seelig G (2017) Scaling single cell transcriptomics through split pool barcoding. bioRxiv 105163. doi: 10.1101/105163

58. Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T (2014) Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Scientific Reports 4:5400. doi: 10.1038/srep05400

59. Shao S, Chang L, Sun Y, Hou Y, Fan X, Sun Y (2017) Multiplexed sgRNA Expression Allows Versatile Single Nonrepetitive DNA Labeling and Endogenous Gene Regulation. ACS Synth Biol

7:176–186. doi: 10.1021/acssynbio.7b00268

60. Sleight SC, Sauro HM (2013) Visualization of Evolutionary Stability Dynamics and Competitive Fitness of Escherichia coli Engineered with Randomized Multigene Circuits. ACS Synth Biol 2:519–528. doi: 10.1021/sb400055h

61. Stevens JT, Carothers JM (2015) Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. ACS Synth Biol 4:107–115. doi: 10.1021/sb400201u

62. Strutt SC, Torrez RM, Kaya E, Negrete OA, Doudna JA (2018) RNA-dependent RNA targeting by CRISPR-Cas9. eLife Sciences 7:e32724. doi: 10.7554/eLife.32724

63. Suástegui M, Guo W, Feng X, Shao Z (2016) Investigating strain dependency in the production of aromatic compounds in Saccharomyces cerevisiae. Biotechnol Bioeng 113:2676–2685. doi: 10.1002/bit.26037

64. Suástegui M, Yu Ng C, Chowdhury A, Sun W, Cao M, House E, Maranas CD, Shao Z (2017) Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in Saccharomyces cerevisiae for high production of polymer and drug precursors. Metab Eng 42:134–144. doi: 10.1016/j.ymben.2017.06.008

65. Tang W, Hu JH, Liu DR (2017) Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. Nat Commun 8:15939. doi: 10.1038/ncomms15939

66. Thompson D, Regev A, Roy S (2015) Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annu Rev Cell Dev Biol 31:399–428. doi: 10.1146/annurev-cellbio-100913-012908

67. Vigouroux A, Oldewurtel E, Cui L, van Teeffelen S, Bikard D (2017) Engineered CRISPR-Cas9 system enables noiseless, fine-tuned and multiplexed repression of bacterial genes. bioRxiv 164384. doi: 10.1101/164384

68. Wang Y-H, McKeague M, Hsu TM, Smolke CD (2016) Design and Construction of Generalizable RNA-Protein Hybrid Controllers by Level-Matched Genetic Signal Amplification. Cell Systems 3:549–562.e7. doi: 10.1016/j.cels.2016.10.008

69. Wen Z, Minton NP, Zhang Y, Li Q, Liu J, Jiang Y, Yang S (2017) Enhanced solvent production by metabolic engineering of a twin-clostridial consortium. Metab Eng 39:38–48. doi: 10.1016/j.ymben.2016.10.013

70. Wong ASL, Choi GCG, Cui CH, Pregernig G, Milani P, Adam M, Perli SD, Kazer SW, Gaillard A, Hermann M, Shalek AK, Fraenkel E, Lu TK (2016) Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc Natl Acad Sci USA 113:2544–2549. doi: 10.1073/pnas.1517883113

71. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The Evolution of Transcriptional Regulation in Eukaryotes. Mol Biol Evol 20:1377–1419. doi: 10.1093/molbev/msg140

72. Wu J, Du G, Chen J, Zhou J (2015) Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in Escherichia coli. Scientific Reports 5:13477. doi: 10.1038/srep13477

73. Wu M-Y, Sung L-Y, Li H, Huang C-H, Hu Y-C (2017) Combining CRISPR and CRISPRi Systems for Metabolic Engineering of E. coli and 1,4-BDO Biosynthesis. ACS Synth Biol 6:2350–2361. doi: 10.1021/acssynbio.7b00251

74. Yamamoto K, Yamanaka Y, Shimada T, Sarkar P, Yoshida M, Bhardwaj N, Watanabe H, Taira Y, Chatterji D, Ishihama A, Traxler MF (2018) Altered Distribution of RNA Polymerase Lacking the Omega Subunit within the Prophages along the Escherichia coli K-12 Genome. mSystems 3:e00172–17. doi: 10.1128/mSystems.00172-17

75. Yao L, Cengic I, Anfelt J, Hudson EP (2015) Multiple Gene Repression in Cyanobacteria Using CRISPRi. ACS Synth Biol 5:207–212. doi: 10.1021/acssynbio.5b00264

76. Yuan Z, Zhao C, Di Z, Wang W-X, Lai Y-C (2013) Exact controllability of complex networks. Nat Commun 4:47. doi: 10.1038/ncomms3447

77. Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160:339–350. doi: 10.1016/j.cell.2014.11.052

78. Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nature Biotechnology 30:354–359. doi: 10.1038/nbt.2149

Figures

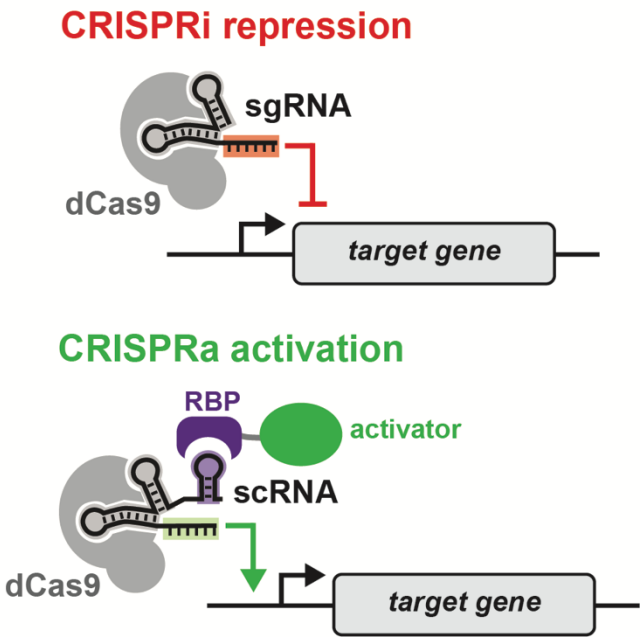


Figure 1. CRISPR-Cas transcriptional regulation in bacteria.

CRISPR-Cas transcriptional regulation is performed by targeting nuclease-deficient Cas9 (dCas9) to target loci using a complementary 20 base target sequence found in the sgRNAs. *S. pyogenes* dCas9 is the most well-characterized CRISPR-Cas protein and can be targeted to any 20 bases followed by a 5'-NGG-3' trinucleotide (protospacer adjacent motif, or PAM site). In bacteria, CRISPRi transcriptional repression is achieved by sterically blocking the RNA polymerase and requires targeting dCas9 near the beginning of the open reading frame of the target gene (top). In eukaryotes, robust CRISPRi repression can be achieved by fusing transcriptional repression domains to dCas9 (not shown). In both bacteria and eukaryotes, CRISPRa is performed by directing a transcriptional activator protein upstream of the target gene using the CRISPR-Cas complex (bottom). This can be achieved by expressing a modified sgRNA (scaffold RNA or scRNA) encoding an additional RNA hairpin at the 3' end that recruits a RNA binding protein fused to the transcriptional activator. Alternatively, using a similar strategy, transcriptional activation proteins can be fused directly to dCas9 (not shown).

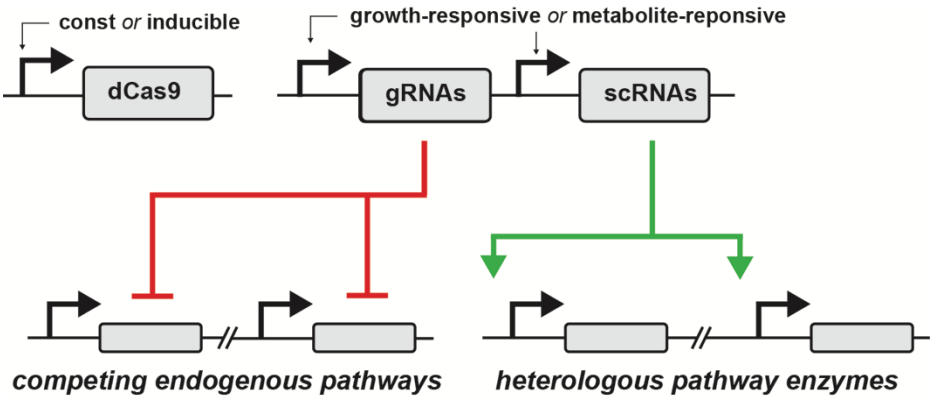


Figure 2. Schematic of dynamic CRISPR-Cas regulation.

Dynamic CRISPR-Cas systems could be constructed by placing the regulated expression of sgRNAs and scRNAs under the control of inducer-, growth-phase-, or metabolite-responsive protein- or RNA-based sensors. By doing so, multi-gene programs could be designed or identified through combinatorial screening to coordinate the timing and expression levels of endogenous and heterologous pathway genes, balance cellular resource demands, and optimize bioproduction titers, rates and yields.

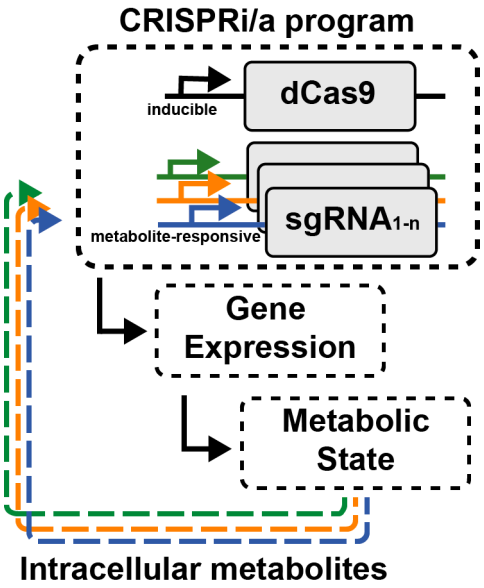


Figure 3. Genetic architecture for dynamic CRISPR-Cas transcriptional control of multiple genes. In this example, a dynamic CRISPR-Cas control system responsive to multiple cellular inputs is implemented by placing separate sgRNAs or scRNAs under the control different metabolite-responsive promoters regulated by dynamic genetic controllers (color-coded promoters). These controllers dynamically titrate the expression of the sgRNAs or scRNAs in response to changes in the intracellular concentration of their cognate metabolites and consequently tune the extent of CRISPRi and CRISPRa applied on the target genes. Colored dashed arrows represent dynamic feedback on each metabolite-responsive promoter by the cognate metabolite. dCas9 is placed under the control of an independent input, such as an inducible promoter, to provide a master switch for the CRISPR-Cas transcriptional regulation program.

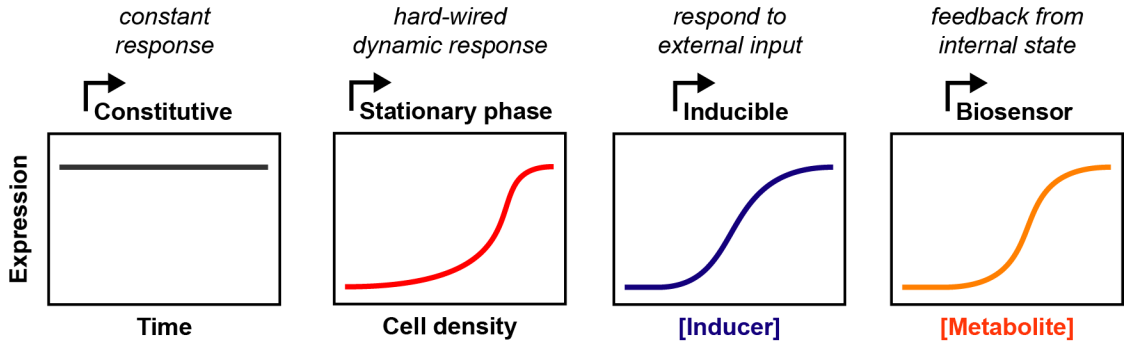


Figure 4. Dynamic genetic controllers produce distinctive outputs in gene expression.
Constitutive promoters are characterized by constant gene expression over time. Stationary phase promoters that begin expressing at set stages in the growth phase can be wired to genes of interest to encode a dynamic response. Inducible promoters can be used to titrate expression in response to an external input. Biosensors use metabolite-responsive genetic controllers to sense the intracellular concentration of their cognate metabolite and titrate expression in response.

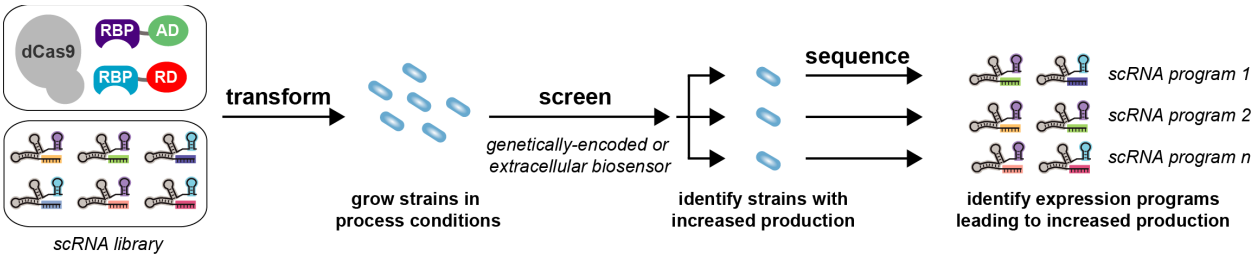


Figure 5. Example of biosensor screen of expression programs for increased production.

The goal of this screen is to identify strains displaying improved production of a target molecule in the desired process conditions. The components necessary for CRISPR-Cas transcriptional regulation are delivered to the bacterial strain of interest. A combinatorial library of scRNAs directs transcriptional regulation to heterologous or endogenous genes. Transcriptional repression can be performed by sterically blocking the RNA polymerase using dCas9 in bacteria, or by recruiting transcriptional repression domain to the CRISPR-Cas complex using scRNAs in eukaryotes. Transcriptional activation can be performed by recruiting transcriptional activator domains to the CRISPR-Cas complex in both bacterial and eukaryotic hosts. Target genes can be selected rationally, or unbiased screens can be performed. The resulting strains are grown in the desired process conditions, after which production is assayed. If a fluorescence-based biosensor for the molecule of interest can be genetically encoded, flow cytometry and cell sorting can be used to rapidly screen strains for programs that increase intracellular production. Conversely, fluorescent biosensors that can detect secreted molecules can be used to screen for programs that increase extracellular production. When high-producing strains are isolated, sequencing of the scRNA target sequences can be used to identify the transcriptional programs responsible for improved production. RBP: RNA-binding protein; AD: activation domain; RD: repression domain.

Chapter 4:

Regulated expression of sgRNAs tunes CRISPRi in *E. coli*

Jason Fontana[1,4], Chen Dong[2,4], Jennifer Y. Ham[3,4], Jesse G. Zalatan[1,2,4], James M. Carothers[1,3,4]

[1] Molecular Engineering & Sciences Institute, University of Washington, Seattle, WA, 98195, USA

[2] Department of Chemistry, University of Washington, Seattle, WA 98195, USA

[3] Department of Chemical Engineering, University of Washington, Seattle, WA, 98195, USA

[4] Center for Synthetic Biology, University of Washington, Seattle, WA, 98195, USA

Published as a research article in the *Biotechnology Journal* on April 10th, 2018.

DOI: 10.1002/biot.201800152

Abstract

Methods for implementing dynamically-controlled multi-gene programs could expand our ability to engineer metabolism for efficiently producing high-value compounds. Working toward this goal, we explored whether CRISPRi repression can be tuned in *E. coli* through the regulated expression of the CRISPRi machinery. We find when dCas9 is not limiting, variations in sgRNA expression alone can lead to CRISPRi repression levels ranging from 5- to 300-fold. We show that titrating sgRNA expression over a 2.5-fold range can lead to 16-fold changes in reporter gene expression. Many different classes of genetic controllers can generate 2.5-fold differences in transcription, indicating they could be integrated in dynamically-regulated CRISPRi circuits. Finally, we observed that CRISPRi cannot be reversed for up to 12 hours by expressing a competing sgRNA later in the growth phase, indicating that CRISPR-Cas:DNA interactions can be persistent *in vivo*. Collectively, our results identify genetic architectures for tuning CRISPRi repression through regulated sgRNA expression and suggest that dynamically-regulated CRISPRi systems targeting multiple genes may be within reach.

Introduction

Naturally-occurring systems respond to rapidly changing cellular conditions by coordinating function within metabolic networks with a degree of precision that cannot be matched using existing approaches for genetic engineering. Optimizing biosynthetic output is now understood to require a complicated balance of: (i) regulating heterologous enzyme function, (ii) redirecting flux through central carbon metabolism, (iii) diverting flux from competing pathways, and (iv) managing cytotoxic intermediates and enzymes whose accumulation disrupts the balance of cellular resources [1, 2]. Notable successes in metabolic engineering, including the synthesis of opioid precursors, the malaria drug artemisinin, and industrial isoprenoids have required extensive efforts to identify and tune the expression levels of multiple heterologous genes while removing competing enzyme activities within the cell [1]. Developing methods to rapidly implement complex multi-gene programs with precise control over gene expression will expand our engineering capabilities and enable discovery-based experiments to uncover genetic programs that optimize metabolic flux and biosynthetic output.

The recently-developed CRISPR-Cas system provides a powerful suite of tools for genome editing [3] and transcriptional control [4, 5]. In particular, CRISPR interference (CRISPRi) can be used to repress genes in a programmable manner by physically blocking transcription [4]. CRISPRi uses a catalytically-inactive Cas9 protein (dCas9) and a single guide RNA (sgRNA) to target arbitrary DNA sequences based on predictable Watson-Crick base pairing. To date, there is a growing number of examples of CRISPRi-based circuits to optimize biosynthesis pathways in diverse bacterial species [6-13], underscoring the potential of the CRISPR-Cas system for creating multi-gene transcriptional control programs.

The ability to tune gene expression using CRISPRi has already proven useful for improving production phenotypes [6-10, 12, 13]. To date, three general strategies have been employed to tune the levels of CRISPRi repression. First, different levels of CRISPRi transcriptional repression can be generated by directing dCas9 to different regions of the target gene [4]. Second, sgRNAs with different degrees of complementarity with the target sequence can be easily designed [4, 14]. Although these two approaches are conceptually straightforward, a major drawback is that the level of CRISPRi transcriptional repression is hard-coded, and new sgRNAs must be introduced to access different levels of repression. Lastly, several groups have shown that transcriptional repression can be tuned by titrating the expression levels of the CRISPRi machinery itself. This approach has included simultaneously varying the levels of plasmid-borne dCas9 and sgRNA [4, 15], and regulating the expression of genomically-integrated dCas9 from chemically-induced promoters [15-17].

Ideally, we would like to implement an alternative approach to tunable CRISPRi through the regulated expression of individual sgRNAs, which has the advantage of allowing transcriptional repression to be separately initiated at different genetic loci. Further, we hope to use dynamically-controlled genetic regulators to sense changes in metabolic state, and in response alter the expression of one or more sgRNAs to tune the output of a multi-gene transcriptional program (Figure 1A). There is growing evidence that dynamically-controlled gene circuits can be employed to solve control problems [18], balance the

supplies and demands for cellular resources [19], and improve product yields in engineered microbes [20-23]. Understanding the design rules for programming variations in CRISPRi through the regulated expression of the CRISPRi machinery could set the stage for rapidly engineering dynamically-controlled multi-gene circuits.

To approach this goal, we investigated genetic mechanisms for tuning CRISPRi in *E. coli*. Using a combination of constitutive and inducible promoters, we quantified the relationships between *S. pyogenes* dCas9 and sgRNA expression and CRISPRi-mediated transcriptional repression (Figure 1B). Our data reveal that 5- to 300-fold CRISPRi repression can be generated through variations in sgRNA transcription rate alone when dCas9 is expressed at high levels. By titrating dCas9 and sgRNA expression, we demonstrate that 16-fold changes in reporter gene expression and up to ~99% transcriptional repression can be generated by 2.5-fold differences in sgRNA transcriptional inputs. These results are compatible with the idea that a number of different classes of genetic controllers could be employed to drive sgRNA transcription and permit tunable CRISPRi repression. Lastly, we also find that CRISPRi is not readily reversed by later expressing a competing sgRNA in our system, consistent with biophysical studies showing that CRISPR-Cas:DNA interactions can be persistent [24]. This property may limit the ability to implement dynamic feedback when CRISPRi repression is already in place unless new strategies for rapidly reversing CRISPRi are devised. Nonetheless, our work suggests that the dynamic control of multiple genes may be achievable through the regulated expression of sgRNAs and elucidates important principles that may enable the construction of CRISPRi programs responsive to changing cellular conditions.

Materials and methods

Bacterial strain construction

Plasmids cloning was performed using standard molecular biology methods. The plasmids used in this study are listed in Table S1. mRFP1 or sfGFP fluorescent reporter strains were generated by bacterial recombineering in *E. coli* cells carrying the lambda-red system [25]. The integrated sequences were then transferred to MG1655 by P1 transduction [26]. The GFP reporter was integrated at the *nfsA* locus, and the RFP reporter was integrated at the *rbsAR* locus [27] (Table S2). sgRNA sequences used in this study are reported in Table S3.

Bacterial cell growth

E. coli cells were inoculated in 500 µL LB supplemented with 100 µg/mL carbenicillin (Sigma-Aldrich) and 25 µg/mL chloramphenicol (Sigma-Aldrich) and grown in 2 mL 96-well plates at 37 °C, 220 RPM overnight. For single-point measurements, cultures were then diluted 1:100 in 500 µL EZ-RDM (Teknova) supplemented with appropriate antibiotics. In induction experiments, anhydrotetracycline (aTc, Sigma) was supplemented upon dilution. For the time-course experiment, cultures were then diluted in 200 µL EZ-RDM

supplemented with appropriate antibiotics and grown for 16 h in flat, clear-bottomed 96-well plates (Corning) at 37 °C, shaking in a Biotek Synergy HTX plate reader. Induction with aTc was performed 4 h after dilution.

Fluorescence measurements

Plate reader measurements were performed on a Biotek Synergy HTX plate reader. sfGFP fluorescence was detected using an excitation wavelength of 485 nm and an emission wavelength of 528 nm; for fluorescence detection of mRFP1, 540 nm and 600 nm were used. Flow cytometry measurements were performed on a MACSQuant VYB flow cytometer (Miltenyi Biotech). 15 µL of overnight cultures diluted 1:40 in Phosphate Buffer Saline (PBS) were injected and a side scatter threshold trigger (SSC-H) was applied to enrich for single cells. Gating for single cells was performed by selecting events found along the diagonal of the SSC-H vs. SSC-A plot. Events with FSC-A and fluorescence values below the detection limit of the instrument were excluded.

Statistical analysis

Statistical analyses were performed using Prism 6.0e. Statistical significance between different samples was evaluated using two-tailed t-tests using $\alpha = 0.05$.

Results

Impact of variations in dCas9 and gRNA expression on CRISPRi repression

The first key step toward implementing dynamically-controlled CRISPRi-based transcriptional programs is to understand how CRISPRi repression responds to changing dCas9 and sgRNA expression levels. High levels of CRISPRi transcriptional repression have commonly been achieved by placing the expression of both dCas9 and sgRNA under the control of strong constitutive promoters [4, 28], such as Spy.pCas9 [29] and Bba_J23119, respectively [parts.igem.org/Part:BBa_J23119 for promoters named Bba_J231**]. We reasoned that it should be possible to tune CRISPRi by expressing limiting amounts of dCas9 or sgRNA when the other is expressed at high levels.

To evaluate the dependence of CRISPRi on the availability of its components and achieve varying levels of repression, we used constitutive promoters of different strengths to titrate the expression of dCas9 and sgRNA. We generated a combinatorial panel of *E. coli* MG1655-derived strains where the CRISPRi components were driven by constitutive promoters from our set, and then measured transcriptional repression of a genomically-integrated fluorescent reporter protein (sfGFP). When dCas9 was expressed at high levels from the Spy.pCas9 promoter, we observed graded increases in reporter expression when using weaker promoters driving the sgRNA (Figure 2). These results indicate that there was a direct correspondence between sgRNA promoter strength and the level of CRISPRi repression observed, ranging from 5- to 300-fold compared to a control not carrying a sgRNA (Figure 2, Figure S1). Similarly, when the sgRNA was expressed at high levels from the strong Bba_J23119 promoter, there was a direct correspondence between dCas9 promoter strength and the level of CRISPRi repression generated, ranging

from 2- to 300-fold (Figure 2, Figure S1). CRISPRi could not be readily titrated through sgRNA expression when dCas9 was expressed from either of two weaker promoters (Bba_J23117 and Bba_J23112; Figure 2, Figure S1), suggesting that dCas9 was a limiting component determining the level of repression in those conditions. It is possible that a narrow window exists where CRISPRi repression can be tuned by further limiting the strength of the sgRNA promoter when dCas9 is controlled by either one of the two weaker promoters (Bba_J23117, Bba_J23112). Regardless, because the upper-limit of transcriptional repression that could be observed in these conditions was lower compared to conditions where dCas9 is controlled by the stronger Spy.pCas9 promoter (Figure S1), smaller changes in target gene expression would be generated.

Collectively, these data demonstrate that a broad range of CRISPRi repression levels can be accessed through regulated changes in sgRNA expression alone, provided an appropriate dCas9 expression regime has been established.

Transcriptional input dynamic ranges required for titratable CRISPRi

We next sought to quantify the sensitivity of CRISPRi repression to changes in the expression of the CRISPRi components. We engineered a set of test strains where either dCas9 or the sgRNA was expressed from a strong constitutive promoter (Spy.pCas9 and Bba_J23119, respectively). To permit facile titratable delivery of the remaining CRISPRi component, we used an anhydrotetracycline (aTc)-inducible promoter (pTet, derived from Tn10 pTet [30]). We varied the transcriptional input of the CRISPRi component expressed from the inducible promoter by titrating the concentration of aTc inducer, and then measured the level of CRISPRi repression of a genomically-integrated reporter gene (mRFP1). When the sgRNA was constitutively expressed and dCas9 was controlled by pTet, even very low levels of aTc resulted in ~99% repression of the reporter compared to a control with no CRISPRi components expressed. This result indicates that CRISPRi repression cannot be readily tuned in this condition (Figure 3A). Conversely, when dCas9 was constitutively expressed, CRISPRi repression could be tuned across a broad range through the induced expression of the sgRNA from the pTet promoter. In this design, we observed 16-fold changes in reporter gene expression (Figure 3A, Figure S2), ranging from 4.6- to 74-fold repression (Figure S3). These data are equivalent to 21% to 1.3% of the reporter expression compared to the control without CRISPRi components. When we placed both dCas9 and sgRNA under inducible pTet control, we observed 30-fold changes in reporter gene expression (Figure 3A), indicating that the obtainable changes in repression increase only by a factor of two when dCas9 and sgRNAs are both regulated compared to regulating the sgRNA alone. This difference arises from constitutive expression of dCas9 and leaky sgRNA expression that result in modest (~4.6-fold) CRISPRi repression when the sgRNA alone is regulated. This leaky CRISPRi effect reduces the overall dynamic range of repression compared to when both CRISPRi components are under inducible control and lower leaky repression is observed in the absence of aTc.

To quantify the dynamic ranges in the CRISPRi components that are required to tune CRISPRi in these conditions, we measured the fluorescent reporter expression generated from pTet when delivered via

the same plasmid vector (ColE1) we used to express the sgRNA in the previous experiments (Figure S4). We varied mRFP1 expression by titrating the concentration of aTc inducer and used the relative changes in mRFP1 output to estimate the changes in sgRNA expression. We assumed similar changes in dCas9 expression, since the dynamic range achievable using pTet is very similar when it is delivered via the plasmid vectors we use for delivering sgRNA expression (ColE1) or dCas9 expression (p15A)[30]. We generated a sgRNA response curve for CRISPRi repression using the relative changes in sgRNA expression we calculated and the respective relative changes in target gene expression they determined (Figure 3B). A ~2.5-fold change in sgRNA input alone was sufficient to traverse the observed 16-fold change in reporter expression (Figure 3B, orange). When both dCas9 and sgRNA were controlled by pTet, so that dCas9 and sgRNA expression both vary by ~2.5-fold, we achieved a 30-fold change in reporter gene expression (Figure 3B, grey). We performed the same experiment at a distinct gene target (sfGFP) and we obtained comparable results (Figure S5), suggesting that the window of sgRNA expression where different repression levels can be achieved is similar when targeting a different gene.

Our analysis suggests that useful changes in target gene expression could be programmed with small variations in CRISPRi components expression. In other words, genetic controllers with dynamic ranges as a low as 2.5-fold may be sufficient for observing meaningful changes in target gene expression, provided the levels of sgRNA expression span the levels needed to generate differences in CRISPRi.

Persistence of CRISPRi repression in the presence of competing sgRNAs

In vitro and *in vivo* studies have demonstrated that Cas9 remains stably associated to its cleaved DNA product [24] and that sgRNA:Cas9 complexes are long lived [31, 32]. Others have shown that when inducing CRISPRi *in vivo* by controlling dCas9 and sgRNA, washing away the chemical inducer leads to complete relief of repression after ~6 h when the cells are actively growing, presumably by diluting the CRISPRi components [4]. We wondered whether the persistence of the repression phenotype typically observed with constitutive expression of the CRISPRi components could be relieved by introducing a competing sgRNA. If not, this would imply that once sgRNAs are expressed in the cell, the resulting CRISPRi repression at targeted loci may persist on the 12-48 h timescales involved in the batch culture conditions often utilized in metabolic pathway engineering. This persistence may be a fundamental constraint that prevents CRISPRi transcriptional repression from being reversed at individual targets.

In order to investigate these questions, we generated test constructs where the strong Bba_J23119 constitutive promoter controlled the expression of a sgRNA targeting an integrated reporter (sfGFP), and the pTet inducible promoter controlled the expression of a second competing sgRNA without a genomic target. Using this system, we evaluated whether inducing the expression of a competing sgRNA was sufficient to observe relief of repression at the reporter gene. To simulate the activation of a new CRISPRi repression program at a late-exponential/stationary phase, mimicking a switch between a growth phase and a production phase, we induced the expression of the second sgRNA 4 h after dilution. We observed persistence of CRISPRi repression on the reporter for up to 12 h after induction of the competing sgRNA

at aTc concentrations approaching saturation (Figure 4). To confirm that the second sgRNA could in principle effectively compete to relieve repression, we also expressed the second sgRNA without including the TetR protein (pTet*), so both the on-target and the competing off-target sgRNA are expressed constitutively. In this situation, repression was markedly weaker compared to expressing the sgRNA targeting sfGFP alone (Figure 4). This result suggests that competition for the available dCas9 can be observed with our system. Our experiments indicate that no switch to a different CRISPRi phenotype could occur in these expression conditions and, more broadly, that CRISPRi programs already established in the cell may not be released when separate CRISPRi programs are initiated later in the growth phase.

Discussion

In this work, we aimed to uncover useful design rules for engineering dynamically-regulated multi-gene CRISPRi-based transcriptional programs. We first investigated our ability to tune CRISPRi through the regulated expression of dCas9 and sgRNAs independently and in combination. Previous efforts where the concentration of active sgRNA was regulated provide a strong case for the importance of sgRNA expression in tuning CRISPRi repression levels [33-35]. Our results using constitutive promoters of different strengths to control the expression of sgRNAs confirm that graded differences in CRISPRi repression can be readily achieved through the regulated expression of sgRNAs in conditions where dCas9 is not limiting. There are many classes of genetic controllers that could be employed to dynamically regulate the expression of sgRNAs in response to changing environments. However, whether the performance characteristics of these controllers are compatible with dynamically tuning CRISPRi repression remains to be demonstrated.

In order to provide a framework for evaluating the performance of dynamic genetic controllers for dynamic CRISPRi*,* we explored the transcriptional requirements for tuning CRISPRi repression by titrating dCas9 and sgRNA expression from inducible pTet promoters. Using this approach, we identified the dynamic range of transcriptional inputs in dCas9 and/or sgRNAs needed to generate targeted levels of repression. We were particularly interested in quantitatively understanding whether a penalty in CRISPRi tunability exists when only the sgRNA level is changed, compared to a condition in which both dCas9 and sgRNA expression levels are varied. We found that regulating the expression of the sgRNA alone was sufficient for generating 16-fold changes in CRISPRi, only two times lower than the 30-fold changes in repression observed when sgRNA and dCas9 expression are both regulated. Our analysis shows that there is a direct correspondence between the level of sgRNA transcriptional input and the level of CRISPRi repression, until the maximum level of repression is achieved at approximately 2.5 times more sgRNA compared to the uninduced control. Our finding that CRISPRi repression can be tuned by regulating the expression of sgRNAs alone is important because it suggests that repression of individual gene targets in response to separate metabolic inputs within dynamically-regulated multi-gene programs can be independently initiated.

In principle, sgRNAs could be wired to dynamic genetic controllers for dynamic control over gene expression in engineered circuits and pathways. Further controlling dCas9 expression via separate

dynamic inputs could provide a master switch for CRISPRi repression [36] and possibly avoid the burdens associated with the untimed expression of heterologous genes in engineered biosynthetic systems [18]. Compared to directly engineering sgRNA with sequences and structures compatible with aptamer domains [33-35], building dynamic CRISPRi control circuits that operate by regulating the expression of sgRNAs, or sgRNAs and dCas9, may be more straightforward to implement. Such circuits could be assembled using many different classes of dynamic genetic controllers with relevance for metabolic engineering, including: (i) carbon-responsive promoters [37], (ii) growth phase-responsive promoters [38], (iii) metabolite-responsive transcriptional regulators [23], (iv) and RNA aptamer-based expression devices responsive to targeted small molecules [39, 40]. These controllers provide dynamic changes in gene expression that are compatible with the requirements for tuning CRISPRi we identified in this study (Table S4). Our results using well-characterized promoters to regulate sgRNAs provide a framework for adjusting the sgRNA expression generated by dynamic genetic controllers to regimes where CRISPRi repression can be tuned by variations in sgRNA levels.

Engineering sophisticated CRISPRi programs that allow the cell to sense its metabolic state and respond with multiple independent adjustments to gene expression will require the ability to dynamically titrate CRISPRi repression. While the genetic designs required for dynamically inducing CRISPRi repression have been studied in this work and elsewhere [4, 15, 16], our ability to dynamically relieve CRISPRi within time-scales relevant for metabolic engineering remains to be addressed. We aimed to elucidate the persistence of CRISPRi repression programs and whether the expression of competing sgRNAs could release repression by inducing a switch to a different phenotype. In particular, we were interested in finding (i) whether releasing CRISPRi repression is achievable with commonly used sgRNA expression conditions and (ii) how long CRISPRi repression persists in the presence of a competing sgRNAs expressed at a later time. Our data show that simply inducing a second, competing sgRNA at standard levels is not sufficient to observe relief of repression of the target gene. More broadly, our results suggest that a previously implemented CRISPRi phenotype may be able to persist for up to 12 h if a second CRISPR transcriptional regulation program is induced later in the growth phase.

In order to develop a dynamic CRISPRi system that can be used to reversibly switch between different genetic programs, additional experiments will be needed to fully investigate the persistence of CRISPRi, its impact on dynamic control, and possible strategies to overcome this limitation. We hypothesize that expressing fast-degrading dCas9 proteins could permit dynamic feedback by increasing turnover of CRISPRi complexes bound to their target and improve the responsiveness of the system to changing conditions [41]. Even in the absence of rapidly-reversible CRISPRi, developing dynamic genetic tools based on the existing CRISPRi system will be useful for implementing persistent responses to changing conditions, for instance when cells shift from rapid growth to stationary phase [21].

Another key feature necessary for implementing dynamic CRISPRi gene expression programs is the ability to predictably target many different genes. While several authors have demonstrated that high levels of CRISPRi repression can be performed on a wide array of different targets in bacteria [17], recent

studies have highlighted how the choice of sgRNA target sequence can strongly affect the performance of CRISPR engineering in *E. coli* [42]. In this work, we focused our investigation on a small test set of genes in the ideal condition where (i) a highly active sgRNA target sequence could be identified, (ii) the target gene was strongly expressed and not essential, (iii) no existing transcriptional regulation is reported, and (iv) no other sgRNAs were competing for dCas9. Additional experiments evaluating how the system performs on other target genes and in non-ideal conditions will be required for identifying a comprehensive set of rules for tuning CRISPRi.

Ultimately, we aim to engineer CRISPR-transcriptional regulation control systems that are straightforward to implement and are capable of dynamically controlling genes in a tunable manner. Such programs, for instance, could be used for discovery-based high-throughput screens to identify regulatory architectures that improve biosynthesis yields. Through combinatorial variations in the promoters driving dCas9 and sgRNAs, the constructs presented in this paper already provide a straightforward way to implement the discrete levels of gene expression needed for the high-throughput screens described above. Taken together with a plethora of results obtained elsewhere, our results underscore the vast potential that could be unlocked with the further development of CRISPRi tools for dynamically tuning the levels and timing of gene expression in engineered bacteria through the regulation of sgRNA expression.

Acknowledgements

We thank Lei (Stanley) Qi, Maureen Thomason, Mary Lidstrom, Willy Voje, Jason Stevens, and members of the Zalatan and Carothers groups for *E. coli* strains, plasmids, technical assistance, advice, and helpful discussions.

This work was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (J.G.Z.), a NSF Award MCB 1517052 and a University of Washington Presidential Innovation Award (J.M.C.).

Conflict of interest

The authors declare no financial or commercial conflict of interest.

References

[1] J. Nielsen, J. D. Keasling, *Cell* **2016**, *164*, 1185.

[2] J. M. Carothers, J. A. Goler, J. D. Keasling, *Curr Opin Biotechnol* **2009**, *20*, 498.

[3] S. H. Sternberg, J. A. Doudna, *Mol Cell* **2015**, *58*, 568.

[4] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, *Cell* **2013**, *152*, 1173.

[5] A. A. Dominguez, W. A. Lim, L. S. Qi, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*.

[6] L. Lv, Y. L. Ren, J. C. Chen, Q. Wu, G. Q. Chen, *Metab. Eng* **2015**, *29*, 160.

[7] J. J. Wu, G. C. Du, J. Chen, J. W. Zhou, *Sci. Rep.* **2015**, *5*.

[8] S. Cleto, J. V. K. Jensen, V. F. Wendisch, T. K. Lu, *ACS Synth. Biol.* **2016**, *5*, 375.

[9] G. C. Gordon, T. C. Korosh, J. C. Cameron, A. L. Markley, M. B. Begemann, B. F. Pfleger, *Metab. Eng* **2016**, *38*, 170.

[10] S. K. Kim, G. H. Han, W. Seong, H. Kim, S. W. Kim, D. H. Lee, S. G. Lee, *Metab. Eng* **2016**, *38*, 228.

[11] L. Yao, I. Cengic, J. Anfelt, E. P. Hudson, *ACS Synth. Biol.* **2016**, *5*, 207.

[12] Z. Q. Wen, N. P. Minton, Y. Zhang, Q. Li, J. L. Liu, Y. Jiang, S. Yang, *Metab. Eng* **2017**, *39*, 38.

[13] M. Y. Wu, L. Y. Sung, H. Li, C. H. Huang, Y. C. Hu, *ACS Synth. Biol.* **2017**, *6*, 2350.

[14] A. Vigouroux, E. Oldewurtel, L. Cui, S. van Teeffelen, D. Bikard, *bioRxiv* **2017**.

[15] M. H. Larson, L. A. Gilbert, X. W. Wang, W. A. Lim, J. S. Weissman, L. S. Qi, *Nat. Protoc.* **2013**, *8*, 2180.

[16] X. T. Li, Y. G. Jun, M. J. Erickstad, S. D. Brown, A. Parks, D. L. Court, S. Jun, *Sci. Rep.* **2016**, *6*.

[17] J. M. Peters, A. Colavin, H. D. Shi, T. L. Czarny, M. H. Larson, S. Wong, J. S. Hawkins, C. H. S. Lu, B. M. Koo, E. Marta, A. L. Shiver, E. H. Whitehead, J. S. Weissman, E. D. Brown, L. S. Qi, K. C. Huang, C. A. Gross, *Cell* **2016**, *165*, 1493.

[18] J. T. Stevens, J. M. Carothers, *ACS Synth. Biol.* **2015**, *4*, 107.

[19] W. J. Holtz, J. D. Keasling, *Cell* **2010**, *140*, 19.

[20] R. H. Dahl, F. Zhang, J. Alonso-Gutierrez, E. Baidoo, T. S. Batth, A. M. Redding-Johanson, C. J. Petzold, A. Mukhopadhyay, T. S. Lee, P. D. Adams, J. D. Keasling, *Nat. Biotechnol* **2013**, *31*, 1039.

[21] A. Gupta, I. M. B. Reizman, C. R. Reisch, K. L. J. Prather, *Nat. Biotechnol* **2017**, *35*, 273.

[22] S. Z. Tan, K. L. J. Prather, *Curr Opin Chem Biol* **2017**, *41*, 28.

[23] F. Z. Zhang, J. M. Carothers, J. D. Keasling, *Nat. Biotechnol* **2012**, *30*, 354.

[24] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, *Nature* **2014**, *507*, 62.

[25] D. G. Yu, H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, D. L. Court, *Proc. Natl. Acad. Sci. U.S.A* **2000**, *97*, 5978.

[26] L. C. Thomason, N. Costantino, D. L. Court, *Curr Protoc Mol Biol* **2007**, *Chapter 1*, Unit 1 17.

[27] S. Sabri, J. A. Steen, M. Bongers, L. K. Nielsen, C. E. Vickers, *Microb Cell Fact.* **2013**, *12*.

[28] D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, L. A. Marraffini, *Nucleic Acids Res* **2013**, *41*, 7429.

[29] W. Y. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, *Nat. Biotechnol* **2013**, *31*, 233.
[30] T. S. Lee, R. A. Krupa, F. Z. Zhang, M. Hajimorad, W. J. Holtz, N. Prasad, S. K. Lee, J. D. Keasling, *J Biol Eng* **2011**, *5*.
[31] V. Mekler, L. Minakhin, E. Semenova, K. Kuznedelov, K. Severinov, *Nucleic Acids Res* **2016**, *44*, 2837.
[32] S. B. Thyme, L. Akhmetova, T. G. Montague, E. Valen, A. F. Schier, *Nat. Commun.* **2016**, *7*.
[33] Y. C. Liu, Y. H. Zhan, Z. C. Chen, A. B. He, J. F. Li, H. W. Wu, L. Liu, C. L. Zhuang, J. H. Lin, X. Q. Guo, Q. X. Zhang, W. R. Huang, Z. M. Cai, *Nat. Methods* **2016**, *13*, 938.
[34] Q. R. V. Ferry, R. Lyutova, T. A. Fulga, *Nat. Commun.* **2017**, *8*.
[35] W. X. Tang, J. H. Hu, D. R. Liu, *Nat. Commun.* **2017**, *8*.
[36] J. G. Zalatan, M. E. Lee, R. Almeida, L. A. Gilbert, E. H. Whitehead, M. La Russa, J. C. Tsai, J. S. Weissman, J. E. Dueber, L. S. Qi, W. A. Lim, *Cell* **2015**, *160*, 339.
[37] W. R. Farmer, J. C. Liao, *Nat. Biotechnol* **2000**, *18*, 533.
[38] G. Miksch, F. Bettenworth, K. Friehs, E. Flaschel, A. Saalbach, T. Twellmann, T. W. Nattkemper, *J. Biotechnol.* **2005**, *120*, 25.
[39] J. M. Carothers, J. A. Goler, D. Juminaga, J. D. Keasling, *Science* **2011**, *334*, 1716.
[40] A. E. Borujeni, D. M. Mishler, J. Z. Wang, W. Huso, H. M. Salis, *Nucleic Acids Res* **2016**, *44*, 1.
[41] D. E. Cameron, J. J. Collins, *Nat. Biotechnol* **2014**, *32*, 1276.
[42] E. A. Moreb, B. Hoover, A. Yaseen, N. Valyasevi, Z. Roecker, R. Menacho-Melgar, M. D. Lynch, *Acs Synth Biol* **2017**, *6*, 2209.

Figures

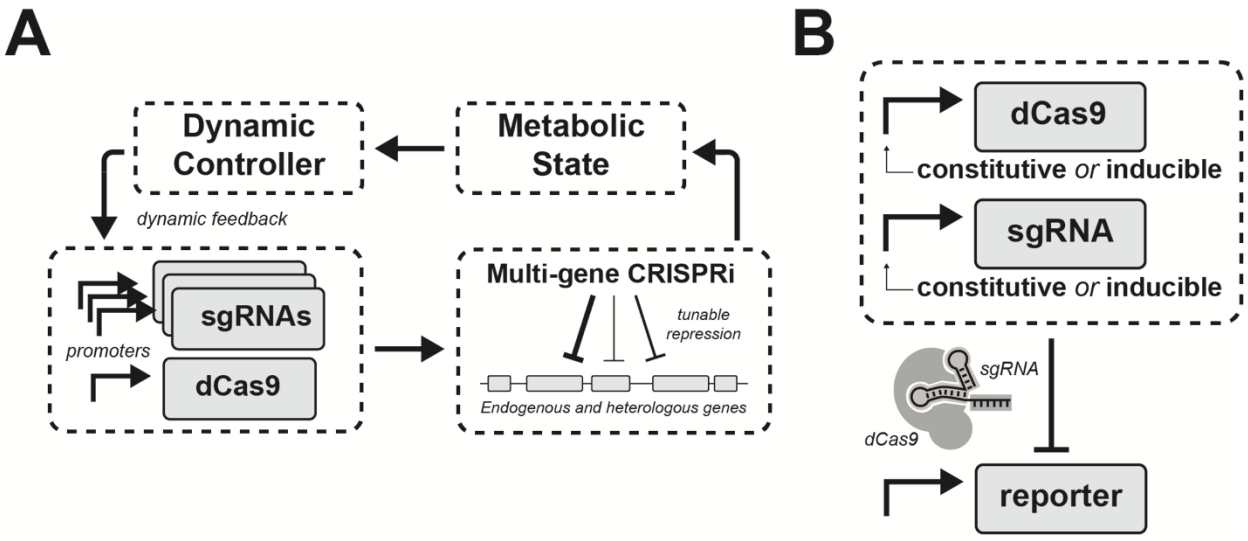


Figure 1. Schematic of dynamically-regulated multi-gene program for control over gene expression. A CRISPRi-based dynamic control system is engineered by linking one or more dynamic genetic controllers to dCas9 and/or sgRNAs. The CRISPRi transcriptional program controls the expression of multiple endogenous and heterologous genes, leading to a change in the metabolic state of the cell. Upon detecting changes in the concentration of their cognate ligand(s), the dynamic genetic controllers in turn regulate the expression of dCas9 and/or sgRNAs, resulting in tuning of the repression applied on the CRISPRi targets. B) Experimental setup used to quantify the relationship between *S. pyogenes* dCas9 and sgRNA expression levels and CRISPRi-mediated transcriptional repression. dCas9 and sgRNA were placed under the control of a combination of engineered constitutive and inducible promoters; CRISPRi repression was evaluated by monitoring the expression of integrated fluorescent reporters.

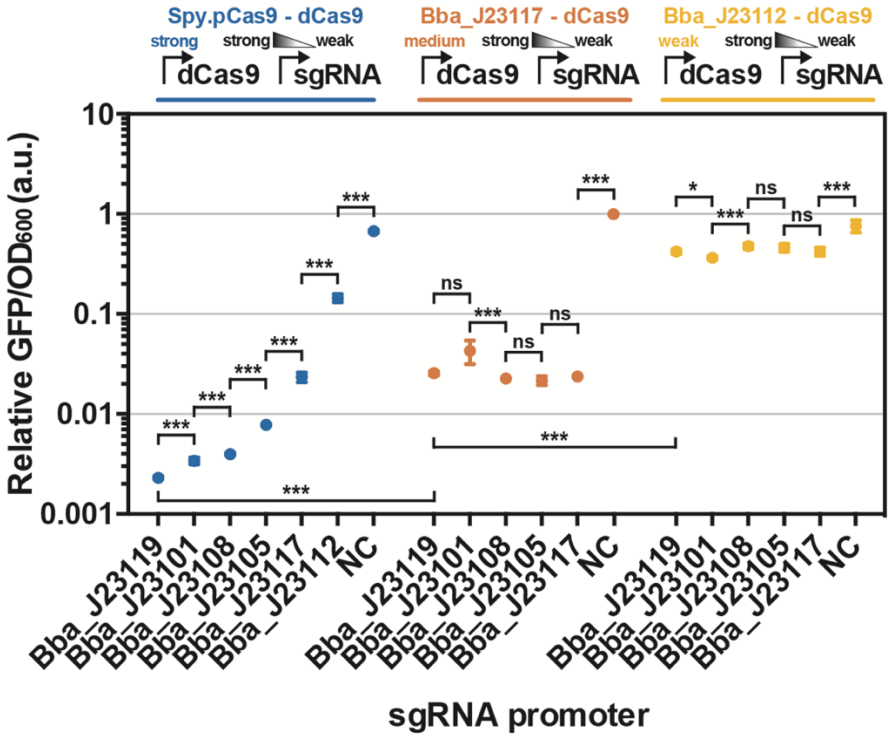


Figure 2. Tuning CRISPRi repression by titrating the expression of dCas9 and sgRNA using combinations of synthetic constitutive promoters.

dCas9 was controlled by one of three constitutive promoters (Spy.pCas9 [29], Bba_J23117, Bba_J23112 [parts.igem.org/Part:BBa_J23119], ordered left to right from strongest to weakest and color-coded; Spy.pCas9 was inferred to be the strongest based on the measured repression values). A sgRNA targeting sfGFP was controlled by one of six constitutive promoters (Bba_J23119 is the parent consensus promoter from which the other five promoters, ordered left to right from strongest to weakest, were derived [parts.igem.org/Part:BBa_J23119]); when we measured the expression of an integrated sfGFP reporter driven by Bba_J23119 and Bba_J23117, Bba_J23119 generated 340-fold more sfGFP expression than Bba_J23117 (not shown). The y-axis is the fluorescent intensity normalized to OD_{600} of stationary-phase *E. coli* cultures expressing an integrated sfGFP reporter and plasmid-borne CRISPRi components. Values are relative to the sample displaying the highest fluorescence and include baseline subtraction of the fluorescence/OD_{600} of *E. coli* MG1655 cultures. Refer to Figure S1 for corresponding fold repression values. NC: negative control not carrying any sgRNA. Data are plotted against the mean of three replicates with standard deviations. * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; ns: not significant (two-tailed t-tests).

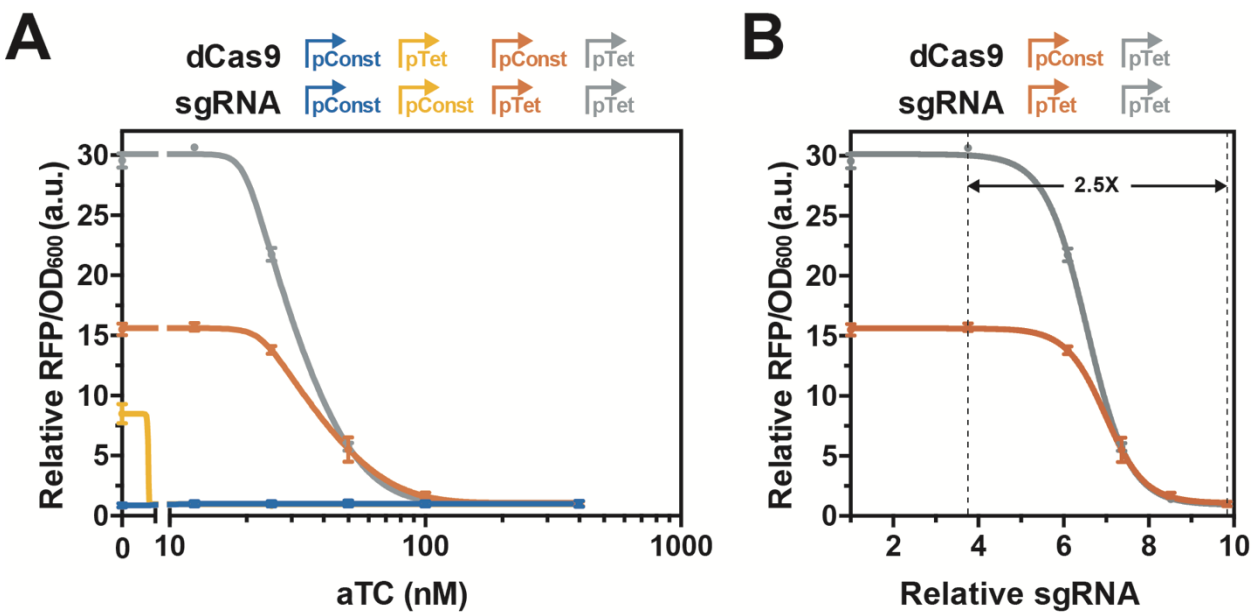


Figure 3. Analysis of the dynamic ranges in dCas9 and/or sgRNA inputs required for tuning CRISPRi.

A) Tuning of CRISPRi repression via inducible promoters. dCas9 was controlled by the constitutive promoter Spy.pCas9 [29] ("pConst") or an aTc-inducible promoter [30] ("pTet"). A sgRNA targeting an integrated mRFP1 reporter was controlled by the constitutive promoter Bba_J23119 [parts.igem.org/Part:BBa_J23119]) ("pConst") or an aTc-inducible promoter ("pTet"). All combinations were tested. Induction was performed at dilution at the final concentrations reported on the x-axis. The y-axis is the relative change in fluorescent intensity normalized to OD_{600} of induced and uninduced stationary-phase *E. coli* cultures expressing the integrated mRFP1 reporter and plasmid-borne CRISPRi components. Values are relative to the fluorescent intensity normalized to OD_{600} of each strain when induced with 400 nM aTc concentrations and include baseline subtraction of the fluorescence/OD_{600} of *E. coli* MG1655 cultures. B) Relationship between relative changes in sgRNA expression and expression of the CRISPRi target when the sgRNA or dCas9 and the sgRNA are regulated. Relative changes in the expression of the CRISPRi target are calculated in Panel A. Relative changes in sgRNA expression are calculated using data found in Figure S4. aTc: anhydrotetracycline. Data are plotted against the mean of three replicates with standard deviations.

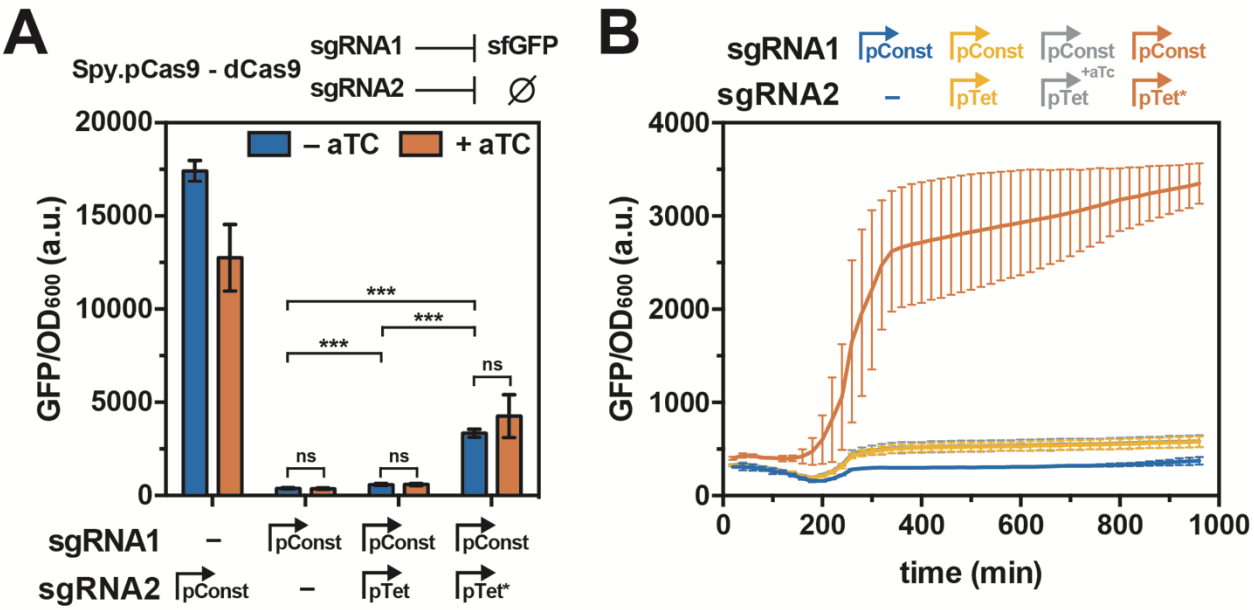


Figure 4. Persistence of CRISPRi repression upon expression of a competing sgRNA.

A) A sgRNA targeting an integrated sfGFP reporter was controlled by the constitutive promoter Bba_J23119 [parts.igem.org/Part:BBa_J23119]) ("pConst"). A second sgRNA without a genomic target was controlled by the constitutive promoter Bba_J23119 ("pConst"), an aTc-inducible promoter [30] ("pTet"), or the same promoter but not including the TetR repressor ("pTet*"). Induction of the second sgRNA was performed 4 h after dilution using a final concentration of 100 nM aTc. The y-axis is the fluorescent intensity normalized to OD_{600} of *E. coli* cultures expressing the integrated sfGFP reporter and plasmid-borne CRISPRi components grown for 12 h. dCas9 was controlled by the constitutive promoter Spy.pCas9 [29]. B) Time-course data of selected samples described in Panel A. aTc: anhydrotetracycline; "–": no sgRNA. Data are plotted against the mean of three replicates with standard deviations. * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; ns: not significant (two-tailed t-tests).

Chapter 4: Supplementary Information

Jason Fontana[1,4], Chen Dong[2,4], Jennifer Y. Ham[3,4], Jesse G. Zalatan[1,2,4], James M. Carothers[1,3,4]

[1] Molecular Engineering & Sciences Institute, University of Washington, Seattle, WA, 98195, USA

[2] Department of Chemistry, University of Washington, Seattle, WA 98195, USA

[3] Department of Chemical Engineering, University of Washington, Seattle, WA, 98195, USA

[4] Center for Synthetic Biology, University of Washington, Seattle, WA, 98195, USA

Supplementary tables

Table S1. *E. coli* plasmids included in this study.

Plasmid	Marker	Origin	Promoter	Gene	Terminator	Figure
pJF001	AmpR	ColE1	Bba_J23119	sgRNA (GFP)	TrrnB	2, 4, S1, S5
pJF002	AmpR	ColE1	1) TetR-pTet 2) Bba_J23119	1) sgRNA (RR2) 2) sgRNA (GFP)	1) TrrnB 2) TrrnB	4
pJF003	AmpR	ColE1	1) TetR-pTet 2) Bba_J23117	1) sgRNA (RR2) 2) sgRNA (GFP)	1) TrrnB 2) TrrnB	4
pJF009	AmpR	ColE1	n/a	n/a	n/a	2, S1
pJF015	AmpR	ColE1	Bba_J23117	sgRNA (GFP)	TrrnB	2, S1
pJF016	AmpR	ColE1	Bba_J23105	sgRNA (GFP)	TrrnB	2, S1
pJF017	AmpR	ColE1	Bba_J23108	sgRNA (GFP)	TrrnB	2, S1
pJF018	AmpR	ColE1	Bba_J23101	sgRNA (GFP)	TrrnB	2, S1
pJF020	AmpR	ColE1	1) pTet 2) Bba_J23119	1) sgRNA (RR2) 2) sgRNA (GFP)	1) TrrnB 2) TrrnB	4
pJF024	AmpR	p15A	TetR-pTet	sgRNA (GFP)	TrrnB	S5
pJF028	AmpR	ColE1	Bba_J23112	sgRNA (GFP)	TrrnB	2, S1
pJF043	ChlorR	p15A	n/a	n/a	n/a	S4
pCD016	AmpR	ColE1	Bba_J23119	sgRNA (RR2)	TrrnB	3
pCD017	ChlorR	p15A	Spy.pCas9	dCas9	db1	2, 3, 4, S1, S2, S3, S5
pCD032	AmpR	ColE1	TetR-pTet	sgRNA (RR2)	TrrnB	3, S2, S3
pCD035	ChlorR	p15A	Bba_J23112	dCas9	db1	2, S1
pCD048	ChlorR	p15A	Bba_J23117	dCas9	db1	2, S1
pSLQ1055[a]	ChlorR	p15A	TetR-pTet	dCas9	db1	3, S5
pBbE2A[b]	AmpR	ColE1	TetR-pTet	mRFP1	db1	S4

a: Reference [4]; b: Reference [30].

Table S2. *E. coli* strains included in this study

Strain	Description	Genotype
MG1655	parent *E. coli* strain	F- λ- ilvG- rfb-50 rph-1
CD02	MG1655/sfGFP	MG1655 *Bba_J23119-sfGFP KanR::nfsA*
JF01	MG1655/mRFP1	MG1655 *Bba_J23119-mRFP1 KanR::rbsAR*

Table S3. sgRNA sequences included in this study.

sgRNA name	Target gene	DNA Sequence	Target Strand
RR2[a]	mRFP1	TGGAACCGTACTGGAACTGC	Non-template
GFP[a]	sfGFP	CATCTAATTCAACAAGAATT	Non-template

a: Reference [4]

Table S4. Dynamic ranges in gene expression generated by selected available dynamic genetic controllers.

Controller	Class	Mode of induction	Developed in	DR[a]
glnAp2[b]	TF and promoter	Accumulation of acetyl-phosphate	*E. coli* BW13711 ΔgnlL	12
S-P promoters[c]	Promoter	Entry to stationary phase	E. coli K12 DH5α	4-6800
FadR-pAR[d]	TF and promoter	Fatty-acid/acyl-coA relieve FadR repression	*E. coli* DH1 ΔfadE	25-60
aREDs[e]	Ligand- dependent ribozymes	Ligand binding decreases RNA degradation	*E. coli* BL21 (DE3)	2
Synthetic riboswitches[f]	Aptamer based translational controllers	Ligand binding activates translation	*E. coli* DH10B	Up to 383

a: Dynamic range; b: Reference [37]; c: Stationary-phase promoters, reference [38]; d: Reference [23]; e: Aptazyme-regulated expression devices, reference [39]; f: Reference [40].

Supplementary Figures

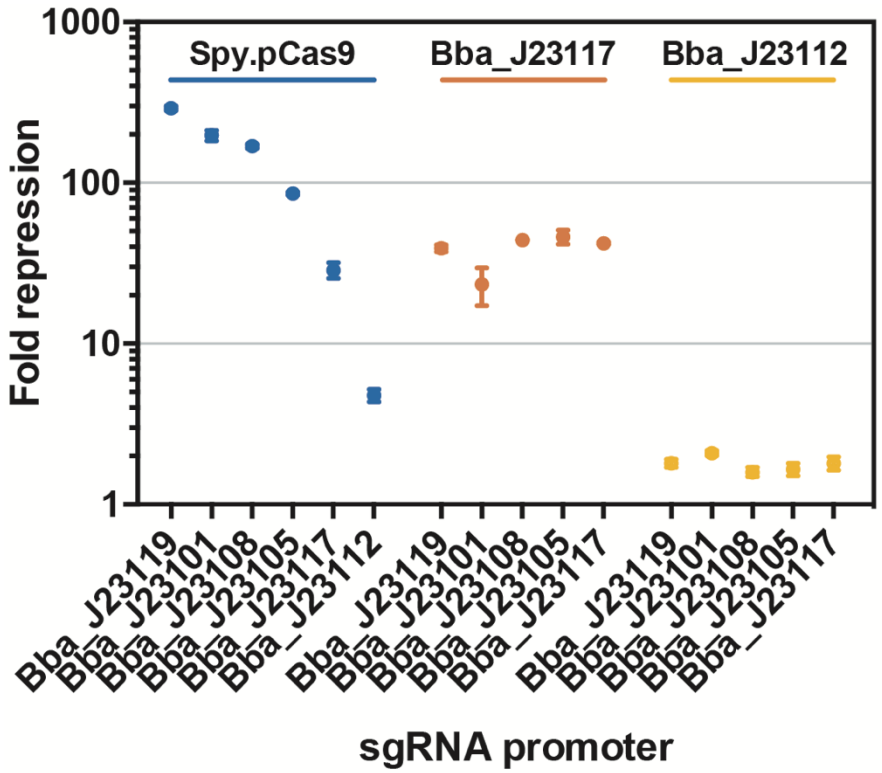


Figure S1. Fold repression when titrating CRISPRi by regulating the expression of dCas9 and sgRNAs using combinations of synthetic constitutive promoters.

dCas9 was controlled by one of three constitutive promoters (Spy.pCas9 [29], Bba_J23117, Bba_J23112 [parts.igem.org/Part:BBa_J23119], ordered left to right from strongest to weakest and color-coded; Spy.pCas9 was inferred to be the strongest based on the measured repression values). A sgRNA targeting sfGFP was controlled by one of six constitutive promoters (Bba_J23119 is the parent consensus promoter from which the other five promoters, ordered left to right from strongest to weakest, were derived [parts.igem.org/Part:BBa_J23119]). On the y-axis is the fold repression of stationary-phase *E. coli* cultures expressing an integrated sfGFP reporter and plasmid-borne CRISPRi components. Fold repression was calculated by dividing the fluorescence intensity normalized to OD_{600} of *E. coli* cultures of each condition by that of a negative control not carrying a sgRNA. The baseline fluorescence/OD_{600} of *E. coli* MG1655 cultures was subtracted to both values prior to division. NC: negative control not carrying any sgRNA. Data are plotted against the mean of three replicates with standard deviations.

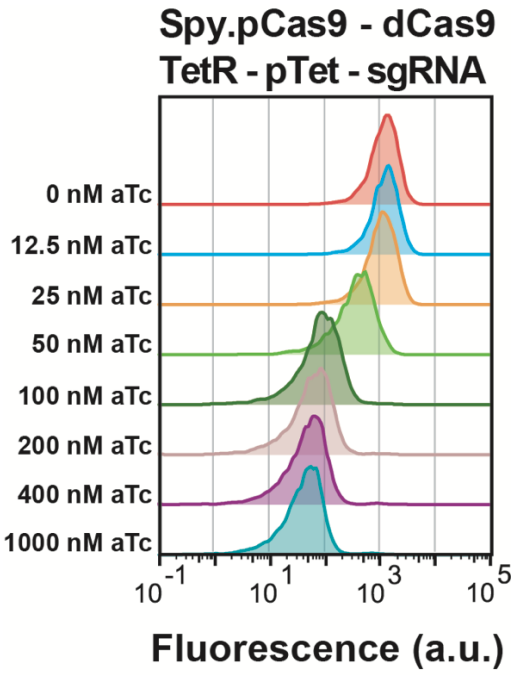


Figure S2. Population distribution of CRISPRi repression upon titration of sgRNA expression. Reported are the flow cytometry traces of induced and uninduced stationary-phase *E. coli* cultures expressing an integrated mRFP1 reporter and plasmid-borne CRISPRi components. dCas9 was controlled by the constitutive promoter Spy.pCas9 [29]. A sgRNA targeting an integrated mRFP1 reporter was controlled by an aTc-inducible promoter [30]. The distribution of red fluorescence intensity of one of three replicates for each condition is shown.

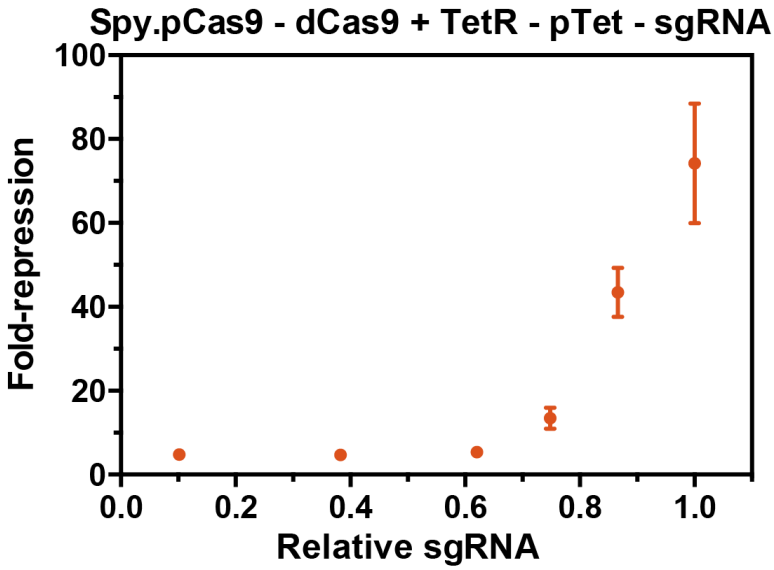


Figure S3. Fold-repression when titrating sgRNA expression a pTet inducible promoter.
dCas9 was controlled by the constitutive promoter Spy.pCas9 and the sgRNA was controlled by a pTet promoter [30]; each data point on the x-axis represents a different inducer concentration, and relative changes in sgRNA expression were estimated in Figure S4. On the y-axis is the fold repression of an integrated fluorescent reporter upon expression of the CRISPRi components, calculated as described in Figure S1. Data are plotted against the mean of three replicates with standard deviations.

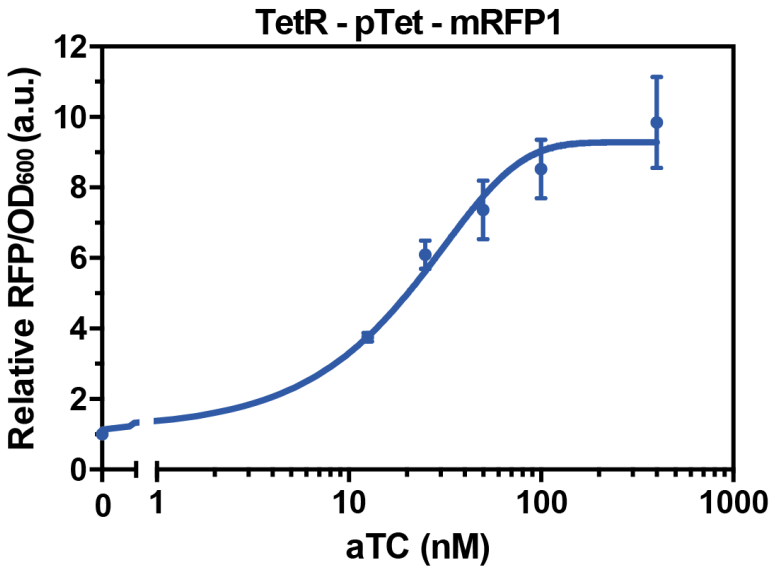


Figure S4. Dynamic range in gene expression provided by the aTc-inducible promoter.

A mRFP1 reporter was placed under the control of a pTet promoter [30]. *E. coli* cultures harboring this construct were induced at dilution with the final aTc concentrations reported on the x-axis. The y-axis is the relative fluorescence intensity normalized to OD_{600} of stationary-phase cultures. The baseline fluorescence/OD_{600} of wild-type *E. coli* MG1655 cultures was subtracted to all values. Values are relative to the uninduced condition. aTc: anhydrotetracycline. Data are plotted against the mean of three replicates with standard deviations.

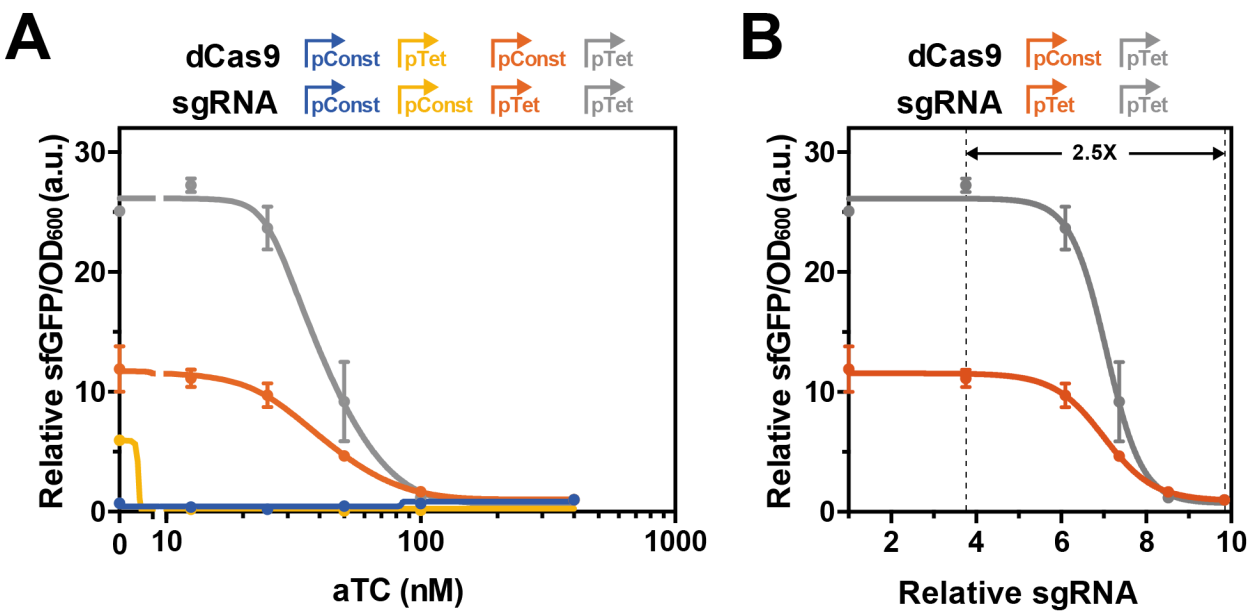


Figure S5. The window of sgRNA expression where repression can be titrated is similar when targeting a different gene.

The experiment described in Figure 3 was repeated when targeting sfGFP instead of mRFP1. A) Tuning of CRISPRi repression via inducible promoters. dCas9 was controlled by the constitutive promoter Spy.pCas9 [29] ("pConst") or an aTc-inducible promoter [30] ("pTet"). A sgRNA targeting an integrated sfGFP reporter was controlled by the constitutive promoter Bba_J23119 [parts.igem.org/Part:BBa_J23119]) ("pConst") or an aTc-inducible promoter ("pTet"). All combinations were tested. Induction was performed at dilution at the final concentrations reported on the x-axis. The y-axis is the relative change in fluorescent intensity normalized to OD_{600} of induced and uninduced stationary-phase *E. coli* cultures expressing the integrated sfGFP reporter and plasmid-borne CRISPRi components. Values are relative to the fluorescent intensity normalized to OD_{600} of each strain when induced with 400 nM aTc concentrations and include baseline subtraction of the fluorescence/OD_{600} of *E. coli* MG1655 cultures. B) Relationship between relative changes in sgRNA expression and expression of the CRISPRi target when the sgRNA or dCas9 and the sgRNA are regulated. Relative changes in the expression of the CRISPRi target are calculated in Panel A. Relative changes in sgRNA expression are calculated using data found in Figure S4. aTc: anhydrotetracycline. Data are plotted against the mean of three replicates with standard deviations.

Chapter 5:
Challenges and opportunities with CRISPR activation in bacteria for data-driven metabolic engineering

Jason Fontana[*,1], David Sparkman-Yager[*,1], Jesse G. Zalatan[2], and James M. Carothers[3]

*: these authors contributed equally

[1] Molecular Engineering & Sciences Institute and Center for Synthetic Biology
[2] Department of Chemistry
[3] Department of Chemical Engineering
University of Washington
Seattle, WA 98195
United States

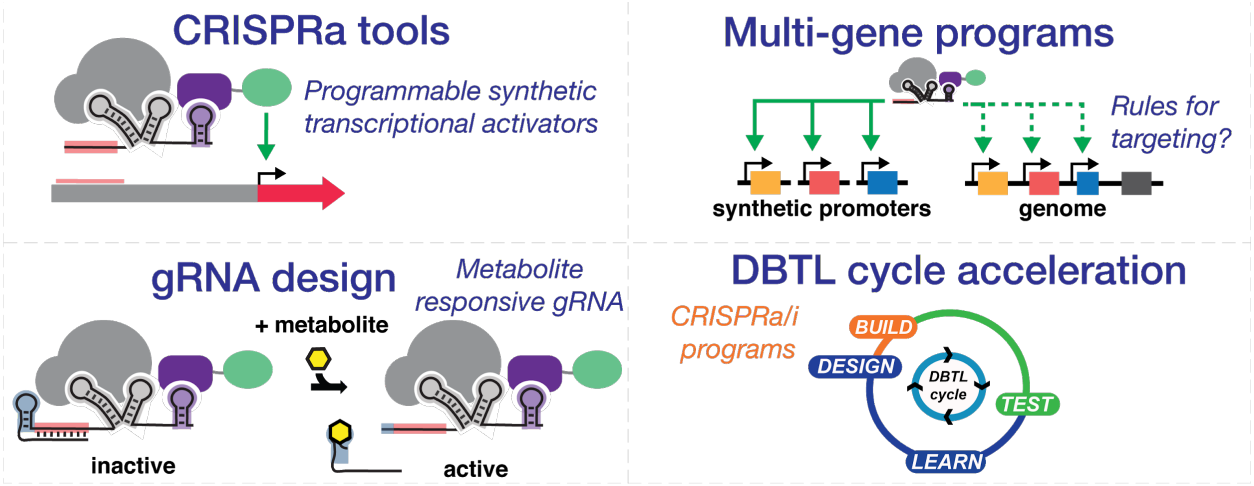
Highlights

- CRISPR activation is an emerging tool for engineering bacterial metabolism.
- The promoter design rules for effective CRISPR activation in bacteria are complex.
- Guide RNAs can be engineered as flexible platforms to program CRISPR activation.
- CRISPR activation may accelerate strain optimization Design-Build-Test-Learn cycles.

Abstract

Creating CRISPR gene activation (CRISPRa) technologies in industrially-promising bacteria could be transformative for accelerating data-driven metabolic engineering and strain design. CRISPRa has been widely used in eukaryotes, but applications in bacterial systems have remained limited. Recent work shows that multiple features of bacterial promoters impose stringent requirements on CRISPRa-mediated gene activation. However, by systematically defining rules for effective bacterial CRISPRa sites and developing new approaches for encoding complex functions in engineered guide RNAs, there are now clear routes to generalize synthetic gene regulation in bacteria. When combined with multi-omics data collection and machine learning, the full development of bacterial CRISPRa will dramatically improve the ability to rapidly engineer bacteria for bioproduction through accelerated design-build-test-learn cycles.

Graphical abstract



Introduction

Bacterial metabolism is made up of complex gene networks that can be engineered to produce medically- and industrially-important chemicals. The complexity of these networks means that sophisticated organism engineering efforts are typically needed to optimize the production of high-value compounds [1,2]. In principle, synthetic multi-gene transcriptional programs could be constructed to engineer metabolic networks for efficient industrial chemical production [3,4]. In practice, however, an incomplete understanding of metabolic networks, combined with a limited ability to predictably control the expression of multiple genes, makes achieving this goal difficult [5]. A recurring challenge when engineering strains for chemical production remains the difficulty of predicting the optimal expression level of pathway and non-pathway genes that will result in optimal yields. To overcome this challenge, there is a need for new technologies for rapidly implementing and analyzing combinatorial multi-gene expression programs. These technologies could be combined with advanced capabilities for multi-omics data collection and machine learning to enable accelerated design-build-test-learn cycles (DBTL) [1,6,7].

CRISPR-Cas tools have changed every aspect of microbial engineering, including the speed with which genomes can be edited and the ability to target specific genes for activation and repression [8,9]. CRISPR-Cas tools for programming gene expression use the catalytically-inactive Cas9 protein (dCas9) along with guide RNAs that recognize DNA targets through predictable Watson-Crick base pairing [9]. A major strength of CRISPR-based synthetic gene regulation is that new combinatorial multi-gene expression programs that include simultaneous transcriptional activation and repression could be rapidly implemented.

There are well-established approaches for repressing genes (CRISPRi) by targeting dCas9 to physically block RNA polymerase and inhibit transcription [10]. In eukaryotic cells, robust transcriptional activation can be applied using CRISPR-Cas to direct activation domains upstream of target genes (CRISPRa) [8,9]. However, the development of CRISPRa in bacteria has been hindered by the lack of effective activation domains. The recent discovery that at least four different bacterial activators can be linked to programmable CRISPR-Cas DNA binding domains has promised to significantly change the outlook for CRISPRa in bacteria [11–14]. Further, new efforts to uncover practical rules for activating transcription with bacterial CRISPRa may make it possible to build complex multi-gene programs that regulate the expression of both heterologous and endogenous genes. By building on recent efforts, the further development of engineered guide RNAs as flexible platforms for programming CRISPRa may create new capabilities for predictable and metabolite-responsive synthetic gene regulation [15,16]. This review focuses on new advancements in bacterial CRISPRa technologies that promise to significantly accelerate strain optimization through data-driven metabolic engineering.

CRISPRa for regulating bacterial transcription

In bacteria, the implementation of complex multi-gene CRISPR-Cas expression programs has been limited by a lack of effective gene activators. To address this problem, new synthetic transcriptional activators have been developed in *E. coli* that link activation domains to programmable CRISPR-Cas DNA binding domains

[11–14]. The resulting CRISPRa tools have proven capable of driving heterologous gene expression at levels suitable for metabolic engineering. Some successes have also been achieved in activating the expression of endogenous genes from genomic loci. The further development of these capabilities may permit the optimization of metabolic production through the construction of multi-gene programs simultaneously targeting heterologous and endogenous genes.

There are two mechanistic approaches that have been employed to link activation domains to CRISPR-Cas DNA binding domains. Activation domains can be (i) directly fused to dCas9, or (ii) recruited to dCas9 using modified sgRNAs (scaffold RNAs or scRNAs) that bind to RNA binding protein-activation domain fusions [17–19]. Using the first approach, CRISPRa has been achieved in *E. coli* by fusing the ω-subunit of RNA polymerase (*rpoZ*) to dCas9 to obtain 23-fold increases in reporter gene expression from synthetic promoters [11]. These fusions have been applied both in *E. coli* and non-model bacteria. In *E. coli*, dCas9-RpoZ was used activate transcription and identify genes that increase tolerance to the monoterpene pinene, as well as new epistatic interactions between antibiotic resistance genes [20,21]. These tools were successfully ported to *B. subtilis* to obtain 3-fold activation of reporter gene expression and applied to systematically improve production of amylase BLA by 260-fold compared to a commonly used strong promoter [22]. In *L. enzymogenes*, dCas9-RpoZ was used to enhance production of anti-MRSA antibiotics up to 9-fold [23]. In *M. xanthus*, dCas9-RpoZ was able to generate 8-fold increases in the expression of the epothilone production gene cluster, leading to a 6.8-fold improvement in epothilone A production [24]. Finally, a new portable CRISPRa system where the activation domain AsiA was fused to dCas9 was recently introduced [14]. Using this system, reporter gene expression could be activated by 135-fold in *E. coli*, ~3-fold in *S. enterica,* and ~12-fold in *K. oxytoca.*

The second approach for bacterial CRISPRa relies on modified gRNAs (scRNAs) that recruit an RNA binding protein fused to an activation domain to the CRISPRa complex. One successful strategy uses the RNA binding protein MCP fused to the SoxS activation domain (MCP-SoxS) (Figure 1) [12]. Using MCP-SoxS and a corresponding MS2 scRNA, 50-fold CRISPRa activation was demonstrated and applied to drive ethanol production in *E. coli* from a *Z. mobilis* gene cluster. MCP-SoxS can activate expression from genes that use σ factors σ^{70} [12] and, at lower levels, σ^{38}, σ^{32}, and σ^{24} [25]. While together these σ^{70}-family promoters cover the majority of the *E. coli* genome, σ^{54} promoters, which drive nitrogen starvation genes, could not be activated by MCP-SoxS. Recently, an alternative bacterial CRISPRa system that is effective at σ^{54} promoters was introduced based on the PspFΔHTH::λN22plus activator [13]. Therefore, PspFΔHTH::λN22plus and MCP-SoxS can be used in combination to target a different, non-overlapping set of promoters in *E. coli*. Further, PspFΔHTH::λN22plus was reported to activate two promoters in the nitrogen fixation pathway of *K. oxytoca* by up to 6-fold [13]. In the dCas9-RpoZ and PspFΔHTH::λN22plus systems, obtaining the highest levels of activation requires knocking out the native copy of *rpoZ* [11,26] or *pspF* [13], respectively, to remove the competing, endogenous functions. It is possible, however, to obtain significant activation without using knockout strains. In contrast, the MCP-SoxS and dCas9-AsiA systems do not require any host engineering to achieve their highest levels of activation.

The available CRISPRa tools are uniquely positioned to rapidly implement combinatorial multi-gene expression programs targeting synthetic promoters and identify optimal expression conditions for metabolite production [27]. These tools were recently applied in a proof-of-concept experiment to tune the expression of three genes in the pathway responsible for producing violacein, a pigment with antitumoral properties [13]. Further improving our ability to predictably tune CRISPRa at multiple sites independently could provide a technology for the rapid combinatorial optimization of multi-gene pathways. Dynamically-controlled CRISPRi was recently shown to improve production of salicylic acid in engineered *E. coli* through the conditional knock-down of essential genes [28]. Developing dynamically-controlled CRISPRa could provide additional avenues to control both the timing and expression levels of multiple genes in engineered metabolic pathways and networks [3].

Promoter design rules improve CRISPRa in bacteria

Recent work has identified multiple features of bacterial promoters that impose stringent requirements on CRISPRa-mediated gene activation [25] (Figure 1). These behaviors suggest an explanation for why CRISPRa and other tools for gene activation in bacteria have lagged far behind comparable tools in eukaryotic systems, where such strict target site requirements are absent. For instance, the activity of CRISPRa using MCP-SoxS is influenced by the strength of the target promoter, the sigma factor regulating the promoter and the sequence composition of the promoter [25]. Most strikingly, when activating synthetic promoters in bacteria, CRISPRa is sensitive to the position and periodicity of the scRNA target site relative to the transcription start site (TSS) [13,25]. Activation can only be performed at precisely defined positions in phase with the transcription start site, which are intervened by regions of lower activity or inactivity [13,25]. These requirements are much more stringent than those for activation in eukaryotic cells [17] and constrain CRISPRa to precisely-positioned PAM sites which may not be found on every gene. Engineered Cas9 variants and alternative Cas proteins have been introduced that expand the range of PAM sequences that can be targeted and increase the density of available PAM sites up to 6 times [25,29–31]. One of the variants, dxCas9(3.7), has been used to demonstrate activation of *E. coli* genes previously inaccessible by dCas9 [13,25]. By combining dxCas9(3.7) and newly defined rules for CRISPRa, 3 out of 7 endogenous *E. coli* genes were successfully activated [25]. However, the field still lacks integrated models for predicting effective CRISPRa target sites for arbitrary genes, and explanations for the failure to activate some genes remain elusive. Genome-wide CRISPRa screens of endogenous promoters could more fully elucidate the requirements for CRISPRa targeting. Once predictive rules for targeting endogenous genes are available, combinatorial multi-gene programs for optimizing bioproduction could be extended to endogenous genes, in addition to synthetic promoters.

gRNAs can be engineered to program CRISPRa responses

gRNA engineering has long been understood to provide routes for tuning CRISPR-Cas functions, and more recently, as a mechanism for encoding dynamic responses to molecular targets. While most of

the gRNA design work to date has been performed on guides used for DNA cleavage or CRISPRi, the principles, whether controlling the stability of the guide-Cas9 complex or the entire DNA-guide-Cas9 complex, may be readily applicable to CRISPRa efforts. Guide RNAs are comprised of two main components: the twenty nucleotide spacer sequence, which hybridizes to the target DNA, and the Cas9-binding handle, which drives the formation of the gRNA-Cas9 complex. Structurally, the only difference between a gRNA and a scRNA used for CRISPRa is the presence of an additional 3' RNA hairpin. This hairpin enables the co-localization of the activation domain by binding its cognate RBP tag. While this motif adds additional complexity, it also provides more opportunities for design. Several alternative architectures have also been described that insert the RNA hairpin motif at multiple points within the guide [32,33]. To date, three cognate pairs of RNA binding protein (RBP) and RNA hairpin have been utilized to implement CRISPRa in bacteria (Figure 2) [12,13]. Other pairs have been demonstrated in eukaryotes and may be functional in bacteria as well [18,34]. These orthogonal pairs provide the opportunity to simultaneously implement multiple activators in the same cell.

While the relative simplicity of gRNAs allows the rational specification of target sites, the sequence identity of a chosen site has been demonstrated to have a significant impact on its CRISPR activity. Both the sequence identity and secondary-structure (base-pairing interactions) of gRNA elements are critical for function [35]. There have been several attempts to predict CRISPR activity for novel target sequences, and while these models can be used to increase the probability of selecting a functional guide, they primarily utilize sequence elements, rather than structural information, and have not been applied directly to CRISPRa [36–38]. One key feature that has been demonstrated to influence CRISPR activity is the secondary structure that the guide RNA adopts [39,40]. The degree of secondary structure, whether internal to the spacer or between the spacer and the rest of the guide, has been observed to reduce gRNA effectiveness (Figure 2) [39]. Even the transiently-stable structures the guides adopt during transcription have been demonstrated to impact their activity [39]. As the guides utilized for CRISPRa are actively transcribed from heterologous promoters inside the cell, avoiding transient misfolding may prove important for achieving predictable activity. Thus, developing tools for screening gRNA co-transcriptional folding pathways may aid in the *a priori* selection of highly functional spacer sequences.

To generate differences in the expression levels of multiple genes, it is necessary to develop a general strategy to fine-tune CRISPRa-mediated gene expression at each promoter. To date, two main strategies have been demonstrated to modulate the CRISPR activity for a given target sequence: spacer truncations and 5' extension. In CRISPRi systems, the level of transcriptional repression applied to target genes has been reduced by truncating the sgRNA target sequence from the 5' end [10,41]. Practically, spacers shorter than 12 nucleotides may increase off-target activity as the first 12 nt, and even more so the first 7 nt known as the 'seed region', have an especially large impact on the activity [10,42]. Alternatively, it has been demonstrated that adding a 5' extension onto the guide, which folds back to occlude the spacer, results in monotonic drops in guide activity with increasing stability of the designed interaction; this correspondence even applies to guides from other Cas proteins with different guide architectures [43]. This

demonstration provides evidence that computational predictions of gRNA structure may be sufficient to predict guide function. In order to improve the forward engineering of CRISPRa systems and accelerate DBTL cycles, developing quantitatively-accurate predictions of CRISPRa activity based on scRNA structure will be essential.

Towards nucleic acid-responsive gRNAs for CRISPRa

In order to implement complex genetic and metabolic circuits, it becomes necessary to be able to link the intracellular concentration of target molecules to the regulatory circuit being implemented. One such implementation would be to use the levels of cellular RNAs to regulate the activity of CRISPR-based transcriptional programs. To that end, there have been several demonstrations that the activity of a gRNA can be regulated by the presence of target nucleic acid sequences that hybridize to the gRNA [44]. While there have been slightly different implementations, the general principle is that a trans-acting 'trigger' strand is able to bind to a gRNA and either occlude or reveal the spacer sequence, modulating the guide's activity. One of the most common mechanisms is inspired by a previously published 'toehold switch', in which a cis-repressed gRNA is activated upon toehold-mediated hybridization (Figure 2) [45]. Several gRNA switches have even demonstrated the ability to respond to RNA trigger strands within a cell to control gene expression levels [46–49]. For example, 'toehold-gated' sgRNAs (thgRNAs) were capable of inducing CRISPRi in response to endogenous small RNAs (sRNAs) and mRNAs in *E. coli*, with repression up to 5-fold [46]. However, half of the thgRNAs responsive to endogenous RNAs resulted in low levels of repression (< 2-fold). Unlike short synthetic RNA trigger strands, longer endogenous RNAs may be less effective as trigger strands due to competition for binding from intramolecular RNA structure and cellular proteins. Improving the activity of thgRNAs responsive to cellular RNAs will require advancements in the *a priori* identification of sites within cellular RNAs that can be utilized as highly-active trigger strands. Furthermore, in order to apply these mechanisms to CRISPRa, it will be necessary to ensure that the interaction between a trigger-responsive scRNA and a large cellular RNA does not itself interfere with the mechanism of CRISPRa activation.

Metabolite-responsive gRNAs for CRISPRa

In addition to regulation by cellular RNAs, the ability to regulate CRISPR activity in response to real-time concentrations of cellular metabolites would provide many opportunities for implementing and accelerating DBTL cycles for metabolic engineering. Metabolite responsive gRNAs could be used for both readouts of intracellular metabolite concentrations to inform machine learning models, or for implementing model-suggested regulation such as feedback or feedforward motifs. While efforts to design metabolite-responsive gRNAs are fairly new, metabolite-responsive RNAs have become useful tools in metabolic engineering [50,51]. By combining a metabolite-binding RNA aptamer with a control structure, the binding state of the aptamer can be converted into a conditional genetic output.

There have been several demonstrations that small molecule responsive gRNA activity can be

dynamically regulated with an aptamer in *cis* [15,16,52,53]. Some demonstrations involve inserting the aptamer at the 3' end of the guide, where it stabilizes the active gRNA structure in a ligand-responsive manner [52,53]. However, adding both an aptamer and a recruitment hairpin to the 3' end of the RNA could interfere with gRNA folding and function. Other strategies, which utilize 5' extension or Cas9 handle insertion, may therefore be preferable. For example, an aptazyme, or ligand-responsive self-cleaving ribozyme, was used to remove a repressive 5' extension from the guide upon the addition of the target ligand (Figure 2) [16]. This resulted in ligand-responsive control over both Cas9-mediated cleavage and CRISPRa in mammalian cells. In another example, aptamers were inserted into the Cas9-binding handle, or one of two other gRNA hairpins, generating ligand-responsive CRISPRi in *E. coli* (Figure 2) [15]. Depending on the aptamer insertion site within the guide, the addition of ligand can either activate or deactivate CRISPRi.

The above successes in identifying ligand-responsive gRNAs provide great confidence that it will be possible to engineer small molecule-responsive scRNAs for conditional CRISPRa. However, there are still hurdles to overcome before metabolite-responsive CRISPRa can be used effectively for metabolic engineering applications. First, design rules that allowing reliable integration of aptamers with diverse sequences and secondary structures into scRNAs must be uncovered. Second, mechanisms for tuning the response to match the desired metabolite concentrations must be developed [54]. Aptamer-regulated kinetic control mechanisms, similar to those found in natural bacterial riboswitches, may provide an approach for engineering metabolite-responsive CRISPRa targeted to specific concentrations of metabolites [55,56]. For example, 10-fold variations in switching concentration among a family of *E. coli* thiamine pyrophosphate (TPP) riboswitches are known to be the result of differences in the amount of time the RNA is available to interact with the ligand [56]. Creating kinetically-controlled aptamer-regulated scRNAs may confer the ability to engineer metabolite-responsive CRISPRa systems functional as feedback controllers, or production biosensors useful for optimizing strain performance.

Conclusions

The relationships between the expression levels and reaction kinetics for enzymes in both endogenous and engineered metabolic networks are poorly understood. This incomplete knowledge constitutes a major limitation for the field of metabolic engineering [1]. Because of these gaps, data-driven methods relying on cycles of genetic engineering, high-throughput production screening, multi-omics analysis, and machine learning have become increasingly central to strain optimization [57,58]. To accelerate data-driven metabolic engineering, methods to independently target and predictably manipulate the expression levels of multiple genes are needed. By coupling new tools for CRISPRa with existing approaches for CRISPRi, it should be possible to more efficiently search gene expression spaces and optimize bioproduction in engineered bacteria through accelerated DBTL cycles (Figure 3).

CRISPRa can now be used to selectively activate synthetic promoters with large dynamic ranges and in a way that is relatively straightforward to implement. Recent advances in gRNA design have enabled

the identification of small-molecule responsive gRNAs able to dynamically-regulate gene expression in *E. coli*, opening the door for the development of CRISPRa-based metabolite biosensors and circuit controllers. The ability to use CRISPRa to activate endogenous genes remains limited by the sequence constraints of the native genomic loci, where less-than-optimal position of PAM site or the inherent features of the promoters can significantly impact the activation that can be achieved. Predictive models are needed to identify which endogenous genes can be activated and which target sites are the most effective. The refinement of sequence and structure-based rules for constructing synthetic promoters and cognate scRNAs for expressing heterologous genes will improve the ability to precisely tune multi-gene pathways. CRISPRa has been demonstrated in *E. coli* [11–14,20,21,25] and other industrially and medically relevant bacteria including *B. subtilis* [22], *K. oxytoca* [13,14], *L. enzymogenes* [23], *M. xhantus* [24] and *S. enterica* [14]. Porting these tools to other non-model bacteria with diverse substrate utilization, a range of metabolic capabilities, and resistance to harsh bioprocessing conditions could accelerate the development of efficient bioproduction processes. Collectively, these strategies lay the groundwork for more widespread use of bacterial CRISPRa in basic research and advanced applications in data-driven metabolic engineering.

Acknowledgments

The authors thank Joely Nelson and members of the Zalatan and Carothers groups for technical assistance and helpful discussions. This work was supported by NSF Award 1817623 (J.G.Z., J.M.C.) and NSF Award 1844152 (J.M.C.).

Reference Highlights

Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG: **Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria**. *Nat Commun* 2018, **9**:2489. This paper introduced a new bacterial CRISPRa system capable of generating high levels of activation at σ^{70} promoters. This work demonstrated the first use of scaffold RNAs in bacteria and their use for simultaneous CRISPRa and CRISPRi.

Liu Y, Wan X, Wang B: **Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria**. *Nat Commun* 2019, **10**:3693. This work describes a bacterial CRISPRa system that activates σ^{54} promoters using a new activation domain based on the PspF activator, expanding the range of promoters that can be targeted with CRISPRa.

Kundert K, Lucas JE, Watters KE, Fellmann C, Ng AH, Heineike BM, Fitzsimmons CM, Oakes BL, Qu J, Prasad N, et al.: **Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs**. *Nat Commun* 2019, **10**:1–11. This paper demonstrates a strategy for isolating ligand-activated and -repressed guide RNAs in bacteria. The authors show that different aptamers and different insertion sites within the guide can be successfully utilized in the same screening methodology. It provides a potential roadmap for designing metabolite-responsive guides for CRISPR activation.

Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF: **Internal guide RNA interactions interfere with Cas9-mediated cleavage**. *Nat Commun* 2016, **7**:1–7. This paper illuminates the significant role of RNA structure in determining guide RNA activity. The authors demonstrate that even nonfunctional guide RNAs can degrade the performance of co-expressed guides, and highlight the rarely-discussed impact that co-transcriptional folding can have on guide activity.

Tian T, Kang JW, Kang A, Lee TS: **Redirecting Metabolic Flux via Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in Escherichia coli**. *ACS Synth Biol* 2019, **8**:391–402. The authors show that a system of multiplexed CRISPRi repression can be applied to endogenous genes to increase isopentanol precursor availability. The framework described in this paper could be integrated with CRISPRa to apply multiplexed up and down-regulation of genes for strain optimization.

References

1. Nielsen J, Keasling JD: **Engineering Cellular Metabolism**. *Cell* 2016, **164**:1185–1197.

2. Lee SY, Kim HU, Chae TU, Cho JS, Kim JW, Shin JH, Kim DI, Ko Y-S, Jang WD, Jang Y-S: **A comprehensive metabolic map for production of bio-based chemicals**. *Nat Catal* 2019, **2**:18–33.

3. Fontana J, Voje WE, Zalatan JG, Carothers JM: **Prospects for engineering dynamic CRISPR–Cas transcriptional circuits to improve bioproduction**. *J Ind Microbiol Biotechnol* 2018, **45**:481–490.

4. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY: **Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering**. *Trends in Biotechnology* 2019, **37**:817–837.

5. Liu R, Bassalo MC, Zeitoun RI, Gill RT: **Genome scale engineering techniques for metabolic engineering**. *Metabolic Engineering* 2015, **32**:143–154.

6. Carbonell P, Jervis AJ, Robinson CJ, Yan C, Dunstan M, Swainston N, Vinaixa M, Hollywood KA, Currin A, Rattray NJW, et al.: **An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals**. *Commun Biol* 2018, **1**:1–10.

7. Brunk E, George KW, Alonso-Gutierrez J, Thompson M, Baidoo E, Wang G, Petzold CJ, McCloskey D, Monk J, Yang L, et al.: **Characterizing Strain Variation in Engineered E. coli Using a Multi-Omics-Based Workflow**. *cels* 2016, **2**:335–346.

8. Wang H, La Russa M, Qi LS: **CRISPR/Cas9 in Genome Editing and Beyond**. *Annual Review of Biochemistry* 2016, **85**:227–264.

9. Dominguez AA, Lim WA, Qi LS: **Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation**. *Nature Reviews Molecular Cell Biology* 2016, **17**:5–15.

10. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA: **Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression**. *Cell* 2013, **152**:1173–1183.

11. 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**:7429–7437.

12. Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG: **Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria**. *Nature communications* 2018, **9**:2489.

13. Liu Y, Wan X, Wang B: **Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria**. *Nature Communications* 2019, **10**:3693.

14. Ho H-I, Fang J, Cheung J, Wang HH: **Programmable and portable CRISPR-Cas transcriptional activation in bacteria**. *bioRxiv* 2020, doi:10.1101/2020.01.03.882431.

15. Kundert K, Lucas JE, Watters KE, Fellmann C, Ng AH, Heineike BM, Fitzsimmons CM, Oakes BL, Qu J, Prasad N, et al.: **Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs**. *Nat Commun* 2019, **10**:1–11.

16. Tang W, Hu JH, Liu DR: **Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation**. *Nature Communications* 2017, **8**:15939.
17. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al.: **Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation**. *Cell* 2014, **159**:647–661.
18. Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, et al.: **Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds**. *Cell* 2015, **160**:339–350.
19. Mali P, Esvelt KM, Church GM: **Cas9 as a versatile tool for engineering biology**. *Nature Methods* 2013, **10**:957–963.
20. Niu F-X, Huang Y-B, Ji L-N, Liu J-Z: **Genomic and transcriptional changes in response to pinene tolerance and overproduction in evolved Escherichia coli**. *Synthetic and Systems Biotechnology* 2019, **4**:113–119.
21. Otoupal PB, Erickson KE, Escalas-Bordoy A, Chatterjee A: **CRISPR Perturbation of Gene Expression Alters Bacterial Fitness under Stress and Reveals Underlying Epistatic Constraints**. *ACS Synth Biol* 2017, **6**:94–107.
22. Lu Z, Yang S, Yuan X, Shi Y, Ouyang L, Jiang S, Yi L, Zhang G: **CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in Bacillus subtilis**. *Nucleic Acids Res* 2019, **47**:e40–e40.
23. Yu L, Su W, Fey PD, Liu F, Du L: **Yield Improvement of the Anti-MRSA Antibiotics WAP-8294A by CRISPR/dCas9 Combined with Refactoring Self-Protection Genes in Lysobacter enzymogenes OH11**. *ACS Synth Biol* 2018, **7**:258–266.
24. Peng R, Wang Y, Feng W, Yue X, Chen J, Hu X, Li Z, Sheng D, Zhang Y, Li Y: **CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in Myxococcus xanthus**. *Microbial Cell Factories* 2018, **17**:15.
25. Fontana J, Dong C, Kiattisewee C, Chavali VP, Tickman BI, Carothers JM, Zalatan JG: **Effective CRISPRa-Mediated Control of Gene Expression in Bacteria Must Overcome Strict Target Site Requirements**. *bioRxiv* 2019, doi:10.1101/770891.
26. Dove SL, Hochschild A: **Conversion of the ω subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target**. *Genes Dev* 1998, **12**:745–754.
27. Tian T, Kang JW, Kang A, Lee TS: **Redirecting Metabolic Flux via Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in Escherichia coli**. *ACS Synth Biol* 2019, **8**:391–402.
28. Dinh CV, Prather KLJ: **Development of an autonomous and bifunctional quorum-sensing circuit for metabolic flux control in engineered *Escherichia coli***. *Proc Natl Acad Sci USA* 2019, **116**:25562–25568.

29. Johnny H. Hu DRL Shannon M Miller, Maarten H Geurts, Weixin Tang, Liwei Chen, Ning Sun, Christina M Zeina, Xue Gao, Holly A Rees, Zhi Lin: **Evolved Cas9 variants with broad PAM compatibility and high DNA specificity**. *Nature* 2018, **556**:57–63.

30. Chatterjee P, Jakimo N, Jacobson JM: **Robust Genome Editing of Single-Base PAM Targets with Engineered ScCas9 Variants**. *bioRxiv* 2019, doi:10.1101/620351.

31. Martella A, Firth M, Taylor BJM, Göppert A, Cuomo EM, Roth RG, Dickson AJ, Fisher DI: **Systematic Evaluation of CRISPRa and CRISPRi Modalities Enables Development of a Multiplexed, Orthogonal Gene Activation and Repression System**. *ACS Synth Biol* 2019, **8**:1998–2006.

32. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, et al.: **Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex**. *Nature* 2015, **517**:583–588.

33. Shechner DM, Hacisuleyman E, Younger ST, Rinn JL: **Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display**. *Nature Methods* 2015, **12**:664–670.

34. Cheng AW, Jillette N, Lee P, Plaskon D, Fujiwara Y, Wang W, Taghbalout A, Wang H: **Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling**. *Cell Res* 2016, **26**:254–257.

35. Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, Barrangou R: **Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality**. *Molecular Cell* 2014, **56**:333–339.

36. Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ: **CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo**. *Nature Methods* 2015, **12**:982–988.

37. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, et al.: **Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR**. *Genome Biology* 2016, **17**:148.

38. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al.: **Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9**. *Nature Biotechnology* 2016, **34**:184–191.

39. Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF: **Internal guide RNA interactions interfere with Cas9-mediated cleavage**. *Nat Commun* 2016, **7**:1–7.

40. Smith JD, Schlecht U, Xu W, Suresh S, Horecka J, Proctor MJ, Aiyar RS, Bennett RAO, Chu A, Li YF, et al.: **A method for high-throughput production of sequence-verified DNA libraries and strain collections**. *Molecular Systems Biology* 2017, **13**:913.

41. Vigouroux A, Oldewurtel E, Cui L, Teeffelen S van, Bikard D: **Engineered CRISPR-Cas9 system enables noiseless, fine-tuned and multiplexed repression of bacterial genes**. *bioRxiv* 2017, doi:10.1101/164384.

42. Cui L, Vigouroux A, Rousset F, Varet H, Khanna V, Bikard D: **A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9**. *Nat Commun* 2018, **9**:1–10.

43. Kocak DD, Josephs EA, Bhandarkar V, Adkar SS, Kwon JB, Gersbach CA: **Increasing the specificity of CRISPR systems with engineered RNA secondary structures**. *Nat Biotechnol* 2019, **37**:657–666.

44. Jin M, Garreau de Loubresse N, Kim Y, Kim J, Yin P: **Programmable CRISPR-Cas Repression, Activation, and Computation with Sequence-Independent Targets and Triggers**. *ACS Synth Biol* 2019, **8**:1583–1589.

45. Green AA, Silver PA, Collins JJ, Yin P: **Toehold Switches: De-Novo-Designed Regulators of Gene Expression**. *Cell* 2014, **159**:925–939.

46. Siu K-H, Chen W: **Riboregulated toehold-gated gRNA for programmable CRISPR–Cas9 function**. *Nat Chem Biol* 2019, **15**:217–220.

47. Li Y, Teng X, Zhang K, Deng R, Li J: **RNA Strand Displacement Responsive CRISPR/Cas9 System for mRNA Sensing**. *Anal Chem* 2019, **91**:3989–3996.

48. Oesinghaus L, Simmel FC: **Switching the activity of Cas12a using guide RNA strand displacement circuits**. *Nature Communications* 2019, **10**:2092.

49. Hanewich-Hollatz MH, Chen Z, Hochrein LM, Huang J, Pierce NA: **Conditional Guide RNAs: Programmable Conditional Regulation of CRISPR/Cas Function in Bacterial and Mammalian Cells via Dynamic RNA Nanotechnology**. *ACS Cent Sci* 2019, **5**:1241–1249.

50. Abatemarco J, Sarhan MF, Wagner JM, Lin J-L, Liu L, Hassouneh W, Yuan S-F, Alper HS, Abate AR: **RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes**. *Nat Commun* 2017, **8**:1–9.

51. Porter EB, Polaski JT, Morck MM, Batey RT: **Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors**. *Nature Chemical Biology* 2017, **13**:295–301.

52. Liu Y, Zhan Y, Chen Z, He A, Li J, Wu H, Liu L, Zhuang C, Lin J, Guo X, et al.: **Directing cellular information flow via CRISPR signal conductors**. *Nat Meth* 2016, **13**:938–944.

53. Lin B, An Y, Meng L, Zhang H, Song J, Zhu Z, Liu W, Song Y, Yang C: **Control of CRISPR-Cas9 with small molecule-activated allosteric aptamer regulating sgRNAs**. *Chem Commun* 2019, doi:10.1039/C9CC05531B.

54. Ricci F, Vallée-Bélisle A, Simon AJ, Porchetta A, Plaxco KW: **Using Nature's "Tricks" To Rationally Tune the Binding Properties of Biomolecular Receptors**. *Acc Chem Res* 2016, **49**:1884–1892.

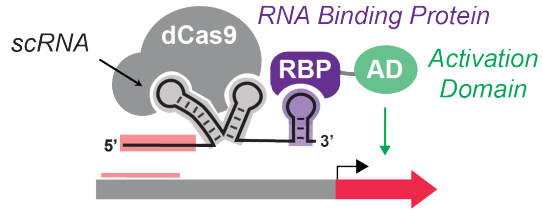
55. Chen X, Ellington AD: **Design Principles for Ligand-Sensing, Conformation-Switching Ribozymes**. *PLOS Computational Biology* 2009, **5**:e1000620.

56. Guedich S, Puffer-Enders B, Baltzinger M, Hoffmann G, Da Veiga C, Jossinet F, Thore S, Bec G, Ennifar E, Burnouf D, et al.: **Quantitative and predictive model of kinetic regulation by E. coli TPP riboswitches**. *RNA Biol* 2016, **13**:373–390.

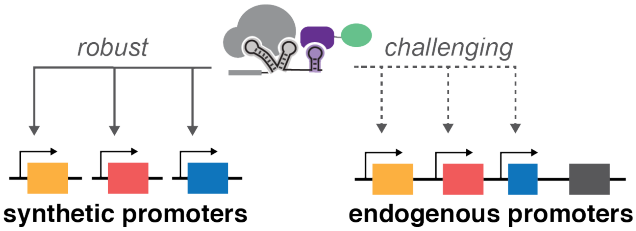
57. Leavell MD, Singh AH, Kaufmann-Malaga BB: **High-throughput screening for improved microbial cell factories, perspective and promise**. *Current Opinion in Biotechnology* 2020, **62**:22–28.
58. Costello Z, Martin HG: **A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data**. *npj Syst Biol Appl* 2018, **4**:1–14.

Figures

CRISPRa recruits an activation domain upstream of target genes.



Multi-gene programs can be implemented by expressing multiple scRNAs.



CRISPRa activity is determined by multiple factors, including:

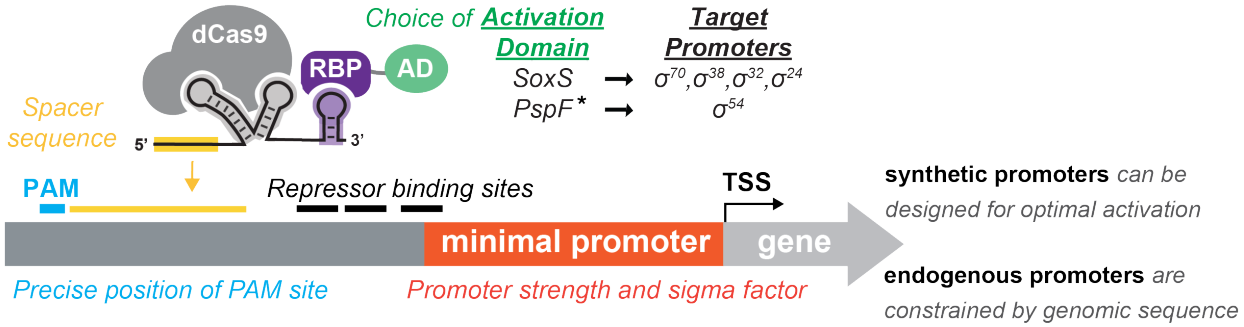


Figure 1. CRISPR activation (CRISPRa) is a powerful tool for programmable activation of genes in bacteria.

A CRISPRa system is shown where an activation domain is recruited to dCas9 using a modified guide RNA (scaffold RNA, scRNA) that binds to an RNA binding protein-activation domain fusion (RBP-AD). While CRISPRa can be used to robustly activate synthetic promoters designed for optimal activation, activating endogenous genes is constrained by the genomic sequence. Factors known to determine CRISPRa activity are indicated. * : PspF activation was demonstrated using a different modified sgRNA design where two BoxB aptamers were incorporated into the Cas9 handle.

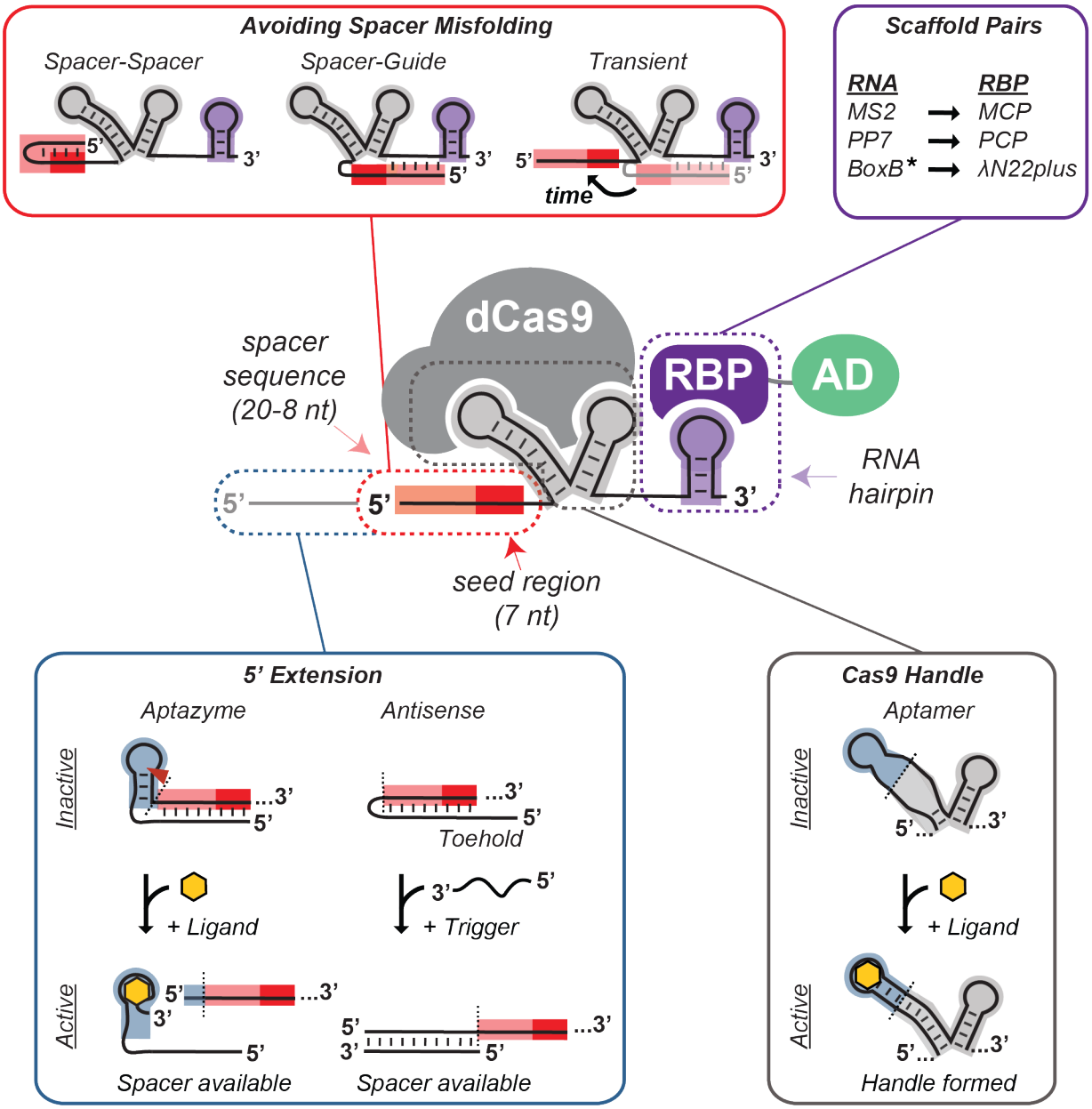


Figure 2. Guide RNA (gRNA) structural determinants of CRISPRa activity.

gRNA structure can be deleterious to function, as in the case of spacer misfolding, or can be useful for programming dynamic responses. Example dynamic gRNA engineering strategies that may apply to bacterial CRISPRa include extending the 5' end of the spacer to respond to ligands or RNA trigger strands. Ligand-responsive CRISPRa activities may also be obtainable by inserting ligand-binding aptamers into the Cas9 binding handle. Several cognate pairs of RNA binding protein (RBP) and RNA hairpin have been demonstrated in bacteria, enabling the simultaneous implementation of CRISPRa with different activation domains (AD). * : CRISPRa using the BoxB:λ22plus pair was demonstrated using two BoxB aptamers incorporated into the Cas9 handle.

CRISPR-Cas engineering can accelerate DBTL cycles.

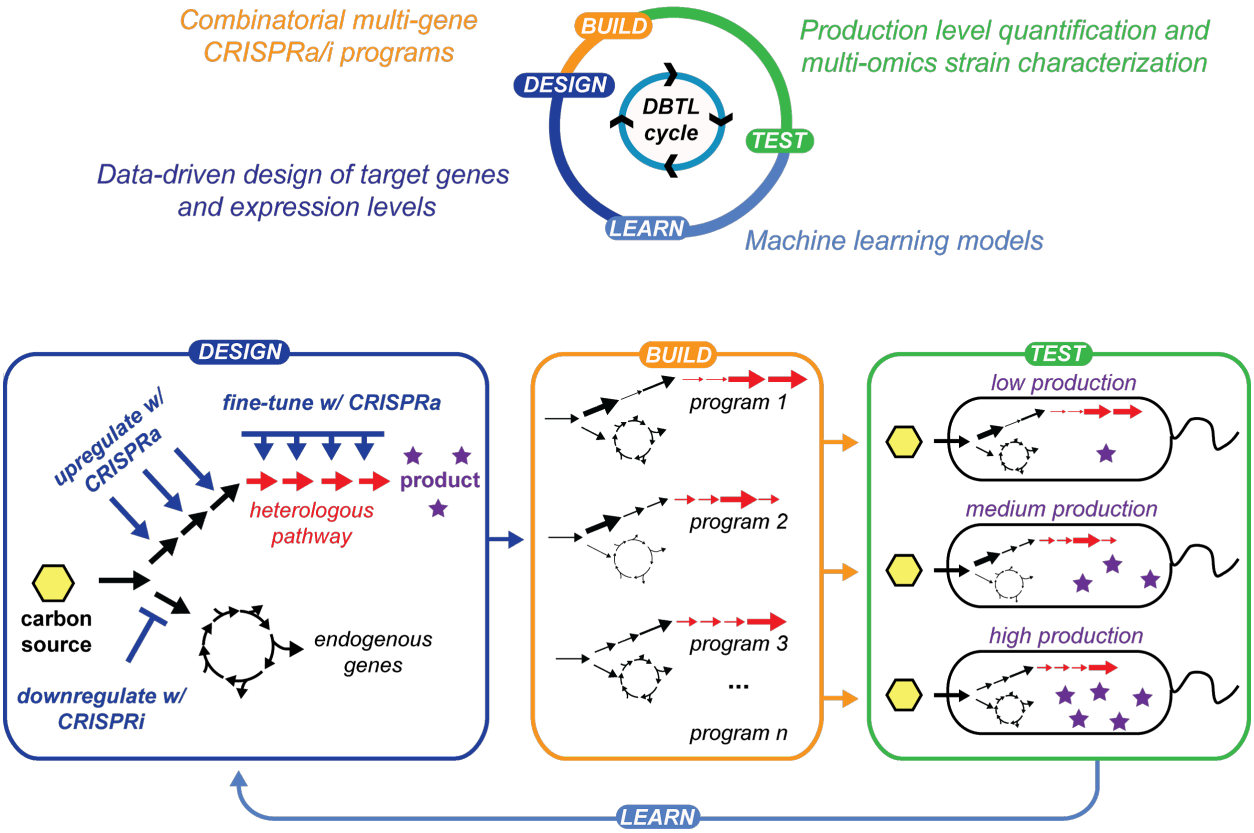


Figure 3. Developing robust workflows to integrate CRISPRa/i engineering into data-driven workflows will create new capabilities for rapidly optimizing chemical production.

In this conception, each Design-Build-Test-Learn (DBTL) cycle uses machine learning and data-driven design to engineer multi-gene CRISPRa/i programs. After each build phase, production titers are measured and the strains are characterized using multi-omics analysis. These data are employed to refine the models and drive the design of CRISPRa/i programs for the next DBTL cycle.

Acknowledgements

This work was only possible thanks to your love, friendship, mentorship, work, advice, help

Adam Moyer
Akemi Pettinà
Alberto Carignano
Andrea Girardi
Andrew Asakawa
Anika Patel
Aria Walls
Augusta Fava
Beatrice Magistro
Ben Tickman
Ben Wiggins
Betty Tonin
Brenna Cherry
Brianne King
Carrie Harwood
Cassandra Burke
Cecilia Ignesti
Chen Dong
Chenggang Xi
Cholpisit Kiattisewee
Christine Savolainen
Chuhern Hwang
Cristina Del Bianco
Dalila Calgaro
Dan Cunningham-Bryant
Darius Izad
Dave Yuan
David Sparkman-Yager
Davide Macchi
Davide Marini
Diego Alba Burbano
Dorothy Bardewyck
Edoardo Mazza
Egidio Fontana
Elise Dibble
Elizabeth Speltz
Elke Zandiri
Emily Cliff
Enrico Smaldone
Eric Klavins
Erin Fagnan
Ethan Pettinà
Evan Zandiri
Ever Pettina
Francesco Guzzonato
Franco Pettinà
Frida Tonin
Gabriel Calgaro
Gianni Bonotto
Giovanni Spillere
Hussna Ali
Ian Faulkner
James Carothers
Jason Stevens
Jennifer Ham
Jesse Zalatan
Jingwen Sun
Joanne Wong
Joely Nelson
Larry Fontana
Laura Bardewyck
Laura Tognacci
Liam Fontana
Lisa Vigolo
Liz Warfield
Luca Maistro
Madelaine Pinkerton
Maire Gavagan
Maitreya Dunham
Maria Vittoria Soldà
Marvyn Pettinà
Mary Lidstrom
Matteo Scollo
Monica Tonin
Neel Shah
Olivier Jousson
Pramod Chavali
Raffaele Zandiri
Robin Kirkpatrick
Semira Beraki
Sevwandi Acqua
Sheref Mansy
Sinduja Marx
Sofia Tonin
Susan Crandell
Tamara Tonin
Tom Haskins
Trent Moriarty
Ugo Calgaro
Vita Romano
Will Chen
William Bardewyck
William Voje, Jr.
Zeudi Pettinà