Programming Bacterial Gene Expression Using Synthetic CRISPR-Cas Transcriptional Regulators

Chen Dong

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Jesse George Zalatan, Chair

James Carothers

Alshakim Nelson

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Programming Bacterial Gene Expression Using Synthetic CRISPR-Cas Transcriptional Regulators

Chen Dong

Chair of the Supervisory Committee:

Jesse George Zalatan

Department of Chemistry

Bacteria play a central role in biosynthesis to produce value-added organic chemicals due to its diverse carbon and energy source preferences. Implementing synthetic transcriptional regulation devices can advance our ability to modify gene expression in bacteria for engineering production strains. The CRISPR-Cas activation (CRISPRa) system, a programmable transcriptional activator with wide applications in eukaryotic organisms, has been under-utilized in bacterial due to the lack of efficient transcriptional activation domains. This work describes our contribution to the development and understanding of bacterial CRISPR-Cas-based transcriptional regulation devices. We screened novel bacterial activation domains to be used as CRISPR-Cas activators in *E. coli*, and optimized our strongest activation domain, SoxS into a programmable CRISPR activator. In addition, we investigated the properties of the well-established CRISPRi repression system and found that partial repression can be achieved by controlling the expression level of the sgRNA. Combining CRISPRa and CRISPRi, we demonstrated inducible simultaneous up- and down-regulation of a dual reporter gene and the CRISPR-mediated regulation of the ethanol bioproduction pathway. Moreover, we also learned important properties of the bacterial CRISPRa system which is much more restrictive than the eukaryotic CRISPRa system. CRISPRa activity is dependent on the sigma factor that the promoter recruits, the baseline strength of the promoter, the sequence composition of the promoter, the presence of nearby transcription factors, and the precise positioning of the scRNA target. We attempted to relieve these restrictions by implementing an engineered Cas protein that can bind to a

wider range of targets. Lastly, we describe our efforts to transport the CRISPRa system into other non-*E.coli* bacteria. CRISPR-SoxS activator proved to be active in *Pseudomonas putida*, but not in the other organisms we tested. Therefore, we propose to characterize host-specific activation domains for bacteria whose cellular machineries are incompatible with SoxS. Together, this work provided a novel programmable gene activation device in bacteria and sets up the foundation for the development of complex, broad-host-range bacterial cellular devices for biosynthetic applications.

Table of Contents

List of publications

Paper title	Journal	Chapter	Contribution	Status	Ref.
Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria	Nature Communications	Ch. 2	First author	Published – 2018	1
Regulated Expression of sgRNAs Tunes CRISPRi in *E. coli*	Biotechnology Journal	Ch. 3	Second author	Published – 2018	2
Systematic Characterization of the Requirements for CRISPR-Cas Activation in Bacteria (tentative title)		Ch. 4	Co-first author	Manuscript in preparation	

Reference

1. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9,** 2489 (2018).

2. Fontana, J., Dong, C., Ham, J. Y., Zalatan, J. G. & Carothers, J. M. Regulated Expression of sgRNAs Tunes CRISPRi in E. coli. *Biotechnology Journal* **13,** 1800069 (2018).

List of figures

Main figures

Supplementary figures

List of tables

Acknowledgments

This work would not have been possible without the constant input of support, resources and idea of many other people, which why I would like to acknowledge everyone who are involved for their tremendous contribution to this work. I would first like to thank my supervisor, Jesse Zalatan for providing me guidance and freedom to explore the world of science and supporting me in critical moments in both my professional and personal life. I would like to thank my advisers from collaborating groups, James Carothers and Mary Lidstrom for inspiring ideas, abundant resources and helpful connections that led to successful projects. I would like to express my gratitude to my committee members Jesse Zalatan, Jennifer Nemhauser, James Carothers, Alshakim Nelson and Dustin Maly for useful discussions and feedback regarding my project. I also would like to thank all our collaborators for generously providing us lab space, instruments, strains and DNA: Maureen Thomason; Carrie Harwood and Amy Schaefer; The Carothers group: Jason Stevens, Willy Voje, Chuhern Hwang, and David Sparkman-Yaeger; The Lidstrom group: Melissa Hendershott, Aaron Puri, Amanda Smith, Frances Chu, Yanfen Fu, and Joseph Groom. I especially would like to thank my friend Jason Fontana, with whom I have been working closely together for 4 years. He has made my work in the lab so much more fun and encouraging. I would also like to thank all the other students involved in this project: My undergraduate mentee Anika Patel, our first-year graduate student Cholpisit Kiattisewee, and our visiting student from Germany, Anja Hofmann. I would like to thank all other members from the Zalatan lab: Brian Fan, Erin Fagnan, Robin Kirkpatrick, Jingwen Sun, Betsy Speltz, Maire Gavagan, Dan Cunningham-Bryant, Emily Cliff and our neighouring group, the Maly lab members for creating a positive and helpful working environment and for all the fun times we had together outside the lab. Finally, I would like to express my love to my wife Guan, my brother Yutian, my parents and grandparents for all the encouragements and support. Also cheers to my dearest friends Yunke, Jianwei, Yi Xie, Chenxi, Yi Li, Ang, Qi, and Xiaochen for all the precious and fun moments together!

Dedicated to my wife, Guan

Chapter 1 | Introduction

1.1 | Background

1.1.1 | Synthetic biology and its application in chemical industry

Rapid improvement in the speed and cost-efficiency of DNA synthesis and sequencing has brought the field of synthetic biology to life[1,2,7]. With the streamlined design and standardization of modular biological entities, the number of genetic tools made available for us to reliably express heterologous gene clusters have vastly expanded. Using well-characterized synthetic promoters, ribosome binding sites, plasmid vectors and genomic integration methods, stable gene editing and heterologous overexpression can be rapidly executed in a variety of organisms, which greatly enhances our ability to engineer cell systems for research and biotechnology applications[3]. An example of engineering artificial gene clusters is the practice of metabolic engineering, where heterologous metabolic enzyme pathways are expressed in model microorganisms to produce value-added chemicals such as opioids precursors and the anti-malarial drug artemisinin[4–6]. Given the rising environmental and sustainability concerns on petroleum-based chemical synthesis, the development of synthetic biology can have strong impacts on the future of chemical industry by providing an alternative bio-manufacturing approach.

1.1.2 | Bacteria as an efficient host organism for microbial production

To re-engineer metabolic pathways for biosynthesis, the following factors need to be taken into consideration: (i) cost-efficiency of starting materials; (ii) choosing a suitable host organism; (iii) combination of gene clusters and metabolic routes. (iv) dynamic control systems for facilitating heterologous expression; (v) debottlenecking approaches to overcome stress and burden; (vi) maximizing yields and productivity[7]. Choosing the most suitable microbial host is a critical first step in designing the most optimal biosynthetic system. An ideal host chassis would utilize cheap and accessible carbon feedstocks, have abundant synthetic biology tools available, possess diverse metabolic routes, grow reasonably fast, and be able to tolerate moderate environmental stresses[3]. The most widely used host chassis for metabolic engineering include *Escherichia coli* and *Saccharomyces cerevisiae*, due to the abundant studies in its genetics and rich accumulation of tools for genome manipulation. Although *E. coli* and *S. cerevisiae* excel at the number of genetic tools available, they fall back in some of the other properties mentioned above. For example, the main

carbon source of *E. coli* and *S. cerevisiae* are sugars that come from agricultural feedstocks, and its increasing cost is making these model microorganisms less ideal for bioproduction. Recently, the unique features of non-model bacteria have become more and more attractive[3,8]. The reason is because many non-model bacteria utilize diverse carbon and energy sources and can tolerate stresses that commonly arise from the accumulation of some chemical products. Considerable progress has been made in the development of genetic tools in non-model bacteria[11]. Noteworthy engineered bacteria include methanotrophs, which utilize methane from natural gas[9,10]; photosynthetic cyanobacteria that utilize easily accessible light and carbon dioxide[12–15]; and halophilic bacteria that can tolerate high pH and salt concentrations, which can potentially remove the sterility restriction by maintaining fermentation processes at harsh conditions[16]. Because non-model bacteria can fill niches in bioproduction, the characterization of synthetic biology devices in non-model bacteria can have great impacts on biotechnological innovations.

1.1.3 | Synthetic pathways need to be regulated for optimal production

Although the process to create heterologous production pathways in engineered bacterial strains has been simplified to some extent, more optimal productivity can only be achieved when these heterologous genes are expressed in a coordinative and regulated manner[17,18]. Due to the lack of host interactions of the heterologous expression cassettes, the uncontrolled overexpression of the heterologous genes can potentially antagonize regular cellular functions and cause toxicity. For example, overexpression of the heterologous enzymes can trigger internal feedback response to divert the metabolic flux away from the product or result in growth defects and selective pressure towards mutants on the heterologous genes[19,20]. To avoid these issues, people have been routinely doing genomic knockouts of genes that can cause divergence in the metabolic flux, but these irreversible deletions of endogenous genes may unpredictably impede native cellular functions. In addition, gene knockouts require extensive cloning and genetic manipulations for every gene to be controlled, which may slow down the engineering workflow. For the heterologous pathway, we still have limited tools for dynamically activating the heterologous genes at a level that minimizes the burden to the internal cell state. Therefore, a robust method for tunable gene regulation that can interact with the internal cell state is highly demanded, so that we can control the expression of the heterologous pathway in order to balance the metabolic flux towards optimal production (A simple

case is depicted in Figure 1.1).

1.1.4 | Transcription control is a widely used methodology for regulating gene expression

Regulation of gene expression can be achieved in different levels. Transcription control regulates the synthesis of the messenger RNA[21,22]; translation control regulates the translation of protein from the mRNA[23]; post-translational modifications controls protein functions by chemically modifying its conformation and folding[24]. While every approach has its own merits, transcription control remains the most versatile and easy-to-use. As discussed in the last paragraph, genomic knockouts and synthetic overexpression are examples of transcription regulation. Other useful genetic tools for transcription control include hybrid proteins consisting of DNA-binding domains (DBD) that can be fused with transcription factors as effector domains (ED). By targeting these hybrid proteins to bind DNA sequence proximal to the gene of interest, the effector domains can then execute the activation or repression of its transcription. Well-established DBD-ED hybrid system include the zinc finger arrays fused with effector domains and transcription activator-like (TAL) effectors fused with effector domains[25]. In these tools, targeted transcription control of genes can be achieved by combining the repeating units that recognize different stretches of DNA sequences together in order to make it match and bind a DNA site at the gene of interest.

1.1.5 | CRISPR-Cas systems and its enhanced utility over traditional tools

The emergence of technologies based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)[26,27,31] has revolutionized the field of transcription regulation. Because target recognition is achieved through DNA-RNA interaction, CRISPR is highly programmable and versatile. The discovery and study of the CRISPR system was the result of decades of work. CRISPR was originally discovered to be a bacterial and archaeal adaptive immune system[28]. The CRISPR locus in the genome stores the immunological memories as short non-repetitive sequences called spacers, typically separated by short, palindromic repeat sequences. The spacer sequences were originated from heterologous invading nucleic acids and were stored in the genome as exact sequence matches. Next to the CRISPR array are genes that expresses proteins that will facilitate adaptive immune response based on the information stored in the CRISPR array, these genes are called CRISPR-associated (cas) genes. During viral infection, the host cell expresses a subset of cas proteins to

insert a stretch of exogenous DNA sequence (protospacer) into the CRISPR array, which becomes a newly added spacer. When the viral infection re-occurs, the spacer sequence will be transcribed into a long RNA strand called the precursor CRISPR RNA (pr-cRNA) and then cleaved into CRISPR RNAs (crRNA), which contains the spacer sequence stored from the invading exogenous DNA. The crRNA then binds to the cas proteins and guides the whole complex to the exogenous DNA at the protospacer sequence and binds it through the base pairing of the spacer sequence on the crRNA and the protospacer sequence on the invading DNA, followed by the cleavage and degradation of the invading DNA. Different types of CRISPR systems have been found in different organisms that have significant variance in mechanisms of action and the cas proteins involved[28]. One type of CRISPR system, the type II CRISPR system has caught the attention of synthetic biologists due to its unique feature. The most well-studied type II CRISPR system is harbored by *Streptococcus pyogenes*. Unlike other types of CRISPR systems that functions as a cascade involving several protein subunits, the type II CRISPR system needs only one cas protein, Cas9 to execute crRNA-guided targeting and cleavage [26,27,31]. Cas9 binds the dual RNA complex of the crRNA and a trans-activating crRNA (tracrRNA) and requires a specific sequence next to the protospacer called protospacer-adjacent motif (PAM) for correct binding. When the Cas9-crRNA-tracrRNA complex finds its target, cas9 will bind the PAM sequence and the crRNA will bind the sequence complementary to the 20-bp spacer through base-pairing interactions. DNA cleavage is facilitated through 2 nuclease domains on the Cas9 protein: the RuvC-like nuclease domain and the HNH nuclease domain, which creates a blunt-ended double-stranded break. The simplicity of the type II CRISPR systems relative to other types has led to a series of engineering efforts to transform it into a gene editing tool[29,30], which then leads to its repurposing into a DNA-binding domain and a transcriptional regulator.

Different components of the CRISPR-Cas9 system have been re-engineered. The combination of the crRNA and the tracrRNA into a chimeric single guide RNA (sgRNA) decreases the CRISPR-Cas9 complex into only 2 components[32]. People have also found two point mutations, D10A on the RuvC I domain and H840A on the HNH domain that destroys Cas9's nuclease activity[30], creating the re-purposed nuclease-defective Cas9 (dCas9)[33]. The dCas9-sgRNA complex can bind DNA at the protospacer sequence with an adjacent PAM without cleaving it, making it become a highly programmable DNA-binding domain (Figure 1.2). It possesses considerable advantage over previous

DNA-binding domains such as zinc fingers and TAL effectors, the main reason being its simplicity to design and program: unlike zinc fingers and TAL effectors that require consideration of complex protein-DNA interfaces when choosing a target-specific DBD, CRISPR-dCas9 only requires changing the 20-bp spacer sequence to its complementary protospacer adjacent to the PAM[25]. Over the last decade, new CRISPR-Cas-based transcription regulation devices have been developed in a rapid pace, where these technologies are mainly utilized in eukaryotic cells. Transcriptional activators and repressors have been fused with dCas9 for targeted gene activation and silencing[32,34]; The sgRNA was also attached with additional functional hairpins structures that recruits its specific RNA-binding protein (RBP). By fusing different effectors with different RBPs we can execute multi-gene ON/OFF control circuits by using sgRNAs attached with different functional hairpins[35–37]. The simplification of the CRISPR-Cas9 complex structure and the addition of novel functional modules has led to development of several well-characterized higher-order CRISPR-based transcriptional regulators[38]. Notable examples of these synthetic regulators include: CRISPR-dCas9 coupled with the SunTag array where the dCas9 is fused with a small peptide epitope that is recognized by a cognate scFV fused with the transcriptional activator VP64; the synergistic tripartite activation domain where the effector is a hybrid protein of activators VP64, p65, and Rta (VPR); and the synergistic activation mediator (SAM) where MS2 hairpins were extended from the sgRNA tetraloop that recruits activator mediators through MCP. These synthetic regulatory devices have been widely used in various applications, including metabolic engineering in eukaryotic organisms such as yeast.

1.1.6 | Accomplishments and technological gaps in bacterial CRISPR-Cas regulation tools

In bacteria, significant breakthroughs have been made in the development of CRISPR transcriptional regulators as well. The most game-changing discovery among them would certainly be the use of the CRISPR-Cas system for gene silencing. People have found that the mutant dCas9-sgRNA complex, when bound at the non-template strand on the RNAP-binding region or the upstream parts of the transcribed regions, can physically block the binding or elongation of the RNAP and silence gene expression. This technology was referred as CRISPR interference, or CRISPRi, and it has revolutionized gene repression methods in bacteria[33] (Figure 1.2). Compared to traditional gene knockout methods, CRISPRi has considerable advantages: As any other CRISPR regulators, CRISPRi is programmable and easy to implement, and allows rapid generation of multiple gene

repression by delivering multiple sgRNAs. In addition, though having tight repression when placed at the correct position, it is also possible to achieve partial repression by changing the expression levels of each CRISPR components[33]. This enables us to repress essential genes that are impossible to delete using genomic knockouts. CRISPRi has been used in many applications including metabolic engineering in *E.coli* and other bacteria strains. In contrast to the extensive use of CRIPSRi repression in bacteria, CRISPR-based transcription activation (CRISPRa) tools have been under-utilized. The main reason for this technological gap is because unlike eukaryotic systems that have efficient modular transcriptional activators such as VP64, there is lack of a well-characterized transcriptional activators available in bacteria. Only limited work has been previously reported involving the construction of interchangeable activation domains in bacteria, where most of these works were focused on using these activation domains as bacterial two-hybrid systems[39–41]. There are some activation domains that emerged from these studies, such as the N-terminal domain of the α-subunit (α-NTD) of RNA polymerase (RNAP)[43], and the ω-subunit of the RNA polymerase (RNAP)[42]. However, there is little follow-up efforts on those activation domains to turn them into modular transcriptional activators. Only 1 study has reported the use of the RNAP ω-subunit fused with dCas9 as a CRISPRa activator[44], but this system requires the host strain to have its endogenous copy of RNAP knocked out. And the limited cases of reports of this CRISPRa system in the following years suggest that it may have limited utility, or we still lack complete understanding in its design rules[45,46]. Therefore, an efficient bacterial CRISPR-Cas-based transcriptional activation system remains to be highly demanded.

1.2 | Overview of thesis work

This thesis summarizes our work in generating and characterizing a novel CRISPRa system in bacteria, our input in the understanding of CRISPRi utility in bacteria, and our efforts to apply CRISPR-Cas transcription regulation in non-*E.coli* bacteria with interesting potential applications. We utilized the CRISPR complex where the DNA-binding domain and the transcriptional activation domain are separated into modular parts by a scaffold RNA (scRNA) that binds the dCas9 and the activation domain with different hairpin structures[35] (Figure 1.2). In chapter 2, we screened bacterial transcriptional activation domains by linking them to CRISPR-Cas DNA-binding domains and found 4 activation domains that can activate reporter gene expression when coupled with the CRISPR-Cas

complex[47]. We did further modification and optimization of our strongest activation domain SoxS and coupled the CRISPR-SoxS activator with the CRISPRi repression system to demonstrate inducible multi-gene control with simultaneous activation and repression. In addition, we found that effective CRISPR activation requires that the target site be positioned in a narrow region, different from the relatively flexible position requirements in eukaryotic CRISPR activators. And last we implemented our CRISPRa system in a simple ethanol biosynthesis pathway in *E.coli*. In chapter 3, we investigated the tunability of the CRISPRi system in *E.coli* by varying the expression level of each CRISPR components. When dCas9 level is in excess, titrating the levels of sgRNA can give varying CRISPRi repression levels between 5- to 300-fold. We also found that CRISPRi repression cannot be undone by overexpressing a competing off-target sgRNA to displace the on-target sgRNA. In chapter 4, we found that CRISPRa activity is dependent on the sigma factor that the minimal promoter recognizes, the baseline strength of the minimal promoter, the sequence composition of the promoter and surrounding transcription factor binding. In addition, the CRISPRa activity is also highly sensitive to single-base movements of the CRISPRa complex along the DNA. We hypothesize that these single base shifts cause the CRISPRa complex to be rotated “out-of-phase” where the activation domain cannot approach the optimal orientation to interact with the RNAP. We attempted to partially relieve these restrictions by using an engineered CRISPR-Cas system having a wider PAM recognition[46]. In chapter 5, we explored the transportability of the CRISPR-Cas system in non-*E.coli* bacteria having unique bio-industrial applications. We selected expression devices for the CRISPR modules in *Bacillus. subtilis*, *Pseudomonas. putida*, and *Methylomicrobium. buryatense* and tested the functionality of CRISPRi repression and CRISPRa activation in these organisms. CRISPRi proved to be functional in all 3 organisms. CRISPRa with the SoxS activator proved to be functional in *P. putida* but not in *B. subtilis* and *M. buryatense*. We hypothesize the incompatibility of SoxS with native cell machinery and the suboptimal expression of the activation domain to be the 2 possible modes of failure in *B. subtilis* and *M. buryatense*. Finally we propose a general strategy of screening host-specific activation domains for developing CIRSPR-Cas transcriptional regulators in non-model bacteria. In conclusion, this thesis laid out the ground work towards engineering dynamic, programmable, and efficient CRISPR-Cas-based transcriptional regulation systems to control complex genetic circuits in diverse bacterial hosts, which could potentially translate into next-level bio-synthetic platforms.

1.3 | References

1. Yeh, B. J. & Lim, W. A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nature Chemical Biology* **3,** 521–525 (2007).

2. Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering* **14,** 189–195

3. Keasling, J. D. Manufacturing Molecules Through Metabolic Engineering. *Science* **330,** 1355–1358

4. Lee, T. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *Journal of Biological Engineering* **5,** 12 (2011).

5. DeLoache, W. C. *et al.* An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nature Chemical Biology* **11,** 465–471

6. Fossati, E., Narcross, L., Ekins, A., Falgueyret, J.-P. & Martin, V. J. Synthesis of Morphinan Alkaloids in Saccharomyces cerevisiae. *PLoS ONE* **10,** e0124459-15

7. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940 (2006).

8. Clomburg, J. M., Crumbley, A. M. & Gonzalez, R. Industrial biomanufacturing: The future of chemical production. *Science* **355,** aag0804 (2017).

9. Czajka, J., Wang, Q., Wang, Y. & Tang, Y. J. Synthetic biology for manufacturing chemicals: constraints drive the use of non-conventional microbial platforms. *Applied Microbiology and Biotechnology* **101,** 7427–7434 (2017).

10. Haynes, C. A. & Gonzalez, R. Rethinking biological activation of methane and conversion to liquid

fuels. *Nature Chemical Biology* **10,** 331–339

11. Kalyuzhnaya, M. G., Puri, A. W. & Lidstrom, M. E. Metabolic engineering in methanotrophic bacteria. *Metabolic Engineering* **29,** 142–152 (2015).

12. Hagemann, M. & Hess, W. R. Systems and synthetic biology for the biotechnological application of cyanobacteria. *Current Opinion in Biotechnology* **49,** 94–99 (2018).

13. Angermayr, S., Rovira, A. & Hellingwerf, K. J. Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends in biotechnology* **33,** 352–61 (2015).

14. Wang, B., Wang, J., Zhang, W. & Meldrum, D. R. Application of synthetic biology in cyanobacteria and algae. *Frontiers in Microbiology* **3,** 344 (2012).

15. Pakrasi, H. B. Synthetic biology of cyanobacteria: unique challenges and opportunities. 1–14 doi:10.3389/fmicb.2013.00246/abstract

16. Yin, J., Chen, J.-C., Wu, Q. & Chen, G.-Q. Halophiles, coming stars for industrial biotechnology. *Biotechnology Advances* **33,** 1433–1442 (2015).

17. Dunlop, M. J. Engineering microbes for tolerance to next-generation biofuels. *Biotechnology for biofuels* **4,** 32 (2011).

18. Stevens, J. T. & Carothers, J. M. Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS Synthetic Biology* **4,** 107–115

19. Carothers, J. M., Goler, J. A. & Keasling, J. D. Chemical synthesis using synthetic biology. *Curr Opin Biotech* **20,** 498–503 (2009).

20. Martínez-Antonio, A., Janga, S., Salgado, H. & Collado-Vides, J. Internal-sensing machinery

directs the activity of the regulatory network in Escherichia coli. *Trends Microbiol* **14,** 22–27 (2006).

21. Lloyd, G., Landini, P. & Busby, S. Activation and repression of transcription initiation in bacteria. *Regulation of Gene Expression* **37,** 17–31

22. Browning, D. F. & Busby, S. J. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2,** 57–65

23. Marintchev, A. & Wagner, G. Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys* **37,** 197–284 (2004).

24. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *Plos Comput Biol* **11,** e1004049 (2015).

25. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31,** 397–405 (2013).

26. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu Rev Biophys* **46,** 1–25 (2015).

27. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278 (2014).

28. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nature Methods* **10,** 957–963

29. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences* **34,** 401–407 (2009).

30. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial

genomes using CRISPR-Cas systems. *Nature Biotechnology* **31,** 233 (2013).

31. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (2012).

32. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31,** 833 (2013).

33. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183

34. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451

35. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160,** 339–350

36. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588

37. Shechner, D., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods* **12,** 664–670 (2015).

38. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature Publishing Group* 1–11 doi:10.1038/nrm.2015.2

39. Joung, K. J., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proceedings of the National Academy of Sciences* **97,** 7382–7387 (2000).

40. Joung, K. J., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proceedings of the National Academy of Sciences* **97,** 1–6 (2000).

41. Dove, S. L. & Hochschild, A. A Bacterial Two-Hybrid System Based on Transcription Activation. *Methods in molecular biology (Clifton, N.J.)* **261,** 1–16

42. Dove, S. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386,** 627–630

43. Dove, S. L. & Hochschild, A. Conversion of the subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes & Development* **12,** 745–754

44. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41,** 7429–7437

45. 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 Synthetic Biology* **6,** 94–107 (2016).

46. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556,** 57 (2018).

47. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9,** 2489 (2018).

1.4 | Figures

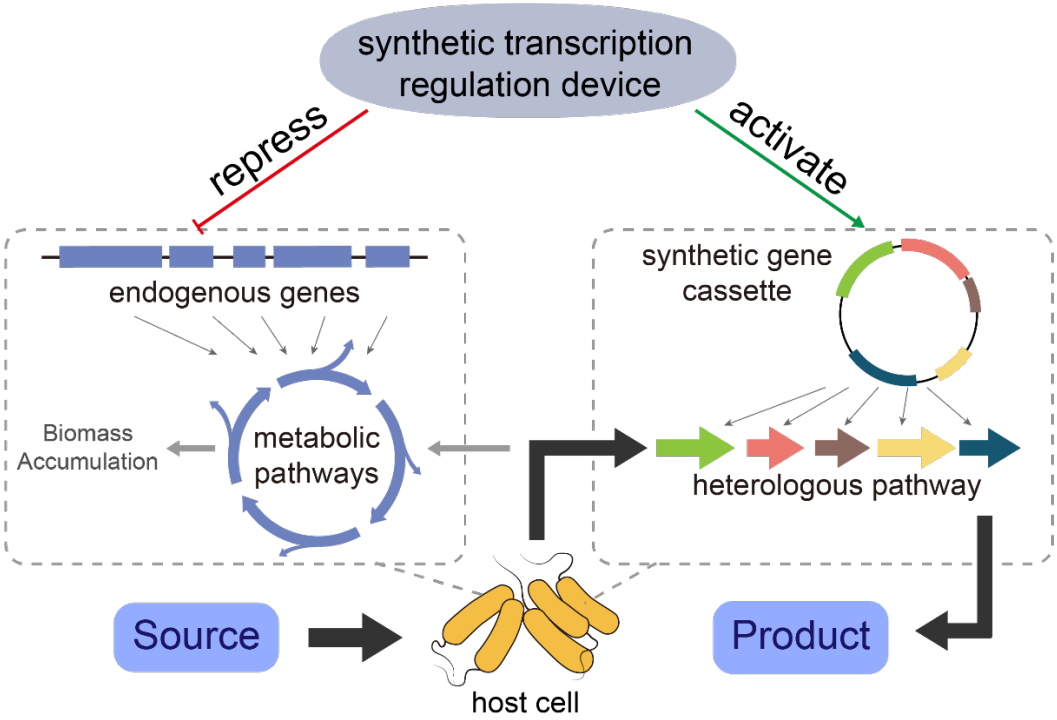


Figure 1.1 Synthetic transcriptional regulation devices in metabolic engineering applications. A simple example for balancing metabolic flux between the heterologous pathways (colored) and the endogenous metabolism (pale blue). When engineered host cells have accumulated enough biomass, the regulation device can execute regulatory programs to repress the expression of the endogenous pathway to minimize the loss of carbon flux towards metabolism and activate the heterologous pathway to increase the rate in which the biosynthetic enzymes are expressed. The two-way multi-gene control allows controlled direction of metabolic flux to increase production titers.

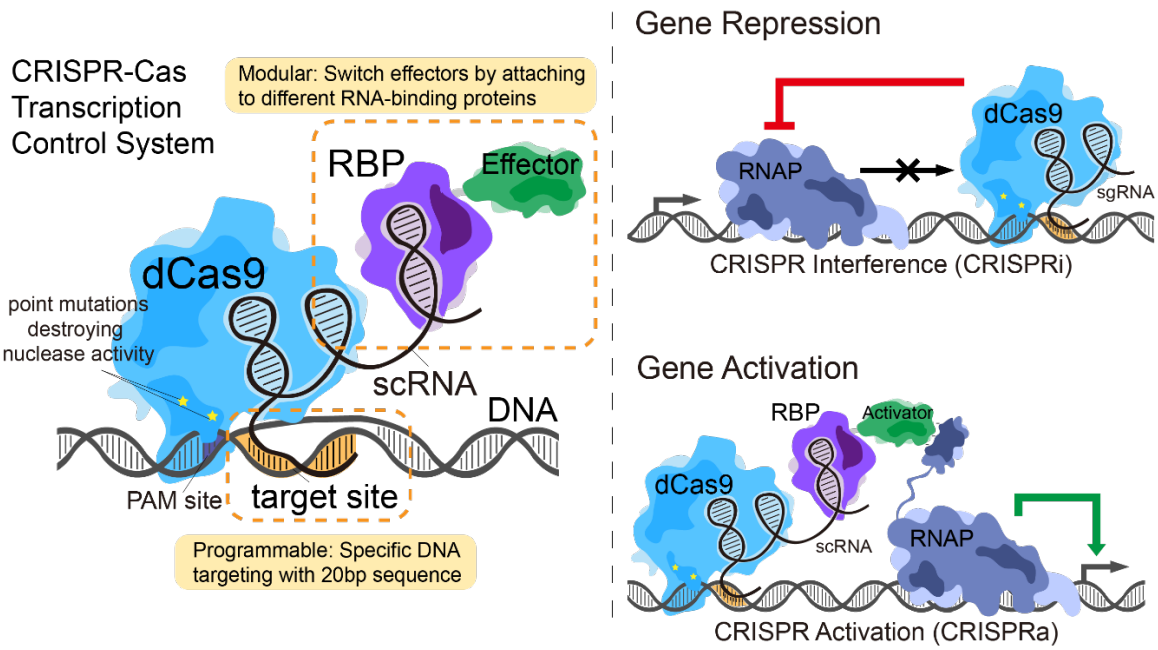


Figure 1.2 The CRISPR-Cas system is an efficient tool for transcriptional regulation.

Left: engineered CRISPR-Cas regulators that consist of a scRNA that has the functional domains split into modular parts: the 20 bp spacer sequence that allows programmable DNA targeting by changing its sequence complementary to the target in the genome, the sgRNA stem loops that binds to the nuclease defective Cas9 (dCas9), and the additional 3' hairpin that recruits its specific RNA-binding protein (RBP) fused with a transcriptional effector. This form of the CRISPR complex is not only programmable, but also modular because different regulatory functions can be executed through attaching different effectors to different RBPs. Right: multigene control can be achieved simultaneously with the CRISPR-Cas system. Gene repression can be achieved by CRISPR interference (CRISPRi) where the CRISPR complex binds to the DNA and physically block the movement of the RNA polymerase (RNAP). Gene activation can be achieved by targeting the CRISPR complex containing an activation domain to the promoter regions of the gene of interest, so that the activation domain can interact with RNAP and recruit it to the promoter to initiate transcription.

Chapter 2 | Characterization of synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria

The work in this chapter was published and reprinted with permission from the journal of initial publication:

Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9**, 2489 (2018).

2.1 | 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.

2.2 | 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. 2.1). 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.

2.3 | Results

2.3.1 | 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. 2.1A)[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. 2.1B)[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. 2.2A). 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. 2.1A)[7,13]. We also obtained significant GFP expression with the T4

bacteriophage activator AsiA (Supplementary Fig. 2.1B) 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. 2.2B). 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. 2.2B). 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].

2.3.2 | 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. 2.3A). For comparison, in eukaryotic cells the same modified scRNA lacking the terminator reduces CRISPRa-mediated reporter gene expression ~2-fold (Supplementary Fig. 2.2A), 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. 2.3B). 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. 2.2B). 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. 2.2C), 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. 2.3C). 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. 2.3D). 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. 2.3), 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. 2.4) compared to the original, unoptimized system (Fig. 2.2B). Using RT-qPCR, we observed a 50-fold increase in GFP mRNA levels relative to negative control strains (Supplementary Fig. 2.5). Importantly, we observed no significant growth burdens associated with expression of the CRISPRa system (Supplementary Fig. 2.6A & 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. 2.6C). 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. 2.3A), 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. 2.7).

2.3.3 | 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. 2.4A). 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. 2.4A). 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. 2.4B), 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. 2.4B). 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. 2.4B). 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. 2.4C). 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. 2.4C).

2.3.4 | 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. 2.5A). 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. 2.5B). To determine if activators with weak activity in our initial assays (Fig. 2.2A) 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. 2.5B), and in no case did we observe dramatic increases in activity relative to our initial assays (Fig. 2.2A). 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.

2.5B). 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. 2.5C). 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.

2.3.5 | 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. 2.6)[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. 2.6). 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.

2.4 | 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. 2.8). 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.

2.5 | Methods

2.5.1 | 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.

2.5.2 | 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. 2.9).

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.

2.5.3 | 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 = $(1000 \times A_{420})/(OD_{600} \times 0.02\ mL \times time)$.

2.5.4 | 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).

2.5.5 | 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.

2.5.6 | 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.

2.5.7 | 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].

2.6 | 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.)

2.7 | 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.

The authors declare no competing interests.

2.8 | 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 PBAD 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).

2.9 | Figures

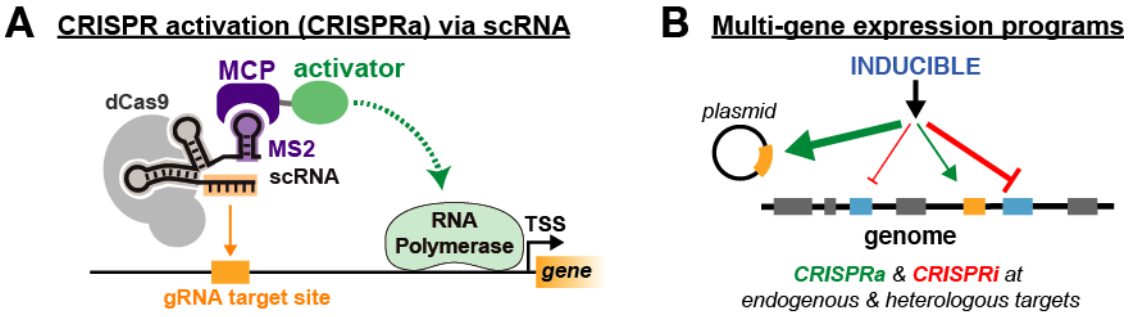
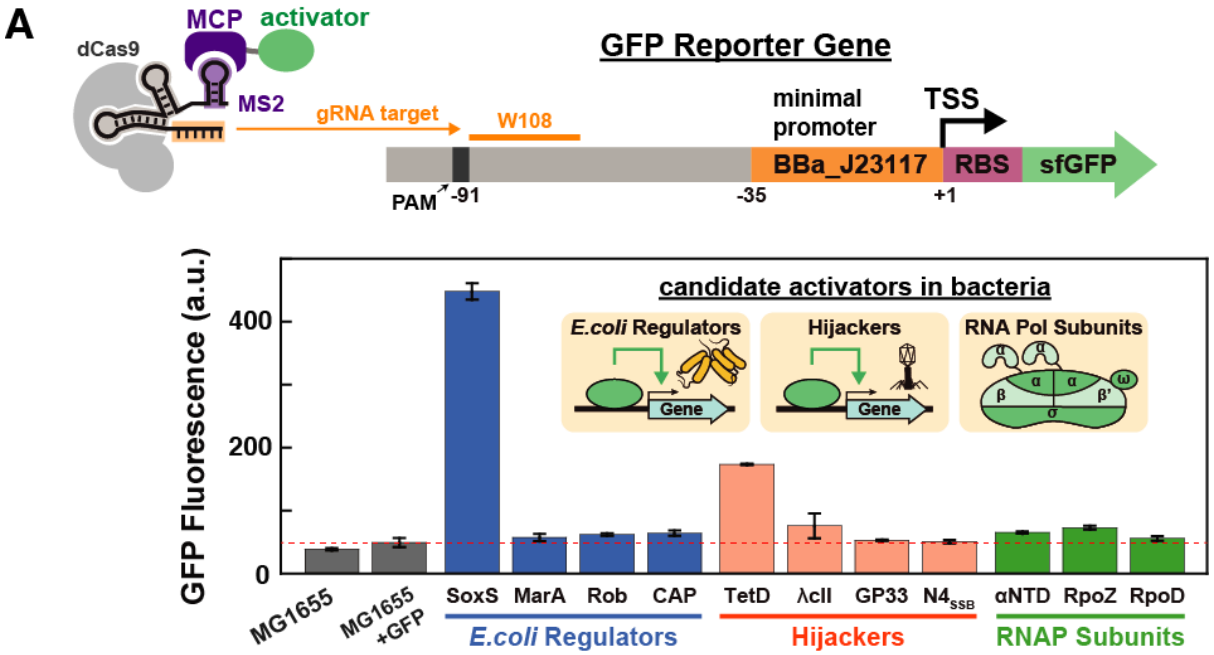


Figure 2.1 CRISPR activation in bacteria enables complex, multi-gene expression programs.

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.


A
MCP
activator
dCas9
MS2
gRNA target
W108
GFP Reporter Gene
minimal promoter
TSS
BBa_J23117
RBS
sfGFP
PAM
-91
-35
+1
candidate activators in bacteria
E.coli Regulators
Hijackers
RNA Pol Subunits
Gene
GFP Fluorescence (a.u.)
400
200
0
MG1655
MG1655 +GFP
SoxS
MarA
Rob
CAP
TetD
λcII
GP33
N4ssb
αNTD
RpoZ
RpoD
E.coli Regulators
Hijackers
RNAP Subunits

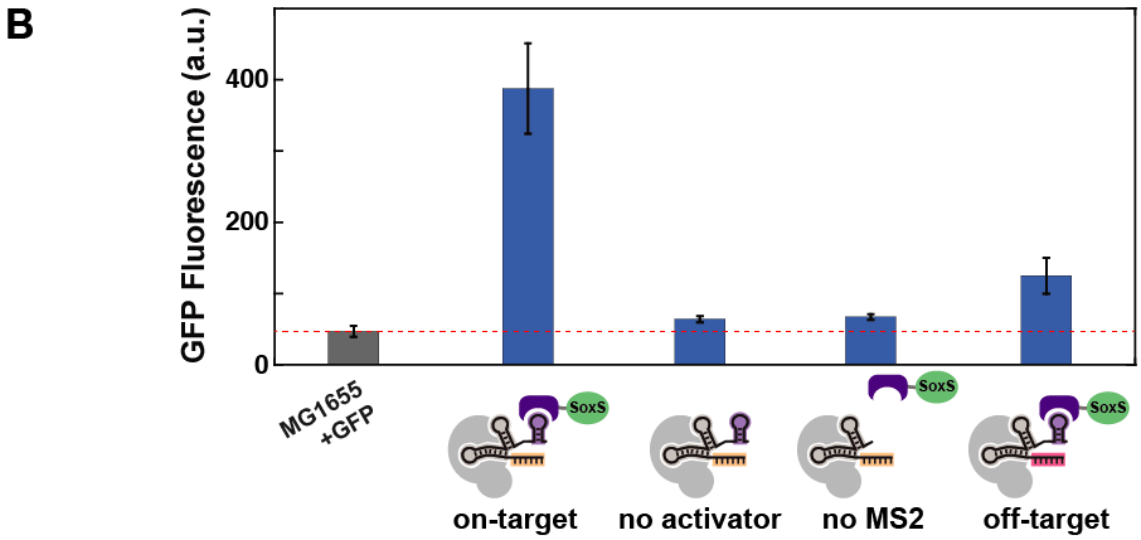

B
GFP Fluorescence (a.u.)
400
200
0
MG1655 +GFP
SoxS
on-target
no activator
no MS2
off-target

Figure 2.2 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

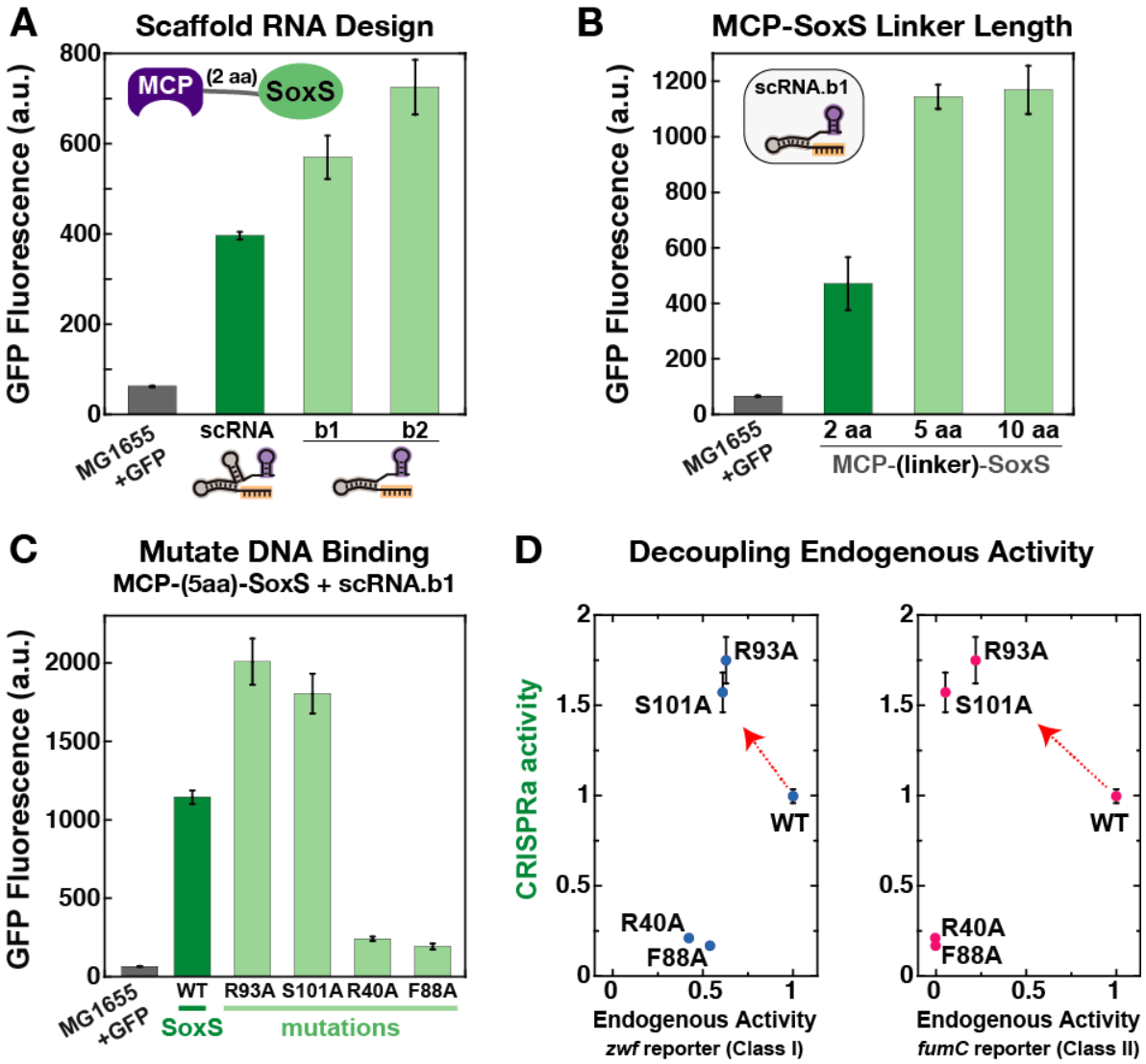


Figure 2.3 Optimization of gene activation.

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. 2.3C) 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. 2.3A & 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. 2.3A). 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. 2.3C).

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).

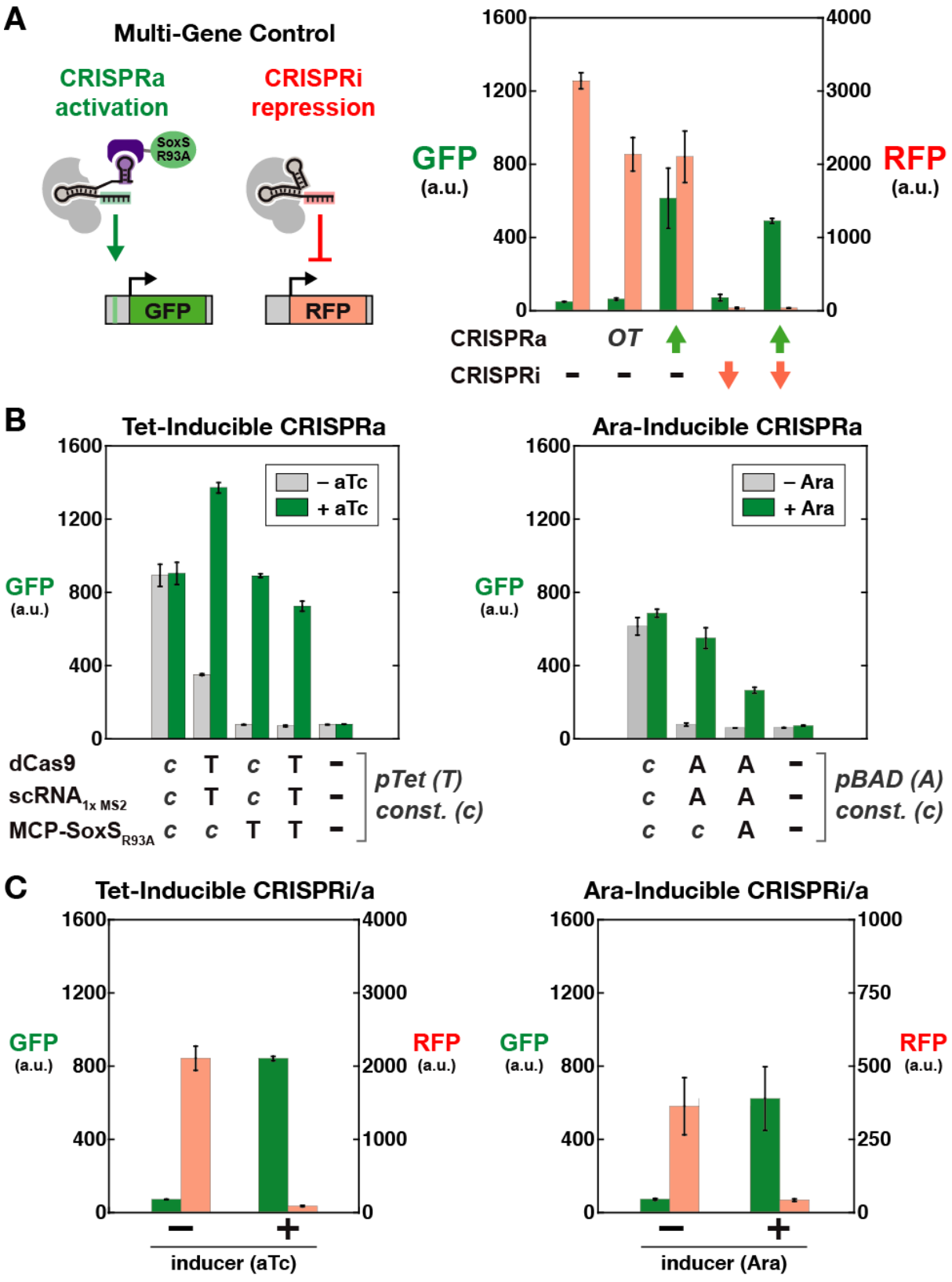


Figure 2.4 Multi-gene, inducible control.

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).

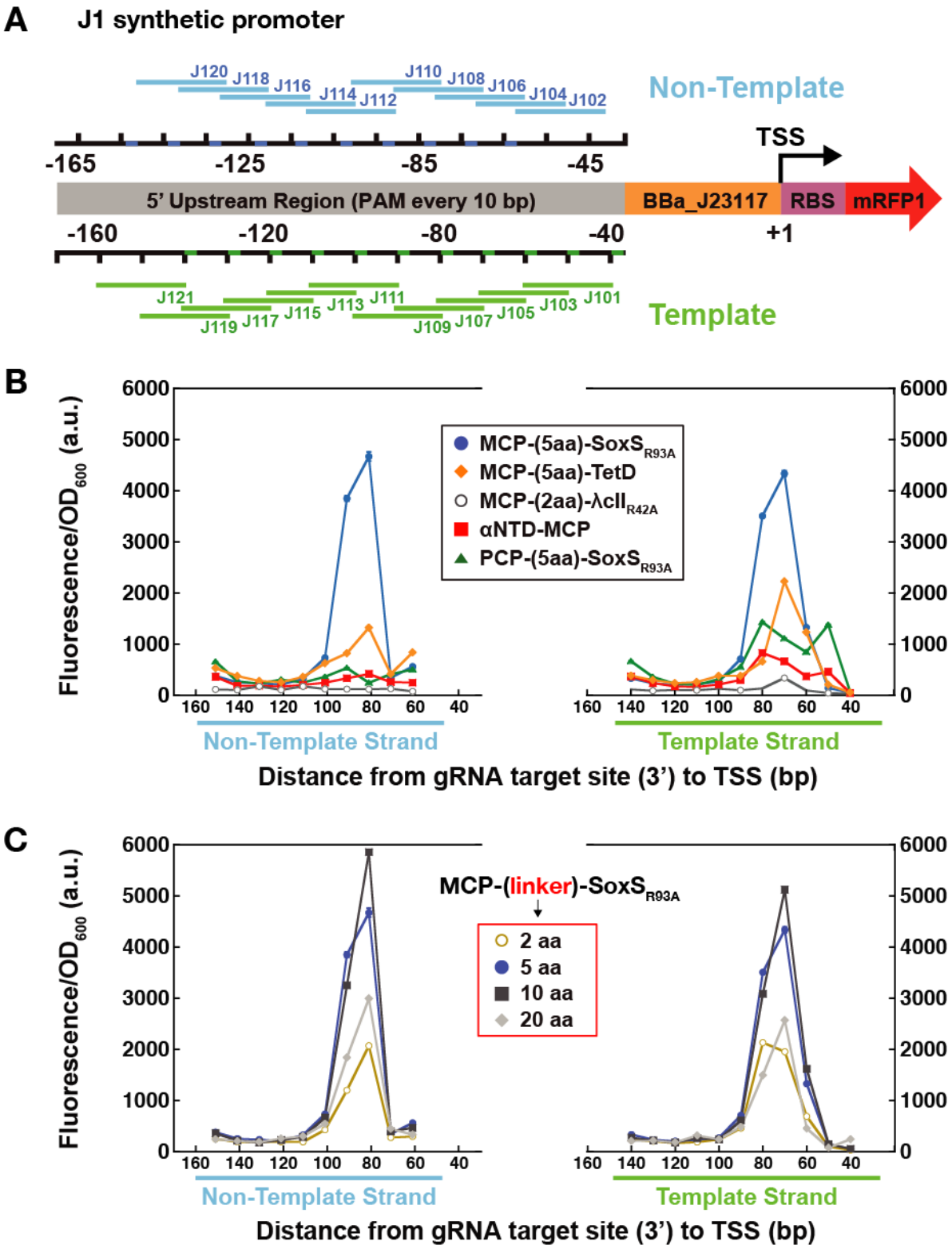


Figure 2.5 Gene activation is highly sensitive to CRISPR target site position.

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. 2.2D), MCP-(2aa)-λcII_{R42A} (see Supplementary Fig. 2.2E), 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 $\pm$ s.d. for at least three biological replicates.

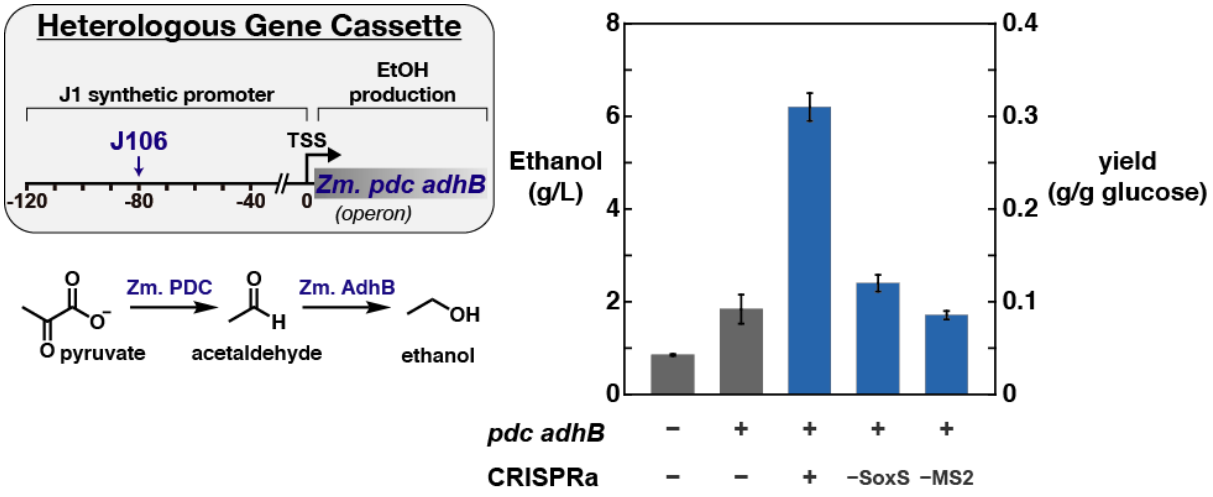
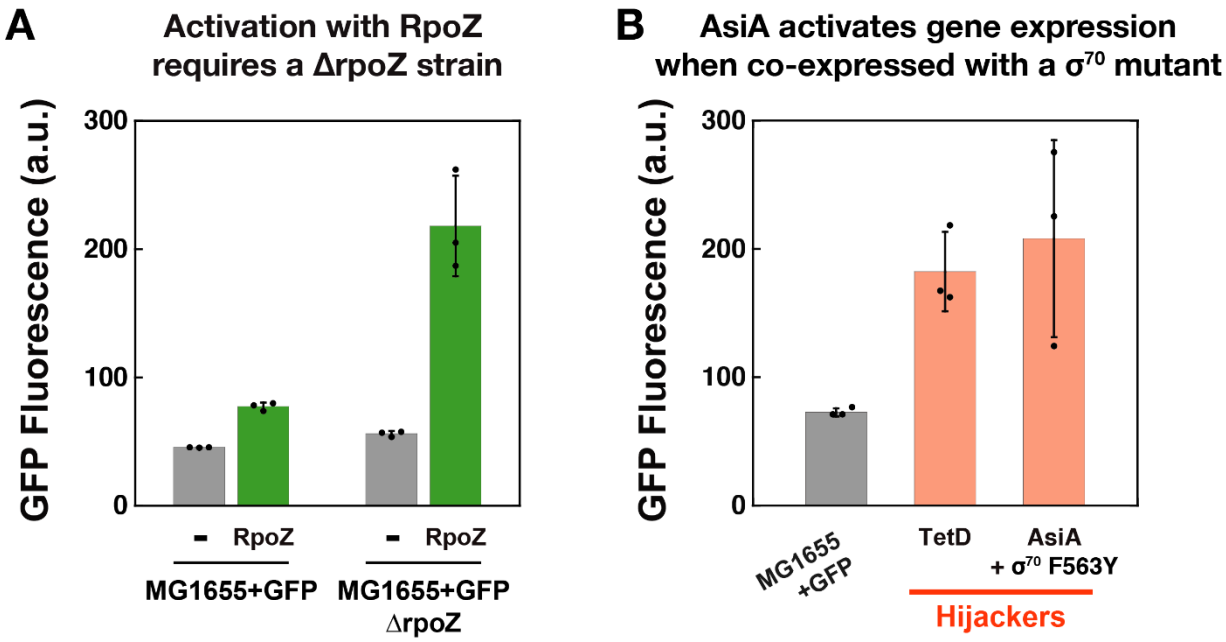


Figure 2.6 CRISPRa-mediated ethanol production.

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).

2.10 | Supplemental information

2.10.1 | Supplementary Figures

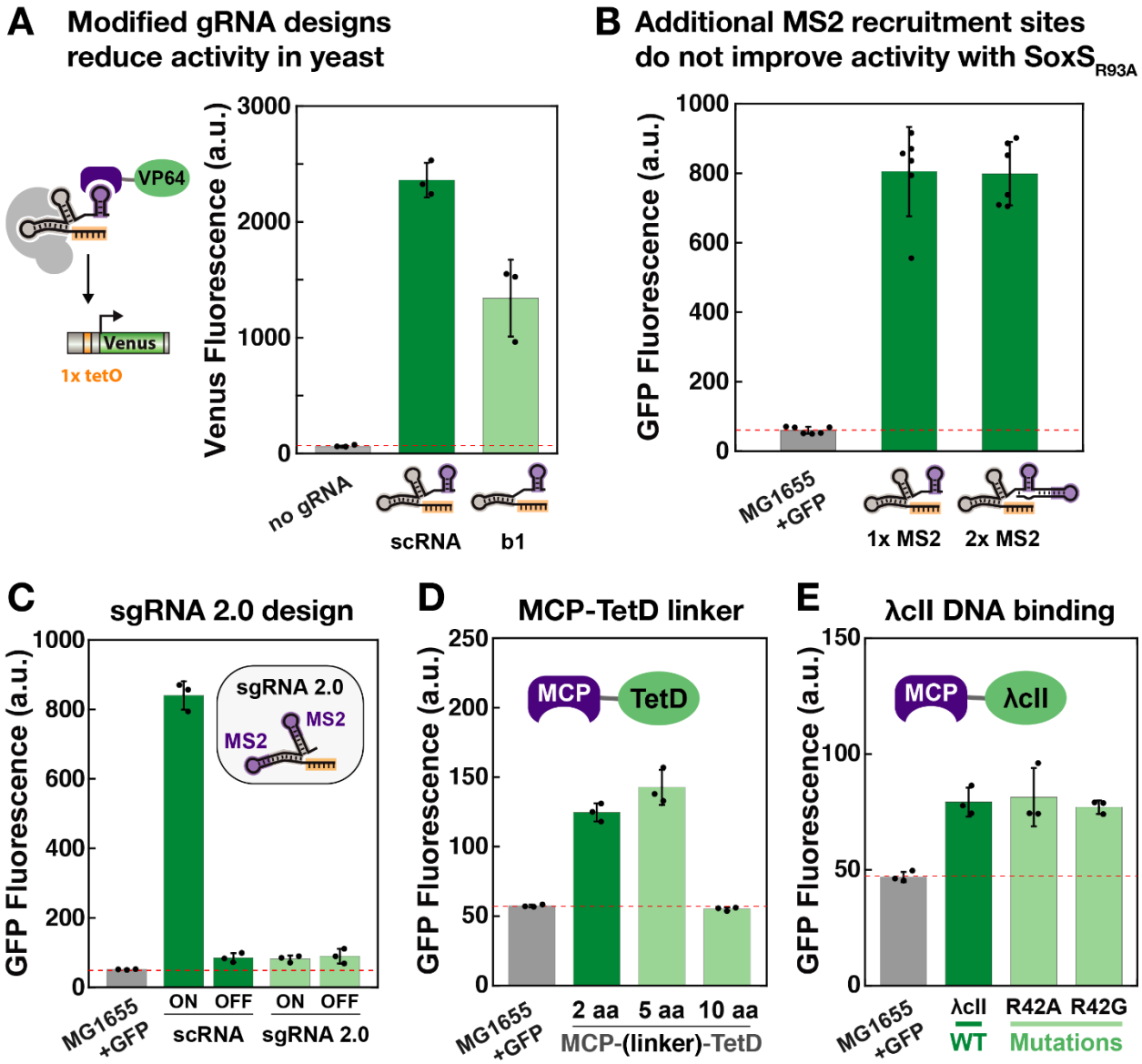


Supplementary Figure S2.1 Some candidate transcriptional activation domains require additional strain modifications for effective reporter gene activation.

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 S2.2 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].

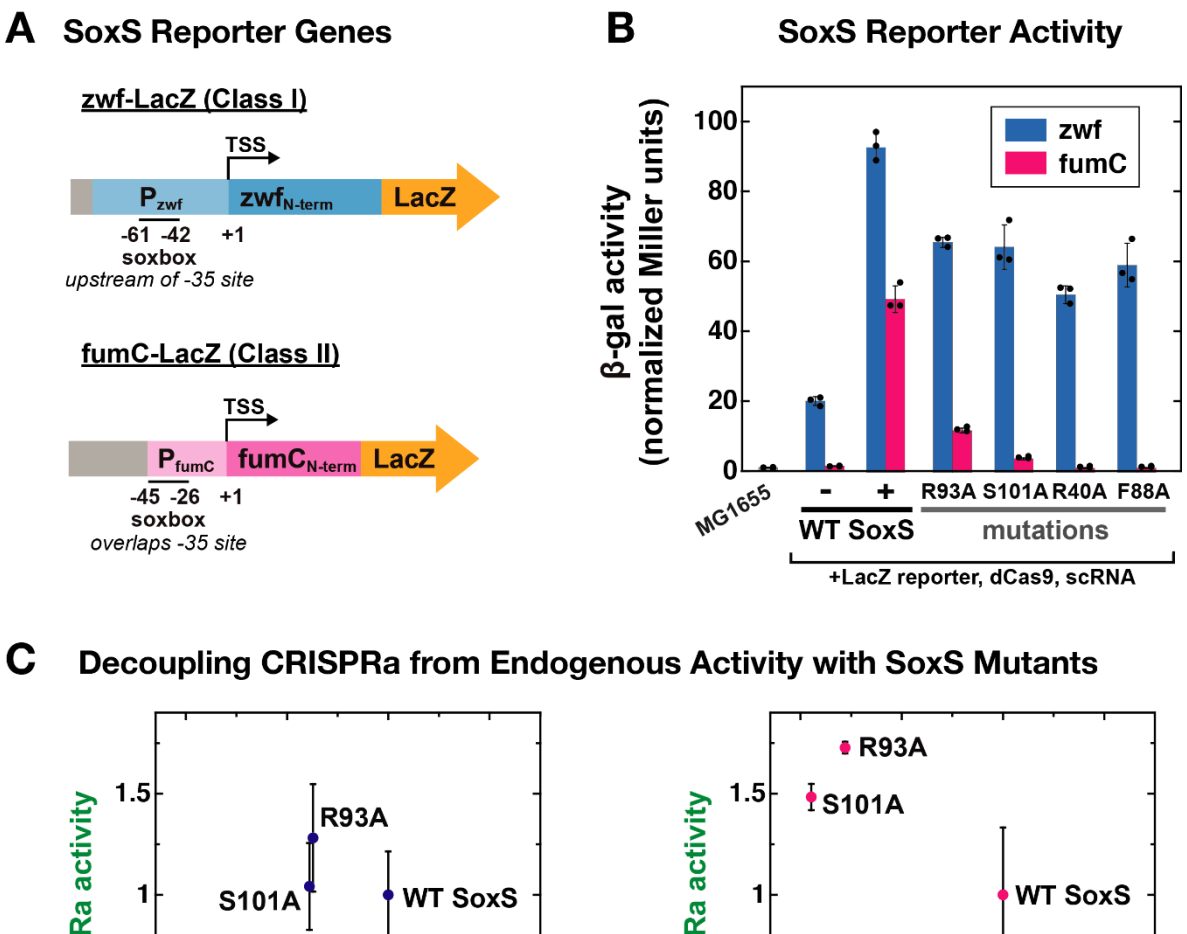
B) A 2x MS2 scRNA, which increases CRISPRa in yeast and human cells relative to a 1x MS2 scRNA[4], does not improve GFP expression with CRISPRa in *E. coli*. The activator in this experiment is MCP-(5aa)-SoxS$_{R93A}$ (Fig. 3).

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).



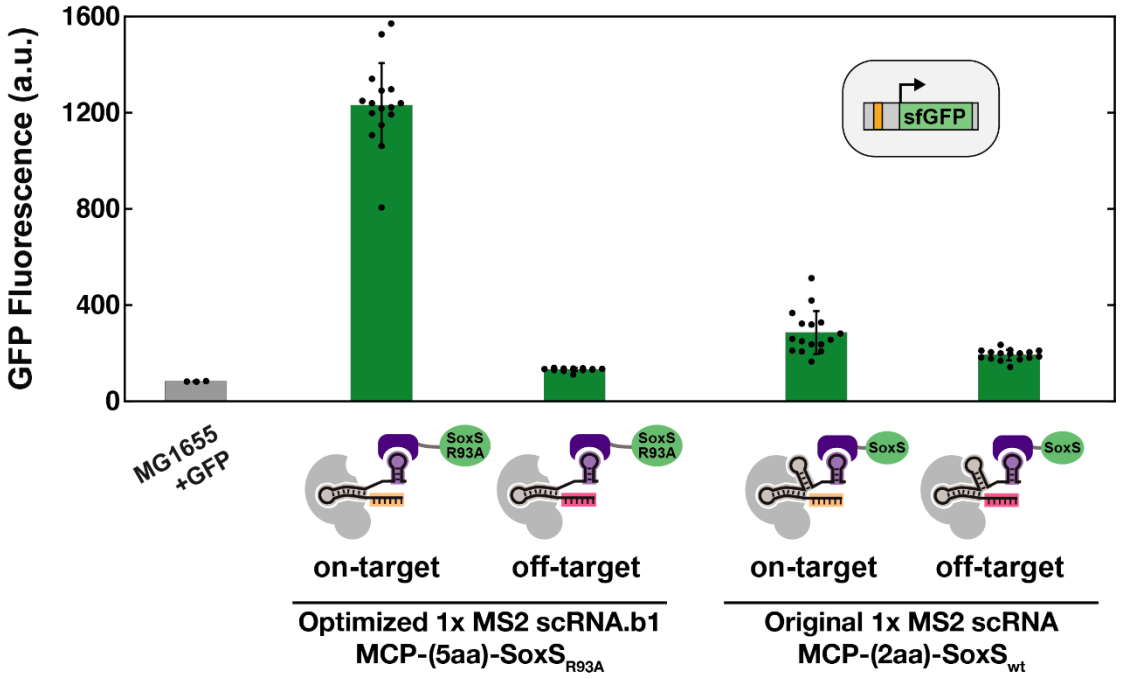
Supplementary Figure S2.3 Mutations in SoxS reduce activity at endogenous SoxS promoters.

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

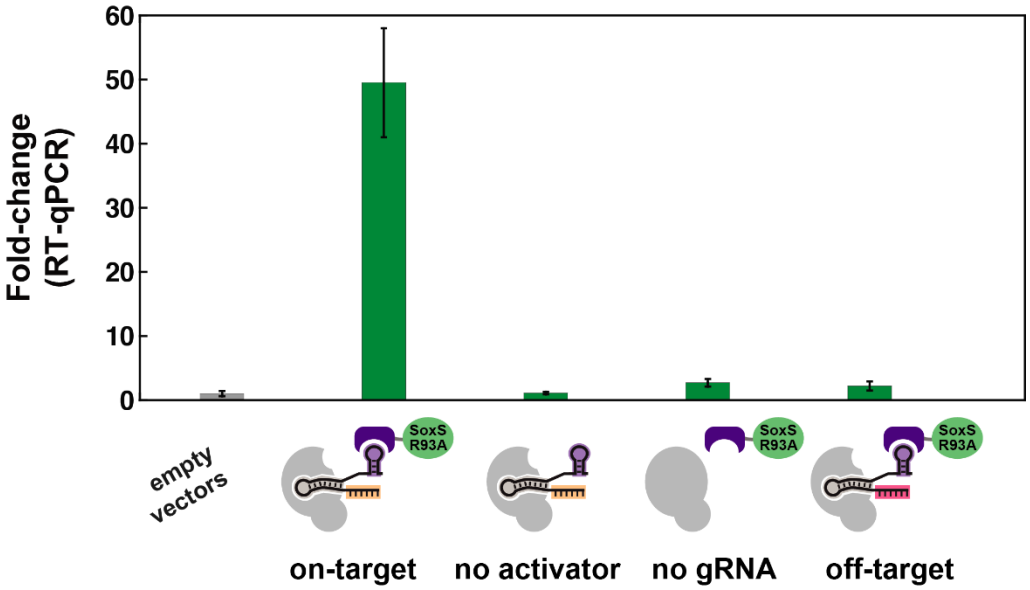
CRISPRa system components dCas9, a 1x MS2 scRNA.b1, and an MCP-SoxS fusion protein. Point mutants of SoxS that disrupt the DNA binding interface lead to reduced LacZ activity, and the results are consistent with previously reported activity trends for these mutants[7]. LacZ activity was measured as previously 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.



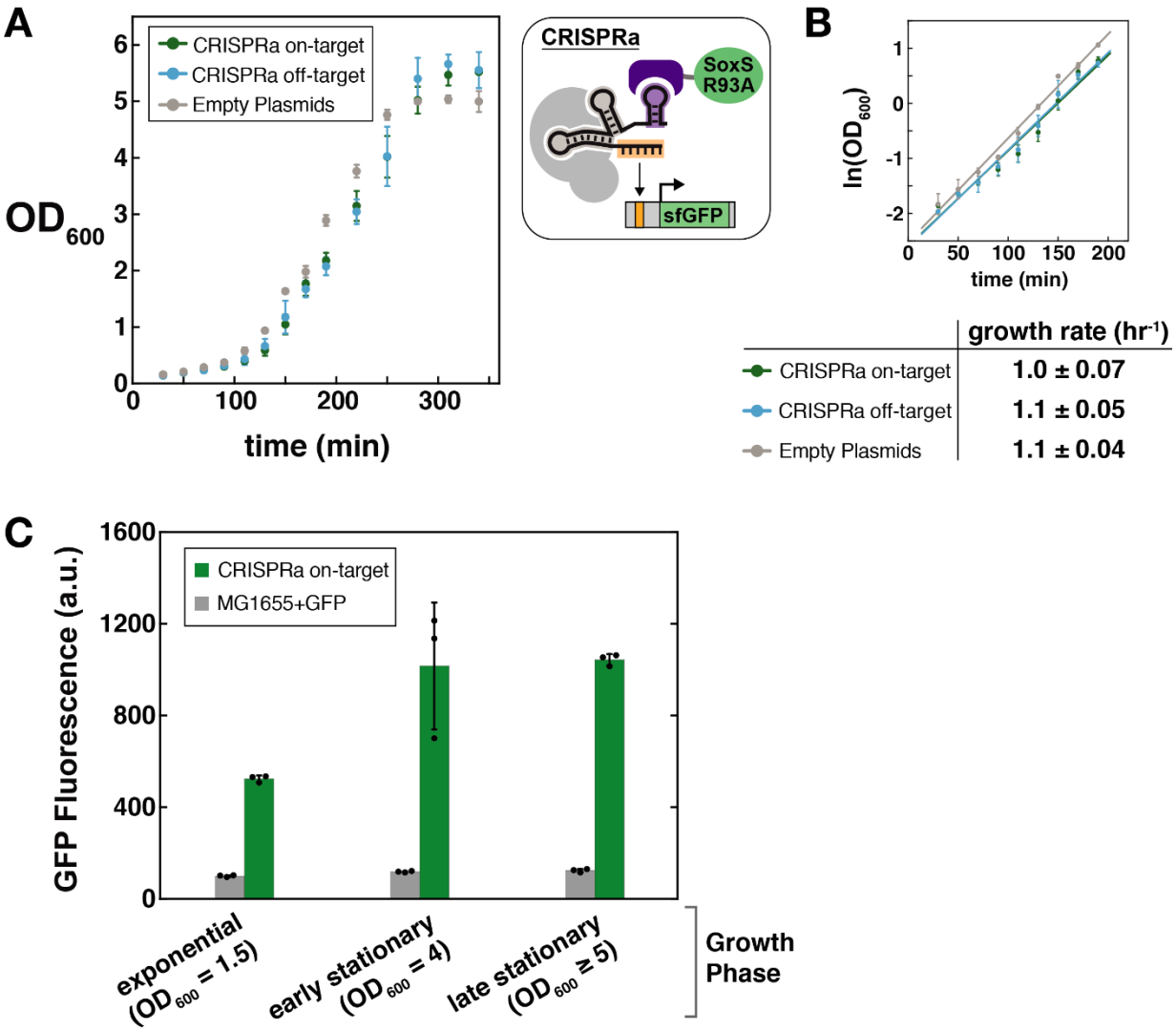
Supplementary Figure S2.4 An optimized CRISPRa system improves activity.

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 S2.5 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.



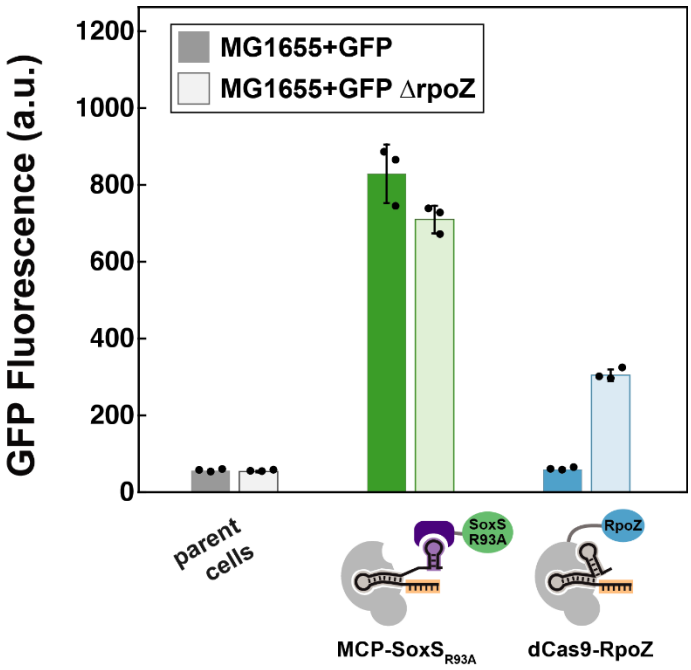
Supplementary Figure S2.6 Relationship between cell growth and CRISPRa.

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.

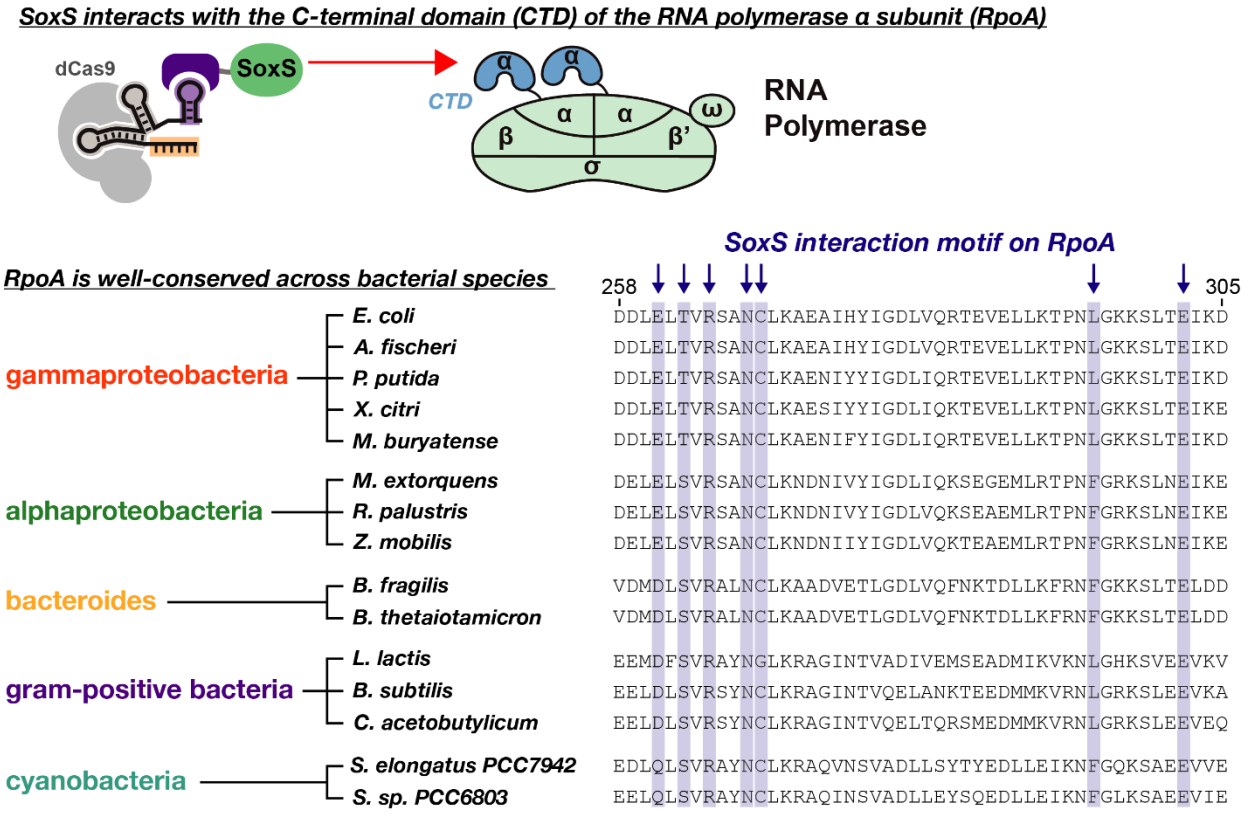
C) CRISPRa-mediated GFP expression increases as cells enter stationary phase, although gene expression is still detectable in exponential phase. Values reported are GFP fluorescence levels

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



Supplementary Figure S2.7 CRISPRa with fully optimized MCP-SoxS$_{R93A}$ is more effective than dCas9-RpoZ, a previously-described bacterial CRISPRa system[1].

MCP-SoxS$_{R93A}$ activates gene expression in an MG1655 strain, where dCas9-RpoZ 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).

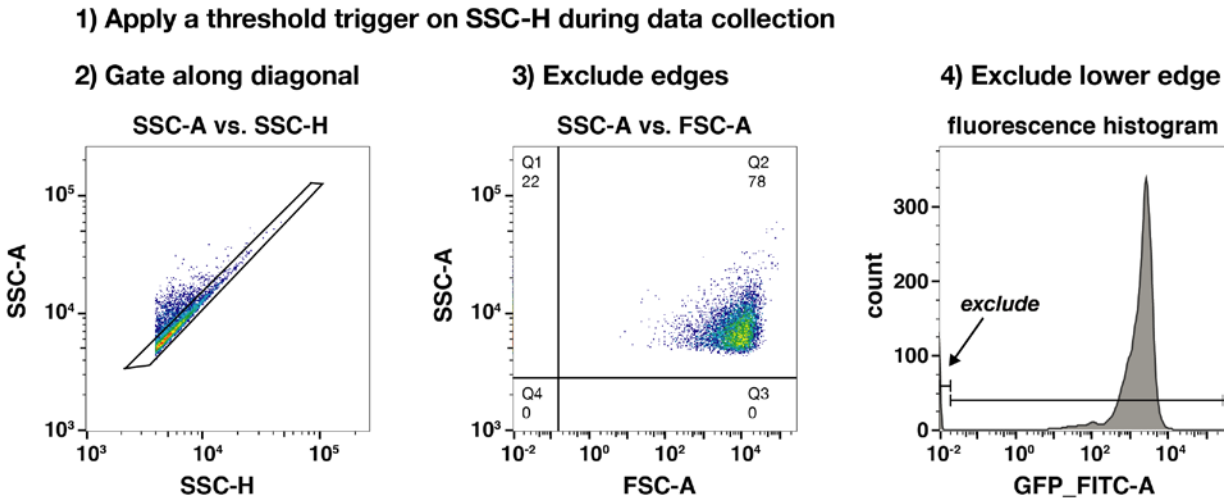


Supplementary Figure S2.8 The SoxS recruitment site on RNA polymerase is highly conserved.

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 S2.9 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.

2.10.2 | Supplementary Tables

Supplementary Table S2.1 *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* Δ*rpoZ*
CD10	MG1655/sfGFP (weak promoter)/ mRFP1 (strong promoter	MG1655 *W1-BBa_J23117-sfGFP::nfsA, BBa_J23119-mRFP1 KanR::rbsAR*

Supplementary Table S2.2 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 S2.3 *E. coli* Expression Plasmids

Plasmid[a]	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) BBa_J23107*	1) dCas9 2) MCP-rpoZ	*1) BBa_B0015* *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* *3) BBa_J23107*	1) dCas9 2) MCP-AsiA 3) rpoD (F563Y)	*1) BBa_B0015* *2) 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*	1) dCas9	*1) BBa_B0015*

			2) BBa_J23107	2) MCP-CAP	*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* *2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (S101A)	*1) 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	*1) BBa_B0015* *2) BBa_B1002*

				(R93A)	
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* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-TetD 3) 1x MS2 scRNA.b2 (J101~121 targets)	*1) BBa_B0015* *2) 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	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*

				(J101~121 targets)	
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 S2.4 Primer Sequences for RT-qPCR

Primer	Sequence
qCD010_sfGFP_f	GAGGGTGAAGGTGATGCTACAA
qCD011_sfGFP_r	GGTCAGAGTAGTGACAAGTGTTGG
qCD012_16S_f	AAAGTTAATACCTTTGCTCATTGACGTT
qCD013_16S_r	GACTACCAGGGTATCTAATCCTGTTT

2.10.3 | 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

GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

CAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

1x MS2 scRNA

GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

CAACTTGAAAAAGTGGCACCGAGTCGGTGCGCGCACATGAGGATCACCCATGTGCTTTTTTT

2x MS2 scRNA

GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

CAACTTGAAAAAGTGGCACCGAGTCGGTGCGGGAGCACATGAGGATCACCCATGTGCCACGAGCGA

CATGAGGATCACCCATGTCGCTCGTGTTCCCTTTTTTT

1x MS2 scRNA.b1

GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

CAACTTGAAAAAGTGCGCACATGAGGATCACCCATGTGCTTTTTTT

1x MS2 scRNA.b2

GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

CAACTTGAAAAAGTGGCACATGAGGATCACCCATGTGCTTTTTTT

1x PP7 scRNA.b1

GAAGATCCGGCCTGCAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT

CAACTTGAAAAAGTAACATAAGGAGTTTATATGGAAACCCTTATGTTTTTTT

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)

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY

LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSMSHQKIIQDLIAWIDEHIDQPLNIDVVAK

KSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFSRVFRRQFDRTPSDYR

HRL

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

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY

LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMSHQKIIQDLIAWIDEHIDQPLNIDV
VAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFSRVFARQFDRTPS
DYRHRL

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

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY
LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMYIEQHSRYQNKANNIQLRYDDKQFH
TTVIKDVLLWIEHNLDQSLLLDDVANKAGYTKWYFQRLFKKVTGVTLASYIRARRLTKAAVELRLTKKTILEIAL
KYQFDSQQSFTRRFKYIFKVTPSYYRRNKLWELEAMH

> MCP-λcII

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY
LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSMVRANKRNEALRIESALLNKIAMLGTEKT
AEAVGVDKSQISRWKRDWIPKFSMLLAVLEWGVVDDDMARLARQVAAILTNKKRPAATERSEQIQMEF

> MCP-RpoZ

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY
LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSARVTVQDAVEKIGNRFDLVLVAARRARQM
QVGGKDPLVPEENDKTTVIALREIEEGLINNQILDVRERQEQQEQEAAELQAVTAIAEGRR

> αNTD-MCP

MQGSVTEFLKPRLVDIEQVSSTHAKVTLEPLERGFGHTLGNALRRILLSSMPGCAVTEVEIDGVLHEYSTKEGVQ
EDILEILLNLKGLAVRVQGKDEVILTLNKSGIGPVTAADITHDGDVEIVKPQHVICHLTDENASISMRIKVQRGR
GYVPASTRIHSEEDERPIGRLLVDACYSPVERIAYNVEAARVEQRTDLDKLVIEMETNGTIDPEEAIRRAATILA
EQLEAFVDLRDVRQPEVKEEKPEASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS
AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY

> MCP-AsiA

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY
LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSMNKNIDTVREIITVASILIKFSREDIVEN
RANFIAFLNEIGVTHEGRKLNQNSFRKIVSELTQEDKKTLIDEFNEGFEGVYRYLEMYTNK

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

MGPSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRLRLTASLRQNGAKTAYRVNLKLDQADVVDSGLPKV

RYTQVWSHDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLVPLGRGGGGSMSHQKIIQDLIAWIDEHIDQP

LNIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFSRVFARQF

DRTPSDYRHRL

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

GCATGCCCGATCAACGTCTCATTTTCGCCAGATATCAAGCAGAGGAGCAAAAGCTCATTTCTGAAGAGGACTTGT
TGCGGAAACGACGAGAACAGTTGAAACACAAACTTGAACAGCTACGGAACTCTTGTGCGTAAGGAAAAGTAAGGA
AAACGATTCCTTCTAACAGAAATGTCCTGAGCAATCACCTATGAACTGTCGACTCGAGCCTCTATGGATTATCAC
CTTGGCTGCAGGCCGGATCTTCCACAACACGCACGGTGTTACATTAGGCATACCGGTCttgacagctagctcagt
cctagggattgtgctagcGAATTCATTAAAGAGGAGAAAGGTACCATGAGCAAAGGAGAAGAACTTTTCACTGGA
GTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGT
GATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTT
GTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAG
AGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCT
GAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAAC
ATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGA
ATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAAT
ACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGAT
CCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAG
CTCTACAAAtaaggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagact
gggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggc
ctttctgcgtttata

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

GCCCTCTAGAGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAAT

ttgacagctagctcagtcctaggtataatagatctGAATTCATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCG

AAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCG

AAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGC

CGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCC

CGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTG

TTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCC

CGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACG

GTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCA

CCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACA

ACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAggatccaaac

tcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttg

tttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata

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

actcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtat
ttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgtggcaattccgacgtcGCC
TACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTGACCTATGG
TGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCGTCCTTTGGGTTCC
ACCGGATACCTCCGGACttgacagctagctcagtcctagggattgtgctagcGAATTCATTAAAGAGGAGAAAGG
TACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGG
TCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTAC
CAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAA
ACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTT
CGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACT
GCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACG
TATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGA
CGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACT
GGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGG
TGCTTAAggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcc
tttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttc
tgcgtttata

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_{1-161}, LacZ, BBa_B0015 terminator. The soxbox is underlined.

GATCAGTGTCAGATTTTTACCCAATGGAAAACGATGATTTTTTTATCAGTTTTGCCGCACTTTGCGCGCTTTTCCCGTAATCG
CACGGGTGGATAAGCGTTTACAGTTTTCGCAAGCTCGTAAAAGCAGTACAGTGCACCGTAAGAAAATTACAAGTATACCCTGG
CTTAAGTACCGGGTTAGTTAACTTAAGGAGAATGACATGGCGGTAACGCAAACAGCCCAGGCCTGTGACCTGGTCATTTTCGG
CGCGAAAGGCGACCTTGCGCGTCGTAAATTGCTGCCTTCCCTGTATCAACTGGAAAAAGCCGGTCAGCTCAACCCGGACACCC
GGATTATCGGCGTAGGGCGTGCTGACTGGGATAAAGCGGCATATACCAAAGTTGTCCGCGAGGCGCTCGAAACTTTCATGAAA
GAAACCATTGATGAAGGTTTATGGGACACCCTGAGTGCACGTCTGGATTTTTGTAATCTCGATGTCAATGACACTGCTGCATT
CAGCCGTCTCGGCGCGATGCTGGATCAAAAAAATCGTATCACCATTAACTACTTTGCCATGCCGCCCAGCACTTTTGGCGCAA
TTTGCAAAGGGCTTGGCGAGGCAAAACTGAATGCTAAACCGGCACGCGTAGTCATGGAGAAACCGCTGGGGACGTCGCTGGCG
ACCTCGCAGGAAATCAATGACATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGG
CGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTT
CCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAG
TGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGT
GACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATG
AAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGG
GTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGT
GATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGT
TGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACTGGAG
GCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAG
CGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAA
ACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAA
GCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCG
AGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATGA
AGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCTG
TATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGC

GATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCAG

GCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGA

GCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATG

GTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTC

TTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAG

TCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTT

CTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGT

TCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGGTG

GCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGA

ACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAGAAGCCG

GGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCAT

CTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACA

GATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCG

TAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTG

TTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATT

TATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGC

ATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAAC

TATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAG

CGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCC

GCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGT

TTCCATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTA

CCAGTTGGTCTGGTGTCAAAAATAAggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaa

agactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctt

tctgcgtttata

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.

ATGGCACGAAAGACCAAACATTTGTTATCAAATGGTAAATAATAAGTGAGCTAAAAGTTGCTTAACGAAAGCAAAACAGAAAG
AAAAAATTAATCAGGTGAGGAGCAGGTCATGAATACAGTACGCAGCGAAAAAGATTCGATGGGGGCGATTGATGTCCCGGCAG
ATAAGCTGTGGGGCGCACAAACTCAACGCTCGCTGGAGCATTTCCGCATTTCGACGGAGAAAATGCCCACCTCACTGATTCAT
GCGCTGGCGCTAACCAAGCGTGCAGCGGCCATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGA
AAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGC
TGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTA
CACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTA
ATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAAC
GGGCGCTGGGTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCG
CCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTG
ACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCT
GTACTGGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCA
GGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGA
ACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACG
CTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTT
GCTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCC
TGCTGATGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGC
TACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTG
GCTACCGGCGATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGA
ATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAA
GGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGT
GCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGG
GTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGACTGG

GTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGA
TCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGT
TTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCAC
TGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGA
ACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGT
CAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCC
ATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTT
TCTTTCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACG
ACATTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAA
GCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAA
AACCTTATTTATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCG
ATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCG
CAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGT
CTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCA
ACATCAGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAAT
ATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCG
CTACCATTACCAGTTGGTCTGGTGTCAAAAATAAggatccaaactcgagtaaggatctccaggcatcaaataaaacgaaaggc
tcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgg
gtgggcctttctgcgtttata

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

gcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctc
atgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtg
ccacctgACGTCGCGGCCGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTG
GTTGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTAC
CAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACttgacagctagctcagtcctagggattgtgctagcGA
ATTCATTAAAGAGGAGAAAGGTACCatgagttatactgtcggtacctatttagcggagcggcttgtccagattgg

tctcaagcatcacttcgcagtcgcgggcgactacaacctcgtccttcttgacaacctgcttttgaacaaaaacat

ggagcaggtttattgctgtaacgaactgaactgcggtttcagtgcagaaggttatgctcgtgccaaaggcgcagc

agcagccgtcgttacctacagcgtcggtgcgctttccgcatttgatgctatcggtggcgcctatgcagaaaacct

tccggttatcctgatctccggtgctccgaacaacaatgaccacgctgctggtcacgtgttgcatcacgctcttgg

caaaaccgactatcactatcagttggaaatggccaagaacatcacggccgccgctgaagcgatttataccccgga

agaagctccggctaaaatcgatcacgtgattaaaactgctcttcgtgagaagaagccggtttatctcgaaatcgc

ttgcaacattgcttccatgccctgcgccgctcctggaccggcaagcgcattgttcaatgacgaagccagcgacga

agcttctttgaatgcagcggttgaagaaaccctgaaattcatcgccnaccgcgacaaagttgccgtcctcgtcgg

cagcaagctgcgcgcagctggtgctgaagaagctgctgtcaaatttgctgatgctcttggtggcgcagttgctac

catggctgctgcaaaaagcttcttcccagaagaaaacccgcattacatcggtacctcatggggtgaagtcagcta

tccgggcgttgaaaagacgatgaaagaagccgatgcggttatcgctctggctcctgtctttaacgactactccac

cactggttggacggatattcctgatcctaagaaactggttctcgctgaaccgcgttctgtcgtcgttaacggcat

tcgcttccccagcgtccatctgaaagactatctgacccgtttggctcagaaagtttccaagaaaaccggtgcttt

ggacttcttcaaatccctcaatgcaggtgaactgaagaaagccgctccggctgatccgagtgctccgttggtcaa

cgcagaaatcgcccgtcaggtcgaagctcttctgaccccgaacacgacggttattgctgaaaccggtgactcttg

gttcaatgctcagcgcatgaagctcccgaacggtgctcgcgttgaatatgaaatgcagtggggtcacattggttg

gtccgttcctgccgccttcggttatgccgtcggtgctccggaacgtcgcaacatcctcatggttggtgatggttc

cttccagctgacggctcaggaagtcgctcagatggttcgcctgaaactgccggttatcatcttcttgatcaataa

ctatggttacaccatcgaagttatgatccatgatggtccgtacaacaacatcaagaactgggattatgccggtct

gatggaagtgttcaacggtaacggtggttatgacagcggtgctggtaaaggcctgaaggctaaaaccggtggcga

actggcagaagctatcaaggttgctctggcaaacaccgacggcccaaccctgatcgaatgcttcatcggtcgtga

agactgcactgaagaattggtcaaatggggtaagcgcgttgctgccgccaacagccgtaagcctgttaacaagct

cctctaacaattcaaaaGATCTAAAGAGGAGAAATCTAGAatggcttcttcaacttttatattcctttcgtcaa

cgaaatgggcgaaggttcgcttgaaaaagcaatcaaggatcttaacggcagcggctttaaaaatgcgctgatcgt

ttctgatgctttcatgaacaaatccggtgttgtgaagcaggttgctgacctgttgaaagcacagggtattaattc

tgctgtttatgatggcgttatgccgaacccgactgttaccgcagttctggaaggccttaagatcctgaaggataa

caattcagacttcgtcatctccctcggtggtggttctccccatgactgcgccaaagccatcgctctggtcgcaac

caatggtggtgaagtcaaagactacgaaggtatcgacaaatctaagaaacctgccctgcctttgatgtcaatcaa

cacgacggctggtacggcttctgaaatgacgcgtttctgcatcatcactgatgaagtccgtcacgttaagatggc

cattgttgaccgtcacgttaccccgatggtttccgtcaacgatcctctgttgatggttggtatgccaaaaggcct

gaccgccgccaccggtatggatgctctgacccacgcatttgaagcttattcttcaacggcagctactccgatcac

cgatgcttgcgctttgaaagcagcttccatgatcgctaagaatctgaagaccgcttgcgacaacggtaaggatat

gccggctcgtgaagctatggcttatgcccaattcctcgctggtatggccttcaacaacgcttcgcttggttatgt

ccatgctatggctcaccagttgggcggttactacaacctgccgcatggtgtctgcaacgctgttctgcttccgca

tgttctggcttataacgcctctgtcgttgctggtcgtctgaaagacgttggtgttgctatgggtctcgatatcgc

caatctcggtgataaagaaggcgcagaagccaccattcaggctgttcgcgatctggctgcttccattggtattcc

agcaaacctgaccgagctgggtgctaagaaagaagatgtgccgcttcttgctgaccacgctctgaaagatgcttg

tgctctgaccaacccgcgtcagggtgatcagaaagaagttgaagaactcttcctgagcgctttctaaggatccaa

actcgagtaaggatctCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGT

TGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA

2.10.4 | 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 3 | Regulated expression of sgRNAs tunes CRISPRi in *E. coli*

The work in this chapter was published and reprinted with permission from the journal of initial publication (License Number: 4574981073453):

Fontana, J., Dong, C., Ham, J. Y., Zalatan, J. G. & Carothers, J. M. Regulated Expression of sgRNAs Tunes CRISPRi in E. coli. *Biotechnology Journal* **13,** 1800069 (2018).

3.1 | 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.

3.2 | 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 3.1A). 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 3.1B). 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.

3.3 | Materials and methods

3.3.1 | 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.

3.3.2 | 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.

3.3.3 | 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.

3.3.4 | Statistical analysis

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

3.4 | Results

3.4.1 | 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 3.2). 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 3.2, Figure S3.1). 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 3.2, Figure S3.1). 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 3.2, Figure S3.1), 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 S3.1), 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.

3.4.2 | 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 3.3A). 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 3.3A, Figure S3.2), ranging from 4.6- to 74-fold repression (Figure S3.3). 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 3.3A), 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 S3.4). 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 3.3B). A ~2.5-fold change in sgRNA input alone was sufficient to traverse the observed 16-fold change in reporter expression (Figure 3.3B, 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 3.3B, grey). We performed the same experiment at a distinct gene target (sfGFP) and we obtained comparable results (Figure S3.5), 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.

3.4.3 | 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 3.4). 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 3.4). 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.

3.5 | 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.

3.6 | 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.).

3.7 | Conflict of interest

The authors declare no financial or commercial conflict of interest.

3.8 | 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.

3.9 | Figures

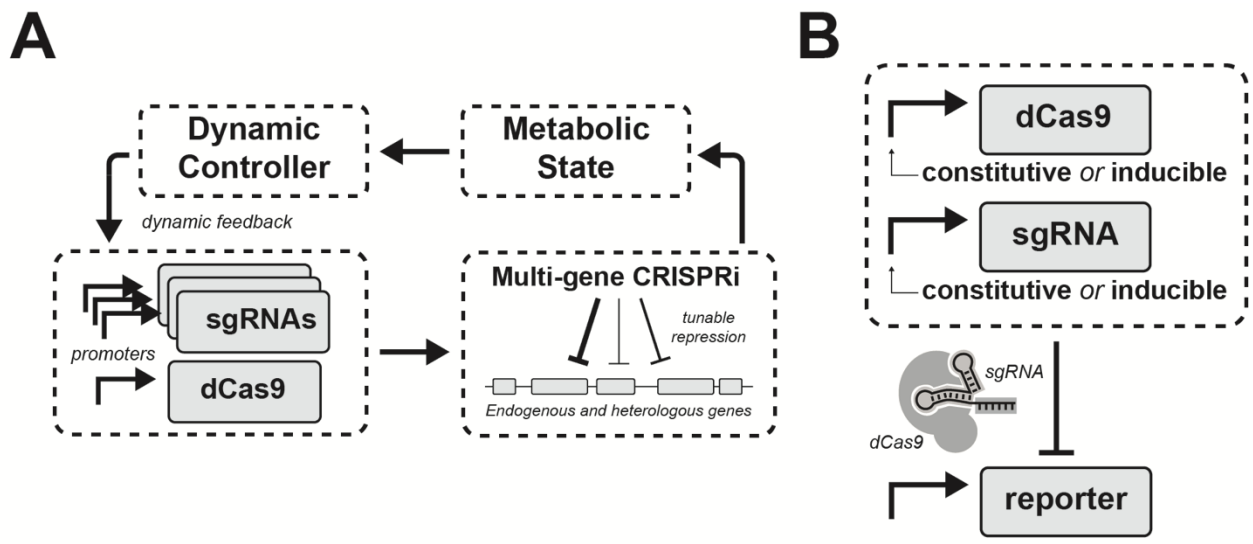


Figure 3.1

A) 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.

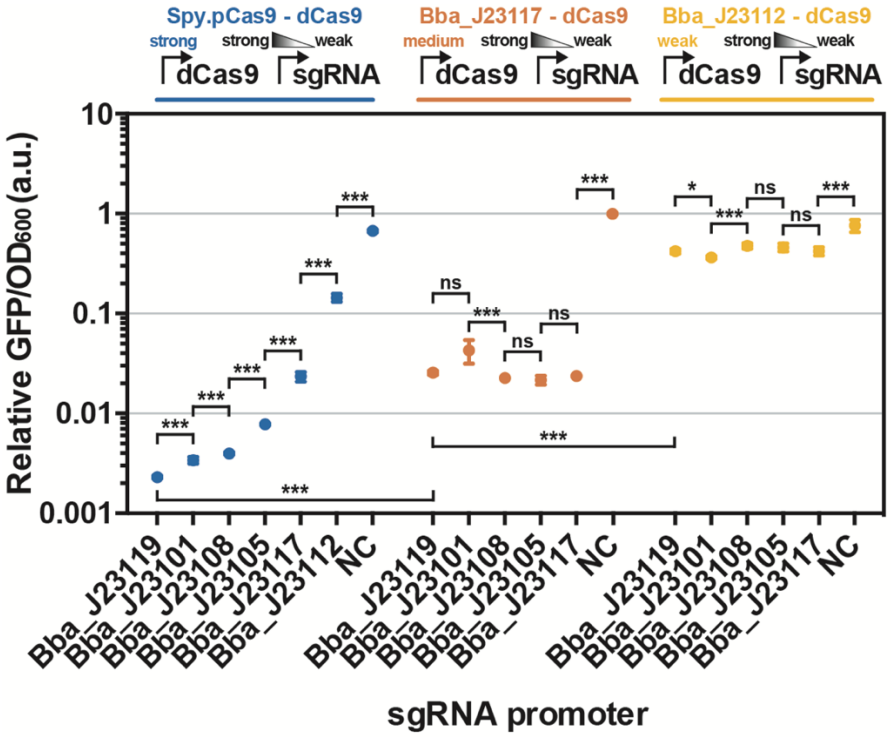


Figure 3.2 Tuning CRISPRi repression by titrating the expression of dCas9 and sgRNA using combinations of synthetic constitutive promoters.

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 S3.1 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).

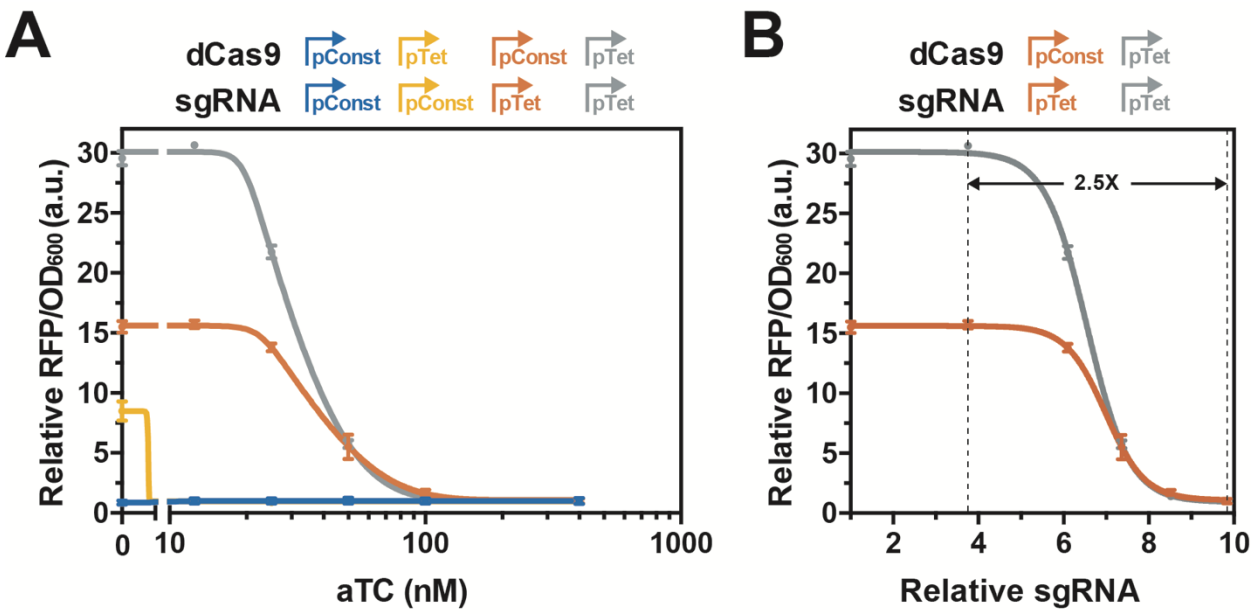


Figure 3.3 Analysis of the dynamic ranges in dCas9 and/or sgRNA inputs required for tuning CRISPRi.

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 S3.4. aTc: anhydrotetracycline. Data are plotted against the mean of three replicates with standard deviations.

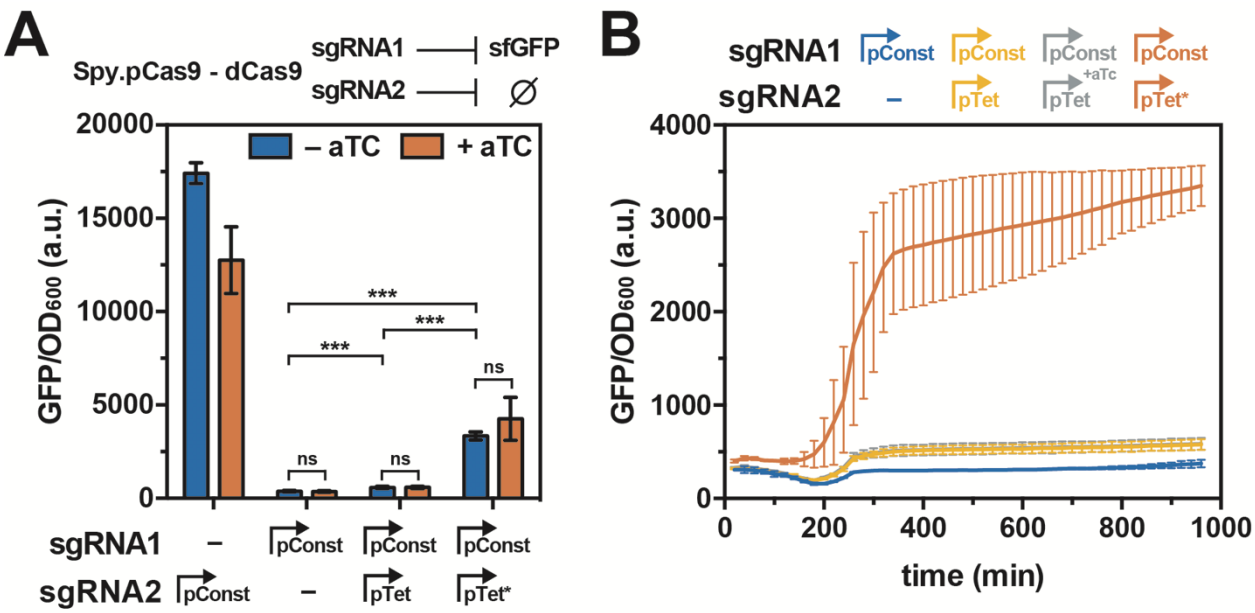


Figure 3.4 Persistence of CRISPRi repression upon expression of a competing sgRNA.

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).

3.10 | Supplementary information

3.10.1 | Supplementary Tables

Supplementary Table S3.1 *E. coli* plasmids included in this study.

Plasmid	Marker	Origin	Promoter	Gene Terminator	Figure
pJF001	AmpR	ColE1	Bba_J23119	sgRNA (GFP) TrrnB	2, 4, S3.1, S3.5
pJF002	AmpR	ColE1	1) TetR-pTet 2)	1) sgRNA (RR2) 1) TrrnB	4

			Bba_J23119	2) sgRNA (GFP) 2) TrrnB	
pJF003	AmpR	ColE1	1) TetR-pTet 2) Bba_J23117	1) sgRNA (RR2) 1) TrrnB 2) sgRNA (GFP) 2) TrrnB	4
pJF009	AmpR	ColE1	n/a	n/a n/a	2, S3.1
pJF015	AmpR	ColE1	Bba_J23117	sgRNA (GFP) TrrnB	2, S3.1
pJF016	AmpR	ColE1	Bba_J23105	sgRNA (GFP) TrrnB	2, S3.1
pJF017	AmpR	ColE1	Bba_J23108	sgRNA (GFP) TrrnB	2, S3.1
pJF018	AmpR	ColE1	Bba_J23101	sgRNA (GFP) TrrnB	2, S3.1
pJF020	AmpR	ColE1	1) pTet 2) Bba_J23119	1) sgRNA (RR2) 1) TrrnB 2) sgRNA (GFP) 2) TrrnB	4
pJF024	AmpR	p15A	TetR-pTet	sgRNA (GFP) TrrnB	S3.5
pJF028	AmpR	ColE1	Bba_J23112	sgRNA (GFP) TrrnB	2, S3.1
pJF043	ChlorR	p15A	n/a	n/a n/a	S3.4
pCD016	AmpR	ColE1	Bba_J23119	sgRNA (RR2) TrrnB	3
pCD017	ChlorR	p15A	Spy.pCas9	dCas9 db1	2, 3, 4, S3.1, S3.2, S3.3, S3.5
pCD032	AmpR	ColE1	TetR-pTet	sgRNA (RR2) TrrnB	3, S3.2, S3.3
pCD035	ChlorR	p15A	Bba_J23112	dCas9 db1	2, S3.1
pCD048	ChlorR	p15A	Bba_J23117	dCas9 db1	2, S3.1
pSLQ1055a	ChlorR	p15A	TetR-pTet	dCas9 db1	3, S3.5
pBbE2Ab .	AmpR .	ColE1 -	TetR-pTet	mRFP1 db1	S3.4

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

Supplementary Table S3.2 *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*

Supplementary Table S3.3 sgRNA sequences included in this study.

sgRNA name	Target gene	DNA Sequence	Target Strand
RR2a	mRFP1	TGGAACCGTACTGGAACTGC	Non-template
GFPa	sfGFP	CATCTAATTCAACAAGAATT	Non-template

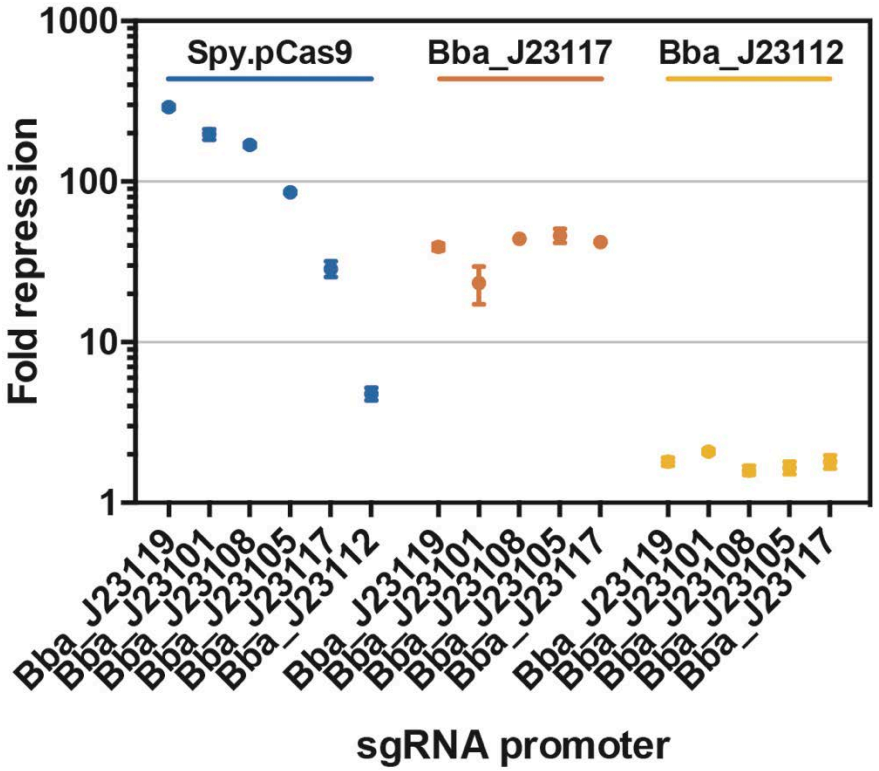
a: Reference [4]

Supplementary Table S3.4 Dynamic ranges in gene expression generated by selected available dynamic genetic controllers.

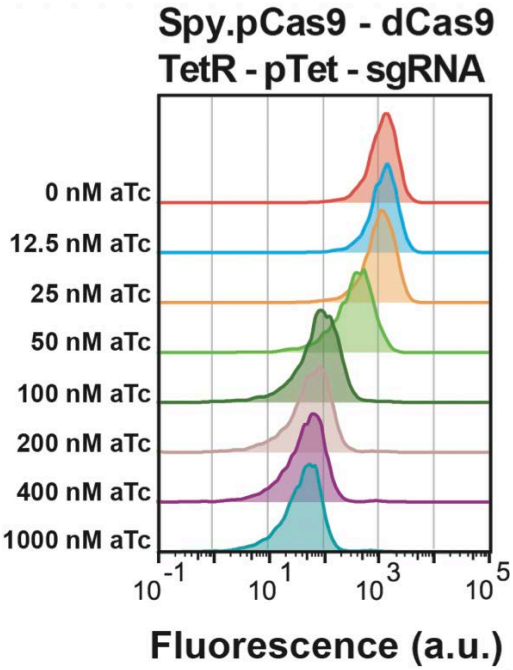
Controller	Class	Mode of induction	Developed in	DRa
glnAp2b	TF and promoter	Accumulation of acetyl-phosphate	*E. coli* BW13711 ΔglnlL	12
S-P promotersc	Promoter	Entry to stationary phase	E. coli K12 DH5α	4-6800
FadR-pARd	TF and promoter	Fatty-acid/acyl-coA relieve FadR repression	*E. coli* DH1 ΔfadE	25-60
aREDse	Ligand- dependent ribozymes	Ligand binding decreases RNA degradation	*E. coli* BL21 (DE3)	2
Synthetic riboswitchesf	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].

3.10.2 | Supplementary Figures



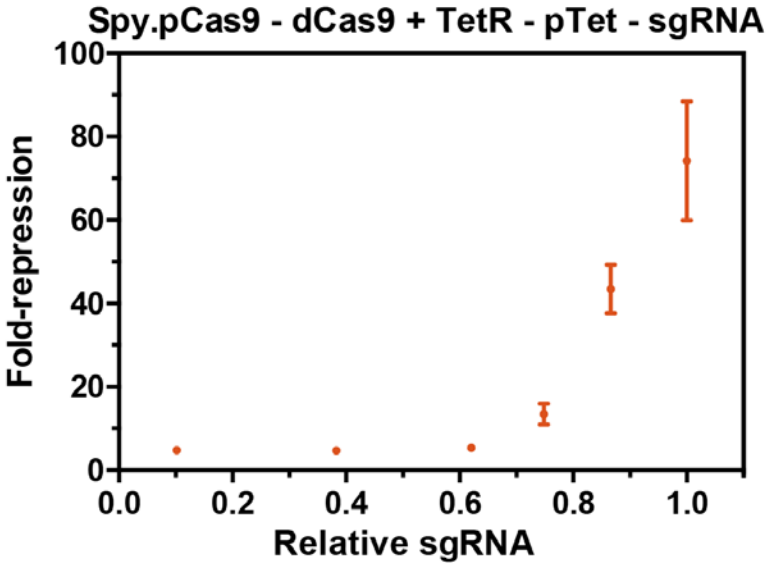
Supplementary Figure S3.1 Fold repression when titrating CRISPRi by regulating the expression of dCas9 and sgRNAs using combinations of synthetic constitutive promoters. dCas9 was controlled by one of three constitutive promoters (Spy.pCas9 [29], Bba_J23117, Bba_J23112 [parts.igem.org/Part:BBa_J23119], 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 OD600 of *E. coli* cultures of each condition by that of a negative control not carrying a sgRNA. The baseline fluorescence/OD600 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.



Supplementary Figure S3.2 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.

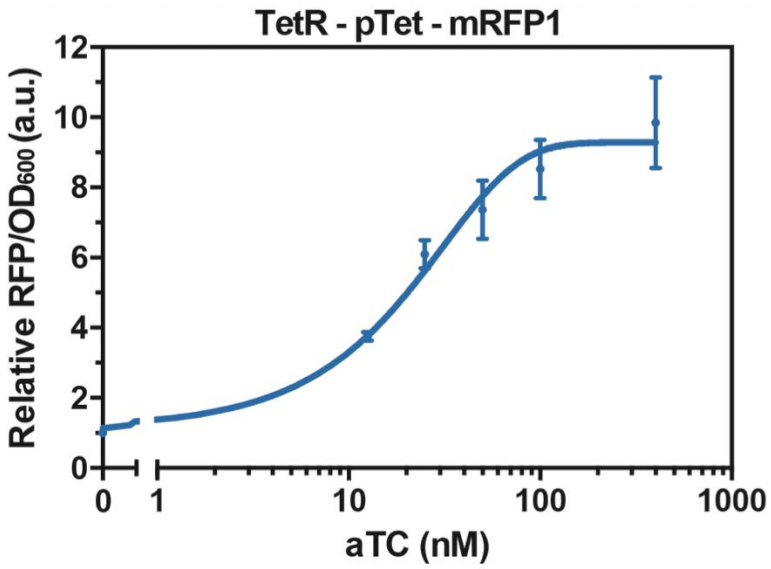
5



Supplementary Figure S3.3 Fold-repression when titrating sgRNA expression a pTet inducible

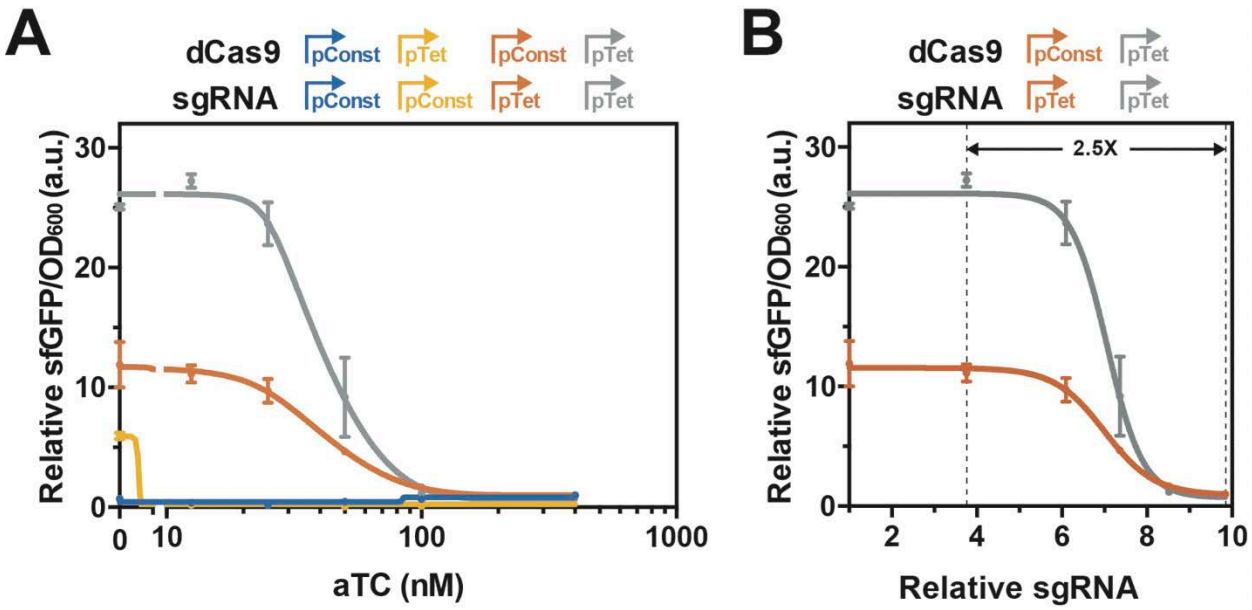
promoter.

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 S3.4. On the y-axis is the fold repression of an integrated fluorescent reporter upon expression of the CRISPRi components, calculated as described in Figure S3.1. Data are plotted against the mean of three replicates with standard deviations.



Supplementary Figure S3.4 Dynamic range in gene expression provided by the aTc-inducible promoter included in this study.

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 OD600 of stationary-phase cultures. The baseline fluorescence/OD600 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.



Supplementary Figure S3.5 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 OD600 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 OD600 of each strain when induced with 400 nM aTc concentrations and include baseline subtraction of the fluorescence/OD600 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 S3.4. aTc: anhydrotetracycline. Data are plotted against the mean of three replicates with standard deviations.

Chapter 4 | Systematic Characterization of the Requirements for CRISPR-Cas Activation in Bacteria

The work in this chapter is part of a manuscript that is currently preparing for submission. Authors of the manuscript:

Jason Fontana[*], Chen Dong[*], Cholpisit Kiattisewee, Venkata P. Chavali, Benjamin I. Tickman, James Carothers[†], Jesse Zalatan[†]

*: these authors contributed equally; †: corresponding authors

4.1 | Abstract

CRISPR-Cas transcriptional activation (CRISPRa) in bacteria has the potential to enable engineering of complex, synthetic, multi-gene expression programs. After limited success in targeting endogenous genes for activation, we identified multiple features of bacterial promoters that impose stringent requirements on effective guide RNA target sites. Most importantly, we found that bacterial CRISPRa is extremely sensitive to both promoter sequence and structure. Shifting the gRNA target site by 2-3 bases along the DNA target causes a nearly complete loss in activity. CRISPRa activity can be rescued upon shifting the target site by a distance corresponding to one full helical turn of the DNA, which is about 10-11 bases. The striking sensitivity to target site position in *E. coli* contrasts sharply with the relatively flexible target site requirements for gene activation in eukaryotic cells and highlights fundamental differences between transcriptional regulation in bacterial and eukaryotic systems. Practically, our results suggest that it will be challenging to find a gRNA target site with an appropriate PAM sequence at precisely the right position at arbitrary genes of interest. To overcome this limitation, we demonstrate that a dCas9 variant with expanded PAM specificity allows 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.

4.2 | 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 new CRISPR-Cas gene activation (CRISPRa) system that is effective in *E. coli*. Our system can be combined with CRISPRi gene repression to programmably and flexibly 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 active target sites at arbitrary promoters in the genome.

To programmably 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 encode CRISPRi and scRNAs targeted upstream of a minimal promoter can encode CRISPRa.

In this work, we demonstrate that the rules for targeting CRISPRa to effective sites in *E. coli* are surprisingly stringent. In prior work, we found that CRISPRa was effective at *E. coli* 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 more strict. We demonstrate that the strength of the promoter, the identity of the sigma factor controlling the target promoter, 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. 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.

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 the NGG PAM sequence that is necessary for *S. py* dCas9 at exactly the right position in an arbitrary gene target. We demonstrate that this challenge can be partially overcome using a re-engineered dCas9 protein that can target an expanded set of PAM sequences[5]. By systematically defining the rules for effective CRISPRa sites, we have identified a clear path for improving and generalizing synthetic gene regulation in bacteria.

4.3 | Results

4.3.1 | A SoxS double mutant reduces off-target activation of endogenous SoxS-responsive promoters

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 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 further reduction in endogenous activity to levels indistinguishable from background (Figure 4.1). 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.

4.3.2 | CRISPRa does not activate endogenous genes at sites predicted from synthetic heterologous promoters

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 with target sites positioned within a 40 bp 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 that could be targets for improving bioproduction[6]. Targeting the CRISPR-Cas complex to

two sites within the optimal 40 bp window resulted in modest increases in expression that were highly variable between replicates, and similar to the results for targeting sites well outside the 40 bp window (Figure 4.2A). Next, we targeted *cysK*, an enzyme involved in cysteine biosynthesis[7]. Similar to what we observed with *aroK-aroB*, targeting a site positioned at -93 from the TSS resulted in expression levels that were highly variable between biological replicates, ranging from 2.4-fold to 0.8-fold compared to a control expressing an off-target scRNA (Figure 4.2B). Targeting two other sites within the 40 bp window, at -79 and -61, resulted in similarly variable and in some cases decreasing levels of expression. Finally, we targeted *ldhA*, an enzyme involved in mixed acid fermentation[8]. We selected 8 sites and observed small increases in gene expression up to 1.5-fold, with no apparent relationship between the position of the target site and *ldhA* expression (Supplementary Figure S4.1). Together, these results suggest that endogenous genes cannot be activated simply by targeting the CRISPR-Cas complex to sites positioned between 60 and 100 bp upstream of the TSS. The levels of activation we achieved were modest compared to the effects seen at heterologous promoters[3]. Even for sites that had modest effects, their positions did not correspond to the predicted optimal sites based on the results on heterologous promoters.

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[9], and it may be difficult to increase the transcription of genes that are already strongly expressed[10]. In addition, the endogenous target genes might require an alternative sigma factor. Our original reporter gene is controlled by the RpoD 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 the presence of 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 reached this conclusion 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, regulator binding sites, and scRNA target site positions.

4.3.3 | 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 4.3A). We observed the most effective gene activation (52-fold) with a moderately weak promoter. With the weakest promoters (2-fold lower basal expression), we could not detect any activation. With stronger promoters, we observed progressively smaller CRISPRa-mediated changes in gene expression; the basal expression level increased, while the maximal, CRISPRa-induced expression remained roughly constant. These results indicate that the CRISPRa activity varies considerably with promoter strength. 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.

4.3.4 | 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[11].The SoxS activator binds directly to the α subunit of RNA polymerase[12], 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 RpoS[13], RpoH[14], RpoE[15], and RpoN[16] to compare with our original housekeeping RpoD promoter (Figure 4.3B). CRISPRa was able to activate reporter gene expression when we targeted sites between -60 and -100 upstream of RpoS, RpoH, and RpoE promoters. CRISPRa was not active on the RpoN (Figure 4.3B), promoter, probably because the RpoN sigma factors interact with DNA in a unique conformation different from the other sigma factors[17]. Activation with promoters regulated by alternative sigma factors was significantly lower that the activation we obtained with a RpoD promoter. Therefore, while activity will depend on the specific characteristics of the target promoter, in principle CRISPRa can activate promoters regulated by non-housekeeping sigma factors such as RpoS, RpoH, and RpoE.

4.3.5 | CRISPRa is sensitive to intervening sequence composition

Our initial model for effective CRISPRa target sites considered the distance of the target site from the TSS but ignored the composition of the intervening sequence upstream of the -35 region. To determine if these sequences affect CRISPRa, we constructed a pooled library of synthetic promoters with

randomized sequences between the scRNA target site and the -35 region. We observed a broad distribution of gene activation over a 27-fold range, even though each sequence in this library contained the same scRNA target site and the same minimal promoter (Figure 4.3C). A simple interpretation of this result is that these randomized sequences contain binding sites for endogenous transcriptional regulators; there is evidence that binding sites can emerge with relatively high frequency from random sequences[18]. These sites could potentially affect the efficacy of CRISPRa by directly blocking access to an 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 model that a bound transcriptional effector can disrupt CRISPRa, we introduced a binding site for the transcriptional repressor TetR upstream of the -35 region[19]. 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 on a promoter that lacks a TetR binding site (Figure 4.3D). Taken together, these experiments suggest that the composition of the intervening sequence between the CRISPR-Cas complex and the minimal promoter has a significant effect on the efficacy of CRISPRa, possibly due to the presence of protein binding sites. Because endogenous genes contain binding sites for a variety of transcriptional activators and repressors upstream of the minimal promoter[20,21], this effect could be contributing to the inconsistent and variable effects we observed when targeting endogenous genes for CRISPRa.

4.3.6 | CRISPRa is sharply dependent on single base shifts in target site position

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 be targeted to any distance in the optimal targeting window. Shifting the target site by 1-3 bases caused significant decreases in activation (Figure 4.4A). Shifting the target site further by 4-9 bases decreased expression further 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 -71 site shifted by 10 and the original -81 site in the unshifted reporter, do not give the same gene expression output (Figure 4.4A). These discrepancies could arise from intrinsic differences in the activity of the 20 base scRNA target sequence (Supplementary Figure S4.2) or from the effect of different intervening sequence composition between the scRNA target site and the minimal promoter (Figure 4.3).

Because we demonstrated that sequence composition can have unexpected effects on CRISPRa (Figure 4.3), we tested whether the periodicity of CRISPRa was similar in different sequence contexts. We obtained a 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 S4.2A). Similar results were also obtained when we performed the base shift experiment with a reporter that has a different 5' upstream sequence (Supplementary Figure S4.2B) or where the minimal BBa_J23117 promoter was replaced by endogenous *aroK* promoter (Supplementary Figure S4.2C). Further, the sharp positioning dependence was observed when targeting the template or non-template strand of the reporter (Supplementary Figure S4.2D). Finally, one possible confounding effect is if inserting bases changes the basal expression from the reporter gene, but we observed that basal expression from the original reporter and the +5 base shifted reporter were indistinguishable (Supplementary Figure S4.2E). Together, these experiments confirm that bacterial CRISPRa is sensitive to periodicity in multiple different sequence contexts.

In each of the experiments described above, inserting bases to shift the scRNA target site means that each site is tested with a different reporter gene construct. To test the positional dependence of CRISPRa at single base resolution in the same 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 S4.2F).

The finding that CRISPRa displays the same ~10 bp 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. Therefore, effective target sites must be located not only at the proper distance, but also 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 are ~1 PAM site every 10 bases in the regions upstream of endogenous promoters in *E. coli* (Supplementary Figure S4.4), and the likelihood that a PAM site will be located at the appropriate phase within a 10 base window is low.

4.3.7 | Modifying the CRISPRa complex structure does not expand the range of effective target sites

If rotating the CRISPRa complex out of phase prevents the SoxS activator from interacting with RNA polymerase, then a longer 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 4.4B). We obtained similar results with a linker with a different amino acid composition (Supplementary Figure S4.5A).

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 constructed multiple alternative scRNA designs that present the MS2 hairpin at different locations. Extending the MS2 step 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 S4.5B). 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 S4.5C).

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 were not targeted to an optimal peak of activation that might be different for each activator. We tested MCP fused to TetD, αNTD, lambda cII, and RpoZ[3], and dCas9 fused to RpoZ[22]; however, none of these constructs produced gene activation at any site that was not already effective with SoxS (Supplementary Figure S4.6).

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[23–25]; we confirmed this result with an endogenous SoxS reporter (Supplementary Figure S4.7). It remains surprising that no structural modifications of the CRISPR complex produced any changes in the phase dependence. If SoxS is simply tethered to the CRISPR 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. Practically, it means that we still lack a way to expand the range of effective CRISPRa target sites.

4.3.8 | An expanded PAM dCas9 variant expands the scope of targetable CRISPRa sites

Because there is likely a limited number of genes with an appropriate NGG PAM site at precisely the optimal position upstream of the promoter, we pursued a strategy to expand the scope of targetable PAM sites. 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[5]. 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 4.5A). Activation levels varied with different PAM sites and correlated well with dxCas9(3.7) activity previously reported in human cells (Supplementary Figure S4.8A)[5]. dxCas9(3.7) showed similar distance and phase dependent target site preferences as dCas9 (Supplementary Figure S4.8B & 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 S4.8D). Accounting for the fact that dCas9 has some activity at non-NGG sites (Figure 4.5A), there are still on average ~2.2-fold more dxCas9(3.7)-compatible PAM sites than dCas9-compatible PAM sites (Supplementary Figure S4.8D).

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 4.5B). We then inserted 5 bp into the reporter to shift the AGT PAM site 5 bases upstream and into the peak activation range. At this reporter, neither dCas9 nor dxCas9(3.7) can activate the NGG PAM site, which is now out of phase. dxCas9(3.7), but not dCas9, was now able to effectively activate the AGT PAM site (Figure 4.5B). This result confirms that dxCas9(3.7) is able to activate optimally-positioned target sites that are inaccessible to dCas9.

4.4 | 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, and (v) the position of the scRNA target site with respect to the TSS at single bp resolution. Some of these factors, such as promoter strength and scRNA target sequence composition, are similar in eukaryotic systems[10,26–28]. 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[11], 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[29], 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[30]. There is precedent for bacterial transcriptional activators that are sensitive to target site periodicity[23–25], but the dramatic changes in activity with only single base shifts was surprising. Moreover, our inability to predictably alter or broaden the range of sites that are effective is puzzling, as it suggests that our basic models for how activators interact with bacterial transcription machinery are 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 heterologous promoters that can be regulated by CRISPRa, which will enable the construction of complex, tunable synthetic multi-gene circuits (Fontana et al., *manuscript in preparation*). To expand the scope of CRISPRa to endogenous target genes, expanded PAM variants like dxCas9(3.7)[5], or orthologous dCas9 proteins with alternate PAM specificities[31,32] make it more likely that target sites can be found that coincide with an optimal position relative to the TSS.

Looking forward, the development of quantitative, predictive models for CRISPRa target site selection on arbitrary promoters would significantly benefit the field. Unbiased high-throughput approaches that rely on scRNA libraries and functional readouts may be critical for developing such models. Once such a general model has been established, high-throughput CRISPRa screens could be applied to discover new gene targets that improve biosynthesis or lead to desired phenotypes. Together with the established tools available for CRISPRi gene repression, our CRISPRa tools provide a powerful avenue for implementing larger and more complex transcriptional programs in bacteria. Ultimately, uncovering generalizable rules for activating genes in bacteria will positively impact a broad range of applications in basic research, biomedicine and metabolic engineering.

4.5 | Methods

4.5.1 | Bacterial Strain Construction and Manipulation

Plasmid cloning and *E. coli* cell culture were conducted following standard molecular biology protocols. Bacterial strains with sfGFP or mRFP reporter strains are described in Table S1. Guide RNA target sequences are described in Table S2. Plasmid containing the reporter genes and the CRISPR components are described in Table S3: *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 Cas protein and the activation domain or in a separate ColE1 plasmid. 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 were 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 standard mixed bases oligos (IDT).

4.5.2 | 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 for 20 hours and then diluted in 1:50 in DPBS and analyzed on a MACSQuant VYB flow cytometer (Miltenyi Biotec) using a previously described gating strategy[3]. A side scatter threshold trigger (SSC-H) was applied to enrich for single cells. 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.

4.5.3 | 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 for 20 hours. 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. 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.

4.5.4 | 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 harvested after 20 hrs of growth and diluted 1:100 into 10 mL EZ-RDM supplemented with appropriate antibiotics and grown at 37 °C and shaking until an OD_{600} of 0.8 (1 cm pathlength) was reached. 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 was provided in Table S4. Expression levels for each gene were calculated by normalizing to the 16S rRNA gene and relative to a negative control carrying an off target-scRNA using the $\Delta\Delta C_T$ method[33].

4.6 | References

1. Brophy, J. A. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11,** 508 (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. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556,** 57 (2018).

6. Rodriguez, A. *et al.* Engineering Escherichia coli to overproduce aromatic amino acids and derived compounds. *Microb Cell Fact* **13,** 126 (2014).

7. Byrne, C., Monroe, R., Ward, K. & Kredich, N. DNA sequences of the cysK regions of Salmonella typhimurium and Escherichia coli and linkage of the cysK regions to ptsH. *J Bacteriol* **170,** 3150–3157 (1988).

8. Jiang, G., Nikolova, S. & Clark, D. P. Regulation of the ldhA gene, encoding the fermentative lactate dehydrogenase of Escherichia coli. *Microbiology+* **147,** 2437–2446 (2001).

9. Silander, O. K. *et al.* A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. *Plos Genet* **8,** (2012).

10. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12,** 326–328 (2015).

11. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu Rev Microbiol* **57,** 441–466 (2003).

12. Shah, I. M. & Wolf, R. E. 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. *J Mol Biol* **343,** 513–532 (2004).

13. Gort, A., Ferber, D. M. & Imlay, J. A. The regulation and role of the periplasmic copper, zinc superoxide dismutase of Escherichia coli. *Mol Microbiol* **32,** 179–191 (1999).

14. Nonaka, G., Blankschien, M., Herman, C., Gross, C. A. & Rhodius, V. A. Regulon and promoter analysis of the E. coli heat-shock factor, σ32, reveals a multifaceted cellular response to heat stress. *Gene Dev* **20,** 1776–1789 (2006).

15. Rhodius, V. A., Suh, W., Nonaka, G., West, J. & Gross, C. A. Conserved and Variable Functions of the σE Stress Response in Related Genomes. *Plos Biol* **4,** (2006).

16. Tian, Z., Li, Q., Buck, M., Kolb, A. & Wang, Y. The CRP–cAMP complex and downregulation of the glnAp2 promoter provides a novel regulatory linkage between carbon metabolism and nitrogen assimilation in Escherichia coli. *Mol Microbiol* **41,** 911–924 (2001).

17. Österberg, S., del Peso-Santos, T. & Shingler, V. Regulation of Alternative Sigma Factor Use. *Annu Rev Microbiol* **65,** 37–55 (2011).

18. Yona, A. H., Alm, E. J. & Gore, J. Random sequences rapidly evolve into de novo promoters. *Nat Commun* **9,** 1530 (2018).

19. Lee, T. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J*

Biol Eng **5,** 12 (2011).

20. Mendoza-Vargas, A. *et al.* Genome-Wide Identification of Transcription Start Sites, Promoters and Transcription Factor Binding Sites in E. coli. *Plos One* **4,** (2009).

21. Lloyd, G., Landini, P. & Busby, S. Activation and repression of transcription initiation in bacteria. *Essays Biochem* **37,** 17–31 (2001).

22. 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).

23. Wood, T. I. *et al.* 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. *Mol Microbiol* **34,** 414–430 (1999).

24. Zhou, Y., Kolb, A., Busby, S. & Wang, Y.-P. Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Res* **42,** 9209–9216 (2014).

25. Müller, J., Barker, A., Oehler, S. & Müller-Hill, B. Dimeric lac repressors exhibit phase-dependent co-operativity11Edited by M. Yaniv. *J Mol Biol* **284,** 851–857 (1998).

26. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583 (2015).

27. Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat Methods* **11,** (2014).

28. 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,** (2017).

29. Rojo, F. Repression of transcription initiation in bacteria. *Journal of bacteriology* **181,** 2987–91 (1999).

30. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159,** 647–661 (2014).

31. Leenay, R. T. & Beisel, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. *J Mol Biol* **429,** 177–191 (2017).

32. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15,** 169–182 (2017).

33. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔC T Method. *Methods* **25,** 402–408 (2001).

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).

4.7 | Figures

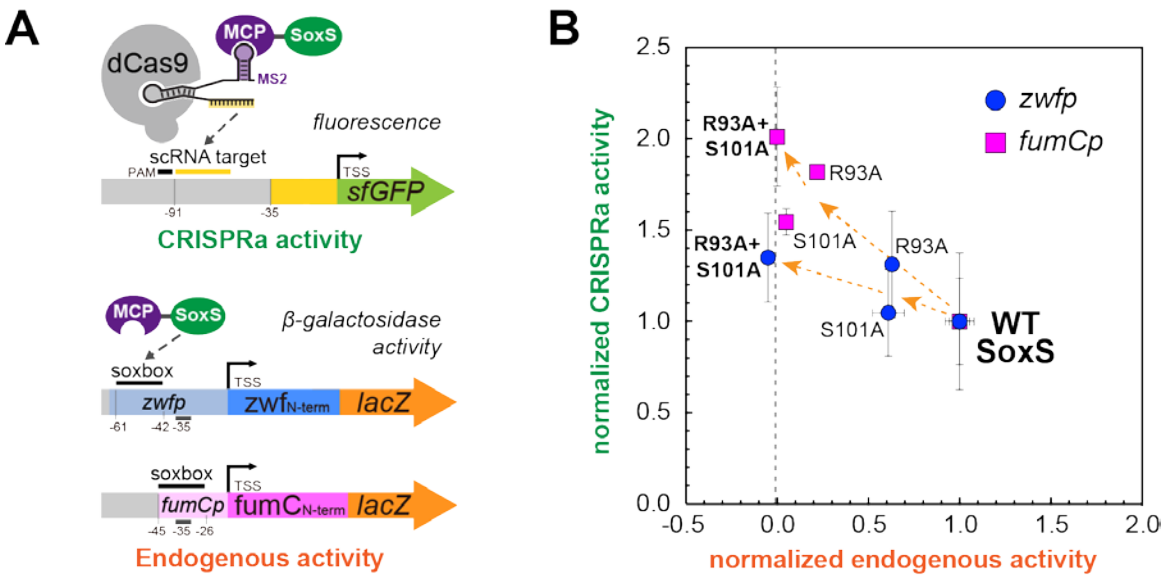


Figure 4.1 The double-mutant SoxS(R93A/S101A) further reduces the activity of SoxS on its endogenous promoters.

A) Reporter constructs for measuring the CRISPRa activity and endogenous activity of wild-type or mutant SoxS activation domains. CRISPRa activity was determined by expressing the CRISPRa components in the *E. coli* CD06 strain where a weakly expressed sfGFP was genomically-integrated (Supplementary Table 1). The CRISPRa components included *Sp*-dCas9, MCP-(5aa)-SoxS(wild-type or mutant), and a scRNA targeting the W108 site -91 bp upstream of the transcription start site (TSS) of sfGFP on the non-template strand (Supplementary Table 2). The scRNA used the 1XMS2 scRNA.b1 design that was characterized previously[3]. The endogenous activity was determined by monitoring *lacZ* expression from reporter plasmids where lacZ was driven by SoxS-regulated promoters *zwfp* and *fumCp*[34]. The *zwfp* represents class I SoxS promoters where the SoxS binding site (soxbox) is upstream of the -35 region; the *fumCp* represents class II SoxS promoters where the soxbox overlaps with the -35 region. Strains were grown overnight to late stationary phase. GFP fluorescence was measured by flow cytometry and *lacZ* activity was measured using a beta galactosidase assay.

B) SoxS(R93A/S101A) displayed no significant endogenous activity on *zwfp* and *fumCp* while retaining the CRISPRa activity of the wild type. Fluorescence and *lacZ* activity values were baseline-subtracted using a strain that does not express a scRNA. CRISPRa activity values are relative to the strain expressing wild-type SoxS. Endogenous activity was calculated by normalizing the lacZ activity unit to the strain expressing wild-type SoxS. Data points represent the endogenous activity (x-axis) and

CRISPRa activity (y-axis) of the strains bearing either the zwfp-lacZ reporter or the fumCp-lacZ reporter. Error bars represent the standard deviations of the normalized CRISPRa activity (vertical) and the normalized endogenous activity (horizontal).

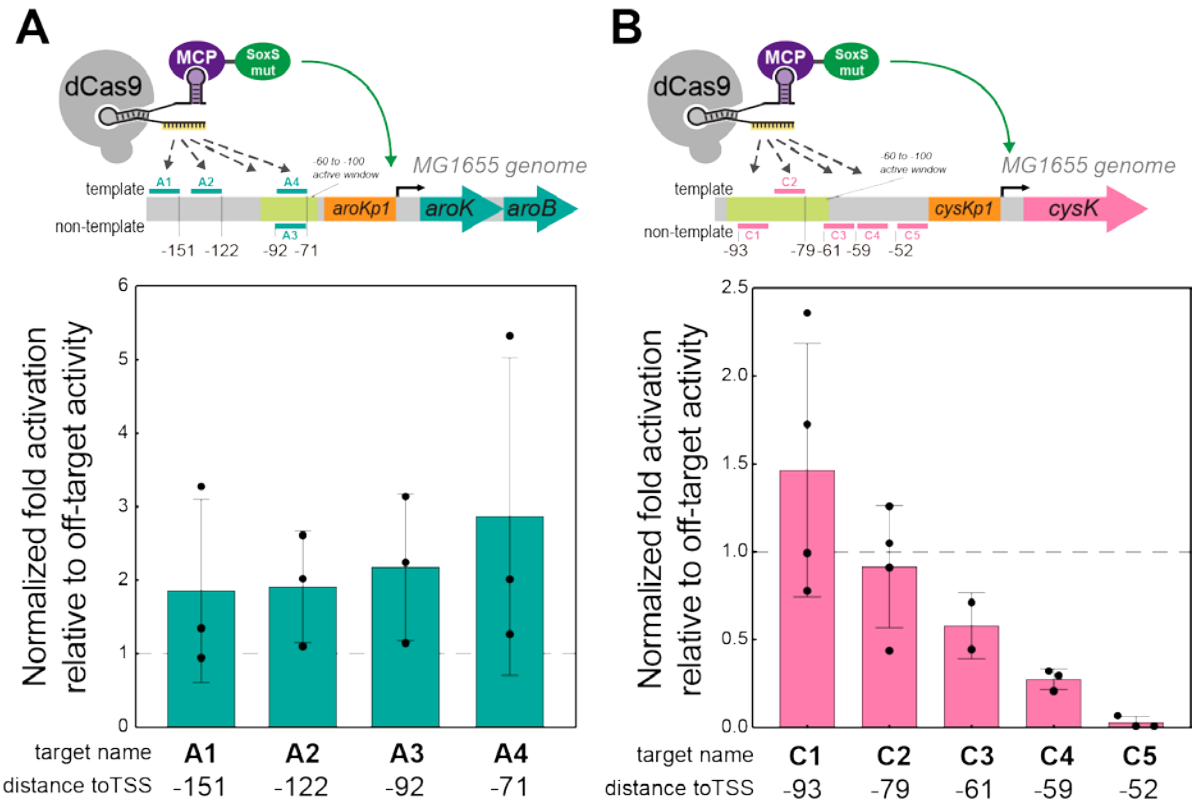
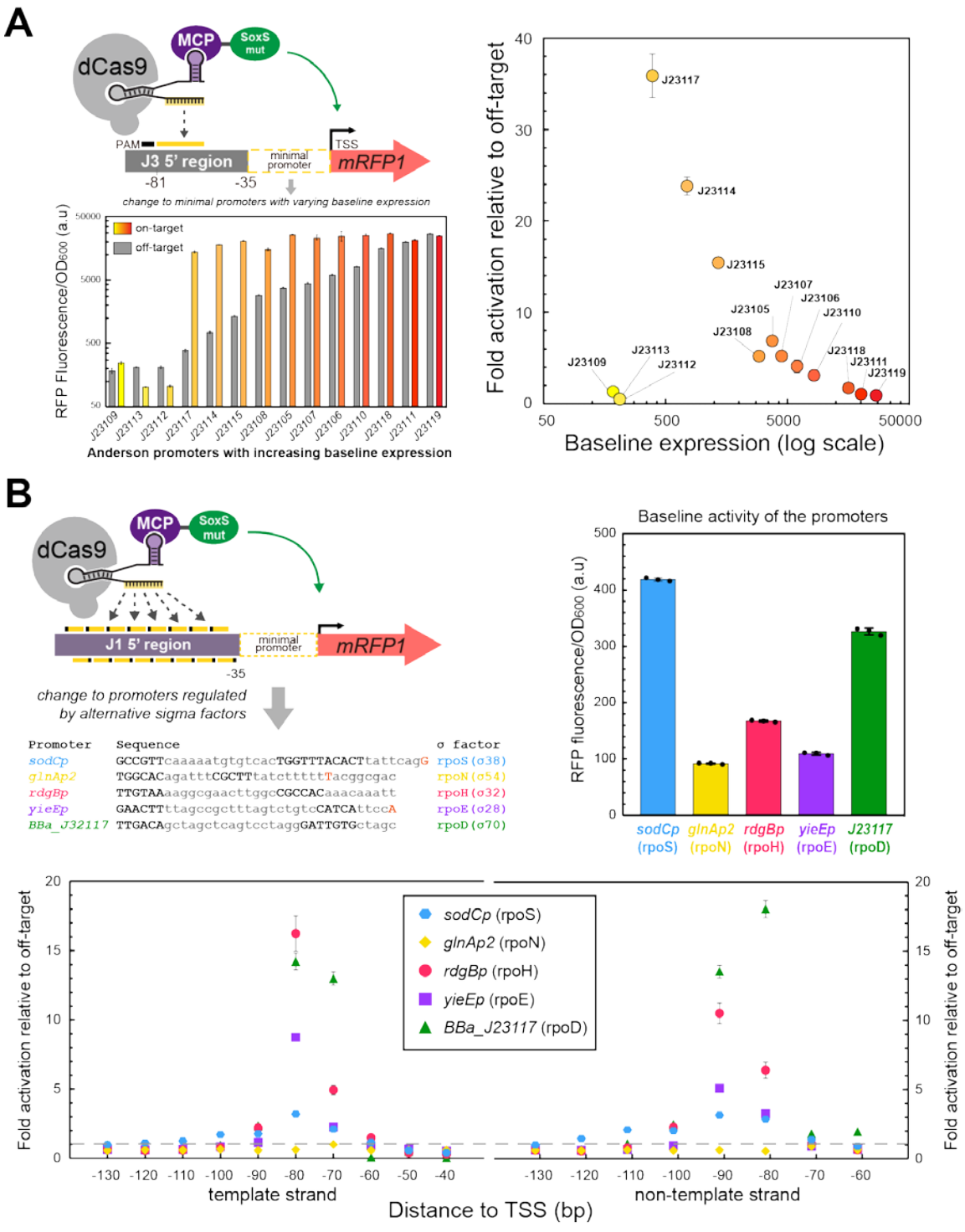


Figure 4.2 CRISPRa does not activate endogenous genes at sites predicted from synthetic heterologous promoters.

A) CRISPRa on the *aroK-aroB* operon. Two scRNA target sites within the 40 bp window where CRISPRa is active (A1-A2) and two sites further upstream (A3-A4) were chosen for the *aroKp1* promoter (Supplementary Table 2). Regardless of the target position, average activation was 2~3-fold, but high variability between biological replicates was observed.

B) CRISPRa on the *cysK* gene. Three scRNA target sites within the 40 bp window where CRISPRa is active (C1-C3) and two sites further downstream (C4-C5) were chosen for the *cysKp2* promoter (Supplementary Table 2). The C1 site displayed the highest average activation (1.5-fold) but also the highest variability among biological replicates. The C4 and C5 sites resulted in significant gene repression, likely because these sites overlap or occlude the core promoter where RNA polymerase binds to DNA. In both experiments, the CRISPRa components (*Sp*-dCas9, activation domain and either an on-target or off-target scRNA) were delivered to *E. coli* MG1655. For the *aroK-aroB* experiment, the activation domain was SoxS(R93A/S101A) and the scRNA was 1XMS2 scRNA.b1; for the *cysK* experiment, the activation domain was SoxS(R93A) and the scRNA was 1XMS2 scRNA.b2. The b2 scRNA design was characterized previously[3]. Total RNA was extracted and cDNA was prepared from liquid cultures grown to late-exponential phase. RT-qPCR was conducted to measure gene expression.

Fold changes were calculated using the ΔΔCT method by normalizing expression to a control expressing an off-target scRNA. Bars indicate the mean normalized expression levels and error bars the standard deviation between biological replicates. Black dots indicate the values of individual biological replicates.


A
dCas9
MCP
SoxS mut
PAM
J3 5' region
minimal promoter
TSS
mRFP1
-81
-35
change to minimal promoters with varying baseline expression
RFP Fluorescence/OD600 (a.u)
on-target
off-target
J23109 J23113 J23112 J23117 J23114 J23115 J23108 J23105 J23107 J23106 J23110 J23118 J23111 J23119
Anderson promoters with increasing baseline expression
Fold activation relative to off-target
Baseline expression (log scale)
J23117
J23114
J23115
J23107
J23106
J23105
J23110
J23108
J23113
J23109
J23112
J23118
J23111
J23119
B
dCas9
MCP
SoxS mut
J1 5' region
minimal promoter
mRFP1
-35
change to promoters regulated by alternative sigma factors
Promoter Sequence σ factor
sodCp GCCGTTcaaaaatgtgtcacTGGTTTACACTtattcagG rpoS(σ38)
glnAp2 TGGCACagatttCGCTTtatcttttttTacggcgac rpoN(σ54)
rdgBp TTGTAAaaggcgaacttggcCGCCACaaacaaatt rpoH(σ32)
yieEp GAACTTttagccgctttagtctgtcCATCAttccA rpoE(σ28)
BBa_J32117 TTGACAgctagctcagtcctaggGATTGTGctagc rpoD(σ70)
Baseline activity of the promoters
RFP fluorescence/OD600 (a.u)
sodCp (rpoS)
glnAp2 (rpoN)
rdgBp (rpoH)
yieEp (rpoE)
J23117 (rpoD)
Fold activation relative to off-target
sodCp (rpoS)
glnAp2 (rpoN)
rdgBp (rpoH)
yieEp (rpoE)
BBa_J23117 (rpoD)
template strand
non-template strand
Distance to TSS (bp)

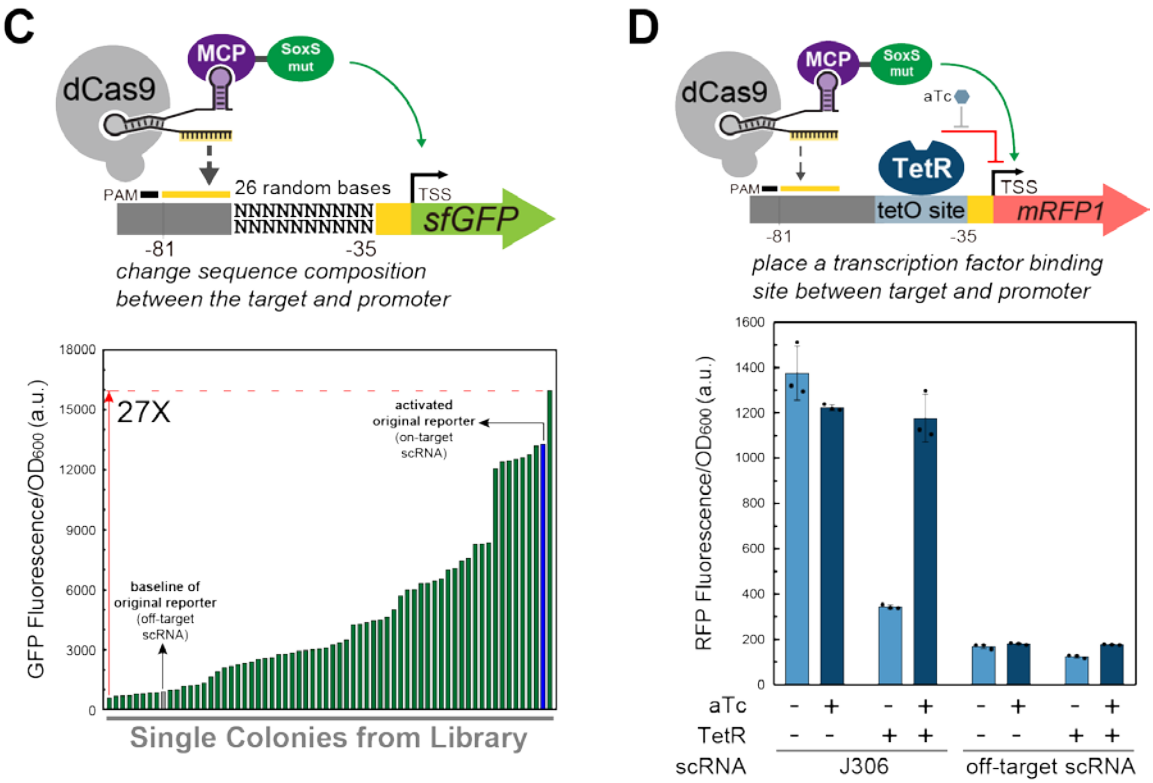


Figure 4.3 CRISPRa activity is dependent on the genomic context of the promoter.

A) CRISPRa activity was not observed at promoters with low baseline expression, but it abruptly increased to the maximum on the BBa_J23117 minimal promoter. Activating promoters with higher baseline expression resulted in steady levels of expression and decreasing CRISPRa activity. Reporter genes were constructed by replacing Bba_J23117 from J3-J23117-mRFP1 reporter (Supplementary methods) with other minimal promoters from the Anderson series (J231NN). All J3-J231NN promoters contain a J306 scRNA target site at the previously characterized optimal position (-81 bp to the TSS on the non-template strand)[3]. These reporter plasmids were co-transformed with the plasmid expressing the CRISPRa components containing *Sp*-dCas9, MCP-(5aa)-SoxS(R93A), and either an on-target (J306) or off-target (J206) 1XMS2 scRNA.b2 into *E. coli* MG1655 (Supplementary Table 2). Cultures were grown overnight to late stationary phase and analyzed on a plate reader. The bottom left plot shows the average RFP fluorescence/OD_{600} values of strains expressing an on-target or off-target scRNA of three biological replicates. Error bars indicate the standard deviation. The plot on the right shows the activation ratio (fluorescence/OD_{600} of the on-target divided by the off-target) of each strain against their baseline expression (off-target fluorescence/OD_{600}). Error bars indicate the standard deviation.

B) CRISPRa can activate promoters regulated by RpoS, RpoH, and RpoE sigma factors when targeting

sites positioned -60 bp to -100 bp to the TSS. The baseline expression level of the promoters tested varied by up to 5-fold. The observed maximum CRISPRa activity for each promoter were: 18-fold for BBa_J23117 (RpoD), 3.2-fold for *sodCp* (RpoS), 16-fold for *rdgBp* (RpoH), and 9-fold for *yieEp* (RpoE). No CRISPRa activity was observed on the *glnAp2* (RpoN) promoter. The J1-J23117-mRFP1 reporter from previous study was used to construct the reporter plasmids. It contains scRNA target sites (J101~J120) situated every 10 bp on both the template and the non-template strand upstream of the minimal promoter[3] (Supplementary Table 2). The minimal promoter from the reporter plasmid was replaced by the *sodCp* (rpoS), *glnAp2* (rpoN), *rdgBp* (rpoH), or *yieEp* (rpoE) minimal promoters. Each reporter was co-transformed with a plasmid containing the CRISPRa components *Sp*-dCas9, MCP-(5aa)-SoxS(R93A/S101A) and one of each of the 1XMS2 scRNA.b2 targeting the sites on the J1 promoter (J101~J120) or an off-target scRNA (hAAVS1) (Supplementary Table 2). Cultures were grown overnight to late stationary phase and analyzed on a plate reader. The plot on the top right shows the baseline fluorescence/OD_{600} of each strain expressing the CRISPRa components and off-target scRNA. Bars indicate the average fluorescence/OD_{600} of three biological replicates and error bars the standard deviation. The bottom plot shows the fold activation (fluorescence/OD_{600} of the on-target divided by the off-target) of each strain versus the position of the scRNA target. Error bars indicate the standard deviation and the grey line represents the 1-fold line.

C) CRISPRa activity differs significantly among promoters with the same scRNA target but different sequence composition between the scRNA target and the -35 region. A pooled reporter library was constructed by replacing the 26 bp of sequence between the scRNA target site and the -35 region on the J3-J23117-sfGFP reporter with random bases. The cloned plasmid library was co-transformed with a plasmid containing the CRISPRa components (*Sp*-dCas9, MCP-(5aa)-SoxS(R93A) and a 1XMS2 scRNA.b2) into *E. coli* MG1655. Single colonies were picked and overnight cultures were grown to late stationary phase and analyzed on a plate reader. CRISPRa activities among single members of the library spanned a 27-fold range. Each green bar represents the fluorescence/OD_{600} value of a culture grown from a single colony of the library. The blue bar represents the fluorescence/OD_{600} value of a strain expressing the original J3-J23117-sfGFP reporter and the CRISPRa components with a scRNA targeting J306. The grey bar represents a negative control expressing the J3-J23117-sfGFP reporter plasmid and the CRISPRa components with an off-target scRNA.

D) CRISPR activity was partially inhibited by the presence of a TetR transcriptional repressor binding

between the scRNA target and the minimal promoter. Releasing TetR from its target DNA upon induction with 1 μM aTc fully restored CRISPRa activity to its original level. The reporter plasmid was constructed by placing a tet operator (tetO) site upstream of the -35 region in the J3-J23117-mRFP1 reporter (Supplementary Information). The reporter plasmid was co-transformed with a CRISPR components plasmid into *E. coli* MG1655. The CRISPR component plasmid contained *Sp*-dCas9, MCP-(5aa)-SoxS(R93A/S101A), a 1XMS2 scRNA.b2 targeting the J306 site or an off-target sequence (J206), and +/- a cassette for expressing TetR. Overnight cultures were grown to late stationary phase in media +/- 1 μM aTc and then analyzed on a plate reader. Bars represent the average fluorescence/OD_{600} value of three biological replicates and error bars the standard deviation. Black dots indicate the values of individual biological replicates.

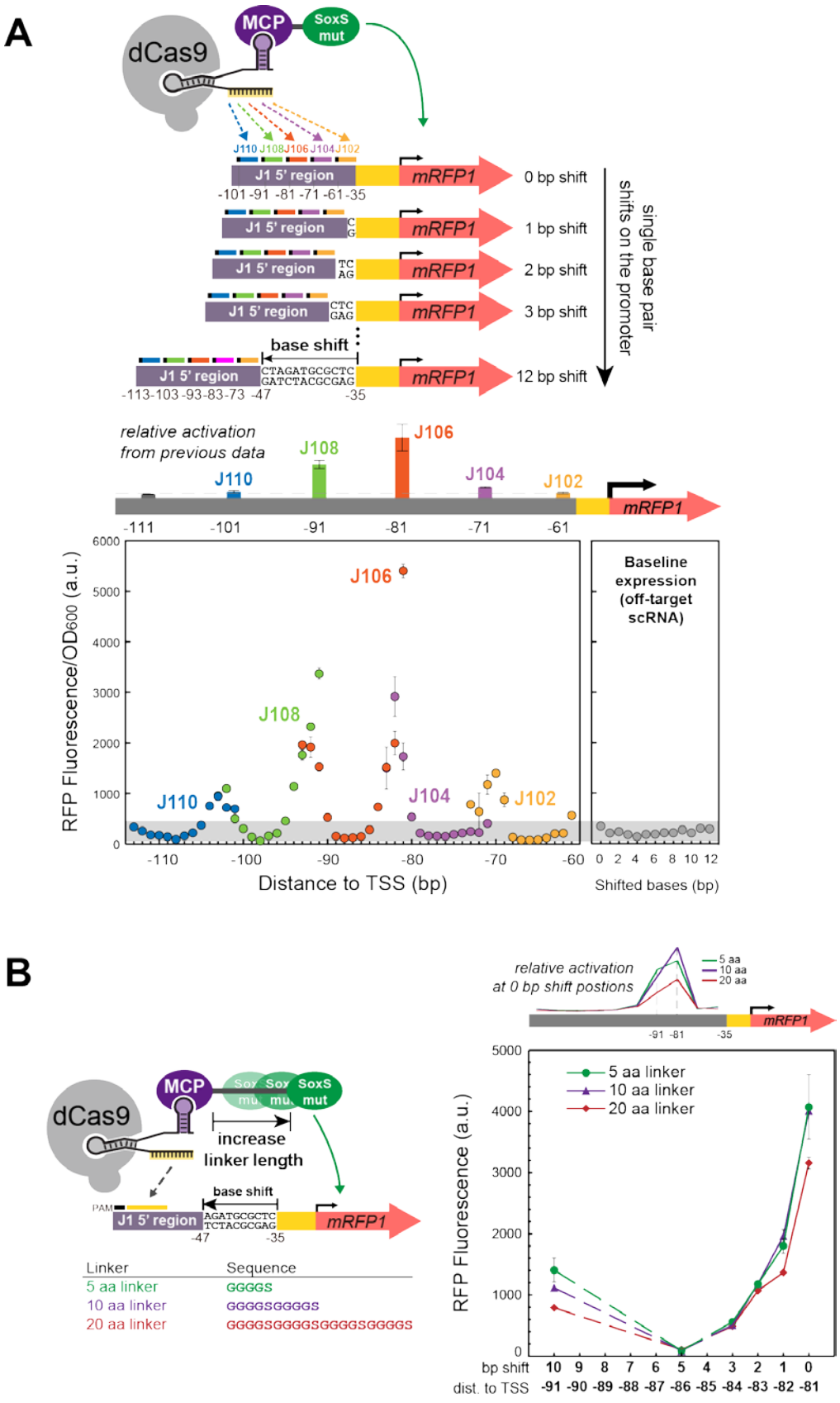


Figure 4.4 CRISPRa activity is dependent on the precise positioning of the scRNA target.

A) CRISPRa displays periodic positioning dependence with peak activities every 10-11 bp between -60

to -100 from the TSS. Peaks in gene expression were observed at the -70, -81, -91, and -102 positions, with -81 and -91 showing high gene expression and -70 and -102 showing low gene expression. Gene expression dropped to baseline levels when the target position was shifted 1-3 bp from the peak position on either side. Reporter genes were constructed by inserting 0~12 bp upstream of the -35 region of the J1-J23117-mRFP1 reporter. 5 scRNA sites were targeted (J102, J104, J106, J108, J110) with positions -61, -71, -81, -91, -101 bp to the TSS on the non-template strand of the original promoter. In this way, the complete -61 to -113 region can be covered at single base resolution. The reporter plasmids were co-transformed into *E. coli* MG1655 together with a plasmid containing the CRISPRa components *Sp*-dCas9, MCP-SoxS(R93A/S101A), and a 1XMS2 scRNA.b2 targeting J102, J104, J106, J108, J110, or an off-target sequence (J206). Cultures were grown overnight to late stationary phase and analyzed on a plate reader. Previous CRISPRa data on the J102, J104, J106, J108, J110 target positions are shown as a reference[3]. Dots indicate average fluorescence/OD_{600} values among three biological replicates and error bars indicate the standard deviation. Strains where CRISPRa is targeted to the same scRNA site that shifted to different positions on the reporter are color coded. The panel on the right shows the baseline expression of each member of the J1-J23117-mRFP1 reporter series with shifted bases. The grey area represents the range of the baseline among the reporter series. C) Extending the linker length between MCP and SoxS does not change the positioning dependence of CRISPRa. Regardless whether a 5 aa, 10 aa, or 20 aa linker was between MCP and SoxS, gene expression decreased from the maximum level at +1, +2, and +3 bp shifts, reached baseline levels at +5 bp shift and partially recovered at +10 bp shift. The J1-J23117-mRFP1 reporter plasmid series with 0, +1, +2, +3, +5, and +10 base shifts were co-transformed with CRISPRa component plasmids expressing *Sp*-dCas9, MCP-SoxS(R93A) with 5aa, 10aa, 20aa linkers, and 1XMS2 scRNA.b2 targeting J306 into *E. coli* MG1655. Cells were grown overnight to late stationary phase and analyzed on a plate reader. Previous CRISPRa data with 5 aa, 10 aa, or 20 aa linker between MCP and SoxS targeting at the -81 and -91 positions are shown as a reference[3]. Dots indicate the average fluorescence/OD_{600} values of three biological replicates and the error bars represent the standard deviation.

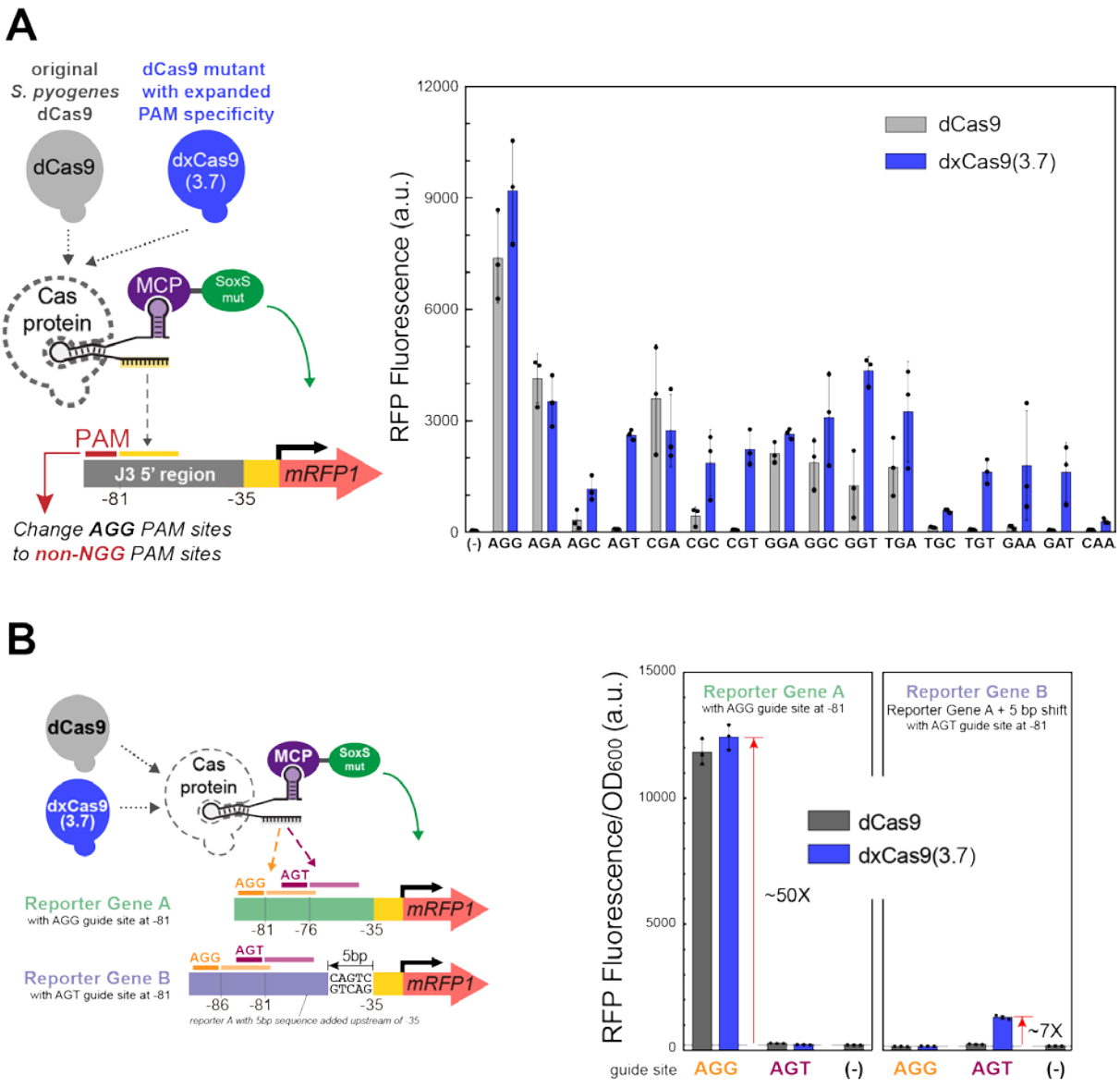


Figure 4.5 dxCas9(3.7) expands the range of active scRNA target sites by recognizing alternative PAMs.

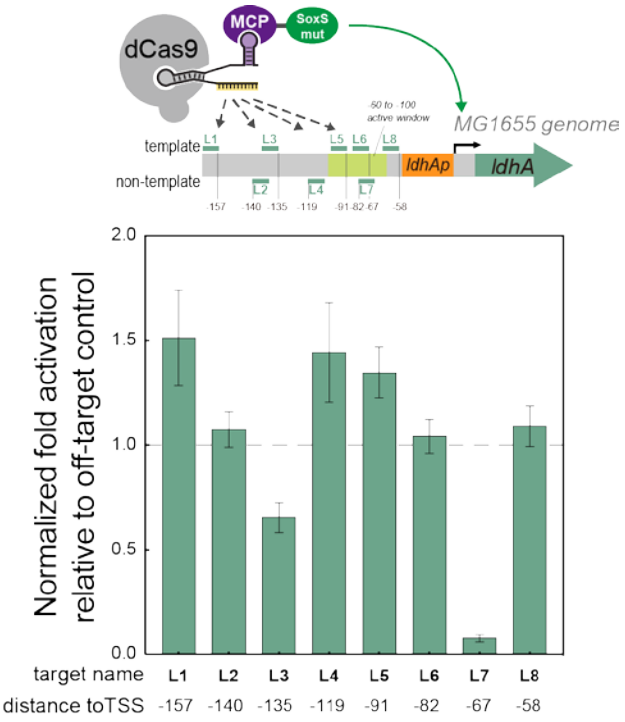
A) CRISPRa with dxCas9(3.7) displayed activity on non-NGG PAM sites with AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA sequences. CRISPRa activity with dxCas9(3.7) on non-NGG PAM sites was generally lower (6-fold to 89-fold relative to a control not expressing a scRNA) compared to the AGG PAM site (188-fold). *Sp*-dCas9 also displayed moderate CRISPRa activity at non-NGG PAM sites with AGA, CGA, GGA, GGC, GGT, TGA sequences. 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[5]. The reporter plasmids were co-transformed into *E. coli* MG1655 together with a CRISPRa component plasmid containing *Sp*-dCas9 or dxCas9(3.7), MCP-SoxS(R93A-S101A),

and a 1XMS2 scRNA.b2 targeting J306. Overnight cultures were grown to late stationary phase and analyzed by flow cytometry. Bars represent the average of the median fluorescence among three biological replicates and the error bars the standard deviation. Black dots represent the median fluorescence values for individual biological replicates. The (-) sign indicates a control expressing the original reporter with the AGG PAM and a CRISPRa component plasmids with *Sp*-dCas9, the activation domain and no scRNA.

B) dxCas9(3.7) can activate promoters with no NGG PAM sites at the optimal position. 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 did not display activity. Two reporter plasmids were constructed that contain the same 2 scRNA targets upstream of the BBa_J23117 minimal promoter driving mRFP1: the M target has an AGG PAM site and the M_2 target has an AGT PAM site (Supplementary Table 2). In reporter gene A, the M target was located -81 to the TSS on the non-template strand and the M_2 target was located -76 to the TSS on the non-template strand. In reporter gene B, 5 bp were inserted upstream of the -35 of reporter A, shifting the locations of the M target and M_2 target to -86 and -81. The sequence of the 5 bp shift was designed in a way so that the sequence between -81 scRNA target and the -35 region on reporter A and B are the same. The reporter plasmids A and B were co-transformed into *E. coli* MG1655 together with a CRISPRa component plasmid containing *Sp*-dCas9 or dxCas9(3.7), MCP-SoxS(R93A-S101A), and a 1XMS2 scRNA.b2 that targeting the M or M_2 site or an off-target sequence (J206). Overnight cultures were grown to late stationary phase and analyzed on a plate reader. Bars represent the average RFP fluorescence/OD_{600} among three biological replicates and the error bars the standard deviation. Black dots represent the RFP fluorescence/OD_{600} values for each biological replicate. 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.

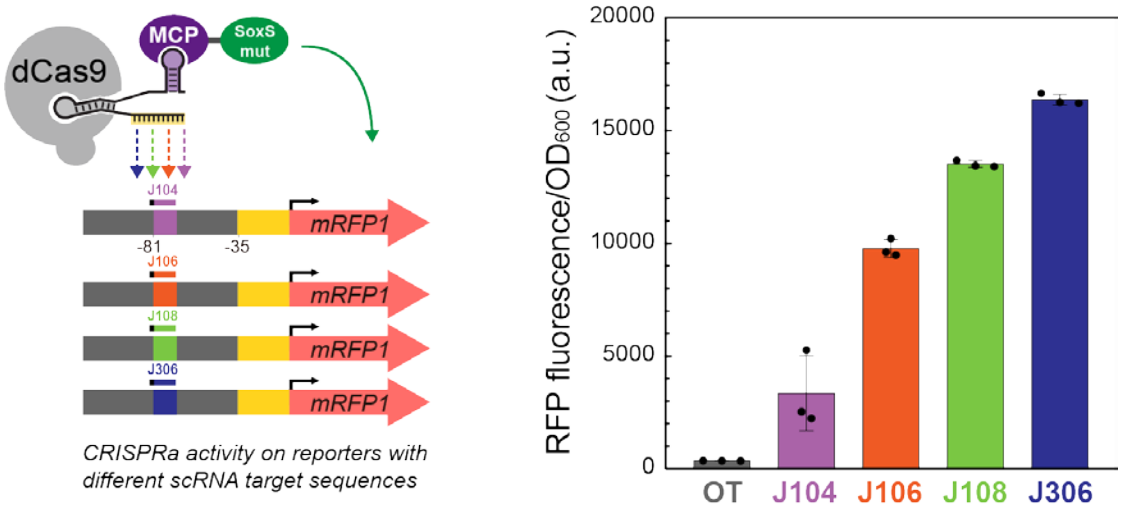
4.8 | Supplemental information

4.8.1 | Supplementary figures



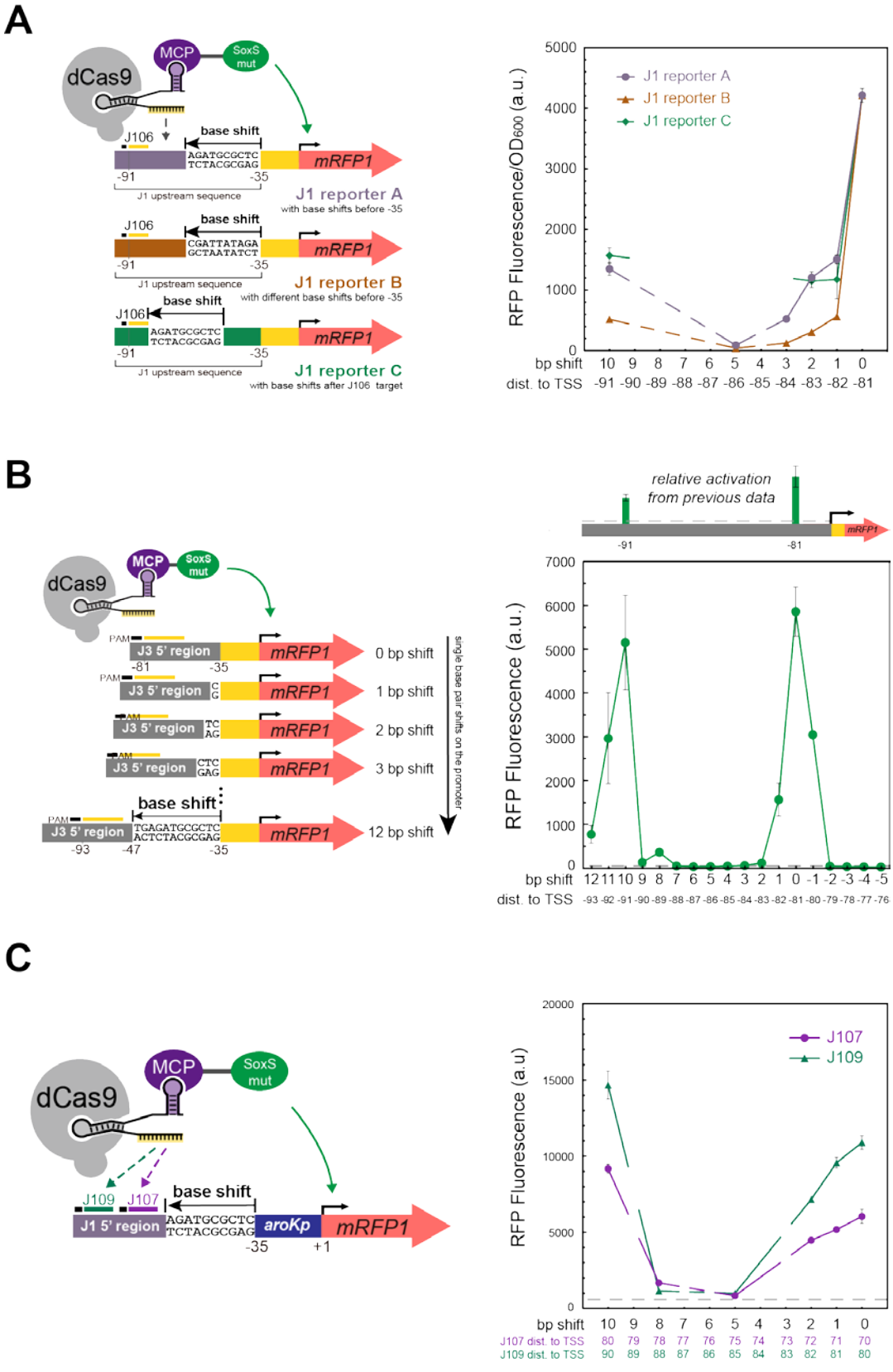
Supplementary Figure S4.1 CRISPRa does not activate *ldhA* at sites predicted from synthetic heterologous promoters.

Eight scRNA target sites (L1-L8) upstream of the *ldhAp* promoter were selected (Supplementary Table 2). Three of the target sites (L5-L7) were within the predicted 40 bp active window. Three of the sites (L1, L4, and L5) displayed weak increase in gene expression up to 1.5-fold and the L3 and L7 sites displayed decreased gene expression. There was no apparent relationship between the position of the sites and *ldhA* expression levels. The CRISPR components (*Sp*-dCas9, MCP-SoxS(R93A), and 1XMS2 scRNA.b2) that contains either an on-target or off-target scRNA were delivered to *E. coli* MG1655. Total RNA was extracted, and cDNA was prepared from liquid cultures grown to late-exponential phase. RT-qPCR was conducted to measure gene expression. Bars indicate the fold changes calculated using the $\Delta\Delta C_T$ method by normalizing the expression of each sample to a control expressing an off-target scRNA. Error bars indicate the standard error of the mean (SEM) between three technical replicates from the same culture.

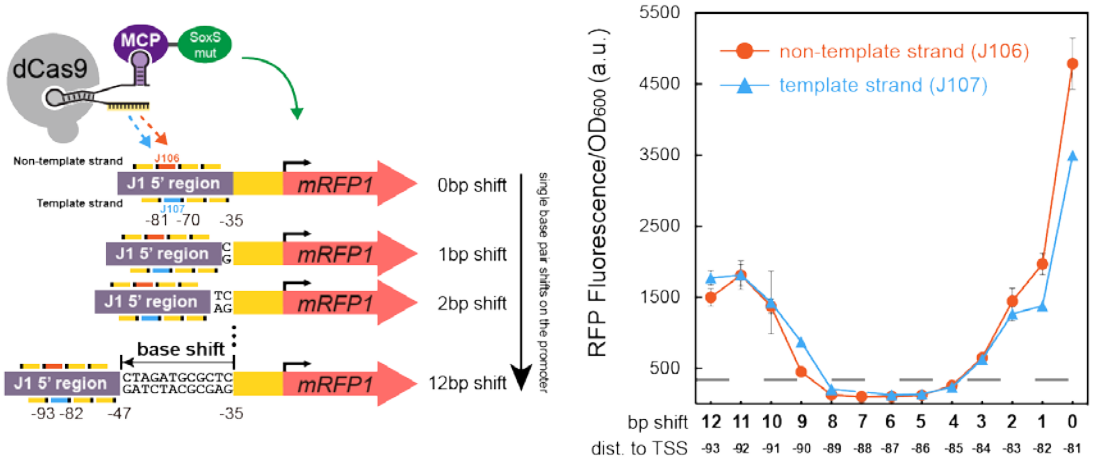


Supplementary Figure S4.2 CRISPRa activity is dependent on the spacer sequence on the scRNA.

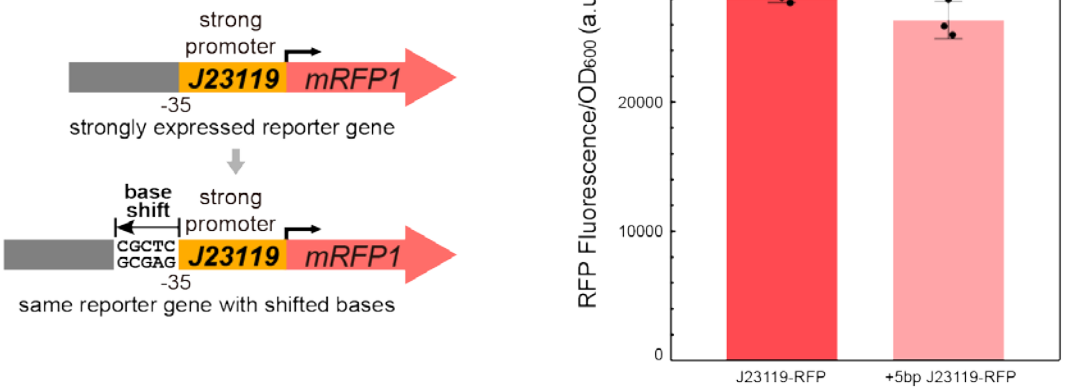
Reporter cassettes that differ only by the sequence of the 20 base scRNA target site at -81 give a broad range of gene expression levels, demonstrating that the sequence of the scRNA target site can have a substantial effect on CRISPRa. Three new reporter plasmids were constructed where the J306 target sites on the J3-J23117-mRFP reporter was replaced by the J104, J106, and J108 sequence (Figure 4.4A). The reporters were co-transformed into *E. coli* MG1655 together with the CRISPRa component plasmid expressing *Sp*-dCas9, MCP-SoxS(R93A/S101A), and 1XMS2 scRNA.b2 with corresponding on-target (J306, J104, J106, and J108) or off-target (J206) spacers. Cells were grown overnight to late stationary phase and analyzed on a plate reader. Bars represent the mean fluorescence/OD_{600} values among three biological replicates and error bars indicate the standard deviation. Black dots indicate the fluorescence/OD_{600} values for each biological replicate.


A
dCas9
MCP
SoxS mut
J106
base shift
AGATGCGCTC
TCTACGCGAG
mRFP1
-91
-35
J1 upstream sequence
J1 reporter A
with base shifts before -35
CGATTATAGA
GCTAATATCT
J1 reporter B
with different base shifts before -35
J1 reporter C
with base shifts after J106 target
RFP Fluorescence/OD600 (a.u.)
J1 reporter A
J1 reporter B
J1 reporter C
bp shift
dist. to TSS
B
relative activation
from previous data
PAM
J3 5' region
-81
0 bp shift
1 bp shift
2 bp shift
3 bp shift
12 bp shift
single base pair shifts on the promoter
TGAGATGCGCTC
ACTCTACGCGAG
-93
-47
RFP Fluorescence (a.u.)
C
J109
J107
J1 5' region
aroKp
+1
J107 dist. to TSS
J109 dist. to TSS

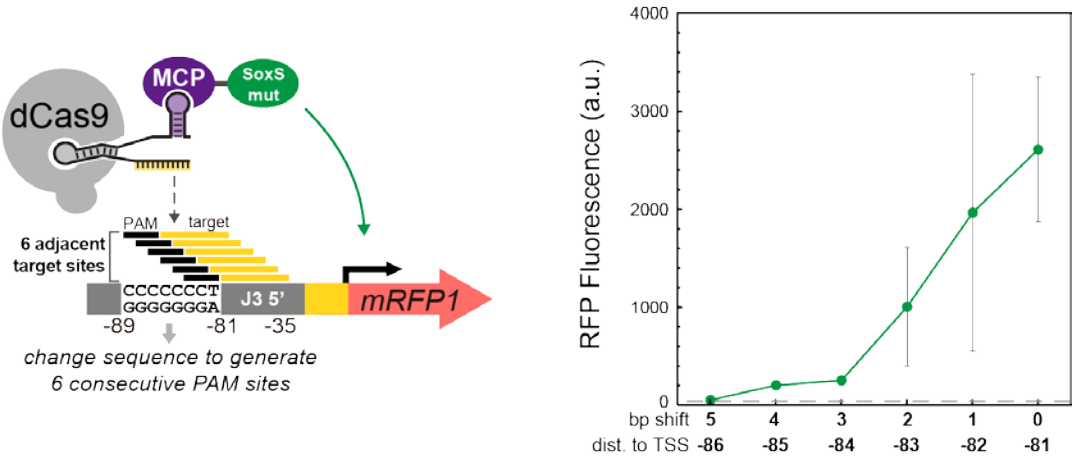
D



E



F



Supplementary Figure S4.3 Sharp positioning dependence of CRISPRa is consistent in multiple different promoters.

A) The sharp positioning requirements of CRISPRa are not affected by the location or identity of the inserted sequence. CRISPRa with three different reporters display similar decreases at 1-2 bp shifts and recovery at 10 bp. Reporters were based on the J1-J23117-mRFP1 (Supplementary Figure S4.3A) with base shifts introduced in different ways. The J1 reporter A series inserts bases upstream of the -35 region. The J1 reporter B series introduces a different sequence at the same site. The J1 reporter C series inserts bases downstream of the J106 target site. There are modest differences between reporter A and reporter B that could indicate a contribution from sequence composition. The reporter plasmids were transformed into *E. coli* MG1655 together with a plasmid expressing dCas9, MCP-SoxS(R93A), and a 1XMS2 scRNA.b2 targeting J106. Cultures were grown overnight to late stationary phase and analyzed on a plate reader. Dots indicate the mean fluorescence/OD_{600} values of 3 biological replicates and error bars indicate the standard deviation.

B) The sharp positioning requirements of CRISPRa was observed on a different heterologous promoter. CRISPRa was targeted at the -81 site (J306) on the J3-J23117 promoter that has a different upstream sequence than J1-J23117. Peaks in gene expression were observed at -81 and -91 to the TSS, displaying a 10 bp periodicity in positioning dependence. The peaks of gene expression on the J3-J23117 promoter is sharper than the J1-J23117 promoter, where gene expression reduces to the baseline after shifting only 2 bp from the peak position. The gene expression on the J3-J23117 promoter after a complete 10 bp-shift period is stronger, displaying a full restoration of the original peak expression. Reporter gene series where 0-12 bp insertions or 1-5 bp deletions were introduced upstream of the -35 of the J3-J23117-mRFP1 reporter were constructed. The reporter plasmids were co-transformed into *E. coli* MG1655 together with a CRISPRa component plasmid containing *Sp*-dCas9, MCP-SoxS(R93A/S101A), and 1XMS2 scRNA.b2 targeting J306. Cultures were grown overnight to late stationary phase and analyzed by flow cytometry. Previous CRISPRa data at -81 and -91 were shown as a reference[1]. Dots indicate the average of median fluorescence values of three biological replicates and the error bars the standard deviation. The grey line represents the baseline activity of the J3-J23117-mRFP1 reporter strain containing an empty vector instead of the CRISPRa component plasmid.

C) The sharp positioning requirements of CRISPRa are not affected by the 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 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. The reporter plasmids were transformed into *E. coli* MG1655 together with a plasmid expressing dCas9, MCP-SoxS(R93A/S101A), and 1XMS2 scRNA.b2 targeting J107 or J109. Cultures were grown overnight to late stationary phase and RFP analyzed by flow cytometry. Dots indicate the values of median RFP fluorescence between 3 biological replicates measured by flow cytometry and error bars indicate the standard deviation.

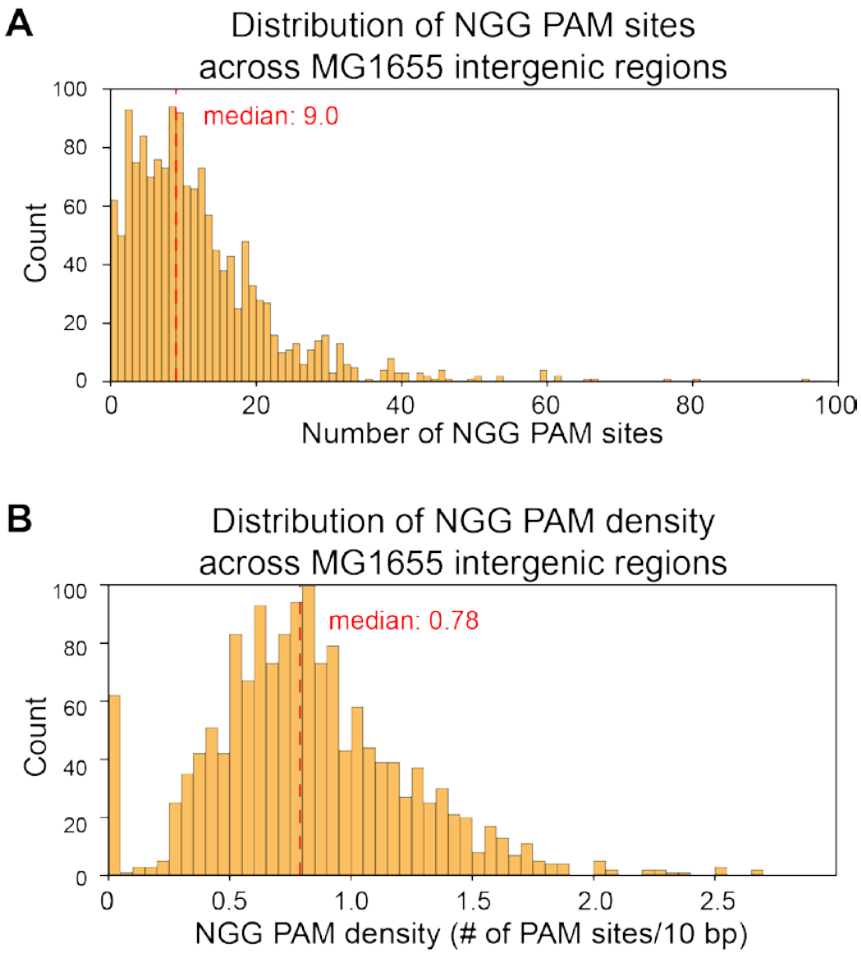
D) The sharp positioning requirements of CRISPRa are observed on both the template and the non-template strand. Both for the J106 target site on the non-template strand and the J107 target on the template strand, the CRISPRa activity drops to the baseline after shifting 3 bp from its original position and recovers after shifting 10 bp. Starting from the J1-J23117-mRFP1 reporter (Supplementary Figure S4.3A), the upstream sequences were shifted by 0~12 bp by adding bases upstream of the -35. The reporter plasmids were transformed into *E. coli* MG1655 together with a plasmid expressing dCas9, MCP-SoxS(R93A) and 1XMS2 scRNA.b2 targeting J106 or J107.

Overnight cultures were grown to late stationary phase and analyzed on a plate reader. Dots indicate the mean fluorescence/OD_{600} values between 3 biological replicates and error bars indicate the standard deviation. The dotted grey line indicates the negative control where an off-target scRNA was co-transformed with the 0 bp-shifted reporter.

E) Adding bases adjacent to the -35 region does not dramatically alter the expression of the promoter. Adding the same 5 bp as in previous figures upstream of the -35 of a strongly expressed mRFP1 does not change its baseline fluorescence. A strong minimal promoter in the Anderson family BBa_J23119 was used to construct the J3-J23119-mRFP1 reporter. 5 bp was then added upstream of the -35. The reporter plasmids were then transformed into *E. coli* MG1655. Overnight cultures were grown to late stationary phase and analyzed on a plate reader. Bars indicate the mean fluorescence/OD_{600} values between 3 biological replicates and error bars indicate the standard deviation. Black dots indicate the fluorescence/OD_{600} values of each biological replicate.

F) The sharp positioning requirements of CRISPRa was observed on a single reporter with consecutive PAM sites. 6 consecutive PAM sites on the non-template strand were engineered by replacing a CCCCCCCT sequence into the region -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 it gradually decreases to one third of the maximum activity after moving 2 bp away (-83). Gene expression reduced to a very weak level when the scRNA target moved 3 bp and 4 bp away (-84 and -85) and reaches the baseline when moved 5 bp (-86). The reporter plasmids were co-transformed into *E. coli* MG1655 together with a CRISPRa component plasmid containing *Sp*-dCas9, MCP-SoxS(R93A/S101A), and 1XMS2 scRNA.b2 targeting J306. Cultures were grown overnight to late stationary phase and analyzed by flow cytometry. Dots indicate the average of median fluorescence values of three biological replicates and the error bars the standard deviation. The grey line represents the baseline activity of the reporter strain containing an empty vector instead of the CRISPRa component plasmid.

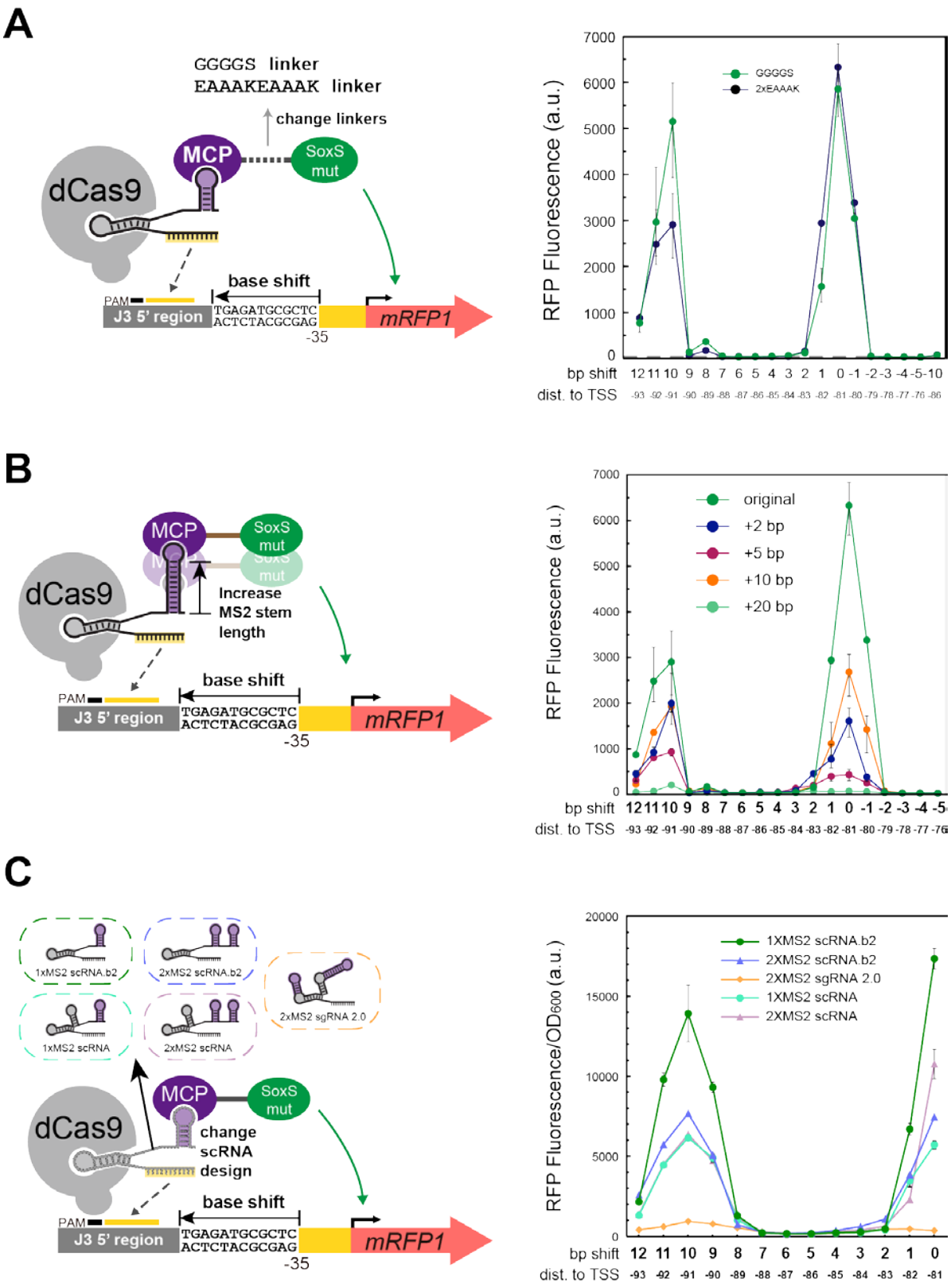


Supplementary Figure S4.4 Computational analysis on the PAM availability in *E. coli* MG1655 intergenic regions.

Vertical dotted red lines indicate the median number of PAM density for each group. 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.

A) Distribution of the number of NGG sequences among *E. coli* MG1655 intergenic regions plotted on a histogram. The median number of NGG PAM sites among *E. coli* MG1655 intergenic regions is 9.0.

B) To better demonstrate the probability of finding PAM sites within a 10 bp window, the term PAM density was defined as the total number of PAM sites found in a single intergenic sequence divided by the length of the intergenic sequence and then multiplied by 10. The median number of NGG PAM density among *E. coli* MG1655 intergenic regions is 0.78.



Supplementary Figure S4.5 Modifying the CRISPRa complex structure does not change its sharp positioning dependence

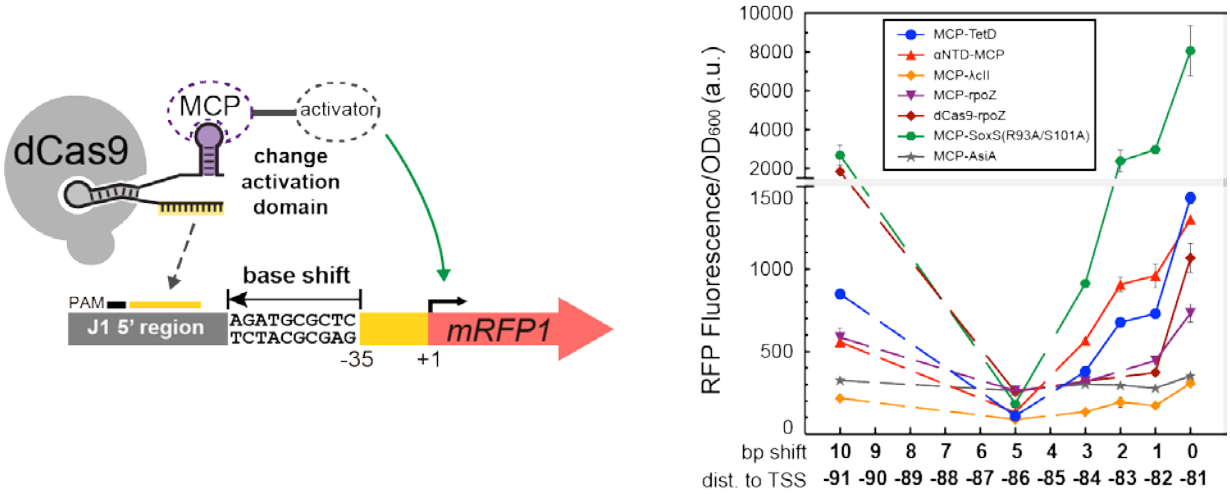
A) Changing the linker between MCP and SoxS does not change the positioning dependence of CRISPRa. A CRISPRa complex with the MCP-(EAAAKEAAAK)-SoxS(R93A/S101A) activation

domain displayed 10 bp periodicity of peak expression similar to that with the MCP-(GGGGS)-SoxS(R93A/S101A) activation domain. The MCP-(EAAAKEAAAK)-SoxS(R93A/S101A) or MCP-(GGGGS)-SoxS(R93A/S101A) activation domains were cloned into the CRISPRa component plasmid that also contains *Sp*-dCas9 and 1XMS2 scRNA.b2 targeting J306, which is then co-transformed into *E. coli* MG1655 together with the J3-J23117-mRFP1 reporter series having 10 bp and 5-1 bp deletions or 0-12 bp insertions upstream of the -35 (Supplementary Figure S4.3B). Cells were grown overnight to late stationary phase and analyzed by flow cytometry. Dots represents the average of median RFP fluorescence among three biological replicates and error bars the standard deviation. 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 has +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 than the original 1XMS2 scRNA.b2. The strain having the scRNA with +20 bp extended MS2 stem did not show any CRISPRa activity. All re-engineered scRNAs target the J306 site and were cloned into the CRISPRa component plasmid that expresses *Sp*-dCas9 and MCP-(EAAAKEAAAK)-SoxS(R93A/S101A). The CRISPRa component plasmid was co-transformed into *E. coli* MG1655 together with the J3-J23117-mRFP1 reporter series having 5-1 bp deletions or 0-12 bp insertions upstream of the -35 (Supplementary Figure S4.3B). Cells were grown overnight to late stationary phase and analyzed by flow cytometry. Dots represents the average of median RFP fluorescence among three biological replicates and error bars the standard deviation.

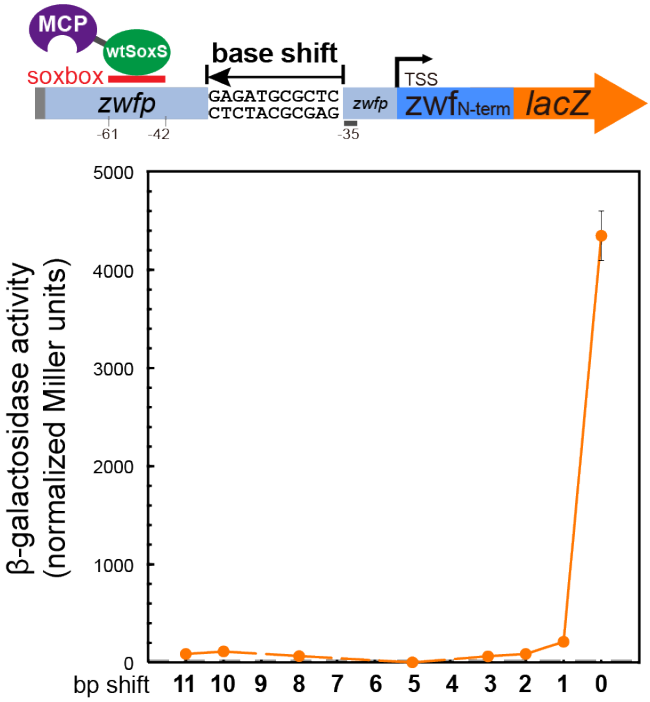
C) No improvements in the effective target range were observed for CRISPRa systems with alternative scRNA designs. The scRNA designs tested are: 1XMS2 scRNA.b2 and 2XMS2 scRNA.b2, the original 1XMS2 and 2XMS2 scRNA design having the tracrRNA hairpin[2], and sgRNA 2.0 where the MS2 hairpins are extended from the RNA stems on the sgRNA[3]. CRISPRa systems having 2XMS2 scRNA.b2, 1XMS2 scRNA, and 2XMS2 scRNA displayed the same 10 bp periodicity but lower peak activity than 1XMS2 scRNA.b2. CRISPRa systems having sgRNA 2.0 did not show any CRISPRa activity. All scRNAs target the J306 site and were cloned into the CRISPRa component plasmid that expresses *Sp*-dCas9 and MCP-(GGGGS)-SoxS(R93A). The CRISPRa component plasmid was co-transformed into *E. coli* MG1655 together with the J3-J23117-mRFP1 reporter series

having 0-12 bp insertions upstream of the -35 (Supplementary Figure S4.3B). Cells were grown overnight to late stationary phase and analyzed on a plate reader. Dots indicate the average fluorescence/OD_{600} values of three biological replicates and the error bars the standard deviation.



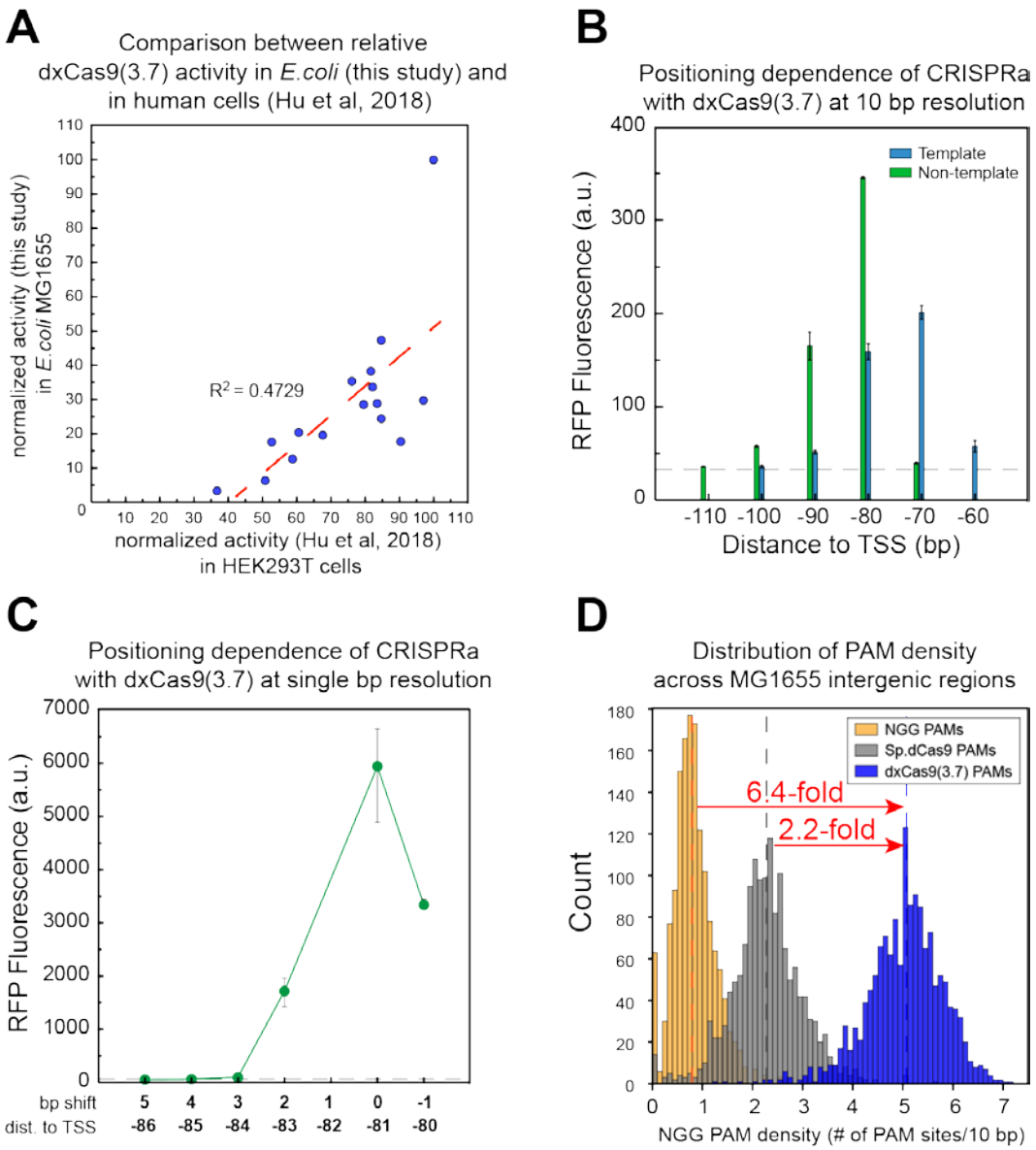
Supplementary Figure S4.6 Similar positioning dependence is observed for CRISPRa with alternative activation domains.

CRISPRa with MCP-TetD and αNTD-MCP activation domains[1] displayed similar positioning dependence as MCP-SoxS(R93A/S101A), where gene expression decreases as the position shifts from 0 to 3 bp, reaching the baseline at 5 bp shift and increases again at 10 bp shift. CRISPRa with MCP-rpoZ[1] and dCas9-rpoZ[4] 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 than MCP-SoxS(R93A/S101A). MCP-λcII and MCP-AsiA[1] 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[1]. The MCP-SoxS(R93A/S101A), MCP-TetD, αNTD-MCP, MCP-λcII and MCP-AsiA plasmids were co-transformed into *E. coli* MG1655 together with the J1-J23117-mRFP1 reporter series having 0, 1, 2, 3, 5, and 10 bp insertions upstream of the -35. The MCP-rpoZ and dCas9-rpoZ plasmids were co-transformed into the *E. coli* strain CD03 (MG1655/ΔrpoZ) (Supplementary Table 1)[1] together with the J1-J23117-mRFP1 reporter series having 0, 1, 3, 5, and 10 bp insertions upstream of the -35. Cells were grown overnight to late stationary phase and analyzed on a plate reader. Dots indicate the average fluorescence/OD_{600} values of three biological replicates and the error bars the standard deviation.



Supplementary Figure S4.7 Wild type SoxS activator displays similar positioning dependence like CRISPRa.

When the wild type SoxS binding site (soxbox) on the endogenous *zwfp* promoter was shifted by 1 bp, gene expression decreases significantly to a very low level. 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 4.1). These reporter plasmids were co-transformed into *E. coli* MG1655 together with a CRISPRa component plasmid that expresses *Sp*-dCas9, MCP-wtSoxS but no scRNAs. Dots indicate the average β-galactosidase activity shown by the normalized Miller unit among three biological replicates and the error bars indicate the standard deviation.



Supplementary Figure S4.8 dxCas9(3.7) displays similar properties to *Sp*-dCas9 and has increased likelihood of finding an active target in the *E.coli* genome.

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.[5] 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). The linear coefficient between the two datasets is $R^2 = 0.4279$

B) CRISPRa with dxCas9(3.7) displayed similar periodic peak activation every 10 bp along the J1-J23117 promoter, with high levels of gene expression at -81 and -91 on the non-template strand and -70 and -80 on the template strand. Plasmids expressing scRNAs targeting -71 to -111 on the non-

template strand (J104, J106, J108, J110, J112) and -60 to -100 TSS on the template strand (J105, J107, J109, J111, J113) of the J1-J23117 promoter was co-transformed with a plasmid expressing dxCas9(3.7) and MCP-SoxS(R93A/S101A) into *E. coli* CD13 strain that has the J1-J23117-mRFP1 reporter integrated into the genome (Supplementary Table 1). Overnight cultures were grown to late stationary phase and analyzed by flow cytometry. Bars represent the average of the median fluorescence among three biological replicates and the error bars the standard deviation. 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 complete reduction of gene expression to the baseline. Reporter gene series where 1 bp deletion or 2-5 bp insertions were introduced upstream of the -35 of the J3-J23117-mRFP1 reporter were constructed. The reporter plasmids were co-transformed into *E. coli* MG1655 together with a CRISPRa component plasmid containing dxCas9(3.7), MCP-SoxS(R93A/S101A), and 1XMS2 scRNA.b2 targeting J306. Cultures were grown overnight to late stationary phase and analyzed by flow cytometry. Dots indicate the average of median fluorescence values of three biological replicates and the error bars the standard deviation. 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).

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 S4.4) among intergenic regions in *E. coli* MG1655 suggests that the average likelihood of finding a dxCas9(3.7)-compatible PAM (blue) is ~6.4-times more than finding an NGG PAM (yellow) and ~2.2-times more than finding an *Sp*-dCas9-compatible PAM (grey). Vertical dotted lines indicate the median number of 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. Vertical dotted lines indicate the median number of PAM density for each group. 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.

4.8.2 | Supplementary tables

Supplementary Table S4.1 *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 ΔrpoZ	[1]
CD06	MG1655/sfGFP (weak promoter)	MG1655 *W1-BBa_J23117-sfGFP KanR::nfsA*	[1]
CD13	MG1655/mRFP (weak promoter)	MG1655 *J1-BBa_J23117-mRFP1::nfsA*	This study, Supplementary Figure S4.8

Supplementary Table S4.2 gRNA target sites

sgRNA target	DNA Sequence	Target Strand[a]	Distance to TSS[b]
W108[c]	GAAGATCCGGCCTGCAGCCA	NT	91
J101[d]	TGGGTTCCACCGGATACCTC	T	40
J103[d]	AGGCGTCCTTTGGGTTCCAC	T	50
J105[d]	CGGTTACCAAAGGCGTCCTT	T	60
J107[d]	CGGTGTCCTGCGGTTACCAA	T	70
J109[d]	AGGTATCCTGCGGTGTCCTG	T	80
J111[d]	GGGCGACCTCAGGTATCCTG	T	90
J113[d]	GGGCCACCACGGGCGACCTC	T	100
J115[d]	TGGTGACCATGGGCCACCAC	T	110
J117[d]	GGGTGACCTATGGTGACCAT	T	120
J119[d]	TGGTTGCCAAGGGTGACCTA	T	130
J121[d]	AGGACACCTTTGGTTGCCAA	T	140
J102[d]	AGGTATCCGGTGGAACCCAA	NT	61
J104[d]	TGGAACCCAAAGGACGCCTT	NT	71
J106[d]	AGGACGCCTTTGGTAACCGC	NT	81
J108[d]	TGGTAACCGCAGGACACCGC	NT	91
J110[d]	AGGACACCGCAGGATACCTG	NT	101
J112[d]	AGGATACCTGAGGTCGCCCG	NT	111
J114[d]	AGGTCGCCCGTGGTGGCCCA	NT	121
J116[d]	TGGTGGCCCATGGTCACCAT	NT	131
J118[d]	TGGTCACCATAGGTCACCCT	NT	141
J120[d]	AGGTCACCCTTGGCAACCAA	NT	151
hAAVS1[d]	GGGGCCACTAGGGACAGGAT	off-target	n/a

J206	TAGTAGCCGAACACGTCCTC	off-target	n/a
J306	TTGTGTCCAGAACGCTCCGT	NT	81
M	AGCAGAAGTGTCAGCAGTGT	NT	81 (reporter A), 86 (reporter B)
M_2	CGACGAGCAGAAGTGTCAGC	NT	76 (reporter A), 81 (reporter B)
aroKB_A1	GGGCAATTATTTCGTCATGA	T	155
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

[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 W108 site for GFP activation was described previously using a 30 base gRNA targeting site (i.e. spacer sequence)[4]. We use a 20 base gRNA targeting site for the same site (i.e. the same PAM site).

[d] 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[1].

Supplementary Table S4.3 Selected *E. coli* Expression Plasmids

Plasmid[a]	**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***	*J3_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*
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 inserted upstream of -35)*	mRFP1	*BBa_B0015*

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[13]. 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 2) used for the J1-mRFP1 reporter (Supplementary Figure S4.4, S5, S7).

[c] Originally described in previous work[1]. Modified versions of this plasmid are available where BBa_J23117 is replaced with minimal promoters regulated by alternative sigma factors (Fig 3B).

[d] Modified version of this plasmid are available where: BBa_J23117 is replaced with Anderson promoters of different strength (Figure 4.3A) and different PAM sites at the -81 J306 site (Figure 4.5A).

Supplementary Table S4.4 Primer sequences for RT-qPCR

Primer	Sequence	Reference
qCD012_16S_f	AAAGTTAATACCTTTGCTCATTGACGTT	[1]
qCD013_16S_r	GACTACCAGGGTATCTAATCCTGTTT	[1]
qCD034_aroK_f	TCTGGTTGGGCCTATGGGTG	This study
qCD034_aroK_r	TACGAATGGTCACGTCGGCA	This study
qCD020_cysK_f	TGCTGAAACCAGGCGTTGAA	This study
qCD021_cysK_r	TCCCAACGCCAGCAATAAATACA	This study
qCD028_ldhA_f	TGGCTGCGAAGCGGTATGTA	This study
qCD029_ldhA_r	GAACGCCAGCAGACGCATAC	This study

4.8.3 | Supplementary methods

Computational analysis of PAM site availability

The intergenic sequences from *E. coli* were obtained from the RegulonDB database[6]. The intergenic sequences between transcriptional units were obtained by discarding the sequences between convergent genes and between intra-operon coding sequences. 5' untranslated regions (UTRs) were removed from sequences using transcription start site (TSS) information from Kim et al.[7] and RegulonDB. Intergenic sequences upstream of genes with no known TSS or where no intergenic sequence remained after removing the UTRs were discarded, yielding 1527 transcriptional units for further analysis. The number of PAM sites was calculated by counting the number of NGG sequences (or alternative PAM sites for dCas9 and dxCas9(3.7)) on both strands. 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.

Sequence of the CRISPR components

Double mutant activation domain MCP-(5aa)-SoxS(R93A/S101A)

The sequence below shows the double mutant MCP-(5aa)-SoxS activation domain that showed a significant reduction in the endogenous activity on 2 SoxS-regulated promoters zwfp and fumCp, while retaining substantial CRISPRa activity on the W108 site of sfGFP (Figure 4.1). The double mutant SoxS(R93A/S101A) was fused with the modified version of the RNA binding protein $MCP_{\Delta FG, V29I}$.

> MCP-(5aa)-SoxS(R93A/S101A) (optimized for minimal endogenous activity, Figure 4.1 and subsequent)

$MCP_{\Delta FG, V29I}$., 5 aa linker, SoxS(R93A/S101A) (underlined are the alanine point mutations)

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY
LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMSHQKIIQDLIAWIDEHIDQPLNIDV
VAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQTFSRVFARQFDRTPA
DYRHRL

dxCas9(3.7) plasmid construction

The plasmid dxCas9(3.7)-VPR was a gift from David Liu (Addgene plasmid # 108383; http://n2t.net/addgene:108383 ; RRID:Addgene_108383)[5]. The dxCas9(3.7) gene PCR amplification from the dxCas9(3.7)-VPR plasmid and then cloned into the Cas protein expression vector (with Sp.pCas9 promoter and BBa_B0015 terminator controlling the Cas protein, and containing an MCP-SoxS(R93A/S101A) activation domain) using the In-Fusion HD cloning kit (Clontech), giving pCD564. Then an scRNA targeting J306 was further inserted into the plasmid by digesting pCD564 with SpeI and annealing a PCR-amplified scRNA fragment via In-Fusion.

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 4.1) and the J1-J23117-mRFP reporter (Figure 4.3, Supplementary Figure S2 S4, S5, S7) for testing the distance dependence property were described in previous study[1].

J3-J23117_mRFP1 reporter (Figure 4.3 & Supplementary Figure S2 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[1].

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

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT
CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTT
CTGGACACAACGTCGTCTTGAAGTTGCGATTATAGATTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCGAAT
TCATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTA
TGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGA
CCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACG
GTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAAT
GGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGT
TCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTT

GGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGA

AAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTT

ACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTG

AAGGTCGTCACTCCACCGGTGCTTAAGGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGC

TCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTC

ACCTTCGGGTGGGCCTTTCTGCGTTTATA

J3(N26)-J23117-sfGFP reporter library (Figure 4.3)

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

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT

CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTT

CTGGACACAANNNNNNNNNNNNNNNNNNNNNNNNNTTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCGAAT

TCATTAAAGAGGAGAAAGGTACCATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAAT

TAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCA

CCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTG

TTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATG

TACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATA

CCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGT

ACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTC

GCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTG

TCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACA

TGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAGGATCCAAAC

TCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTG

TTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA

Modified synthetic promoters

Promoters with different basal expression levels (Figure 4.3)

Annotations:

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

BBa_J23105

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG
ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtactatgctagcGA
ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23106

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG
ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtatagtgctagcGA
ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23107

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG
ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagccctaggtattatgctagcGA
ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23108

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG
ACACAACGTCGTCTTGAAGTTGCGATTATAGActgacagctagctcagtcctaggtataatgctagcGA
ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23109

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC
TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG
ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacagctagctcagtcctagggactgtgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23110

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtacaatgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23111

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtatagtgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23112

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGActgatagctagctcagtcctagggattatgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23113

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGActgatggctagctcagtcctagggattatgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23114

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttatggctagctcagtcctaggtacaatgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23115

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGAtttatagctagctcagcccttggtacaatgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23118

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtattgtgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

BBa_J23119

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGC

TGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG

ACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctaggtataatgctagcGA

ATTCATTAAAGAGGAGAAAGGTACCATG

Promoters regulated by alternative sigma factors (Figure 4.3)

Annotations:

J1 upstream region, variable promoter (color coded), Bujard RBS, Start codon of mRFP. The J106 target site is underlined

J1-sodCp promoter

TGGCAATTCCGACGTCGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACC

TTTGGTTGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTC

CTGCGGTTACCAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACGCCGTTcaaaaatgtgtcacT

GGTTTACACTtattcagGGAATTCATTAAAGAGGAGAAAGGTACCATG

J1-glnAp2 promoter

TGGCAATTCCGACGTCGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACC
TTTGGTTGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTC
CTGCGGTTACCAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACTGGCACagatttCGCTTtatc
tttttTacggcgacGAATTCATTAAAGAGGAGAAAGGTACCATG

J1-rdgBp promoter

TGGCAATTCCGACGTCGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACC
TTTGGTTGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTC
CTGCGGTTACCAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACTTGTAAaaggcgaacttggcC
GCCACaaacaaattGAATTCATTAAAGAGGAGAAAGGTACCATG

J1-yieEp promoter

TGGCAATTCCGACGTCGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACC
TTTGGTTGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTC
CTGCGGTTACCAAAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACGAACTTtagccgctttagtc
tgtcCATCAttccAGAATTCATTAAAGAGGAGAAAGGTACCATG

J3(tetO)-J23117_mRFP reporter (Figure 4.3)

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

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT
CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTT
CTGGACACAACGTCGTCACTCTATCGTTGATAGAGTttgacagctagctcagtcctagggattgtgctagcGAAT
TCATTAAAGAGGAGAAAGGTACCATG

J1-J23117 promoter with shifted bases (Figure 4.4, S5, S7, S9)

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.

TGGCAATTCCGACGTCGCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGT
TGCCAAGGGTGACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCA
AAGGCGTCCTTTGGGTTCCACCGGATACCTCCGGACCTAGATGCGCTttgacagctagctcagtcctagggattg
tgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

J3-J23117 promoter with shifted bases (Supplementary Figure S4.3 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.

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT
CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTT
CTGGACACAACGTCGTCTTGAAGTTGCGATTATAGATGAGATGCGCTCttgacagctagctcagtcctagggatt
gtgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

Modified J3-J23117 promoter with 6 adjacent PAMs around -81 bp to TSS (Supplementary Figure S4.3)

J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP. Region with additional PAM sites inserted. The J306 target site is underlined.

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT
CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACCCCCCCTACGGAGCGTT
CTGGACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAAT
TCATTAAAGAGGAGAAAGGTACCATG

Modified J3-J23117 promoter with non-NGG PAMs around -81 bp to TSS (Figure 4.5)

J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP, region with modified PAMs. The J306 target site is underlined.

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT
CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTNNNACGGAGCGTT
CTGGACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAAT
TCATTAAAGAGGAGAAAGGTACCATG

Promoter to demonstrate dxCas9(3.7) versatility (Figure 4.5)

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

TGGCAATTCCGACGTCAGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCAT
CGCTACGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACACTGCTGA
CACTTCTGCTCGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAAT
TCATTAAAGAGGAGAAAGGTACCATG

4.8.4 | Supplementary references

1. 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).

2. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160,** 339–350 (2015).

3. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583 (2015).

4. 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).

5. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556,** 57 (2018).

6. Santos-Zavaleta, A. *et al.* RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res* **47,** (2018).

7. Kim, D. *et al.* Comparative Analysis of Regulatory Elements between Escherichia coli and Klebsiella pneumoniae by Genome-Wide Transcription Start Site Profiling. *PLoS Genetics* **8,** e1002867-15

Chapter 5 | Exploring the Portability of CRISPR-Cas Transcriptional Regulators in Alternative Bacteria Species

5.1 | Introduction

In metabolic engineering, biosynthetic platforms have been limited to a few organisms that have well-studied genetics and physiology. While bioengineers are enjoying the benefits of these standardized organisms with their rich variety of available genetic tools, non-standardized bacteria have recently caught more and more attention[1–3]. Bacteria have evolved to occupy unique niches in the biosphere; hence their physiological properties are diverse and many among them have potential applications in bio-industry. For example, methanotrophic bacteria that can utilize 1-carbon sources such as methane are a potential host for biosynthesis from natural gas or refinery waste gas[4]. There is also a large group of bacteria including *Pseudomonas* that degrade toxic pollutants[5,6], and they could become platforms for bioremediation-coupled synthesis. Cyanobacteria uses carbon dioxide in the atmosphere as the carbon source and light as the energy source[2,7–11], which will make biosynthesis potentially more cost-efficient and environmentally friendly. Halophilic bacteria are tolerant to high pH and high salt conditions and could possibly eliminate sterility restrictions in fermentation plants[12]. Given the unique productive features of those organisms discussed above, the value for developing the necessary synthetic biology devices in these hosts are remarkable. However, due to the lack of predictable and effective genetic tools available, many non-standardized bacteria are still under-utilized for metabolic engineering.

Unlike in *E. coli*, tools for programmable and multi-gene transcriptional regulation have been deficient in non-standardized bacteria. However, because bacteria share a common scheme for transcription initiation and regulation, it is reasonable to believe the regulation devices developed in model organisms such as *E. coli* could be functional in other bacteria as well. In this chapter, to explore the transportability of the CRISPRa system, we describe the efforts to characterize it in 3 non-*E. coli* bacteria: the gram-positive firmicute *Bacillus subtilis*, the gram-negative gammaproteobacteria *Pseudomonas putida*, and the haloalkaliphilic gammaproteobacterial methanotroph *Methylomicrobium buryatense* (Figure 5.1A). *B. subtilis* is known for its highly potent secretion machinery, making it a promising host for protein manufacturing[13]. It is also well-studied and has a considerable number of

basic genetic tools[14–17]. *P. putida* is a non-pathogenic bacterium that is capable for degrading toxic aromatic compounds such as toluene and styrene, which could become a bio-platform for both bioremediation and biosynthesis[5,6,18,19]. *M. buryatense* is a methanotroph that utilizes 1-carbon compounds such as methane and methanol [20–23], which could become a promising platform for biosynthesis from natural gas and oil refinery waste gas. Because these organisms have unique metabolic capabilities, developing a CRISPRa system in these hosts will not only allow us to learn about CRISPRa's host-compatibility, but also add a useful regulation device in their synthetic biology toolbox.

Previous publications have validated the utility of CRISPRi in *B. subtilis*[24,25] and *P. putida*[26,27]. In order to test the functionality of CRISPRi in *M. buryatense* and CRISPRa in all three organisms, we adapted simple host-specific assays established in previous studies[22,25,28]. In our initial experiments, we have validated CRISPRi in *M. buryatense* on an endogenous metabolic gene. For testing CRISPRa, we built host-specific reporters that are all driven by the previously characterized J1-J23117 promoter (Figure 2.5A)[29] and selected scRNA target sites based on our predictions in *E. coli*. We observed CRISPRa activity in *P. putida* but not in *B. subtilis* and *M. buryatense*. Further characterization of the CRISPRa system in *P. putida* shows that it behaves similarly compared to its equivalent in *E. coli* in terms of distance and phase dependence properties. We then analyzed the failure modes of CRISPRa-SoxS system in *B. subtilis* and *M. buryatense* and hypothesized that the MCP-SoxS activation domain might not be functioning properly due to reasons such as low expression and stability or inability to interact with the native transcription machinery. Lastly, we propose a scheme for selecting host-specific activation domains for importing CRISPRa to novel bacteria.

5.2 | Results

5.2.1 | Perspectives on the transportability of CRISPRa into the selected organisms

In chapter 1, we briefly discussed the possibility of transporting the CRISPRa-SoxS system into non-*E. coli* bacteria. SoxS natively recruits the RNA polymerase (RNAP) by interacting with the C-terminal domain of RNAP α-subunit. As shown in chapter 2, we aligned the sequence of the rpoA gene that encodes the RNAP α-subunit from a variety of bacteria including *B. subtilis*, *P. putida*, and *M. buryatense* and analyzed the sequence conservation of the specific residues that has been previously

shown to interact with SoxS (Figure 5.1B)[29]. We found that the sequence of the SoxS-interacting regions is well-conserved across different phyla such as Proteobacteria, Firmicutes, Bacteroidetes, and Cyanobacteria[29] (Supplementary Figure 2.8). The notion that the CRISPR-Cas system is universally functional across many different bacterial organisms has been widely accepted[30,31]. Since the development of CRISPR interference, it has been implemented in many non-*E. coli* bacteria including *B. subtilis* [24,25] and *P. putida* [26,27] with high effectivity and reproducibility. Together, these two evidence point towards the possibility to engineer the CRISPRa-SoxS system as a transcriptional regulation tool in non-*E. coli* bacteria.

5.2.2 | CRISPRa-SoxS is functional in *P. putida* having similar properties compared to in *E. coli*

Pseudomonas putida is a Gram-negative, rod-shaped, saprotrophic soil bacterium that belongs to the Gammaproteobacterial class[32]. Unlike its pathogenic relative in the same family, *Pseudomonas aeruginosa*, it is *generally recognized as safe* (GRAS). The isolation and study of *P. putida* has spanned more than 50 years where people have discovered its unique properties such as its ability to degrade toxic aromatics through the TOL pathway[33–35]. Nowadays, *P. putida* has been recognized by its high potential in bioremediation and biosynthesis applications. After decades of engineering, people have developed biosafety laboratory working strains of *P. putida* such as KT2440[36] and F1[37] and added a comprehensive toolbox for genome manipulations and gene expression, including methods for genomic integration[38], gene delivery by broad-host-range plasmids[39], and gene expression devices[40].

To set up an assay to test CRISPRa-SoxS in *P. putida*, we took our previous CRISPRa-SoxS(R93A-S101A) cassette with their original expression devices from *E. coli*. We also added a weakly expressed reporter gene using J1-J23117 promoter to drive the sfGFP. Both the CRISPRa cassette and the reporter gene cassette were moved into a mini-Tn7 vector and integrated into the genome of *KT2440* via transposon-mediated site-specific insertion[38]. We cloned the scRNAs into the broad-host-range vector pBBR1MCS and transformed it into the engineered CRISPRa + reporter strain via electroporation[39]. The newly constructed strains were then grown in liquid culture until late stationary phase and their fluorescence were analyzed by flow cytometry.

We first tested CRISPRa on the J109 scRNA target, which is -80 bp to the transcription start site (TSS) on the template strand. We observed a 8.5-fold CRISPRa activity relative the off-target negative control (Figure 5.2A). We then transformed plasmids having the complete set of the scRNAs targeting every site on the J1 promoter into our CRISPRa + reporter strain, so that we can learn about its distance rules in *P. putida* (Figure 5.2A). We observed active CRISPRa in *P. putida* KT2440 with the distance dependence behaving similarly to what has been found in *E. coli* with minor differences: high CRISPRa activity was observed at -81 bp and -91 bp relative to the TSS on the non-template strand, which is the same as in *E. coli*. On the template strand, the maximum activity is at -80 bp, while in *E. coli* high CRISPRa activities were observed at -80 bp and -70 bp (Figure 5.2B). In *P. putida*, the decay of CRISPRa activity when the target moves out from the optimal range is a bit slower: weak CRISPRa activity is observed at -101 bp and -111 bp on the non-template strand and -60 bp and -90 bp on the template strand. Apart from these trivial differences, the distance requirements for effective CRISPRa targeting in *P. putida* are very similar compared in *E. coli*, and the effective CRISPRa target range is from -81 bp to -111 bp on the non-template strand and from -60 bp to -100 bp on the template strand.

Next, we explored the phase dependence properties of CRISPRa-SoxS in *P. putida*. We designed an assay that is similar to what has been described in chapter 4: We built a series of integration plasmids from the CRISPRa + reporter construct where we consecutively inserted single base pairs between the J1 upstream region and the J23117 minimal promoter. This generated a set of reporter genes that has 1 bp ~ 11 bp shifts (Figure 5.3A). We integrated these cassettes into the KT2440 genome and then transformed a scRNA targeting the site (J106) where maximum activity was observed on the 0 bp shifted reporter. As a result, we observed the same trend observed in *E. coli*, where CRISPRa activity begin to decrease to two-third of the maximum activity at 1 bp and 2 bp shifts. At 3 bp shift, the activity reaches one third of the maximum and reduces to the baseline after 4 bp shift. CRISPRa activity begins to reappear at 9 bp shift and is increased to about half of the maximum value at 11 bp shift (Figure 5.3B). Putting the data from the distance and the phase assay together, the properties of CRISPRa-SoxS in *P. putida* are similar to *E. coli*, which is not surprising given that the two species are relatively closely related and have a well-conserved SoxS-interaction surface on the RNAP.

5.2.3 | CRISPRi activity was observed in *M. buryatense*, but CRISPRa-SoxS did not show activity

Methylomicrobium buryatense is an aerobic gammaproteobacterial type I methanotroph that was originally isolated from natural gas-containing soda lakes[20]. Its natural adaptability to the harsh environment in the soda lake makes it tolerant to high pH levels, high salt concentrations, and fluctuating temperatures. This makes *M. buryatense* a good platform for methane-derived biosynthesis because it is tolerant to fermentation conditions that excludes growth of contaminating microorganisms[21,22]. Efforts have been made on enabling genetic engineering in *M. buryatense*, where people have developed a genome-sequenced laboratory strain 5GB1. And by curing its native plasmid to allow the delivery of IncP-based broad-host-range plasmids, people generated the engineered 5GB1C strain, where stable integration sites, replicable plasmids, and constitutive promoters have been validated[22]. Although there are some reported usages of CRISPR technologies in methanotrophs[41], the CRISPR-Cas system has never been systematically tested in *M. buryatense*. Here we present our work in attempting to test the functionality of CRISPRi and CRISPRa in *M. buryatense* 5GB1C.

To construct a strain to test CRISPRi or CRISPRa, dCas9 was cloned into a delivery vector that allows gene inserts to be integrated at the MBURv2_130895 locus[22]. The dCas9 was expressed by the ptet promoter and the whole cassette was integrated into *M. buryatense* 5GB1C through bacterial conjugation followed by sucrose counter-selection to eliminate the antibiotic resistance marker. Then we transformed a sgRNA expressed by the J23119 promoter in a broad-host range IncP-based vector[50] into 5GB1C through conjugation[22]. The sgRNA targets the mmoY gene, which is the first gene of the operon that encodes part of the soluble methane monooxygenase (sMMO), one of the enzymes in methanotrophs to catalyze the first step of methane metabolism, converting methane to methanol[43]. Natively, the sMMO genes are also regulated by the extracellular presence of Cu^{2+}, where sMMO is turned on in the absence of Cu^{2+} and turned off in the presence of Cu^{2+}. We constructed 2 strains that contain different sgRNAs targeting different sites on the mmoY gene, grew the strains in liquid culture with no Cu^{2+} together with base strains that are grown in media supplied with or without trace amount of Cu^{2+}, and then performed real-time quantitative PCR (RT-qPCR) to determine the expression of the sMMO genes. Our results show that both guide sites had around 26-fold repression of mmoY compared to the strain without the sgRNA. And our positive control where the no sgRNA

strain was grown in media containing Cu^{2+} showed no expression in mmoY (Figure 5.4A). This result suggests that CRISPRi works the way we expected in *M. buryatense* 5GB1.

To construct the strain for testing CRISPRa, dCas9 and MCP-SoxS(R93A-S101A) were integrated at the MBURv2_130895 locus. The dCas9 and MCP-SoxS(R93A-S101A) was expressed by the ptet promoter. Another reporter cassette was constructed that includes the catechol dioxygenase gene (xylE) driven by the J1-J23117 promoter. This cassette was cloned into a delivery vector and integrated into the genome at the MBURv2_60220 locus. Based on the *E. coli* predictions, scRNAs targeting the J1 promoter that are supposed to be the active (J106 and J108) and inactive (J103) were selected and cloned into the broad-host range IncP-based vector and subsequently transformed into the *M. buryatense* 5GB1C bearing the dCas9, activation domain and the reporter gene. CRISPRa activities were then evaluated by the xylE assay[28,44]. The result shows no observable increase xylE activity on any of the guide targets compared to the control (Figure 5.4B). In conclusion, although with the same expression devices validated by CRISPRi for expressing dCas9 and the guide RNAs, CRISPRa-SoxS did not give any observable transcriptional activation in *M. buryatense* 5GB1 at predicted active sites based on previous *E. coli* data.

5.2.4 | Efforts to transport the CRISPRa-SoxS system into *B. subtilis*

Bacillus subtilis is one of the most well-characterized Gram-positive bacteria. Due to its ability to form endospores upon nutrient fluctuation, it has been studied as a model organism for bacterial cell differentiation and competence[45–47]. Its high efficiency in protein secretion has also made it a promising host for the biomanufacturing of proteins[13]. *B. subtilis* has natural competence to uptake DNA during late logarithmic growth phase, and the uptake of DNA is tightly associated with homologous recombination[45–47]. This greatly simplified the genomic integration of synthetic DNA and led to the characterization of multiple stable integration vectors for synthetic biology[48,49]. In addition, other devices for building and testing transcription units were developed in *B. subtilis*, such as replicable plasmids[15], constitutive and inducible promoters[14], and reporter genes[15]. CRISPRi has also been widely applied in *B. subtilis* for pathway regulation in metabolic engineering or phenotypic screens to analyze gene functions[24,25].

We first validated the functionality of CRISPRi in B. subtilis by targeting the CRISPRi complex to a strongly expressed GFPmut1 gene (Figure 5.5A). In order to test the activity of CRISPRa-SoxS in *B. subtilis*, we built a reporter strain from the wild-type laboratory strain 168[50] using two integration plasmids. The first plasmid contains the dCas9 and the MCP-SoxS(R93A-S101A) activation domain, where the dCas9 and the activation domain were expressed by the pveg promoter. The whole gene cassette was cloned into the pBs1C vector that will integrate into the amyE locus in the genome[16]. The second plasmid contains a GFPmut1 fluorescence reporter, which is expressed by the J1-J23101 promoter and cloned into the pBs4S vector that will integrate into the thrC locus in the genome[15,16]. After the reporter strain was successfully constructed, a set of replicable plasmids derived from the pBs0E ori1030 vector[15] containing an scRNA targeting at sites on the J1-J23101 promoter (J103, J106, J108, J109, and J112) was transformed into the reporter strain and CRISPRa activity was assessed by the GFP fluorescence levels using flow cytometry. Positive controls where the GFPmut1 was expressed by a strong pveg promoter showed an increase in fluorescence over the baseline, suggesting that the reporter strain and the fluorescent assay worked properly. We observed no detectable CRISPRa activity either on sites that should be active (J106, J108, and J109) or should not be active (J103 and J112) based on the *E. coli* predictions (Figure 5.5B).

5.2.5 | Comparison of the expression levels of the activation domains among 4 bacterial organisms

Since CRISPRi has been validated in all the organisms discussed in this chapter, the lack of CRISPRa activity detected in *B. subtilis* and *M. buryatense* is unlikely to stem from the malfunction of dCas9 or the guide RNAs. One possible mode of failure could be the insufficient expression of the activation domain. Our previous study to explore the dynamics of CRISPRa on the expression of each components suggested a strong dependence on the levels of the activation domain, suggesting that the activation domain is the limiting component in the system[29]. Thus, we hypothesize that insufficient expression of the activation domain could be one of the reasons for the inactivity of CRISPRa in *B. subtilis* and *M. buryatense*.

In order to test the expression levels of the MCP-SoxS(R93A-S101A) activation domain, we used MCP-specific antibodies to do a quantitative western blot. The band intensities generated from densitometry analysis were normalized to the total amount of cells (OD_{600} × mL) from which the

lysates were prepared (Figure 5.6A). The following strains were compared in the analysis: our original *E. coli* MG1655 strain containing MCP-SoxS(R93A-S101A) expressed by the J23107 promoter; the *P. putida* KT2440 strain that contains MCP-SoxS(R93A-S101A) expressed by the J23107 promoter; the two *B. subtilis* 168 strains containing the activation domain expressed by either pliaG or pveg promoter; and the *M. buryatense* 5GB1C strains containing the activation domain expressed by either J23107 or ptet promoter. Among the strains tested, CRISPRa activity was only observed in the *E. coli* and the *P. putida* strain. Western blot analysis shows that the *E. coli* MG1655 strain had the highest level of expression. The *P. putida* KT2440 strain has an expression level about half of that in *E. coli*. In *B. subtilis* 168 strain, the pliaG promoter did not give any detectable level of MCP-SoxS(R93A-S101A), while the pveg promoter showed an expression level comparable to that in *P. putida*. The fact that the strain bearing pveg-MCP-SoxS(R93A-S101A) did not display detectable CRISPRa activity suggest that expression is less likely to be the cause of failure in *B. subtilis*. In *M. buryatense* 5GB1C, both the J23107 promoter and the ptet promoter showed barely detectable expression levels of MCP-SoxS(R93A-S101A) around 10-fold lower than that in *P. putida*, which could be at an insufficient level for proper CRISPRa activity (Figure 5.6B).

5.3 | Discussion

In this study, the functionality of the CRISPR-Cas system has been tested in 3 non-*E. coli* bacteria. While *B. subtilis* and *P. putida* already have well-established CRISPRi systems, there is lack of precedence in CRISPR-Cas-based transcription regulation in *M. buryatense*. We resolved this uncertainty by demonstrating CRISPRi-mediated repression on the sMMO gene in *M. buryatense*. Given that the CRISPRi system is applicable in all three organisms, we can conclude that the dCas9 and the sgRNA are functioning normally in them. If the activation domain can interact with the RNAP from all these species, which is not an unreasonable assumption based on the phylogenetic analysis of the SoxS-interacting surface of the RNAP α subunit (Figure 5.1), we would expect that the CRISPRa system is also transportable into these bacteria species. Our attempts to transport the CRISPRa system into those bacteria resulted in detectable CRISPRa activity in one of the species, *P. putida*, but not *B. subtilis* and *M. buryatense*. We then checked the expression levels of the activation domain that is expressed in each of the organisms. Compared to the expression level in *P. putida* where CRISPRa activity was detected, one of the *B. subtilis* strains had comparable expression level

while still did not show detectable CRISPRa activity. Both of the *M. buryatense* strains had much weaker expression levels of the activation domain than the one in *P. putida*.

In less well-characterized organisms like *M. buryatense*, our engineering power will be limited by the availability of predictable expression devices. No detectable CRISPRa activity was observed in strains where the MCP-SoxS(R93A-S101A) activation domain was expressed ptet promoter (Figure 5.4B) and the J23107 promoter (data not shown). According to our quantitative western blot data, these promoters gave only weak expression to the MCP-SoxS(R93A-S101A) activation domain (Figure 5.6) thus it could be one of the reasons why CRISPRa did not show any activity in *M. buryatense*. To overcome the expression problem, more efficient and robust expression devices need to be developed. If expressing the activation domain by a strong constitutive promoter does not render a stable integration, toxicity issues of MCP-SoxS to the host cells need also to be considered. Our recent work on building an arabinose-inducible pBAD promoter can potentially provide an option for increased expression of the activation domain and at the same time avoiding potential toxicity issues (work in progress).

The fact that in *B. subtilis* the pveg promoter gave expression levels of MCP-SoxS(R93A-S101A) comparable to the one in *P. putida*, but still did not display any detectable CRISPRa activity suggests that the expression level of MCP-SoxS(R93A-S101A) might not be the only problem. This made us revisit our previous assumption that SoxS is universally a functional activation domain across bacterial species. Even if SoxS interacts well with the α subunit in all the species, it might not be the only contributing factor for positive transcription initiation. The assembly of other RNAP components and the protein-protein and protein-DNA interaction within the transcription unit might be significantly different among these organisms. A recent study that mapped the transcription profiles of a library containing thousands of regulatory elements originating from diverse bacterial species showed substantially differentiated behaviors of regulatory elements in different host organisms[51]. When the study compared the activity of the regulatory elements in 3 different bacteria organisms: *E. coli*, *B. subtilis*, and *P. aeruginosa*, they only found 16.9% of the library are universally active across all three organisms. In addition, the study shows that among the 3 organisms, *B. subtilis* is the least tolerant to exogenous regulatory elements, where only 10.8% of the Enterobacteracea and 3.2% of the

Pseudomonaceae regulatory elements were active[51]. This evidence suggests the possibility that the SoxS transcription factor that is originated from *E. coli* might not adapt to the native genomic context in *B. subtilis*. Furthermore, similar problems might be also present in *M. buryatense*. In addition to activator-transcription unit interactions, the composition and biochemical environments in the cytoplasm might also differ greatly between bacteria species. For example, *M. buryatense* has a higher intracellular lipid content than *E. coli* and grows at high pH[22], which could create discrepancies in the cytoplasmic biochemical compositions that affect the proper folding and function of the SoxS activation domain. Together, this evidence along with our data suggest that SoxS might not be universally applicable in CRISPRa systems in all bacteria.

In summary, CRISPRa-SoxS can be applied in other non-*E. coli* bacterial systems, but the SoxS activation domain is not universally transportable to all bacterial species. Low activator expression from poorly characterized genetic tools, incompatibility of the activation domain to the native machinery, and divergent cytoplasmic environments could all contribute to the non-universality of CRISPRa-SoxS. However, the fact that CRISPRi works well in most bacteria and that mechanisms for transcription activation are mostly generalizable among bacteria re-enforces our belief that CRISPRa should in principle work in any bacteria if we can find a working activation domain that expresses well in the cell. Therefore, we propose to screen host-specific activators when transporting the CRISPRa system. Similar to our work in *E. coli*, we can rationally select transcription factors or proteins that interact with the host-specific transcription machinery and perform reporter screen assays to select for host-specific activation domains. We can also perform a genome-wide screening where MCP is fused with a cDNA library from the host genome and screen for proteins that demonstrated CRISPRa activity. In addition, predictable, dynamic and efficient expression devices need to be developed in bacteria that are not well-studied. With the growing understanding on the physiology in non-standard bacteria and the expansion of synthetic biology devices in them, the development of host-compatible programmable CRISPRa-based transcription regulation system should be moving towards a positive direction.

5.4 | Methods

5.4.1 | Strain construction and manipulation

All plasmid constructions and preparations were conducted in *E. coli* NEB turbo. *E. coli* MG1655 was used for testing the expression of the activation domain. Plasmid and strain construction for in *E. coli* systems were described in previous chapters. All *P. putida* strains and plasmid vector described above were gifts obtained from the Harwood lab at University of Washington. The dCas9 and activation domains were cloned into the pminiTn7 integration vector for miniTn7-mediated transfer[38]. To briefly describe the conjugation protocol: The *E. coli* S17 pir+ strain harboring the pTNS1 transposon plasmid were mixed with *E. coli* DH5α strain harboring the pminiTn7 donor plasmid with the CRISPR cassettes, the *E. coli* DH5α strain harboring the pRK2013 helper plasmid, and the acceptor strain *P. putida* KT2440. The mixed biomass was then spread onto an LB plate and grown overnight at 30 °C, after which the biomass was transferred onto a plate with selective Pseudomonas isolation agar (PIA) (BD Difco) containing 30 μg/mL gentamycin to select for the positive conjugants. PCR-verified colonies were then transformed with pflpZ plasmid to cure the gentamycin resistance marker via the flippase mechanism. The pBBR1MCS vector was used to clone the scRNA construct[39]. Electrocompetent cells of the reporter strain were prepared by spinning down a 6 mL 20-hour overnight culture, washing the pellet twice with 8mL sterile 300 mM sucrose, and resuspending with 200 μL sterile 300 mM sucrose. The mixture of 50 ng DNA and 100 μL competent cells was then electroporated in a 2-mm gap electro cuvette and electroporated at 2.5 kV in a MicroPulser electroporator (Biorad).

All *M. buryatense* strains and plasmids were gifts from the Lidstrom lab at University of Washington. Integration plasmids were cloned from the delivery vectors pAWP49 and pMH15 and the scRNA gene was cloned into the broad-host range IncP-based vector pAWP78[22,42]. The plasmids were transformed into *E. coli* S17-1 conjugation strain, which becomes to donor strain. The *M. buryatense* 5GB1C acceptor strains were spread onto an NMS2 mating plate[22] and mixed with the donor strain. After 2 days of incubation in 25% methane in air at 30 °C, biomass from the mating plate was then then transferred into NMS2 plates[22] containing 50 μg/mL of kanamycin and incubated for another 4 days. Colonies were then patched onto NMS2 plates containing kanamycin (50 μg/mL) and rifampicin (50 μg/mL) to select out the donor strains. The resulting biomass were then resuspended in 1 mL

NMS2 and plated onto NMS2-rifampicin (50 μg/mL) plates containing 7.5% sucrose to counter-select inserts that have the kanamycin marker. Positive integration strains that has the marker successfully flipped out was confirmed by patching colonies onto NMS2 and NMS2-kanamycin plates and verifying with colony PCR.

All *B. subtilis*-related strain*s and plasmid vectors were obtained from the Bacillus Genetic Stock Center (BGSC* http://www.bgsc.org/*). Two plasmids vectors were used for integrations: The* pBs1C vector has the dCas9 and the activation domain and integrates into the amyE locus and the pBs4S vector has the reporter genes and integrates into the thrC locus[16]. The replicable plasmid pBs0E with the ori1030 origin was used as the vector to clone the scRNAs[15]. The protocol adapted from previous publication was used to deliver the plasmids[25], *B. subtilis* strain 168 was grown overnight in 3 mL MC medium[25] and then inoculated 1 OD×mL into 10 mL BMK medium[25] and grown at 37 °C until the culture reached OD_{600} = 1.5. Then 150 μL of the culture were mixed with 1 μL of ~250 ng/μL DNA and incubated at 37 °C without shaking for 10 min, followed by incubation at 37 °C with shaking for 2 hours. The cells were then plated on LB plates containing the appropriate antibiotics and grown at 37 °C overnight. Positive integrations were verified by colony PCR.

5.4.2 | Flow cytometry

P. putida strains were grown from single colonies in 500 μL cultures of LB supplied with the appropriate antibiotics in 96-deep-well plates at 30 °C with 225 rpm shaking for 20 hours. *B. subtilis* strains were grown from single colonies in 500 μL cultures of LB supplied with the appropriate antibiotics in 96-deep-well plates at 37 °C with 200 rpm shaking for 20 hours. The liquid cultures were then diluted 1:50 in 200 μL DPBS and transferred into a clear round-bottom 96-well plate and analyzed on a MACSQuant VYB flow cytometer (Miltenyi Biotec). Gating and data analysis methods were described in chapter 2[29].

5.4.3 | xylE assays

Gene expression was analyzed via the xylE assay that has been described previously[28,44]. *M. buryatense* strains were first spread on NMS2 plates containing appropriate antibiotics and incubated in 25% methane in air at 30 °C. Then they were inoculated into 5 mL liquid cultures of NMS2 with

antibiotics and shaken at 220 rpm in sealed glass tubes containing 25% methane in air at 30 °C. After 20 hrs, the cultures were propagated into a fresh 5 mL culture and grow in the same way until the OD_{600} reaches 0.5 ~ 0.8. Cells were then spun down and resuspended in 50 mM Tris pH=7.5 to a final OD_{600} of 0.5. Then 100 µL of resuspended cells were then mixed with 10 µL 10 mM freshly prepared catechol solution and immediately analyzed in a SpectraMax 190 plate reader (Molecular Devices). Absorbance at 375 nm was measured every 20 seconds for 10 min, the slope of the curve in the linear interval was compared to determine the xylE activity.

5.4.4 | Quantitative RT-qPCR

M. buryatense colonies were inoculated into 5 mL liquid cultures of NMS2 with antibiotics and shaken at 220 rpm in sealed glass tubes containing 25% methane in air at 30 °C. After 20 hrs, the cultures were propagated into a fresh 5 mL culture that contains appropriate antibiotics with or without 7.04 µM Cu^{2+}. Cells grown in media containing 7.04 µM Cu^{2+} were harvested at OD 600 = 0.7 ~ 0.8, Cells grown in media without 7.04 µM Cu^{2+} were harvested at OD 600 = 0.4 ~ 0.5. 2 ~ 3 OD × mL of culture was spun down and the pellet was proceeded for RNA extraction. RNA extraction, reverse transcription and quantitative RT-qPCR protocols were described in chapter 2[29]. The rpoD gene was used as the reference gene.

5.4.5 | Quantitative western blot

Strains of 4 different bacteria all containing the dCas9, MCP-SoxS(R93A-S101A), and reporters were grown the same way as described in the methods above for 20 hrs. Cells were then harvested, spun down and resuspended in 1 mL DPBS. The resuspended cells were then lysed at 100 kPa in a French Pressure Cell Press (SIM-AMINCO). The lysate was then mixed 4:1 with 5XSDS loading dye (225 mM Tris pH = 6.8, 50% glycerol, 5% SDS, 0.25 M DTT, and 0.75 mM bromophenol blue) and boiled at 95 °C for 15 min. The samples were then run on a Mini Protean TGX Precast SDS-PAGE gel (Biorad) and subsequently transferred onto a nitrocellulose membrane. The membrane was then placed into a BlotCylcer Touch (Precision Biosystems) for automated blocking and antibody incubation. The cycle was set to overnight blocking (12 hrs) ; 5 hrs primary antibody incubation; 5 × 5 min washes; 1 hr secondary antibody incubation; followed by 3 × 5 min washes. The Odyssey blocking buffer (LI-COR) was used for the blocking and washes. Anti-Bacteriophage MS2 Coat

Protein [3H4] Antibody (Kerafast) was used as the primary antibody in a 1:5000 solution in Odyssey blocking buffer. IRDye 800CW Donkey anti-Mouse IgG secondary antibody (LI-COR) was used as a secondary antibody in a 1:10000 diluted solution in Odyssey blocking buffer. The membrane was imaged on the Odyssey IR imager (LI-COR) and densitometry analyses was conducted using the Image Studio Lite software (LI-COR). Normalized expression levels for each samples was calculated by dividing the band intensities from the densitometry analysis to the total amount of cells used to prepare the lysates (OD_{600} × mL).

5.5 | Acknowledgements

We would like to thank Carrie Harwood and Amy Schaefer for providing the *P. putida* KT2440 strain, the integration plasmids pminiTn7, pTNS1, pRK2013 and pflpZ, broad-host-range vector pBBR1MCS and helpful advice regarding *P. putida*. We would like to thank Mary Lidstrom, Melissa Hendershott, Aaron Puri, Amanda Smith, Frances Chu, Yanfen Fu, and Joseph Groom for providing the *M. buryatense* 5GB1C strain, the integration vectors pAWP49 and pMH15, the replicable IncP-based vector pAWP78, lab space, supplies, and guidance for all the *M. buryatense* experiments. We also would like to thank Maureen Thomason and Jason Peters for helpful guidance regarding *B. subtilis*.

5.6 | Reference

1. Clomburg, J. M., Crumbley, A. M. & Gonzalez, R. Industrial biomanufacturing: The future of chemical production. *Science* **355,** aag0804 (2017).

2. Czajka, J., Wang, Q., Wang, Y. & Tang, Y. J. Synthetic biology for manufacturing chemicals: constraints drive the use of non-conventional microbial platforms. *Applied Microbiology and Biotechnology* **101,** 7427–7434 (2017).

3. Keasling, J. D. Manufacturing Molecules Through Metabolic Engineering. *Science* **330,** 1355–1358

4. Kalyuzhnaya, M. G., Puri, A. W. & Lidstrom, M. E. Metabolic engineering in methanotrophic bacteria. *Metabolic Engineering* **29,** 142–152 (2015).

5. Molina-Santiago, C. *et al.* Pseudomonas putida as a platform for the synthesis of aromatic compounds. *Microbiology (Reading, England)* **162,** 1535–1543 (2016).

6. Loeschcke, A. & Thies, S. Pseudomonas putida—a versatile host for the production of natural products. *Applied Microbiology and Biotechnology* **99,** 6197–6214 (2015).

7. Woo, H. M. & Lee, H. J. Toward solar biodiesel production from CO2 using engineered cyanobacteria. *FEMS microbiology letters* **364,** (2017).

8. Angermayr, S., Rovira, A. & Hellingwerf, K. J. Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends in biotechnology* **33,** 352–61 (2015).

9. Wang, B., Wang, J., Zhang, W. & Meldrum, D. R. Application of synthetic biology in cyanobacteria and algae. *Frontiers in Microbiology* **3,** 344 (2012).

10. Heidorn, T. *et al.* Chapter Twenty-Four Synthetic Biology in Cyanobacteria Engineering and

Analyzing Novel Functions. *Methods in Enzymology* **497,** 539–579 (2011).

11. Pakrasi, H. B. Synthetic biology of cyanobacteria: unique challenges and opportunities. 1–14 doi:10.3389/fmicb.2013.00246/abstract

12. Yin, J., Chen, J.-C., Wu, Q. & Chen, G.-Q. Halophiles, coming stars for industrial biotechnology. *Biotechnology Advances* **33,** 1433–1442 (2015).

13. van Dijl, J. & Hecker, M. Bacillus subtilis: from soil bacterium to super-secreting cell factory. *Microbial Cell Factories* **12,** 1–6 (2013).

14. Guiziou, S. *et al.* A part toolbox to tune genetic expression in Bacillus subtilis. *Nucleic Acids Research* **44,** 7495–7508 (2016).

15. Popp, P. F., Dotzler, M., Radeck, J., Bartels, J. & Mascher, T. The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with Bacillus subtilis. *Scientific Reports* **7,** s41598-017-15107-z (2017).

16. Radeck, J. *et al.* The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with Bacillus subtilis. *Journal of Biological Engineering* **7,** 1–17 (2013).

17. Song, Y. *et al.* Promoter Screening from Bacillus subtilis in Various Conditions Hunting for Synthetic Biology and Industrial Applications. *PLoS ONE* **11,** e0158447-18

18. Nikel, P. I. & de Lorenzo, V. Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metabolic Engineering* (2018). doi:10.1016/j.ymben.2018.05.005

19. Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V. & Wittmann, C. Industrial biotechnology of Pseudomonas putida and related species. *Applied Microbiology and Biotechnology* **93,** 2279–2290

(2012).

20. Kaluzhnaya, M. *et al.* Taxonomic Characterization of New Alkaliphilic and Alkalitolerant Methanotrophs from Soda Lakes of the Southeastern Transbaikal Region and description of Methylomicrobium buryatense sp.nov. *Systematic and Applied Microbiology* **24,** 166–176 (2001).

21. Gilman, A. *et al.* Bioreactor performance parameters for an industrially-promising methanotroph Methylomicrobium buryatense 5GB1. *Microbial Cell Factories* **14,** 182 (2015).

22. Puri, A. W. *et al.* Genetic Tools for the Industrially Promising Methanotroph Methylomicrobium buryatense. *Applied and Environmental Microbiology* **81,** 1775–1781

23. Torre, A. *et al.* Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in Methylomicrobium buryatense strain 5G(B1). *Microbial Cell Factories* 1–15 doi:10.1186/s12934-015-0377-3

24. Wu, Y. *et al.* CRISPRi allows optimal temporal control of N-acetylglucosamine bioproduction by a dynamic coordination of glucose and xylose metabolism in Bacillus subtilis. *Metabolic Engineering* (2018). doi:10.1016/j.ymben.2018.08.012

25. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165,** 1493–1506

26. Tan, S., Reisch, C. R. & Prather, K. L. A Robust CRISPR Interference Gene Repression System in Pseudomonas. *Journal of Bacteriology* **200,** e00575-17 (2018).

27. Kim, S. *et al.* CRISPR interference-mediated gene regulation in Pseudomonas putida KT2440. *Microbial Biotechnology* (2019). doi:10.1111/1751-7915.13382

28. Chu, F. & Lidstrom, M. E. XoxF Acts as the Predominant Methanol Dehydrogenase in the Type I

Methanotroph Methylomicrobium buryatense. *Journal of Bacteriology* **198,** 1317–1325

29. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9,** 2489 (2018).

30. Choi, K. & Lee, S. CRISPR technologies for bacterial systems: Current achievements and future directions. *Biotechnology Advances* **34,** 1180–1209 (2016).

31. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183

32. Regenhardt, D. *et al.* Pedigree and taxonomic credentials of Pseudomonas putida strain KT2440. *Environmental Microbiology* **4,** 912–915 (2002).

33. Silva-Rocha, R., Pérez-Pantoja, D. & de Lorenzo, V. Decoding the genetic networks of environmental bacteria: regulatory moonlighting of the TOL system of Pseudomonas putida mt-2. *The ISME journal* **7,** 229–32 (2013).

34. Bertini, L. *et al.* Deepening TOL and TOU catabolic pathways of Pseudomonas sp. OX1: cloning, sequencing and characterization of the lower pathways. *Biochimie* **95,** 241–50 (2013).

35. Franklin, F., Bagdasarian, M., Bagdasarian, M. & Timmis, K. Molecular and functional analysis of the TOL plasmid pWWO from Pseudomonas putida and cloning of genes for the entire regulated aromatic ring meta cleavage pathway. *Proceedings of the National Academy of Sciences* **78,** 7458–7462 (1981).

36. Belda, E. *et al.* The revisited genome of Pseudomonas putida KT2440 enlightens its value as a robust metabolic chassis. *Environmental Microbiology* **18,** 3403–3424 (3403).

37. Zylstra, G., McCombie, W., Gibson, D. & Finette. Toluene degradation by Pseudomonas putida

F1: genetic organization of the tod operon. *Applied and environmental microbiology* **54,** 1498–503 (1988).

38. Choi, K.-H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nature Protocols* **1,** nprot.2006.24 (2006).

39. Kovach, M. E. *et al.* Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166,** 175–176 (1995).

40. Cook, T. B. *et al.* Genetic tools for reliable gene expression and recombineering in Pseudomonas putida. *Journal of Industrial Microbiology & Biotechnology* **45,** 517–527 (2018).

41. Tapscott, T., Guarnieri, M. T. & Henard, C. A. Development of a CRISPR/Cas9 System for Methylococcus capsulatus In Vivo Gene Editing. *Appl Environ Microb* **85,** (2019).

42. Marx, C. J. & Lidstrom, M. E. Development of improved versatile broad-host-range vectors for use in methylotrophs and other Gram-negative bacteria. *Microbiology* **147,** 2065–2075 (2001).

43. Lawton, T. J. & Rosenzweig, A. C. Methane-Oxidizing Enzymes: An Upstream Problem in Biological Gas-to-Liquids Conversion. *Journal of the American Chemical Society* **138,** 9327–9340

44. Puri, A. W. *et al.* Quorum Sensing in a Methane-Oxidizing Bacterium. *Journal of Bacteriology* **199,** e00773-16 (2017).

45. Anagnostopoulos, C. & Spizizen, J. Requirements for transformation in bacillus subtilis. *Journal of bacteriology* **81,** 741–6 (1961).

46. Dubnau, D., Davidoff-Abelson, R. & Smith, I. Transformation and transduction in Bacillus subtilis: Evidence for separate modes of recombinant formation. *Journal of Molecular Biology* **45,** 155–179 (1969).

47. Dubnau, D. Genetic competence in Bacillus subtilis. *Microbiological reviews* **55,** 395–424 (1991).

48. Guérout-Fleury, A.-M., Frandsen, N. & Stragier, P. Plasmids for ectopic integration in Bacillus subtilis. *Gene* **180,** 57–61 (1996).

49. Härtl, B., Wehrl, W., Wiegert, T., Homuth, G. & Schumann, W. Development of a new integration site within the Bacillus subtilis chromosome and construction of compatible expression cassettes. *Journal of bacteriology* **183,** 2696–9 (2001).

50. Zeigler, D. R. *et al.* The origins of 168, W23, and other Bacillus subtilis legacy strains. *Journal of bacteriology* **190,** 6983–95 (2008).

51. Johns, N. I. *et al.* Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nature Methods* **15,** 323 (2018).

5.7 | Figures

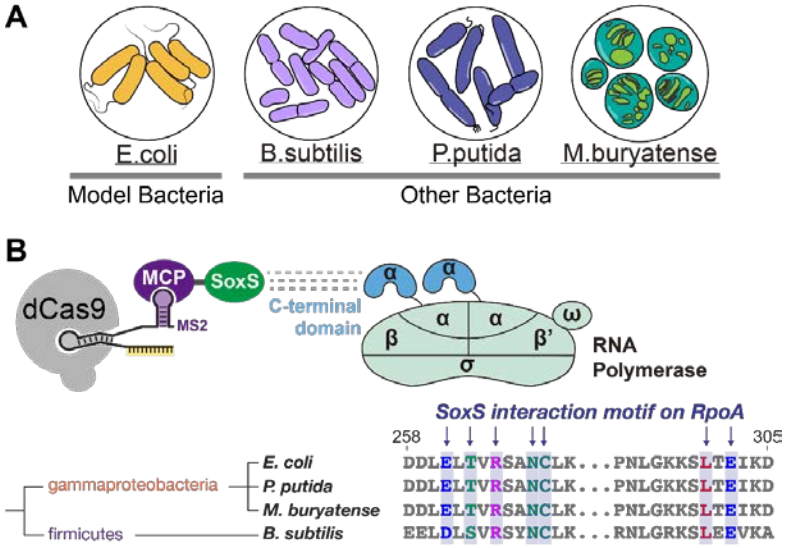


Figure 5.1 Transporting the CRISPRa system into non-*E. coli* bacteria

A) *Bacillus subtilis*, *Pseudomonase putida*, and *Methylomicrobium buryatense* were chosen as non-*E. coli* host organisms for testing CRISPRa activity. B) Phylogenetic analysis of the SoxS-interacting region on the RNAP α subunit (RpoA) in the 4 bacterial organisms tested. The residues that have been shown to interact with SoxS are shaded in grey and color coded. The SoxS interactions site is well conserved in all four organisms. This illustration is reprinted from part of the work done in chapter 2[29].

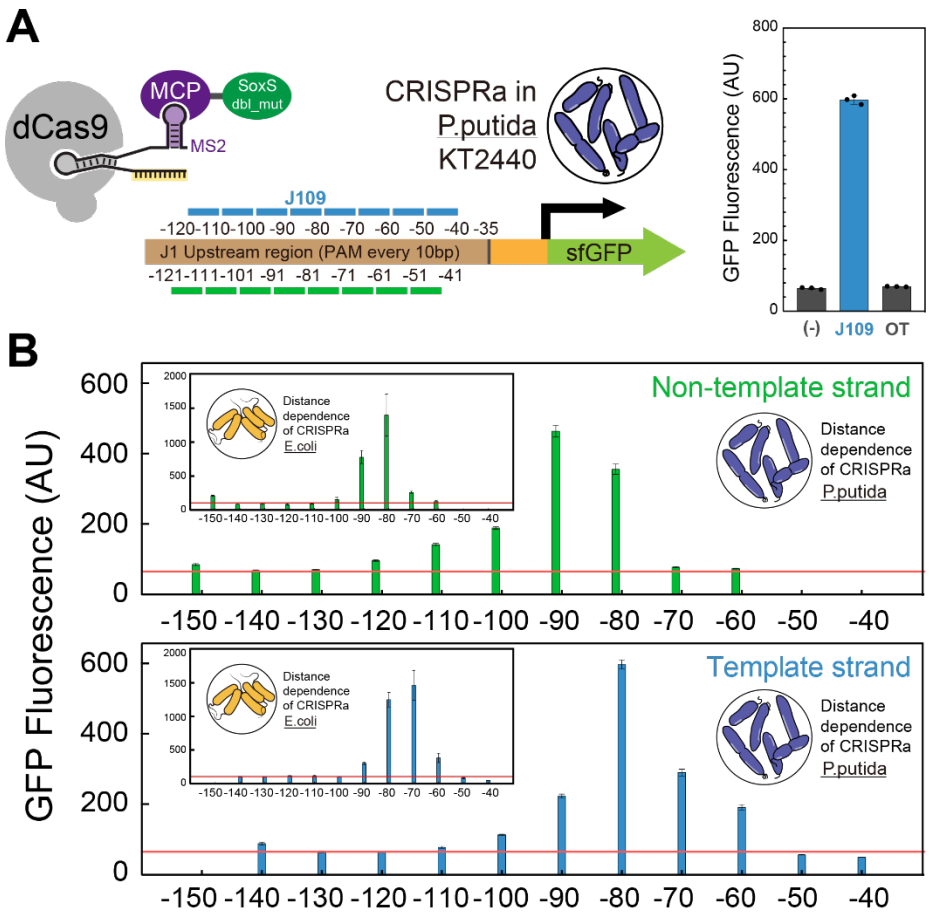


Figure 5.2 CRISPRa in *Pseudomonas putida* and its distance dependence properties

A) Left: The J1-J23117-sfGFP was constructed and integrated into *P. putida* KT2440 together with pCas9-dCas9 and J23107-MCP-SoxS(R93A-S101A). scRNAs that target sites along the J1 promoter (J101-J121) were expressed on a broad-host-range plasmid. Cells were grown to late stationary phase and CRISPRa activities were measured by GFP fluorescence using flow cytometry. Right: At the J109 site where the scRNA is -80 bp to the TSS on the template strand, CRISPRa was active giving an 8.5-fold increase in fluorescence. (-) represents a strain where no CRISPRa components are expressed and OT represents a strain where CRISPRa components are expressed with an off-target scRNA. Bars represent the average of the median fluorescence values measured by flow cytometry among 3 biological replicates and error bars represents the standard deviation. Black dots represent the median fluorescence values for each individual replicate. B) CRISPRa activities in *P. putida* along the J1 promoter in a 10 bp interval. Top plot shows CRISPRa activity on the J1 target sites on the non-template strand and bottom plot shows CRISPRa activity on the J1 target sites on the template strand. Subplots shows the CRISPRa activity on the J1 target sites in *E. coli* along the corresponding strands. CRISPRa displayed similar distance dependence properties in *P. putida* than

in *E. coli*, with high activities centered around -81 ~ -91 bp on the non-template strand and -70 ~ -80 bp on the template strand. P. putida CRISPRa displayed a slightly wider targetable range than E. coli were weak CRISPRa was observed at -100 on the non-template strand and -90 on the non-template strand. Bars represent the average of the median fluorescence values measured by flow cytometry among 3 biological replicates and error bars represents the standard deviation. The red line represents the negative baseline with a strain where the CRISPRa components are expressed with an off-target scRNA.

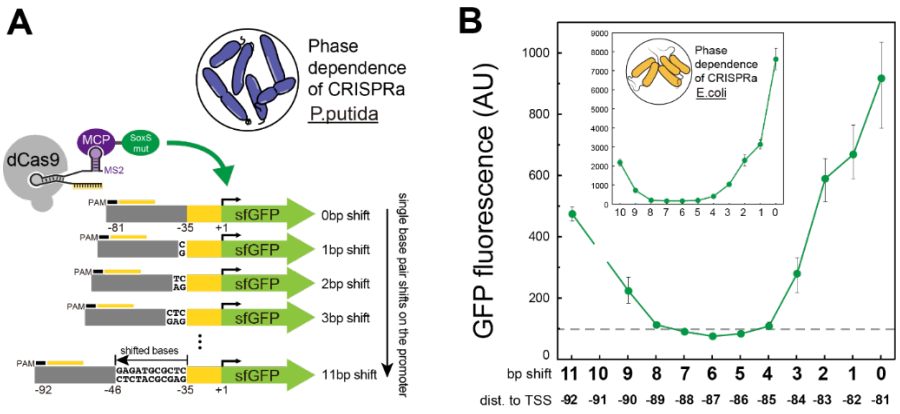


Figure 5.3 Phase dependence properties in *Pseudomonas putida.*

A) The experimental design was adopted from the phase dependence experiment in *E. coli* in chapter 4: a series of J1-J23117-sfGFP reporter genes were generated that contain 1~11 added base pairs before the -35 site of the J23117 promoter (the +10 bp reporter was attempted but did not generate positive clones). The reporter genes were integrated into the genome of KT2440 together with pCas9-dCas9 and J23107-MCP-SoxS(R93A-S101A). Cells were grown to late stationary phase and CRISPRa activities were measured by GFP fluorescence using flow cytometry. B) CRISPRa activity in *P. putida* targeting the reporters with shifted scRNA target positions from the original maximum site. CRISPRa activity drops to the baseline when the scRNA target is shifted 4 bp from its original maximum. The CRISPRa activity begins to come back starting from 9 bp shifts. Subplot shows the data for the same assay on J1-J23117-mRFP in *E. coli* obtained in chapter 4: X-axis on the subplot indicate the number of bases added upstream of the -35 region. Y-axis on the subplot indicate the RFP fluorescence value measured by flow cytometry. Comparison between the data in *P. putida* and *E. coli* reveals similar phase dependence properties in these two organisms. Dots represent the average of the median fluorescence values measured by flow cytometry among 3 biological replicates and error bars represents the standard deviation.

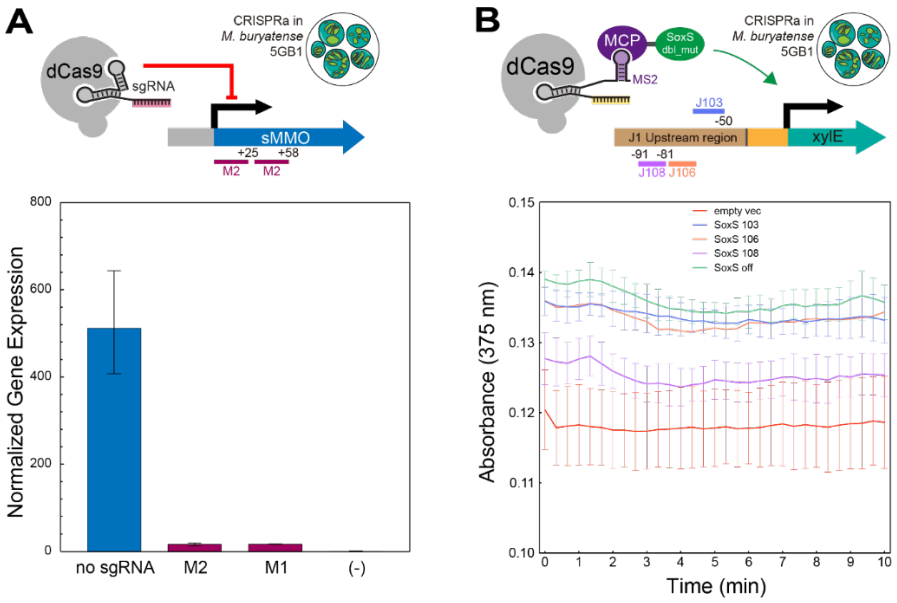


Figure 5.4 Characterization of CRISPRi and CRISPRa in *M. buryatense*.

A) CRISPRi repressing the sMMO operon. The CRISPRi components including ptet-dCas9 and the J23119-sgRNAs were delivered into *M. buryatense* 5GB1C through mating. Two sgRNA targets were selected to target the smmoY gene at +25 and +58 bp to the start codon (M1 and M2). RT-qPCR were performed on cDNA prepared from cultures harvested at 0.4 ~ 0.5 OD using primers that amplify off of the smmoY gene. Compared to the negative control where all the dCas9 is expressed but the sgRNA is not, the CRISPRi strains that target both M1 and M2 displayed significant repression of the smmoY gene. The negative control where the cells were grown in media containing 7.04 μM Cu^{2+} also showed high reduction in smmoY gene expression. Bars indicate the normalized expression values where the amount of smmoY gene expression calculated by the ΔΔCt method was normalized to the negative control where the cells were grown in media containing 7.04 μM Cu^{2+}. Error bars indicate the standard deviation among 3 technical replicates. B) CRISPRa activity was not observed in *M. buryatense* 5GB1C on the J1 target sites tested by the xylE assay. J1-J23117-xylE reporter genes and the CRISPRa components ptet-dCas9 and ptet-MCP-SoxS(R93A-S101A) was integrated into the genome of *M. buryatense* 5GB1C. Plasmids expressing scRNAs that target 3 sites on the J1 promoter (J103, J106, and J108) were transformed into the CRISPRa strain. Negative control trains containing ptet-dCas9 and ptet-MCP-SoxS(R93A-S101A) having an empty vector (empty vec) or a plasmid expressing off-target scRNA (SoxS off) were included. The cells were grown overnight and CRISPRa activity was measured by the xylE assay. Curves indicate the absorbance values at 375 nm measured every 20 sec over a 10 min time period. Error bars indicate the standard deviation among 3

biological replicates. The slope of the curve indicates the rate of the conversion of the substrate by catechol dioxygenase, which reflects the amount of expression of xylE. There are no changes in absorbance in 375 nm observed in any of the strains tested, indicating that no detectable xylE expression was present in these strains. This suggest that CRISPRa was not active in *M. buryatense* 5GB1C at the J103, J106, and J108 sites.

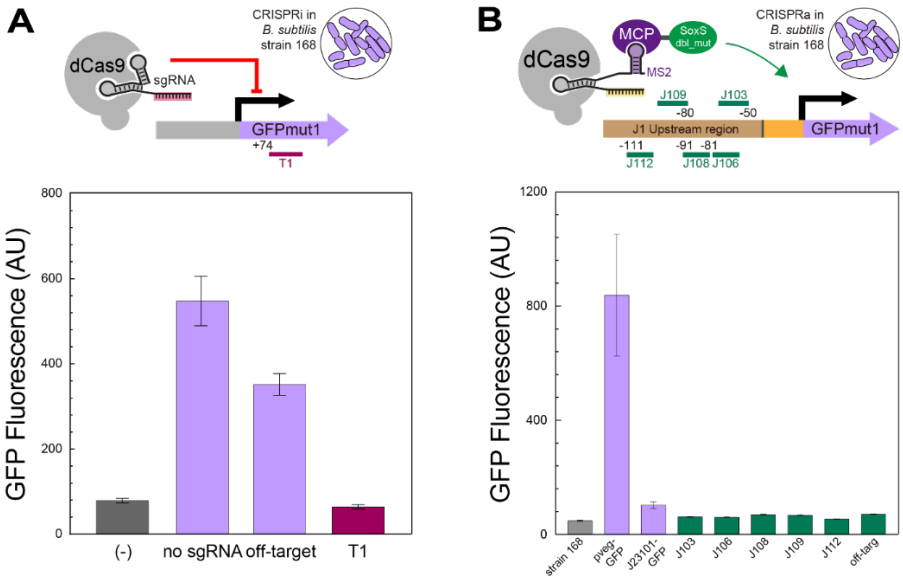


Figure 5.5 Characterization of CRISPRi and CRISPRa in *B. subtilis*.

A) CRISPRi repression on heterologous reporter gene in *B. subtilis*. A strongly expressed pveg-GFPmut1 reporter and the pveg-dCas9 was integrated into the genome of *B. subtilis* strain 168. The pveg-sgRNA targeting a site +74 bp (T1) to the start codon of GFPmut1 was delivered into the cell via a replicable plasmid. CRISPRi activity was determined by the GFPmut1 level measured by flow cytometry. (-) indicates the wild type strain 168. GFPmut1 expression was fully repressed to the basal level (empty strain 168) when the CRISPR complex was targeted at T1. The strain where an off-target scRNA was expressed displayed slightly reduced fluorescence level compared to the negative control where no scRNA plasmid is present. This is likely due to the added antibiotic burden that is necessary to maintain the off-target scRNA plasmid. Bars represent the average of the median fluorescence values measured by flow cytometry among 3 biological replicates and error bars represents the standard deviation. B) CRISPRa was not detected in *B. subtilis* along the J1-J23117 promoter. A weakly expressed J1-J23101-GFPmut1 reporter and the CRISPR components containing pveg-dCas9 and pveg-MCP-SoxS(R93A-S101A) was integrated into the genome of *B. subtilis* strain 168. Pveg-scRNAs targeting 5 different sites (J103, J106, J108, J109, J112) along the J1 promoter were cloned into the replicable plasmid and transformed into the strains containing the other CRISPRa components. A strongly expressed pveg-GFPmut1 reporter strain was used as the positive control. All the strains tested displayed fluorescence levels close to the base strain (J23101-GFP) or the negative control expressing an off-target scRNA (off-targ), suggesting that none of the sites brought significant CRISPRa activities in *B. subtilis*. Bars represent the average of the median fluorescence values

measured by flow cytometry among 3 biological replicates and error bars represents the standard deviation.

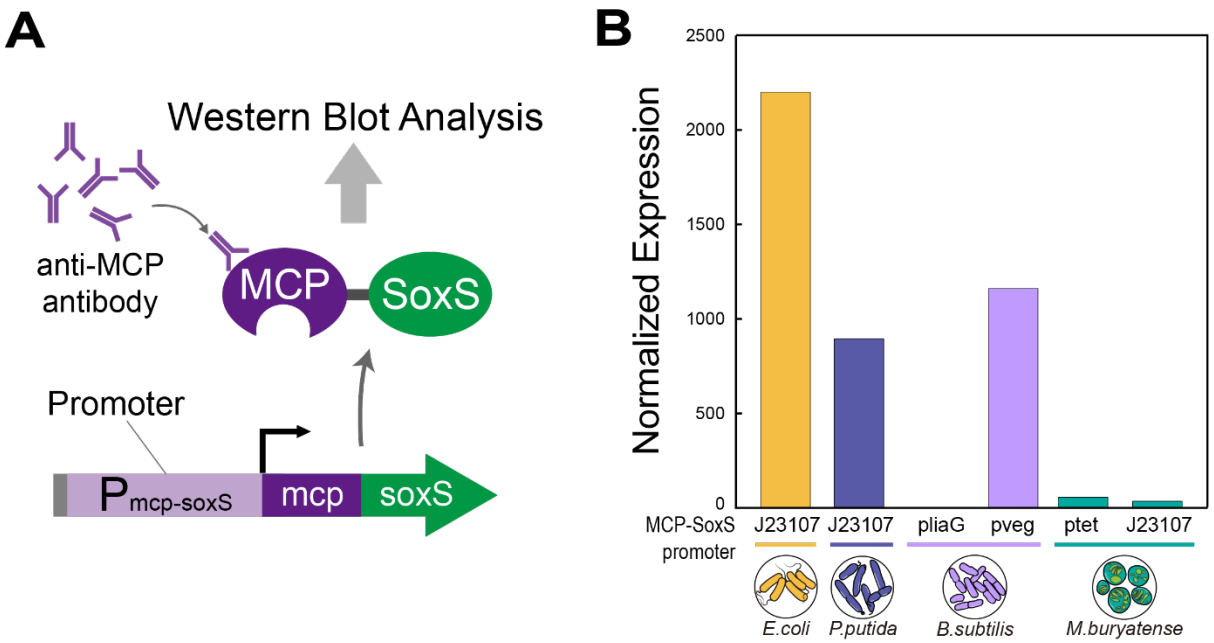


Figure 5.6 Expression levels of the MCP-SoxS activation domain in *E. coli*, *P. putida*, *B. subtilis*, and *M. buryatense*.

A) Different expression devices were used to express the MCP-SoxS activation domain in the four different bacterial organisms. The cells were grown up to stationary phase, lysed and the levels of expressions were analyzed by western blot using anti-MCP antibodies to quantify the amount of activation domain. B) Normalized values of MCP-SoxS activation domain expression measured by western blot. The normalized expression was calculated by dividing the band intensity from densitometry analysis to the total amount of cells (OD_{600} × mL) used to prepare the lysate. Between the 2 strains that displayed detectable CRISPRa activity, *P. putida* has activation domain expression half of that in *E. coli*. In *B. subtilis*, the pliaG promoter displayed no detectable expression of the activation domain and the pveg promoter displayed similar levels of activation domain expression compared to *P. putida*. In *M. buryatense*, both the ptet promoter and the J23107 promoter displayed only weak expression level of the activation domain.

Chapter 6 | Concluding Remarks

6.1 | Summary and future directions

CRISPR-Cas regulators are one of the most programmable gene regulation tools available. Enabling its utility in bacteria could have far-reaching impact on synthetic biology and metabolic engineering, especially given the high potential of non-model bacteria that utilize alternative carbon/energy sources. In this thesis, we have established a working set of CRISPR-Cas activators with different activation domains in bacteria, including our most-characterized MCP-SoxS and others such as MCP-TetD, MCP-AsiA and αNTD-MCP. We have demonstrated the utility of the MCP-SoxS CRISPRa system in several proof-of-concept experiments. We have further explored the dynamics of the widely utilized CRISPRi repression system and learned its properties for tunable repression. We also learned about the restrictive target positioning requirements for efficient activity and found a way to partially relieve the positioning requirements by implementing an evolved Cas protein with broader PAM recognition. Finally, we tested CRISPRa with the MCP-SoxS activation domain in 3 non-*E. coli* bacteria organisms and observed activity on one of them, *P. putida*. In this work, we have comprehensively studied bacterial CRISPR-Cas systems in different aspects, which generated a well-characterized synthetic transcriptional regulation device setting up the foundation towards bioengineering applications in bacteria.

There are questions that remains to be answered to fully understand the utility of bacterial CRISPR-Cas regulators. Therefore, future work should start by addressing the remaining problems: First, our attempts to learn the rules were focused on systematically testing the properties of CRISPRa in a controlled setup on artificially designed synthetic promoters. Whether our rules could be applied on endogenous promoters still remains to be answered. High-throughput screening approach can be applied by generating a scRNA library targeting endogenous promoters in *E. coli*, sorting the cells that were delivered with the library and all other CRISPR components by the activation of target genes. The readout could be either fluorescence from reporter libraries expressed by endogenous transcription units[1] or fitness gains caused by the change in expression of endogenous genes when cells are exposed to stress. Active scRNA target sites can then be analyzed to elucidate any predictable rules for endogenous CRISPRa targeting (Figure 6.1A).

To truly demonstrate its capability in strain optimization, more sophisticated prototyping of CRISPRa in complex biosynthetic pathways is required. The heterologous metabolic pathway for synthesizing p-amino phenylalanine is a multi-step process that has been shown to require precise regulation to avoid cellular intermediate toxicity and increase production titer[2], making it an ideal platform to test the capacity and potency of multi-gene control with CRISPRa and CRISPRi (Figure 6.1B).

We proposed in chapter 5 the method of selecting host-specific activation domains in order to maximize chances of success when transporting CRISPRa into non-model bacteria. Since the transportability into industrially-relevant bacterial host is a major factor for evaluating bacterial regulation devices, future work should prioritize developing a general strategy for transporting CRISPRa to non-model bacteria. A recent study has validated CRISPRa using direct fusions of dCas9-ω and dCas9-α in *B. subtilis*[3]. However, great significance still remains in finding functional activation domains for troubleshooting CRISPRa in *M. buryatense*. Possible alternative candidates for CRISPRa include AraC homologs in *M. buryatense*. Future experiments could involve high-throughput methods that fuse a host-specific cDNA library with MCP and perform a library screening for CRISPRa activity (Figure 6.1C). In conclusion, future goals should be set on deepening the understanding the CRISPRa system and extending its utility, which can take a step forward towards complex regulatory applications in bioindustry.

6.2 | References

1. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods* **3,** nmeth895 (2006).

2. Stevens, J. T. & Carothers, J. M. Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS Synthetic Biology* **4,** 107–115

3. Lu, Z. *et al.* CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in Bacillus subtilis. *Nucleic Acids Research* gkz072- (2019). doi:10.1093/nar/gkz072

6.3 | Figures

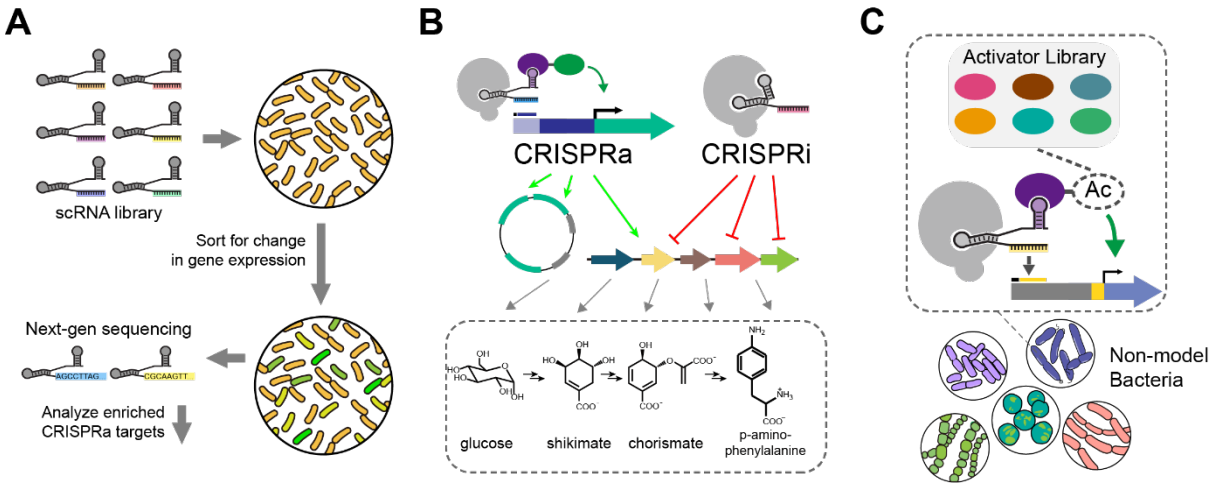


Figure 6.1 Future directions.

A) high-throughput screening using scRNA library to analyze active target sites and derive possible rules for optimal targeting. B) Further demonstration of the capability of the CRISPR-Cas system in regulating the p-amino phenylalanine production. C) Screening host-specific transcriptional activation domains for transporting CRISPRa into non-model bacteria.

~ The End ~