Dynamic CRISPRa/i regulation of Gene expression in CFS and E.coli

Benjamin Ilya Tickman

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

James M. Carothers, Chair

Jesse G. Zalatan

Eric Klavins

Herbert Sauro

Program Authorized to Offer Degree

Molecular Engineering

University of Washington

Abstract

Dynamic CRISPRa/i regulation of Gene expression in CFS and E.coli

Benjamin Ilya Tickman

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

Effective control of gene expression underlies many modern biotechnological applications from metabolic engineering to diagnostics and bio-computation. Historically, a paucity of orthogonal and engineerable regulators has stymied efforts to increase the scale and complexity of gene regulatory networks. The development of CRISPR-based transcriptional regulators has enabled the generation of increasingly sophisticated and practical regulatory networks. The advent of transcriptional networks operating through the regulated expression of guideRNAs has further expanded the composability of CRISPR-regulation, allowing guideRNAs to serve as a standardized information carrier, greatly simplifying the level matching process in mult-layer circuits. Despite these rapid advances, until recently effective CRISPR-regulation in prokaryotes has been limited to CRISPR inhibition due to the lack of versatile activating domains. To meet the needs of increasingly ambitious undertakings we have developed a prokaryotic CRISPRa/i control system programmable through the regulated expression of guideRNAs. In this work we first establish design principles allowing the formation of multi-layer CRISPRa/i regulatory circuits to provide complex and dynamic control of gene expression. To improve upon CRISPRa/i network function we subsequently engineered expression characteristics of CRISPRa/i nodes to provide increased output dynamic ranges, enabling formation of CRISPRa/i expression programs with increased complexity. Finally, we discuss the application of this modular control system to provide dynamic regulation of gene expression in CFS and E.coli towards optimization of bioproduction. The dynamic regulatory capabilities afforded by the CRISPRa/i control system greatly expands the design space of genetically encoded expression programs. This expanded set of capabilities will enable the rapid generation of genetically encoded, dynamic, multi-gene programs providing access to new avenues for the optimization of metabolic engineering as well as complex signal processing in biological systems.

Table of contents

Introduction

In nature cells dynamically regulate gene expression to efficiently respond to environmental stimuli (Zaslaver et al., 2004). In contrast, many metabolic engineering efforts to date have focused on tuning the levels of constitutively expressed heterologous pathway genes to maximize titers (Cress et al., 2015; Jeschek et al., 2017; Xu et al., 2017; Zhao et al., 2015). While conceptually straightforward, static strategies often fail to maximize product titers, especially for heterologous pathways which are overly burdensome, produce toxic intermediates (Stevens and Carothers, 2015), or compete for resources with native cellular processes (Moon et al., 2009). Recent pathway optimization efforts in E.coli have demonstrated that multi-gene dynamic transcriptional regulatory strategies can further increase titers (Doong et al., 2018; Gupta et al., 2017; Xu et al., 2014). In spite of these advances, rational transcriptional engineering efforts in cells are confounded by the fact that host organisms generate both transcriptional and post transcriptional responses to perturbations imposed by metabolic engineers in an attempt to maintain homeostasis (Paddon et al., 2013). This cellular response can result in redirection of flux, minimizing pathway burden, often at the expense of production. Broadly, the work herein describes the development of transcriptional circuits providing dynamic control of gene expression. These circuits opperate through the use of dCas9 based controllers enabling independent targeting of activation (CRISPRa) and repression (CRISPRi)(Dong et al., 2018; Fontana et al., 2020). Combining CRISPRa with CRISPRi permitted the formation of CRISPRa/i circuits operating through the regulated expression of guideRNAs (Tickman & Alba Burbano et al.,2021). These CRISPRa/i

circuits form the basis of a scalable and dynamic control system acting in Cell free systems and *E. coli*. When applied to metabolic engineering efforts, we envision that these regulatory capabilities will generate improved titers, rates, and yields of biosynthetic products. More broadly, the ability of dynamic CRISPRa/i circuits to execute logical operations that are fast, efficient, and complex will streamline the development of biocomputational systems.

Introductory references

Cress, B.F., Toparlak, Ö.D., Guleria, S., Lebovich, M., Stieglitz, J.T., Englaender, J.A., Jones, J.A., Linhardt, R.J., and Koffas, M.A.G. (2015). CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in E. coli. ACS Synth. Biol. *4*, 987–1000.

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

Doong, S.J., Gupta, A., and Prather, K.L.J. (2018). Layered dynamic regulation for improving metabolic pathway productivity in Escherichia coli. Proc. Natl. Acad. Sci. *115*, 2964–2969.

Dudley, Q.M., Karim, A.S., and Jewett, M.C. (2015). Cell-free metabolic engineering: Biomanufacturing beyond the cell. Biotechnol. J. *10*, 69–82.

Fontana, J., Dong, C., Kiattisewee, C., Chavali, V.P., Tickman, B.I., Carothers, J.M., and Zalatan, J.G. (2020). Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. Nat. Commun. *11*, 1618.

Gupta, A., Reizman, I.M.B., Reisch, C.R., and Prather, K.L.J. (2017). Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. Nat. Biotechnol. *35*, 273–279.

Jeschek, M., Gerngross, D., and Panke, S. (2017). Combinatorial pathway optimization for streamlined metabolic engineering. Curr. Opin. Biotechnol. *47*, 142–151.

Moon, T.S., Yoon, S.-H., Lanza, A.M., Roy-Mayhew, J.D., and Prather, K.L.J. (2009). Production of Glucaric Acid from a Synthetic Pathway in Recombinant Escherichia coli. Appl. Environ. Microbiol. *75*, 589–595.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature *496*, 528–532.

Stevens, J.T., and Carothers, J.M. (2015). Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. ACS Synth. Biol. *4*, 107–115.

Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. Proc. Natl. Acad. Sci. *111*, 11299–11304.

Xu, P., Rizzoni, E.A., Sul, S.-Y., and Stephanopoulos, G. (2017). Improving Metabolic Pathway Efficiency by Statistical Model-Based Multivariate Regulatory Metabolic Engineering. ACS

Synth. Biol. *6*, 148–158.

Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., and Alon, U. (2004). Just-in-time transcription program in metabolic pathways. Nat. Genet. *36*, 486–491.

Zhao, S., Jones, J.A., Lachance, D.M., Bhan, N., Khalidi, O., Venkataraman, S., Wang, Z., and Koffas, M.A.G. (2015). Improvement of catechin production in Escherichia coli through combinatorial metabolic engineering. Metab. Eng. *28*, 43–53.

Chapter 1:

Multi-Layer CRISPRa/i Circuits for Dynamic Genetic Programs in Cell-Free and Bacterial Systems

Below is provided our manuscript submitted to *Cell Systems* in its entirety.

Multi-Layer CRISPRa/i Circuits for Dynamic Genetic Programs in Cell-Free and Bacterial Systems

Benjamin I. Tickman[+,1], Diego Alba Burbano[+,1,2], Venkata P. Chavali[1], Cholpisit Kiattisewee[1], Jason Fontana[1], Aset Khakimzhan[3], Vincent Noireaux[3], Jesse G. Zalatan[*1,4], James M. Carothers[*,1,2]

1: Molecular Engineering & Sciences Institute
and Center for Synthetic Biology
University of Washington
Seattle, WA 98195
United States

2: Department of Chemical Engineering
University of Washington
Seattle, WA 98195
United States

3: School of Physics and Astronomy
University of Minnesota
Minneapolis, MN 55455, USA

4: Department of Chemistry

University of Washington

Seattle, WA 98195

United States

+: These authors contributed equally

*: Corresponding authors

zalatan@uw.edu

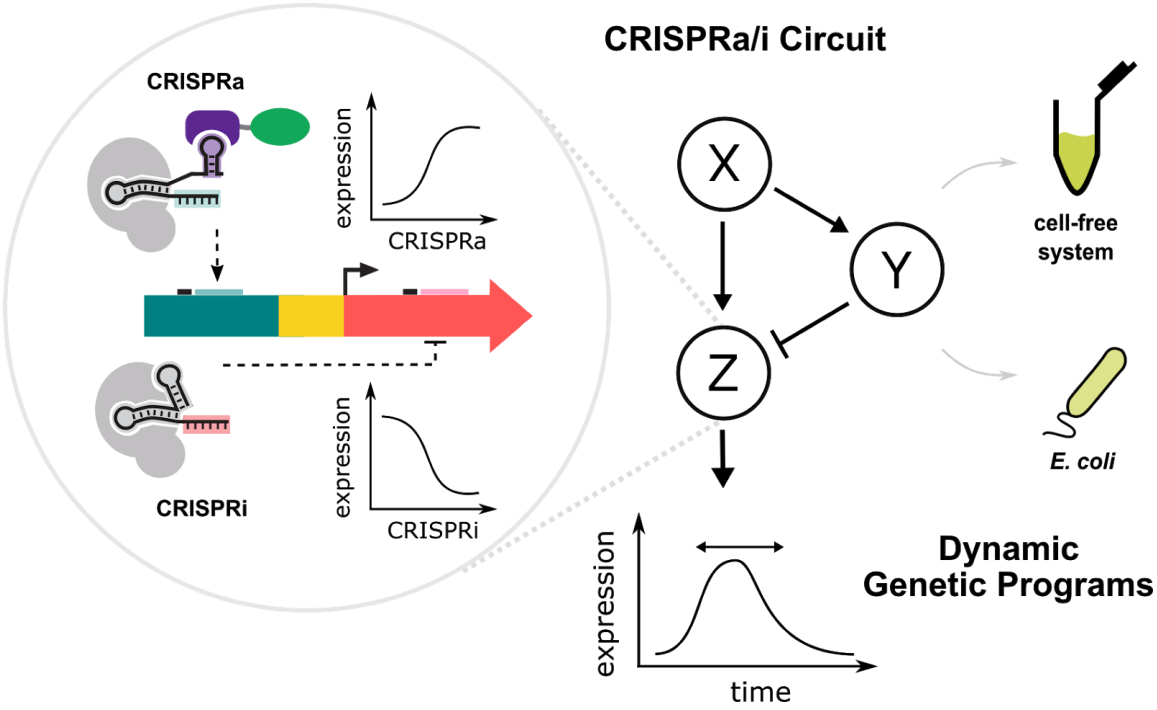
206-543-1670

jcaroth@uw.edu

206-221-4902

In preparation as an article for ***Cell Systems***

Graphical Abstract



Highlights

- Integration of CRISPRa with CRISPRi greatly expands genetic circuit design space
- Level-matching between stages in multi-layer CRISPRa/i circuitry can be accomplished through the regulated expression of guide RNAs
- Rational tuning of guide RNA expression levels programs distinct gene expression dynamics
- Multi-guide circuit functions highlight the potential for scalable circuit design

Summary

CRISPR-Cas transcriptional circuits hold great promise as platforms for engineering metabolic networks and information processing circuits. Historically, CRISPR control systems in prokaryotes have been limited to CRISPRi repression. Creating approaches to integrate CRISPRa for transcriptional activation in these systems would greatly expand CRISPR circuit design space. Here, we develop design principles for engineering prokaryotic CRISPRa/i genetic circuits with network topologies specified by guide RNAs. We demonstrate that multi-layer CRISPRa/i cascades and feed-forward loops can operate through the regulated expression of guide RNAs in cell-free expression systems and *E. coli*. We show that CRISPRa/i circuits can program complex functions by designing type 1 incoherent feedforward loops to serve as fold-change detectors and tunable pulse-generators. The correspondence between predicted effects of CRISPRa/i circuit tuning actions and measured functions suggests that these circuits could be assembled into larger and more complex networks. By investigating how component characteristics relate to network properties such as depth, width, and speed, this work establishes a framework for building scalable CRISPRa/i circuits as regulatory programs in cell-free expression systems and bacterial hosts.

Keywords

CRISPRa, CRISPRi, Transcriptional Circuits, I1-FFL, *E. coli*, Cell-Free, Expression Dynamics

1. Introduction

Inspired by nature, synthetic biologists seek to dynamically regulate gene expression in biological systems to conserve resources, respond to stimuli, and generate complex, time-dependent behavior (Brockman and Prather, 2015; Dinh and Prather, 2020; Fontana et al., 2018a; Santos-Moreno and Schaerli, 2020). However, there are limited examples of synthetic, dynamic transcriptional regulatory networks capable of complex, multi-gene regulation. This paucity can be attributed to the limited number of suitable components for implementing scalable regulatory networks (English et al., 2021; Jeong et al., 2019; Nielsen et al., 2016), and to the difficulty of sequentially combining components into multi-layered operations (Brophy and Voigt, 2014; Gander et al., 2017; Lucks et al., 2008; Qian et al., 2017). Hence, a scalable framework enabling rational design and tuning of dynamic regulatory programs would constitute a significant advance.

CRISPR-Cas transcriptional controls have emerged as a promising route for building gene regulatory networks enabling programmable and orthogonal control at many loci simultaneously (Banerjee et al., 2020; Landberg et al., 2020; Reis et al., 2019; Tian et al., 2019). In these systems, nuclease defective Cas proteins such as *S. pyogenes* dCas9 are combined with guide RNAs (gRNAs) that specify DNA targets. Targeting of this complex to promoters or open reading frames generates gene repression (CRISPRi). Scalable multi-gene circuits can thus be implemented simply through the programmed expression of multiple gRNAs (Gander et al., 2017; Huang et al., 2021; Nielsen and Voigt, 2014; Santos-Moreno et al., 2020). Recent efforts have demonstrated the construction of CRISPRi circuits capable of performing a diverse set

of Boolean logic evaluations (Gander et al., 2017; Nielsen and Voigt, 2014; Tan and Ng, 2021; Xiang et al., 2018), and dynamic expression programs (Dinh and Prather, 2019; Tian et al., 2020; Westbrook et al., 2019; Wu et al., 2020). The recent discovery of new transcriptional activators and promoter design rules for effective CRISPRa in bacteria raised the possibility of circuits combining CRISPRa and CRISPRi to form dynamic gene regulatory networks in prokaryotic systems (Dong et al., 2018; Fontana et al., 2020; Kiattisewee et al., 2021; Liu et al., 2019). Such circuits would enable network topologies and functional capabilities not possible with CRISPRi alone. However, to date there are no examples of prokaryotic CRISPRa being applied in genetic circuits beyond elementary operations in a single layer (Table S2) (Bikard et al., 2013; Dong et al., 2018; English et al., 2021; Fontana et al., 2020; Kiattisewee et al., 2021; Liu et al., 2019).

In this work, we develop genetic components and design strategies allowing CRISPRa to be combined with CRISPRi and generate a CRISPRa/i transcriptional control system operating in *E. coli* and an *E. coli*-derived cell-free expression system (CFS). We show that the strength and dynamics of control actions can be tuned through the regulated expression of guide RNAs, with CRISPR-activation ratios as high as 25 fold in CFS and 40 fold in *E. coli*. We combine components into multi-layered operations by level-matching the output expression levels of upstream components to the acceptable input range of downstream components (McDaniel and Weiss, 2005; Wang et al., 2013). We report the successful construction and tuning of multi-guide CRISPRa/i cascades and type 1 incoherent feed forward loops (I1-FFLs) in CFS and *E. coli* to programmably achieve complex behaviors such as pulse generation and fold-change

detection. Together, a set of generalizable design rules and an easily expandable toolbox of orthogonal components provide a framework for rapid and scalable implementation of higher order CRISPRa/i regulatory networks. We envision that these capabilities will prove useful for the next generation of dynamically-regulated metabolic engineering efforts, input processing for multiplexed biosensors, and biocomputation in living organisms and cell-free systems.

2. Results

2.1 CRISPRa/i Circuits in CFS

2.1.1 Bacterial CRISPRa is Functional in *E. coli* CFS

Cell-free systems have become an attractive platform for prototyping of genetic circuits, construction of synthetic cells and engineered biosynthetic pathways (Adamala et al., 2017; Dudley et al., 2015; Garamella et al., 2016; Karim et al., 2016; Marshall and Noireaux, 2018). However, there are limited examples of genetic circuits capable of dynamic, multi-gene regulation in CFS. While CRISPRi-based genetic control is well established in CFS, there have been no published reports of CRISPRa in CFS. Incorporating CRISPRa into CFS could enable more facile circuit engineering by increasing the number of realizable network topologies (Figure S1), and could overcome challenges that limit the utility of multi-layer CRISPRi repression circuitry in CFS.

A unique feature of CFS is that component turnover is greatly diminished compared to *in vivo* systems. CFS do not undergo dilution due to cell division and experience characteristically low protein and RNA turnover rates compared to cellular systems (Garamella et al., 2016). While component turnover can be accelerated via the addition of degradation machinery, this approach is inefficient and consumes valuable, finite resources (Garamella et al., 2016). In practice, this limited turnover makes repression circuits difficult to implement because even if transcription is halted the gene product is already present. Thus, our first challenge was to adapt the CRISPRa system developed in *E. coli* for use in CFS. In this system, CRISPRa is applied using a

3’-extended guide RNA (scRNA) to direct dCas9 upstream of σ^{70} promoters. The 3’ extension of the guide RNA contains an RNA hairpin (MS2) which binds an RNA binding protein (MCP) fused to a transcriptional activator (SoxS) (Figure 1A). Importantly, in this system scRNAs encode information for targeting of dCas9 to precise locations along DNA as well as recruitment of a functional effector (Dong et al., 2018; Fontana et al., 2020; Kiattisewee et al., 2021; Zalatan et al., 2015). These scRNAs, J106, J206, and J306, are targeted via the spacer sequence directing CRISPRa to an expandable set of orthogonal synthetic promoters J1, J2, and J3 (Fontana et al., 2020).

To understand the portability of the CRISPRa system between E. *coli* and CFS, we tested whether basal expression levels and gene activation in CFS corresponded to previously-observed trends in E. *coli*. In cells CRISPRa can produce high levels of gene expression from a broad range of promoter strengths, but the fold-activation decreases as basal promoter strength increases. We tested a set of synthetic minimal promoters (BBa_J231XX) (Figure S2) of varying strength in CFS. We observed a high correlation between CFS and E. *coli* for both basal promoter strength and fold-activation by CRISPRa, providing Spearman correlation coefficients of 0.91 and 0.88, respectively (Figure S3). This observed correspondence between component function in *E. coli*-derived CFS and *E. coli* is consistent with previous reports (Garamella et al., 2016; Marshall et al., 2018; Shin and Noireaux, 2012), allowing exchange of individual genetic components between the two systems.

Next, we sought to formalize a framework for the construction of higher order CRISPRa/i circuits operating through interconnected CRISPRa/i nodes (Figure 1B). In this framework, CRISPRa/i nodes are discrete transcriptional units containing target

sequences for CRISPRa- and/or CRISPRi-directed transcriptional regulation (Figure 1A). To characterize CRISPRa/i nodes, we isolated dCas9, sc/sgRNAs, and the MCP-SoxS activator onto individual plasmids (Figure 1C), allowing independent titration of expression levels for all CRISPRa/i components. CRISPRa/i node characterization is conducted by measuring the output response of each node to varying levels of component transcriptional inputs provided by titrating component plasmid concentrations. We found that increasing the concentration of sc/sgRNA expressing plasmid resulted in higher overall levels of activation and repression (Figure 1D, 1E), as well as faster control (Figure S4). Titrations of dCas9 expressing plasmid revealed no significant differences in the strength of CRISPRa across a 40-fold range of dCas9 expression levels generated by 0.05 nM to 2 nM dCas9 expressing plasmid (Figure S5, left). For all levels of dCas9 expression, we observed a ~40 min delay between initiation of the cell-free reaction and the onset of CRISPRa/i control, consistent with previously reported time of dCas9 maturation and CRISPR complex formation in CFS (Marshall et al., 2018; Westbrook et al., 2019). Titrations of plasmid expressing MCP-SoxS revealed a relatively wide region between 1 nM and 24 nM over which no significant differences in endpoint measurements of CRISPRa mediated outputs were observed (Figure S5, right). Expression levels for dCas9 and MCP-SoxS activator were held constant throughout the work at 2nM and 4nM plasmid concentrations respectively.

2.1.2 Tuning CRISPRa/i through the Regulated Expression of Guide RNAs in CFS

The ability to easily vary plasmid concentration in CFS, combined with the multi-component nature of CRISPRa/i regulatory complexes (Figure 1C), enables tuning of all component expression levels independently. Some tuning actions are global, for instance varying dCas9 expression levels impacts both CRISPRa and CRISPRi. Other tuning actions, such as varying the level of activator protein, are expected to influence activation but not repression. Likewise, output levels for individual nodes in a circuit can be linearly scaled by changing the concentration of plasmid at that node (Figure 1B). To provide simultaneous and independent control over both timing and expression levels of multiple target genes, we tuned CRISPRa/i control actions through the regulated expression of guideRNAs. Here, the specificity provided by guideRNA targeting allows tuning actions to be applied locally to individual CRISPRa/i nodes.

To quantify time-varying CRISPRa/i-directed changes in gene expression, we calculate production rates of CRISPR-regulated RFP expression relative to unregulated, basal expression of RFP. (relative RFP production rate). At saturating levels of scRNA expression, CRISPRa achieved constant levels of activation over the course of 4-6 hrs providing a ~20-fold ± 2 fold increase in RFP production rate relative to an off-target control (Figure 1D, right). Likewise, relative production rates from CRISPRi with saturating levels of sgRNA (Figure 1E, right) achieved steady state levels of repression by 3 hours. Guide RNA titrations revealed that increasing levels of scRNA decreased the time to 2-fold activation by CRISPRa by up to ~5 hours, and increasing levels of

sgRNA decreased the time to 50% repression by CRISPRi by up to ~10 hours (Figure S4, left). Surprisingly, only sgRNA titrations were able to significantly affect the overall timing of gene expression as determined by the time to half maximum endpoint RFP values, with strong CRISPRi providing a ~3 hour shift to earlier time points as compared to a no sgRNA control (Figure S4, right). Qualitatively, we observed that increasing sgRNA expression levels resulted in a higher fraction of total expression occurring at early time points. In contrast, scRNA titrations primarily provided a scaling factor to CRISPRa output levels without greatly affecting the timing of expression (Figure 1D, 1E, right; Figure S4, right). These data suggest that under these conditions CRISPRa kinetics may be dominated by the time required for MCP-SoxS expression and maturation.

2.1.3 Level-Matching of Multi-Layer CRISPR Circuitry in CFS

To enable the construction of multi-layer circuits, we built activation and activation-repression cascades by level-matching the input/output dynamic ranges between sequential CRISPRa/i nodes. That is, we matched the output transcription levels of an upstream node encoding scRNA to the relevant transcriptional input range of a downstream node encoding another sc/sgRNA. From sc/sgRNA plasmid titrations, we observed that both CRISPRa and CRISPRi respond to changes in input guide RNA expression levels spanning approximately 2 orders of magnitude (Figure 2A, 2C). Across this responsive range of sc/sgRNA inputs, CRISPRa controlled outputs vary by ~24-fold at endpoint. While the dynamic range of CRISPRa generated outputs does not fully span the dynamic range of sc/sgRNA inputs, this characterization suggests that

CRISPRa/i nodes can be sequentially combined by careful matching of upstream output ranges to downstream input ranges to form layered operations.

By tuning the concentration of plasmid expressing sc/sgRNA in the second layer of the cascades we were able to control the degree of overlap between response curves of upstream and downstream layers in the circuit. Based on the scRNA dose-response curves for CRISPRa by two different scRNAs in isolation (Figure 2A), we decided to build CRISPR activation-activation cascades to probe the composability of CRISPRa circuits from components characterized in isolation. Using these curves, we predicted how a 24-fold increase in transcription provided by CRISPRa in the first layer of the cascade affects cascade output. We made this prediction for four different concentrations of plasmid expressing scRNA in the second layer (Figure 2B). Upon construction of these CRISPRa cascades, we observed excellent agreement (R^2 = 0.985) between measured and predicted fold increases in outputs of CRISPRa cascades (Figure 2B). As expected, overlap between layers was maximized at 2 nM scRNA plasmid in the second layer, with the cascade providing a 16.3 ± 3.0 fold increase in measured RFP at endpoint compared to CRISPRa alone. Both lower, 0.5 nM, and higher, 4 nM, concentrations of scRNA expressing plasmid in the second layer of the cascade resulted in decreased fold changes in cascade output, at 7.6 ± .6 fold and 15.2 ± 2.3 fold respectively. From these data, we calculated the efficiency of signal propagation through the activation cascade by comparing the observed fold change in cascade output to the fold change provided by CRISPRa in the input layer. At present, given optimal level-matching we observe 67.9% ± 18.1% signal propagation for the two layer CRISPRa cascade. Together, these results suggest that we can predictably tune

the degree of overlap between layers of CRISPRa/i circuits to propagate signals, satisfy input requirements of potential downstream layers, and tailor absolute gene expression levels.

Next, we constructed an activation-repression cascade with CRISPRa in the input layer activating transcription of RR2 sgRNA which targets the coding sequence of *mRFP1* for CRISPRi in the second layer of the cascade (Figure 2D, top). As in construction of the activation-activation cascade, level-matching was informed by the sc/sgRNA dose-response curves for CRISPRa and CRISPRi obtained in isolation (Figure 2C). As expected, when the overlap between layers was maximized, the CRISPRa/i cascade generated 4.6-fold ± 0.7 more repression than CRISPRi alone (Figure 2D, right). The importance of matching upstream outputs to the responsive range of downstream inputs was illustrated by overexpression of sgRNA in the second layer of the activation-repression cascade, intentionally minimizing the overlap between the upstream and downstream layers of the cascade. Under these conditions, the activation-repression cascade reduced RFP expression by 1.4-fold ± 0.3 compared to CRISPRi (Figure 2D, left).

2.1.4 CRISPRa/i Circuits Encode Dynamic Gene Expression Programs

Next, we sought to investigate the ability of multi-layer CRISPRa/i circuitry to encode dynamic gene expression programs inaccessible to simpler single layer controllers. As a first step, we explored the influence of level-matching on CRISPRa/i cascade dynamics by comparing relative RFP production rates arising from an activation-repression cascade to those generated by CRISPRi. When there is significant

overlap between the response curves of the layers in the cascade, large changes in the timing of gene expression are observed (Figure 2E, right; 2F). For instance, an activation-repression cascade with 0.5 nM of sgRNA plasmid results in repression of the RFP output at a comparable rate, but delayed onset compared to that of CRISPRi repression alone with 4 nM of sgRNA plasmid. This delay is interpreted as the time required for CRISPRa to activate sgRNA expression in the second layer of the circuit. We identify a ~10-fold range of sgRNA plasmid concentrations over which a CRISPR activation-repression cascade can generate significant differences in expression dynamics compared to single-layer CRISPRi (Figure 2F).

As expected, when sgRNA expression levels are mismatched we observe negligible differences in expression dynamics. If the concentration of plasmid expressing sgRNA in the second layer of the activation-repression cascade is too low, e.g. 0.01 nM, no repression is observed. Conversely, high concentrations of sgRNA expressing plasmid in the second layer of the cascade effectively result in CRISPRi applied in a single layer, producing expression dynamics identical to that of the CRISPRi control (Figure 2E, left). Above 2 nM of sgRNA expressing plasmid, we observed no difference in the time to 50% repression for the CRISPRa/i cascade as compared to single-layer CRISPRi (Figure 2F). Together, these results underscore that gene expression dynamics can be tuned by multi-layer CRISPRa/i circuits when there is sufficient overlap between the response curves of the CRISPRa/i circuit components.

After establishing the rules governing construction of layered CRISPRa/i circuitry, we endeavored to create more complex transcriptional programs to explore the scalability and composibility of CRISPRa/i regulatory networks. We combined the

CRISPR activation-repression cascade with CRISPRa to form an incoherent type 1 feed forward loop (I1-FFL), a classic pulse generating circuit that is significantly overrepresented in natural systems (Alon, 2007; Kaplan et al., 2008; Mangan and Alon, 2003; Shen-Orr et al., 2002). When level-matching is taken into consideration, we see that circuit topology determines the timing of gene expression (Figure 3A). As expected, we observe no difference in expression dynamics between CRISPRa and CRISPRa+CRISPRi (CRISPRa+i) at low concentrations of sgRNA expressing plasmid (Figure 3A, blue, orange). When expression of sgRNA at node Y is activated by CRISPRa to form an I1-FFL (Figure 3A, red), we observe a gene expression pulse, qualitatively different from expression generated by CRISPRa or CRISPRa+i. Upon addition of an orthogonal I1-FFL controlling expression of GFP to the same reaction (Figure 3A, green), no differences in the timing of gene expression are observed at the output of the RFP I1-FFL. This result indicates that we can operate multiple circuits simultaneously without compromising the respective expression dynamics.

We were able to tune the timing of the gene expression pulse generated by the I1-FFL by varying the concentration of sgRNA expressing plasmid. The maximum RFP production rate occurred ~110 min earlier in the cell-free reaction when we increased the sgRNA expressing plasmid concentration 10-fold from 0.1 nM to 1 nM (Figure 3B). More generally, we observed that the time of the maximum gene expression pulse could be continuously tuned over a 4-fold change in sgRNA plasmid levels, shifting expression maxima earlier by up to 2 hours compared to unregulated expression (Figure 3C). To capture the I1-FFL expression dynamics and evaluate the feasibility of rationally tuning CRISPRa/i circuits *in silico,* we constructed a coarse-grained

mechanistic model of CRISPRa/i gene regulation. We defined first order chemical reactions for protein and guide RNA production, CRISPR complex assembly, and DNA targeting (Figure 3B, bottom; Table S3). When an initial experimental observation was provided, the model was capable of predicting the effects of tuning actions applied to the I1-FFL on gene expression dynamics (Figure 3B, fit and simulate). Here, the model is fit to the experimental data for an I1-FFL with 0.1 nM sgRNA expressing plasmid and used to predict the expression dynamics for an I1-FFL with 1 nM sgRNA expressing plasmid. Similar results were obtained when fitting to the 1 nM condition and predicting the 0.1 nM condition (Figure S8). We observed a ~10 min difference between measured and predicted timing of maximum RFP production rate. Notably, this difference corresponds to the time resolution of our measurements. Combined with the observed predictability of level-matching in CRISPRa/i cascades, these results suggest that high-fidelity CRISPRa/i circuits could be designed and tuned *in silico* given component characterization data.

2.2 CRISPRa/i Circuits in *E. coli*

While both CRISPRa and CRISPRi have been previously demonstrated in *E.coli*, to the best of our knowledge, there are no reports combining the two to form dynamic, multi-layer circuitry. In *E. coli,* CRISPRa/i circuits are encoded on two plasmids. One plasmid contains dCas9, MCP-SoxS, and scRNAs acting as inputs to the first layer of a circuit, while the second plasmid contains a fluorescent reporter as well as sc/sgRNAs acting in the second layer of a circuit (Figure 4A, 5A). Unlike cell-free systems, in which the expression level of every circuit component can be precisely titrated, gene

expression in cells is constrained by the expression levels achievable given different combinations of plasmid copy number and genetic parts. Level-matching of multi-layer CRISPRa/i circuits in cells is therefore more challenging, and requires attention not only to the dynamic range of components, but also the absolute expression levels and activities.

2.2.1 Level-Matching in Multi-Layer CRISPRa Circuits

To understand the level-matching requirements of scRNAs in multi-layer CRISPRa cascades in *E. coli*, we engineered the basal expression characteristics of CRISPRa nodes in the second layer of a 2 layer activation cascade. In this circuit, scRNA expressed at node X targets CRISPRa to a promoter at node Y, activating expression of a second scRNA, targeting a fluorescent reporter at node Z for CRISPRa. Tuning of basal expression levels was accomplished through the use of synthetic minimal promoters (BBa_J231XX), as well as modifications to the 5' sequence proximal to the minimal promoter at node Y, driving scRNA expression in the second layer (Figure 4B). At the highest basal levels of scRNA expression in the second layer, CRISPRa cascades yielded 1.4x higher output levels than a comparable single-layer circuit. Decreasing basal scRNA expression levels in the second layer of the cascade by ~10x increased the output dynamic range of the CRISPRa cascade to 5.9x. Decreasing basal scRNA expression levels by a further 3.4x increased the output dynamic range by an additional 2x, resulting in an overall activation ratio of 12.3x for the CRISPRa cascade as compared to the single-layer circuit. Output levels of CRISPRa cascades at all tested scRNA expression levels were comparable to output of

single-layer CRISPRa with saturating levels of scRNA expression. We observed that CRISPRa cascades were highly sensitive to scRNA expression, with 32% compression of the output dynamic range observed even at the lowest basal expression level of scRNA at node Y. Compression of the output dynamic range in cascades can be attributed to basal scRNA expression in the second layer of the circuit. These results suggest that engineered genetic parts capable of lower basal scRNA expression levels would minimize compression of activation cascade dynamic ranges.

2.2.2 Inducible CRISPRa by expressing MCP-SoxS from an inducible promoter

To provide an input for dynamic CRISPRa/i circuitry in *E.coli*, we chose to apply control over CRISPRa through inducible expression of the MCP-SoxS activator protein (Figure 5B). We observed that output levels generated by CRISPRa were titratable through aTc induction of MCP-SoxS activator (Figure 5B, right) and that these output levels were similar to CRISPRa employing constitutive expression of MCP-SoxS (Figure 4B, left). Compared to CRISPRa with constitutively expressed activator and off/on target scRNA, the aTc-inducible system provided 40% lower basal levels and 16.5% lower activated levels of reporter expression (Figure 5B). A 40.4 ± 0.77 fold increase in expression was observed for the constitutive CRISPRa system supplied with on versus off-target scRNAs, whereas aTc induction of MCP-SoxS with on-target scRNA yielded a 56.3 ± 0.65 -fold increase. Together, these results establish aTc-inducible expression of the MCP-SoxS activator as a viable means of generating titratable levels of activation.

2.2.3 Level-Matching of Multi-Layer CRISPRa/i Circuits in *E. coli*

To understand the level-matching requirements of sgRNAs in multi-layer CRISPRa/i circuits in *E. coli*, we constructed an inducible CRISPR activation-repression cascade. Here, we titrated the CRISPRa input in the first layer and tuned both the expression characteristics of the promoters and the activities of sgRNAs in the second layer of the cascade. Tuning of CRISPRa inputs in the first layer of the cascade was provided by the previously described inducible MCP-SoxS activator system. In the second layer of the cascade, expression characteristics of promoters were tuned via modifications to the 5' sequence proximal to the minimal promoter, while sgRNA activities were modified through the use of 5' spacer truncations (Fontana et al., 2018b; Qi et al., 2013) (Figure 5C). When the J2 promoter was used to express RR2 sgRNA targeting RFP (Fontana et al., 2018b), we observed 70% repression in the absence of activation (Figure 5C, left). A ~20-fold increase in sgRNA expression provided by CRISPRa (Figure S9) resulted in an output dynamic range of 4-fold, spanning 23% of accessible expression levels. Decreasing the strength of CRISPRi via truncation of the RR2 sgRNA spacer to 14 nucleotides decreased repression in the absence of activation of sgRNA expression to 20% as compared to an off-target control. However, truncated guide RNAs were not able to achieve high levels of repression at maximal levels of activation (Figure 5C, center) resulting in a compressed output dynamic range of 2.5-fold, spanning 48% of accessible expression levels. Tuning of sgRNA expression levels via modifications to the 5' sequence proximal to the minimal promoter resulted in 17% and 92% repression in the absence and presence of CRISPRa applied to RR2 sgRNA respectively, yielding an output dynamic range of ~10-fold, spanning 76% of the

accessible expression space (Figure 5C, right). Taken together, the inducible CRISPRa/i cascade and the CRISPRa cascade paint a cohesive picture indicating that guide RNA expression levels produced by CRISPRa are more than sufficient to saturate downstream layers of CRISPRa/i circuits, and that both CRISPRa and CRISPRi are highly sensitive to basal expression of sc/sgRNAs.

2.2.4 Tunability of CRISPRa/i Enables Interrogation of Complex Behavior in *E. coli*

The ability to tune CRISPRa/i circuits through both promoter engineering and guide RNA truncations allows control over the abundance and strength of individual sc/sgRNAs. Such control enables independent tuning of nodes in multi-layer CRISPRa/i circuitry. Paired with the ease of circuit construction, the CRISPRa/i system is suited for rapid circuit function interrogation. To showcase multi-guide tuning and circuit function exploration, we constructed and characterized three different I1-FFLs in which application of repression by node Y is varied.

To enable observation of dynamic circuit behaviors and provide titratable levels of input activation, we chose to use inducible expression of the activator protein MCP-SoxS (Figure 5A, top). To understand the effect of tuning actions on I1-FFL output, we constructed three different network topologies (Figure 5A), an I1-FFL, CRISPRa+i, and CRISPRa with an off-target sgRNA. We compared the response of these three circuits to increasing levels of MCP-SoxS induction for three different tunings of sgRNA expression (Figure 5A, bottom). To determine how much of the repression in the I1-FFL could be attributed to the basal expression from node Y, we compared the CRISPRa+i

circuit to CRISPRa with off-target sgRNA (Figure S10). We observed 7500 RFU from CRISPRa+i with off-target sgRNA. CRISPRa+i expressing full-length on-target sgRNA from the J2 promoter provided 2200 and 71 RFU at 200 nM and 0 nM aTc, respectively. Thus, basal expression of the sgRNA from node Y has the effect of reducing, or compressing, the output range of the MCP-SoxS CRISPRa titration by 70%. In an I1-FFL with the same sgRNA expression tuning, the output dynamic range was only 3-fold (Figure S10). We reasoned that decreasing the strength of repression at node would reduce compression in the CRISPRa+i circuit and increase the dynamic range of the I1-FFL. Tuning repression through sgRNA truncation to 14nt decreased compression of the output range by the CRISPRa+i circuit from 70% to 24%, and increased the output dynamic range of the I1-FFL from 3 fold to 15 fold (Figure S10). Tuning repression through modifications to the sequence 5' of the minimal promoter expressing sgRNA resulted in only 4% compression of the output range by CRISPRa+i. However, counterintuitively the I1-FFL output dynamic range was only 2 fold with the same modifications to the sequence 5' of the minimal promoter expressing sgRNA (Figure S10). While the latter result was unexpected, the current suite of tuning actions available to the CRISPRa/i system in *E. coli* nonetheless does allow independent tuning of interactions between nodes in multi-guide circuits.

I1-FFLs are used in many naturally-occurring sensory systems as fold-change detectors to generate dynamic outputs determined by relative, as compared to absolute, differences in inputs to the system (Adler and Alon, 2018). Formally, fold-change detection (FCD) can be defined as a logarithmic relationship between inputs "I" and outputs "O", i.e., an input/output response curve, satisfying the equation

$O = a \cdot ln(I) + b$.Theoretical work has shown that a transcriptional I1-FFL is capable of fold-change detection only under specific ratios of component expression levels and strengths (Goentoro et al., 2009). Experimentally, we can test for fold-change detection in these circuits by evaluating the variance explained by a logarithmic fit to outputs taken as a function of aTc-induced MCP-SoxS inputs. Consistent with expectation, we observed that only I1-FFLs with specific sgRNA tunings were capable of detecting fold changes of aTc over the linear range of MCP-SoxS induction (Figure 5B, right). Over this linear range, we observed an R^2 of 0.975 for a logarithmic fit between the inputs and outputs, as compared to an R^2 of 0.853 for a linear fit to the data for the I1-FFL with 20nt RR2 sgRNA expressed from the J2 promoter (Figure 6A, bottom right). We can extend the test for fold-change detection beyond the linear range of aTc-induced MCP-SoxS inputs by converting these inputs into the corresponding CRISPRa responses. We linearized the CRISPRa response to aTc induction by dividing CRISPRa output levels at a given aTc induction level by the CRISPRa output at saturating concentrations of aTc. Plotting I1-FFL outputs against this percent induction of the CRISPRa response provided an R^2 of 0.989 for a logarithmic fit, as opposed to only 0.896 for a linear fit. By comparison, the corresponding CRISPRa+i circuit exhibited an R^2 for the input-output relationship of only 0.852 for a logarithmic fit, but 0.997 for a linear fit (Figure S11). Taken together, these data show that CRISPRa/i circuits assembled into I1-FFLs can be tuned to achieve fold-change detection.

To investigate the composability of CRISPRa/i controlled gene expression dynamics, we tested three inducible circuits under continuous dilution: CRISPRa, an activation-repression cascade, and an I1-FFL tuned for fold-change detection (Figure

6B, left). We observed increases in RFP/OD_{600} roughly 1 hr post induction for both the I1-FFL and CRISPRa corresponding to the action of the first layer in both circuits (Figure 6B, right). For both the I1-FFL and the activation repression cascade, repression onset was observed at ~5 hours, corresponding to the action of the second layer of each circuit. While I1-FFLs are recognized as a classic pulse generating circuit, to achieve fold-change detection gene expression pulses must display perfect adaptation (Adler and Alon, 2018; Goentoro et al., 2009), meaning gene expression must return to the basal level after completing the pulse. Here we observed an adaptive pulse of gene expression from the I1-FFL that starts at the baseline and ends at the baseline, further corroborating the ability of the CRISPRa/i I1-IFFL to function as a fold-change detector. Overall, these results indicate that CRISPRa/i circuits are composable in that the observed dynamic behaviors can be understood from the functions of the parts.

3. Discussion

We have developed a set of components and a unifying framework for building dynamic CRISPRa/i gene regulatory networks that are scalable, composable, and tunable. These networks are built from CRISPRa/i nodes, which we define as transcriptional units that can be targeted for regulation by both CRISPRa and CRISPRi. The CRISPRa/i framework leverages an expandable set of synthetic promoters and orthogonal guide RNAs to specify arbitrary transcriptional regulatory topologies. The characteristics and limitations of the network are therefore determined by the properties of the constituent components. Particularly, promoter dynamic range and guide RNA function specify the transcriptional input-output relationship of each node. Understanding these relationships will be fundamental for building deep, wide, and fast regulatory networks.

We can estimate the upper bound for the maximum depth of activation cascades based on the observed cascade performance in this work. At present we observe 68% ± 18% signal propagation in a two layer activation cascade (Figure 2B). The total fraction of signal propagated in deeper cascades can be calculated by raising the fraction of signal propagated between two layers to the total number of internal layers in the cascade. This calculation indicates the current CRISPRa/i system in CFS can support cascades up to 6 layers deep before output activation ratios fall below 2.5 fold. Similarly, we achieve 68% ± 2.7% signal propagation in *E. coli* (Figure 4B), suggesting activation cascades of up to 6 layers could be built with the current implementation of the CRISPRa system *in-vivo*. With the components presented In this work, transcriptional

activation generated by CRISPRa does not fully span the input dynamic range of sc/sgRNA expression in downstream layers (Figure 2), resulting in degradation of signals as they are propagated through multi-layer CRISPRa/i circuits. The most general solution to increase fidelity of signal propagation in both CFS and *E. coli* is through engineering improved system components. Promoters with lower basal expression, leading to larger output dynamic ranges, would span a higher fraction of the input dynamic range of downstream nodes, resulting in less signal degradation between layers and deeper CRISPRa/i circuitry. We estimate that engineered promoters with a mere 5-fold increase in output dynamic range would allow CRISPRa-directed outputs to fully span the input dynamic range of sc/sgRNAs in downstream layers. In this system, modest improvements in signal propagation efficiency between layers would enable drastically deeper CRISPRa/i networks. For example, increasing the fraction of signal propagated between layers by 12%, and output dynamic ranges by two fold would, in theory, enable cascades up to 14 layers deep before output activation ratios fall below 2.5 fold. Such a network would be twice as deep as the largest engineered CRISPRi transcriptional cascades to date (Gander et al., 2017). While deeply layered cascades remain beyond the scope of current engineered regulatory networks, the large activation ratios and high fidelity of signal propagation observed in the CRISPRa/i system contribute to the robust operation of shallower networks.

Natural systems coordinate the expression of many outputs with few internal layers of computation using wide, highly-interconnected networks such as dense overlapping regulons (Rosenfeld and Alon, 2003; Shen-Orr et al., 2002; Thieffry et al., 1998). Our results indicate the CRISPRa/i system is well suited to design of wide control

circuits for simultaneous and independent multi-gene regulation. In CFS, CRISPRa levels and kinetics are unchanged with respect to scRNA expression across at least an order of magnitude (Figure S6). Additionally, RFP expression levels are unchanged over a ~40-fold range of dCas9 plasmid concentration (Figure S5). This indicates that the cell-free reaction has the resources to express high levels of scRNA and dCas9 without hindering system performance. Construction of two orthogonal I1-FFLs in the same reaction showcases the ability of CFS to harbour large circuits expressing many different sc/sgRNAs to execute multiple independent programs simultaneously. Likewise, *E. coli* are capable of expressing high levels of sc/sgRNA without experiencing growth defects or retroactivity due to guide RNA competition for dCas9 (Huang et al., 2021). We observe that modest overexpression of off-target scRNA has minimal effects on CRISPRa levels (Figure S13), consistent with recent modeling work suggesting favorable scaling for CRISPRa networks as compared to CRISPRi (Clamons and Murray, 2019). Taken together, these results suggest that the CRISPRa/i system could support the operation of programs containing up to 20 independent sc/sgRNAs with minimal impact to system performance in both CFS and *E. coli*.

In both natural and engineered systems RNA-based regulatory approaches provide a means for fast and metabolically-efficient control of gene expression (Bobrovskyy and Vanderpool, 2013; Chappell et al., 2017; Stevens and Carothers, 2015; Takahashi et al., 2015; Westbrook et al., 2019). Our analysis of relative production rates shows that guide RNA mediated information propagation through internal layers of CRISPRa/i circuits can be fast (~30 mins/layer) (Figure S12) compared to the time required for initial expression of functional CRISPRa components.

Under saturating expression of dCas9, sgRNA titrations reveal an initial 40-50 min delay to the onset of CRISPRi, which could be attributed to maturation of dCas9 and formation of active CRISPR complexes (Figure 1E, right). Likewise, at high expression levels of dCas9, we observe small differences in the timing of gene expression across a wide range of scRNA expression levels (Figure 1D; Figure S5, right). In CFS, CRISPRa controlled production rates reach steady state with respect to constitutive expression over the course of several hours (Figure 1D, right). In contrast, CRISPR activation generated by cascades can experience minimal additional delays compared to CRISPRa in a single layer (Figure S12). For many applications in biocomputing and metabolic control, successful operation is determined not only by the fidelity but also the speed at which information is propagated through the regulatory network. CFS containing pre-expressed dCas9 and activator protein could accelerate the onset of CRISPRa/i regulation in these systems.

Given the previously demonstrated orthogonality of CRISPR-based regulation and the independence of CRISPRa/i nodes observed in this work, we find the CRISPRa/i system to be readily composable into larger motifs. CRISPRa/i circuits can be built by level-matching the response curves of different nodes, without complications arising from retroactivity and crosstalk. Indeed, upon construction of two orthogonal I1-FFLs in the same CFS reaction (Figure 3A), expression dynamics of the first I1-FFL were independent from both the presence and action of the second I1-FFL, and nearly identical to expression dynamics observed in isolation. Such orthogonality enables design of circuits with deterministic functions based solely on proper implementation of network topologies. In CFS, we built circuits capable of generating distinct gene

expression profiles determined by the specific network topologies (Figure 3A). We showed that expression dynamics are tunable through component expression levels within a given circuit topology when there is overlap between upstream and downstream circuit layers (Figure 2E, 3B). In *E. coli*, dilution experiments revealed I1-FFL gene expression dynamics to be an almost perfect superposition of the dynamics of CRISPRa and an activation-repression cascade (Figure 6B). This composability is also captured by the relatively simple CFS CRISPRa/i model, which only specifies CRISPR complex assembly and node targeting reactions for each sc/sgRNA (Table S3). Paired with advancements in high-throughput component characterization in CFS and state-of-the-art modeling frameworks (Lehr et al., 2019; Moore et al., 2018; Poole et al., 2020), the CRISPRa/i system presents a route towards scalable computer-aided design and implementation of dynamic gene regulatory networks in CFS.

Overall, this work establishes a paradigm in which CRISPRa/i system components can be easily combined to form scalable, dynamic gene regulatory networks in CFS and *E. coli*. The CRISPRa/i system has proven capable of building layered operations and simultaneously executing multiple regulatory programs without compromising guideRNA-encoded expression dynamics. The dynamic gene expression profiles arising from CRISPRa/i regulation are composable, in that network expression dynamics can be understood as the aggregate of the constituent components in both CFS and *E. coli* (Figure 3B, Figure 6B). These attributes allow rational design of CRISPRa/i circuits to tailor expression dynamics of multiple genes independently and simultaneously. We anticipate broad-ranging applications in engineered bacterial hosts, cell-free systems, and the next generation of artificial cells. Specifically, we foresee

applications in metabolic engineering, with feedforward motifs providing time-ordered enzyme expression, and tunable delays enabling phenotype switching in multi-phase reactions. For biosensing applications, the scalable and versatile nature of the CRISPRa/i system will allow combinatorial logical responses and complex input-output relationships to be specified, increasing the ease of connecting sensing to reporting stages. Finally, as synthetic biology efforts transition from repurposing natural systems to the bottom-up construction of fully-artificial cells, the CRISPRa/i system could be a foundational technology capable of implementing the complex dynamic control of gene expression observed in nature, while remaining compact, robust, and engineerable.

Acknowledgments

We thank members of the Carothers and Zalatan groups for advice, materials, and comments on the manuscript. This work was supported by US National Science Foundation (NSF) Award CBET 1844152 (to J.M.C. and V.N.), NSF Award MCB 1817623 (to J.G.Z. and J.M.C.) and NSF Award EF-1935087 (to J.M.C.).

Author Contributions

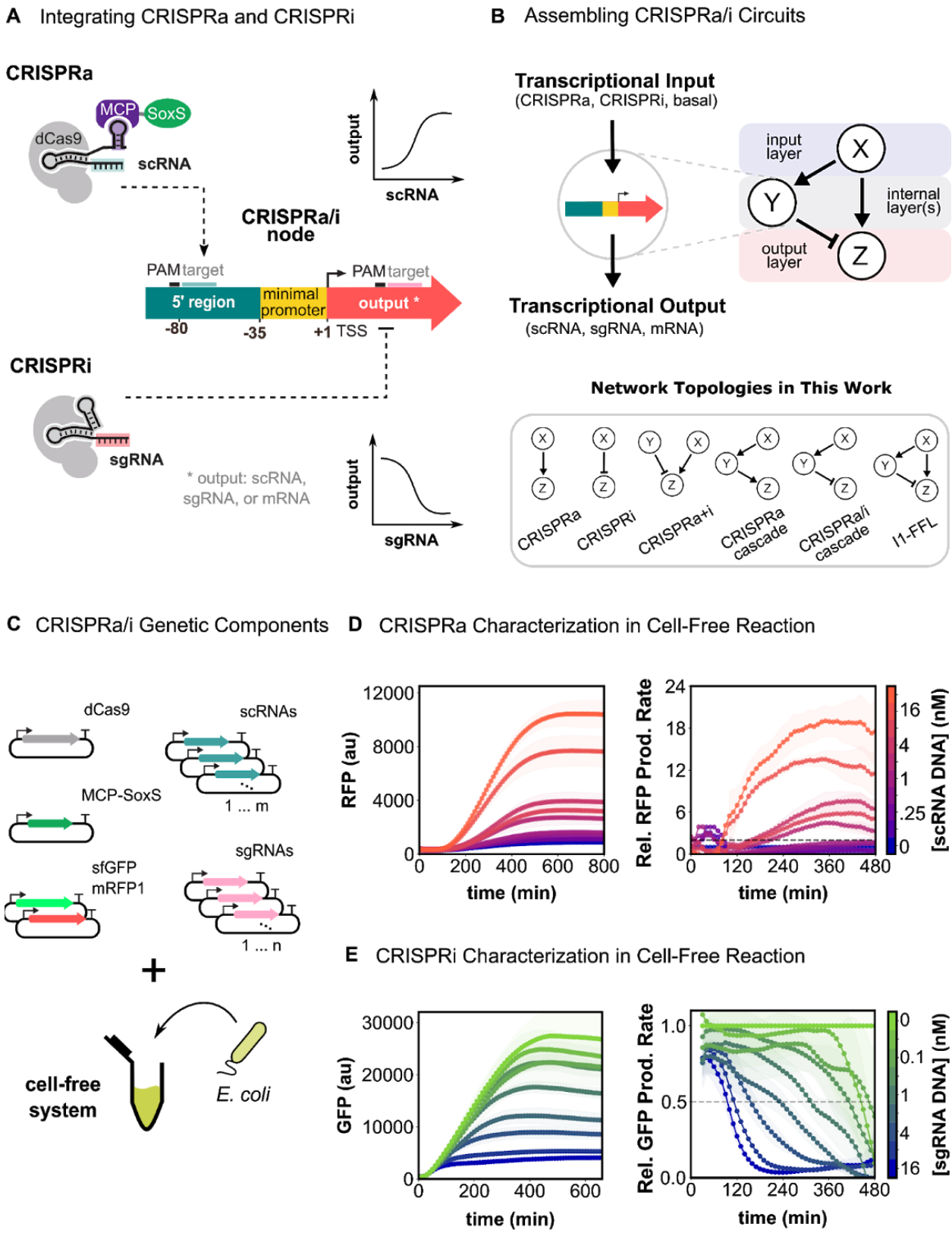
B.I.T., D.A.B., V.P.C. and J.M.C. designed the research, B.I.T., D.A.B., V.P.C., and A.K. performed experiments, B.I.T., C.K., and J.F. engineered components for CRISPRa, B.I.T., D.A.B., V.P.C., C.K. and A.K. analyzed the data. B.I.T., D.A.B., J.G.Z., and J.M.C. wrote the manuscript with input from all of the authors.

Declaration of Interests

The Noireaux laboratory receives research funds from Arbor Biosciences, a distributor of the myTXTL cell-free protein synthesis kit.

Figures, Titles and Legends

Figure 1: Combining CRISPRa with CRISPRi expands CRISPR circuit design space



A. Schematic of CRISPRa/i nodes. Modified guide RNAs (scaffold RNAs, or scRNAs) include a 3' MS2 hairpin to recruit a transcriptional activator fused to the MS2 coat protein (MCP). The expression of scRNAs targeted to an appropriate site upstream of the promoter results in CRISPR transcriptional activation (CRISPRa) from the node. The expression of a small guide RNAs targeted to the coding sequence (CDS) results in CRISPR transcriptional repression (CRISPRi) from the node.

B. Schematic of CRISPRa/i circuits. CRISPRa/i nodes can be combined to form multi-layer CRISPRa/i networks when the guide RNA output generated by one node directs CRISPRa or CRISPRi at one or more other nodes in the network.

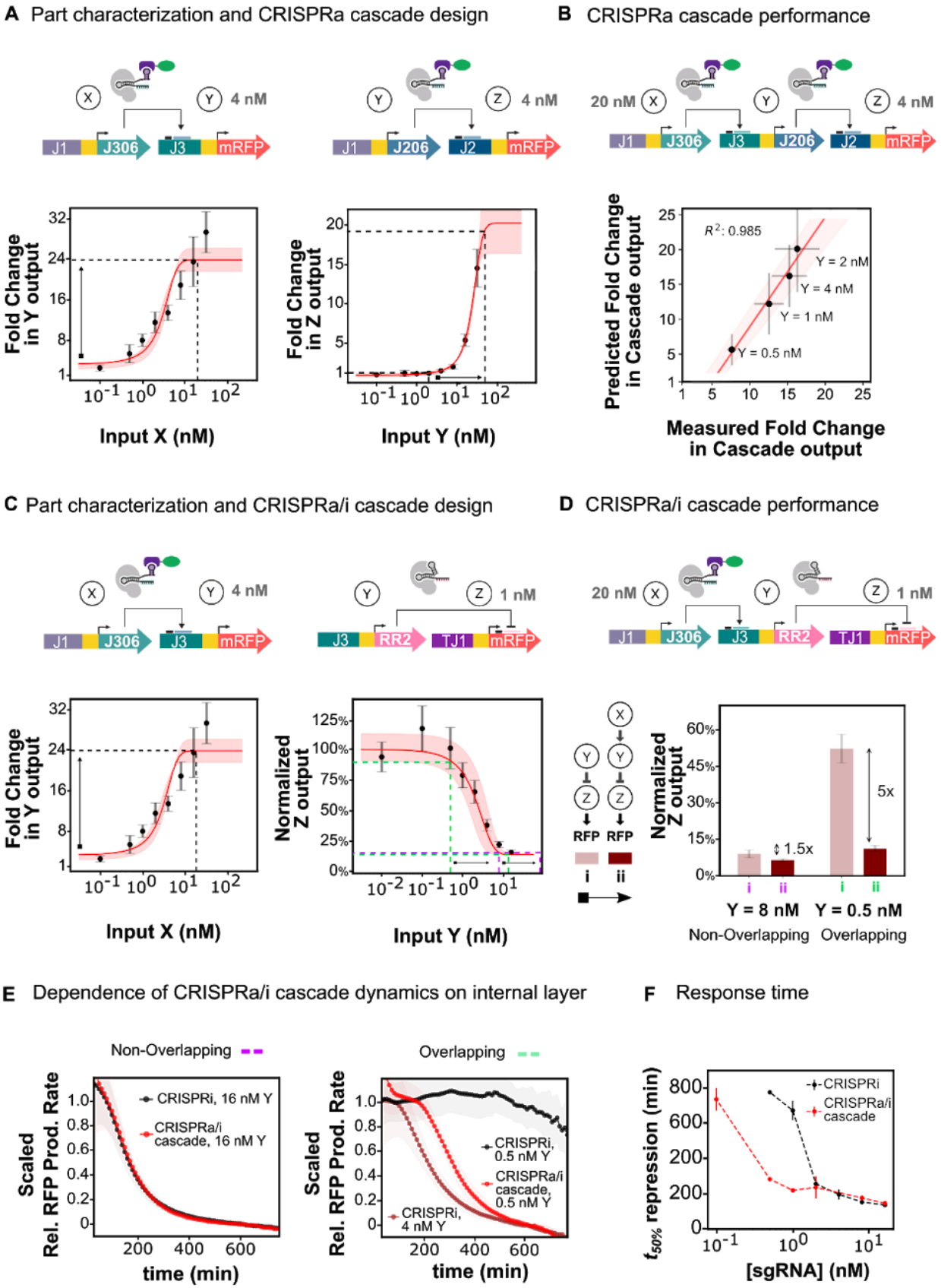
C. Schematic presentation of CRISPRa/i genetic components for use in a Cell-Free System (CFS). CRISPRa/i system components are encoded on individual plasmids and assembled into networks by mixing with *E. coli*-derived CFS.

D. Time course of CRISPRa in CFS. **Left:** CRISPRa-directed red fluorescent protein expression levels (mRFP1) from the J3 promoter are plotted as a function of time and J306 scRNA plasmid concentration. **Right:** Relative CRISPRa-directed RFP production rates are plotted as a function of time and J306 scRNA plasmid concentration. Dashed line represents 2-fold activation compared to the basal expression control (0 nM J306 scRNA plasmid). Values represent the mean ± standard deviation of three technical replicates.

E. Time course of CRISPRi in CFS. **Left:** CRISPRi-directed repression of green fluorescent protein (sfGFP) expression is plotted as a function of time and SF1 sgRNA plasmid concentration. SF1 targets the sfGFP CDS. **Right:** Relative GFP

production rates are plotted as a function of time and SF1 sgRNA plasmid concentration. Dashed line represents 50% repression compared to the basal expression control (0 nM SF1 sgRNA plasmid). Values represent the mean ± standard deviation of three technical replicates.

Figure 2: Level-matching enables construction of multi-layer CRISPRa/i circuits in CFS



A. CRISPRa transcriptional input-output response curves. **Top:** Expression of J306 scRNA (X) directs CRISPRa from the J3 promoter (Y) producing an RFP output. Expression of J206 scRNA (Y) directs CRISPRa from the J2 promoter (Z), producing an RFP output. **Bottom:** Fold change in transcriptional output is plotted as a function of transcriptional input, specified by the scRNA plasmid concentration. Fold change is calculated as the ratio of RFP expression level at the reaction endpoint in the presence versus the absence of scRNA plasmid. Red line indicates a logistic fit to the data. **Bottom Right:** Pink and green dashed lines are guides showing the predicted effect of 25-fold increases in scRNA transcriptional input generated by CRISPRa in the first layer on the output of CRISPRa in the second layer of a two-layer activation cascade. Values represent the mean ± standard deviation of three technical replicates.

B. Two-layer CRISPRa cascade. **Top:** Expression of J306 scRNA at node X directs CRISPRa from the J3 promoter at node Y. Expression of J206 scRNA from the J3 promoter (Y) directs CRISPRa from the J2 promoter, which expresses mRFP1 (Z). **Bottom:** Scatter plot comparing measured and predicted fold change in cascade output. The measured fold change in cascade output is calculated as the ratio of measured RFP outputs with and without CRISPRa in the first layer of the circuit (Methods, 5.1), and are presented as the mean ± standard deviation of three technical replicates. Predicted cascade fold changes and uncertainties are calculated from the fits to the scRNA plasmid titrations shown in 2A.

C. CRISPRa/i transcriptional input-output response curves. **Top:** Expression of J306 scRNA (X) directs CRISPRa from the J3 promoter (Y), producing an RFP output.

Expression of the RR2 sgRNA (Y) directs CRISPRi of mRFP1 (Z). **Bottom:** Fold change in transcriptional output is plotted as a function of transcriptional input, specified by the scRNA and sgRNA plasmid concentration. Red line indicates a logistic fit to the data. **Bottom left:** Response curve for J306 scRNA presented as in 2A **Bottom Right:** CRISPRi data are represented as percent expression of a no repression control (Methods, 5.1). Pink and green dashed lines show the expected effect of a 25-fold increase in sgRNA transcription provided by CRISPRa in the first layer on the CRISPRi-directed output in the second layer of a two-layer CRISPR activation-repression cascade. Values represent the mean ± standard deviation of three technical replicates.

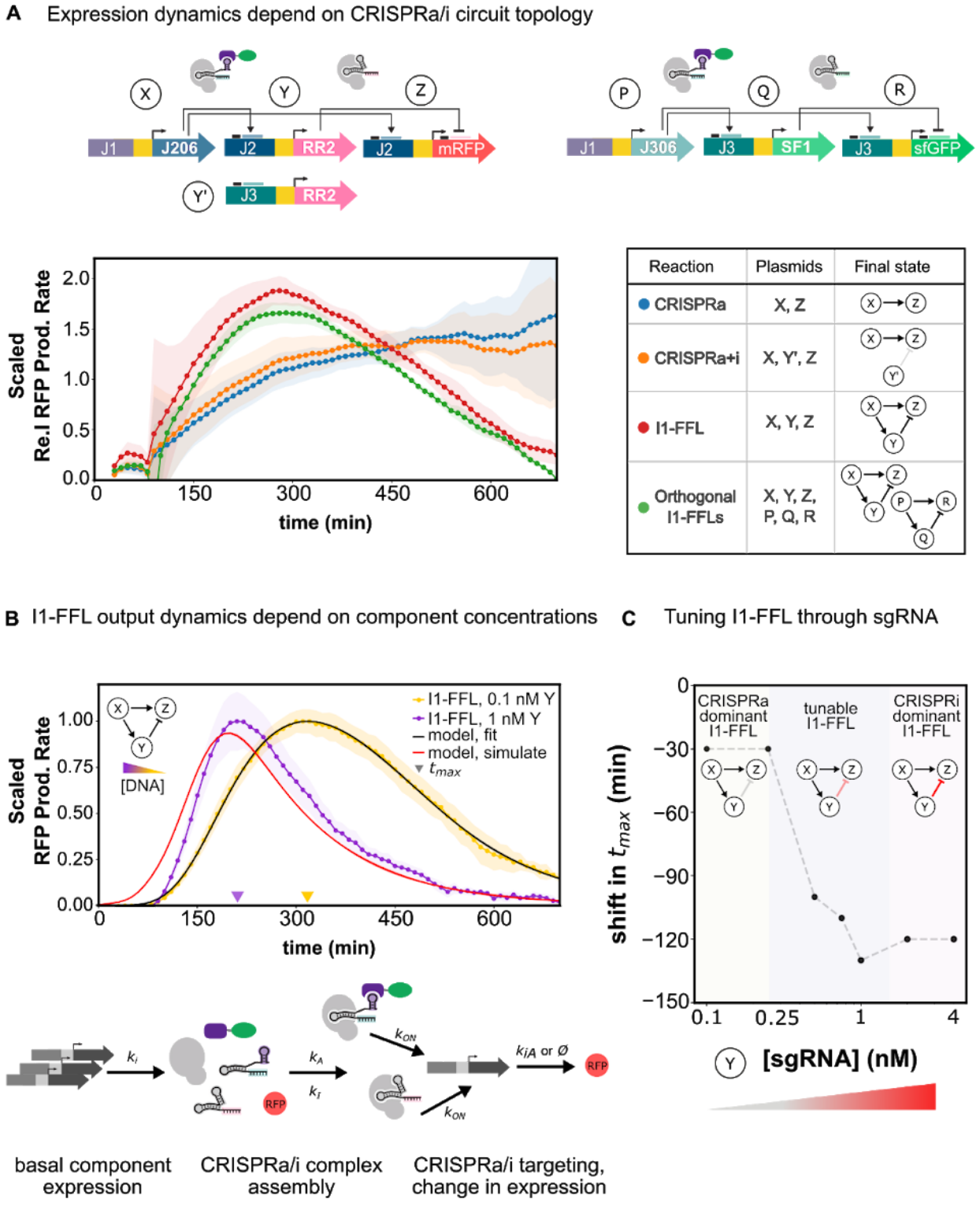
D. Two-layer CRISPRa/i cascade. **Top:** Expression of J306 scRNA (X) directs CRISPRa from the J3 promoter. RR2 sgRNA expression from the J3 promoter (Y) directs CRISPRi of mRFP1 (Z). **Bottom:** Percent of maximum expression is calculated as in Fig 2C comparing CRISPRi in one layer (I, light red) to two-layer CRISPR activation-repression cascades (II, dark red) at two different concentrations of sgRNA in the second layer. Values represent the mean ± standard deviation of three technical replicates.

E. CRISPRi and CRISPRa/i circuit dynamics. **Left:** Gene expression over time for RFP controlled by CRISPRi and CRISPRa/i cascade at a concentration of plasmid Y that falls outside the range where input-output levels between the first and second layers overlap (no level-matching). Black line represents the scaled relative RFP production rate due to CRISPRi with 16 nM of plasmid encoding RR2 sgRNA while the red line represents the scaled relative RFP production rate generated when CRISPRa is

applied to the same concentration of sgRNA plasmid. **Right:** Comparison of CRISPRi and CRISPRa/i cascade at a concentration of plasmid Y that permits level-matching between the first and second layers of the circuit. Black line represents the scaled relative RFP production rate generated by CRISPRi with 0.5 nM of plasmid encoding RR2 sgRNA while the red line represents expression generated when CRISPRa is applied to the same concentration of sgRNA plasmid. The dark red line shows the scaled relative expression generated by CRISPRi with 4 nM of sgRNA plasmid and is provided as a point of reference. Data are presented as mean ± standard deviation of three technical replicates.

F. Time to 50% repression is plotted against the concentration of sgRNA plasmid for both CRISPRi (black line) and a CRISPRa/i cascade (red line). Showing that when output/input ranges of the first and second layer overlap, multi layer CRISPRa/i circuits can be used to tune the timing of gene expression. Data are presented as mean ± standard deviation of three technical replicates.

Figure 3: Programming distinct multi-layer CRISPRa/i circuit dynamics in CFS

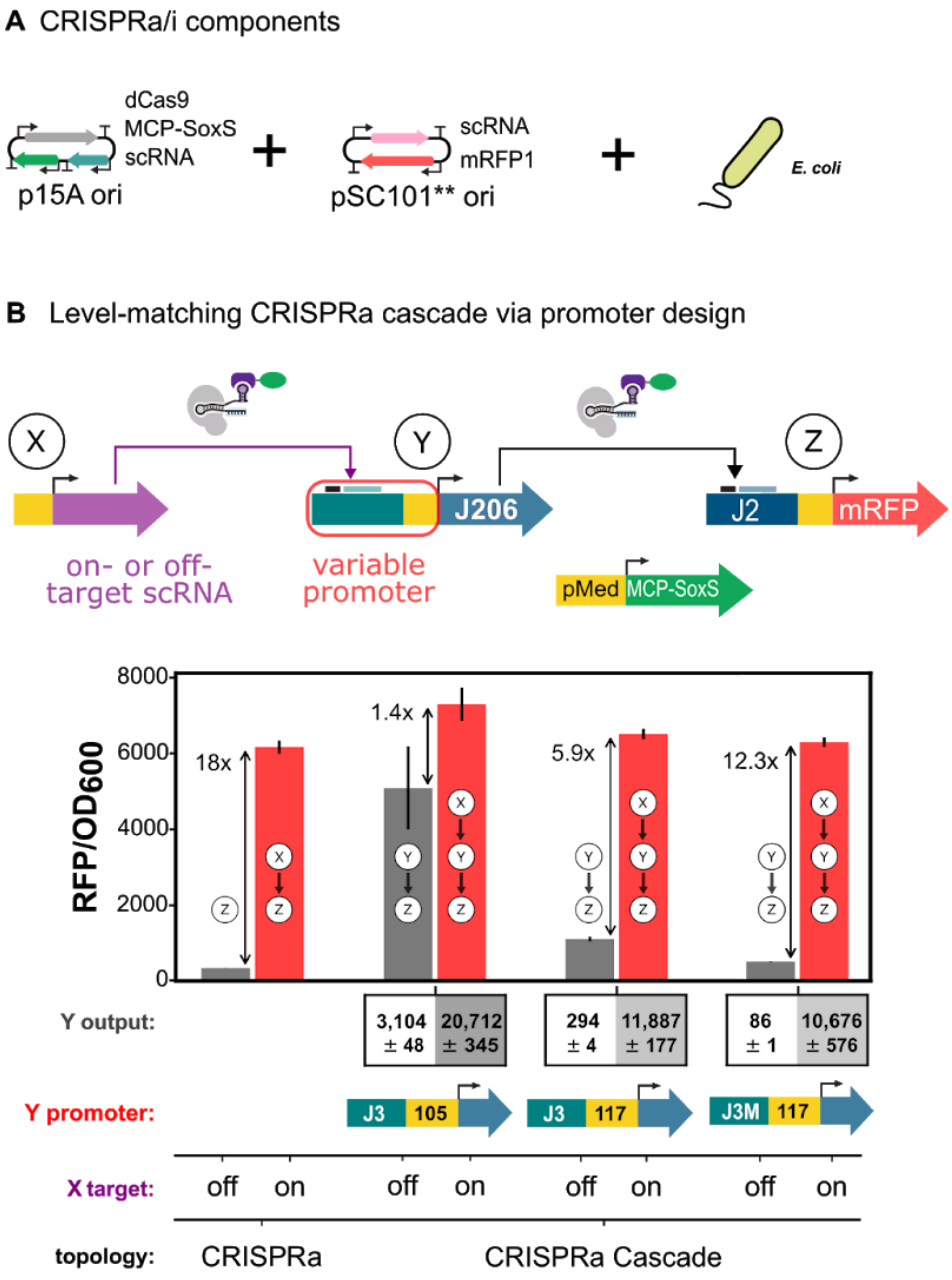


A. Time course for four different CRISPRa/i networks including orthogonal type 1 incoherent feed forward loops (I1-FFL) operating in CFS. Normalized relative RFP production rates are plotted as a function of network topology, as indicated. Outputs are scaled by the respective endpoints to fit on a common axis. Values represent the mean ± standard deviation of three technical replicates.

B. Time course for I1-FFL variants shows dependence of output expression dynamics on sgRNA concentration. Scaled RFP production rates are plotted as a function of time for two different RR2 sgRNA plasmid concentrations; 0.1 nM and 1 nM shown in yellow and purple respectively. Fits to the illustrated I1-FFL are used to predict expression dynamics as a function of sgRNA concentration. The model fit to the measured 0.1 nM data is shown in black, and the model prediction for expression dynamics from an I1-FFL with 1 nM sgRNA expressing plasmid is shown in red. Yellow and purple triangles along the x-axis denote the time of maximal expression of the pulse generated by the I1-FFL. Values represent the mean ± standard deviation of three technical replicates.

C. Shift in time of the I1-FFL expression maxima depends on sgRNA plasmid concentration, which controls the strength of the CRISPRi connection between nodes Y and Z. The time at which the maximum output rate is achieved (t_{max}) is plotted as a function of RR2 sgRNA plasmid concentration, which is expressed from node Y and acts on node Z.

Figure 4: Level-matching of CRISPRa cascades in *E. coli*

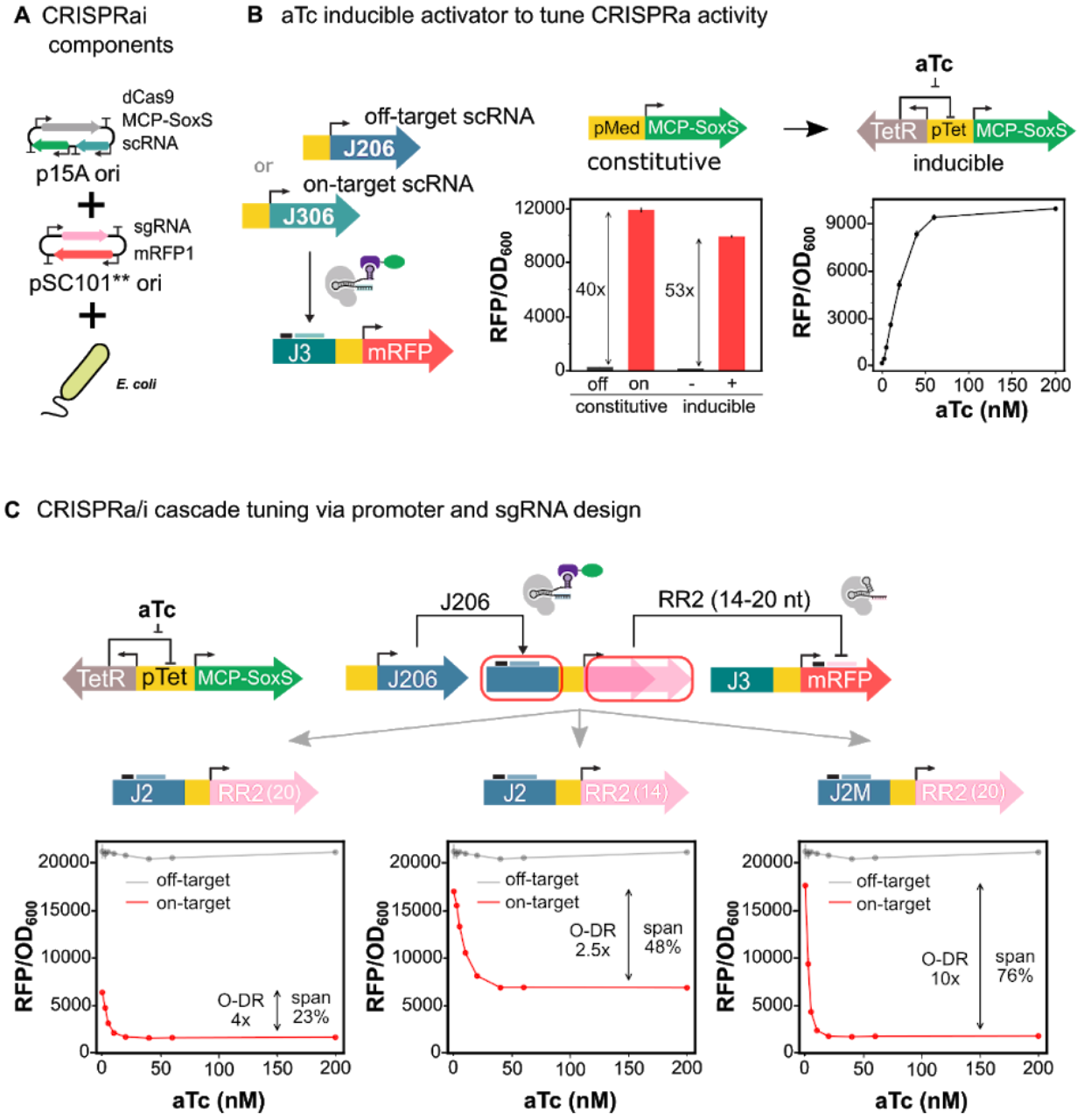


A. Schematic of CRISPRa genetic components and assembly to form CRISPRa circuits in *E. coli*. Circuits are assembled by transformation of different combinations of plasmids into *E. coli*. Activator, dCas9, and scRNAs in the first layer of a circuit are expressed from a p15A plasmid while reporters and scRNAs in the second layer of a circuit are expressed from a pSC101** origin of

replication plasmid. Data are collected in *E. coli* MG1655 grown overnight at 37 °C with shaking in EZ MOPS with 0.2% glucose and appropriate antibiotic selection.

B. **Top:** schematic of CRISPRa cascade. Tuning actions are applied by changing expression characteristics of the activatable promoter in the second layer of the circuit. **Bottom:** CRISPRa on the J2 promoter is compared to the output at node Z of the activation cascade with (red bars) and without (grey bars) input provided by node X. Y output denotes the expression levels obtained at node Y in the presence and absence of input activation from X for the tuning variant provided by Y promoter. Data are represented as mean ± standard deviation of RFP/OD_{600}.

Figure 5: Level-matching of CRISPRa/i cascades in *E. coli* with titratable input



A. Schematic of CRISPRa/i genetic components and assembly to form CRISPRa/i circuits in *E. coli*. Circuits are assembled by transformation of different combinations of plasmids into *E. coli*. Activator, dCas9, and scRNAs in the first layer of a circuit are expressed from a p15A plasmid while reporters, and

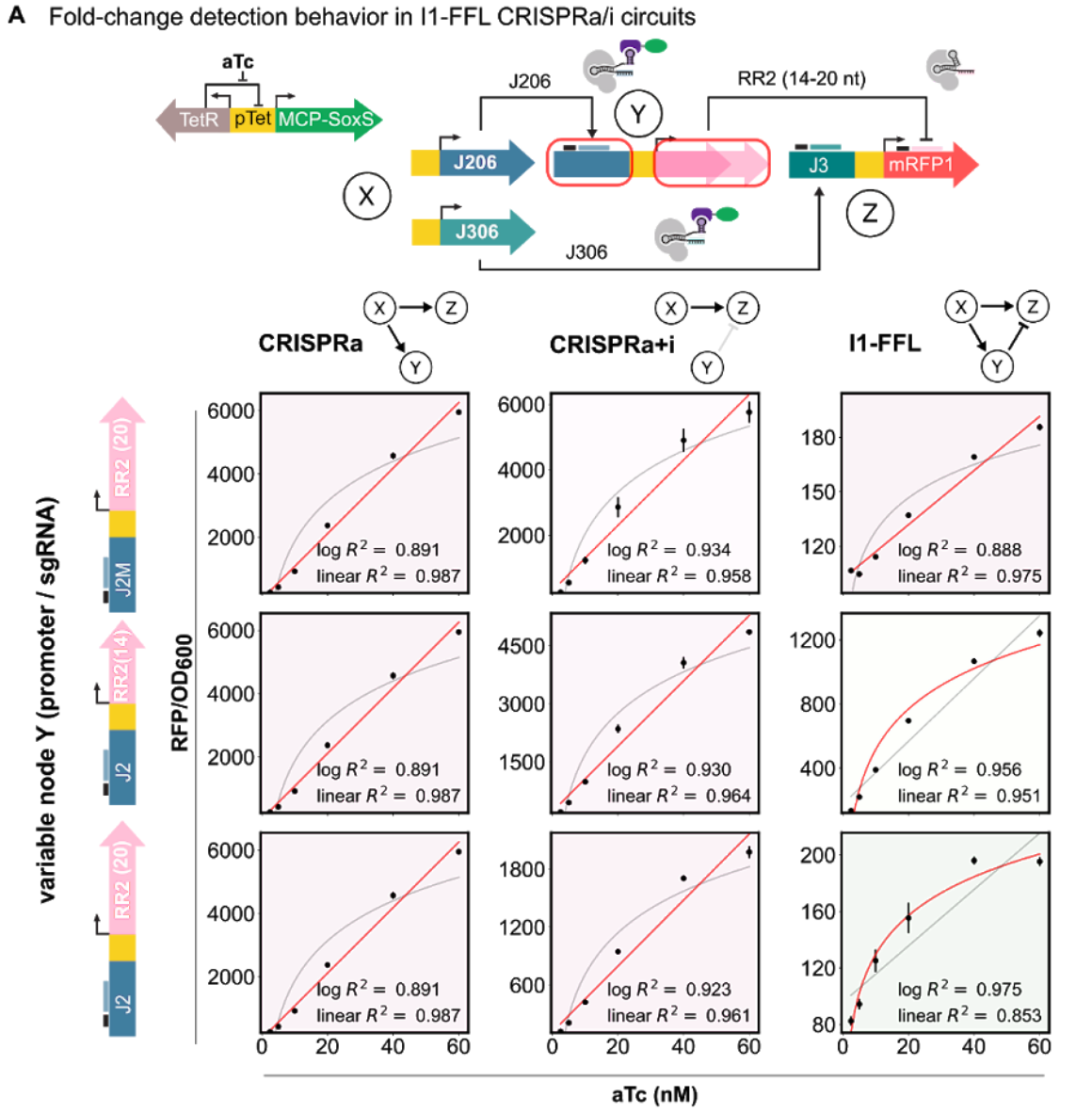
sc/sgRNAs in the second layer of a circuit are expressed from a pSC101** origin of replication plasmid. Inducible CRISPRa/i data are collected in *E. coli* MG1655 grown O/N, diluted 1:40, grown for 3 hours, and diluted 1:40 into inducing media, all at 37 °C with shaking in EZ MOPS with 0.2% glucose and appropriate antibiotic selection.

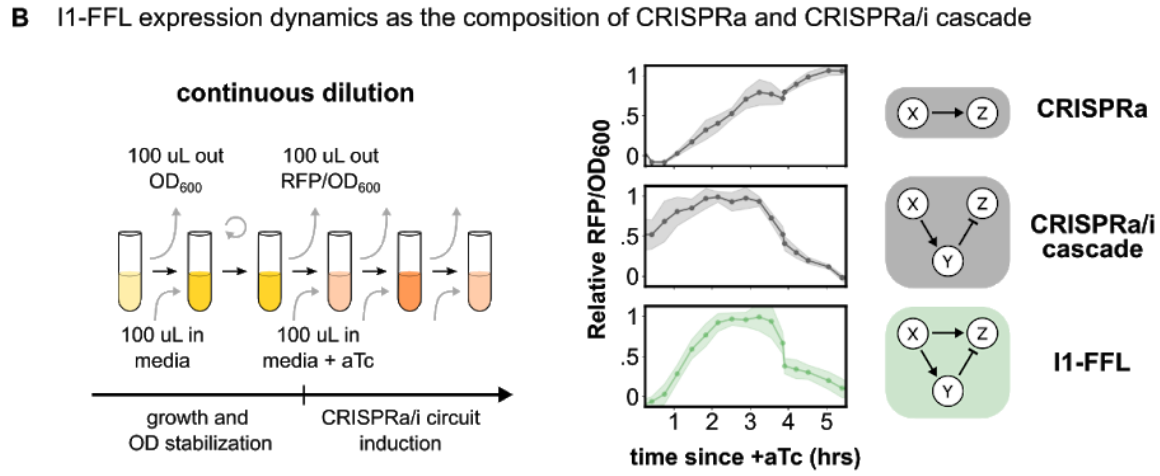
B. Dependence of CRISPRa on aTc-induced MCP-SoxS expression. **Left:** Circuit schematic showing how inducible CRISPRa is provided by pTet controlled expression of MCP-SoxS activator. For both constitutive and inducible CRISPRa, J306 scRNA directs dCas9 to the J3 promoter. Constitutive CRISPRa is compared with on/off target scRNA while inducible CRISPRa is compared ± inducer with on-target scRNA. **Center:** mRFP1 output of aTc-inducible CRISPRa ± aTc is compared to constitutive CRISPRa with on/off-target scRNA. Data are represented as mean of measured RFP/OD_{600} ± standard deviation of three biological replicates. **Right:** aTc induction of activator protein provides titratable levels of activation. Data are represented as the mean ± standard deviation of measured RFP fluorescence/OD_{600} for three biological replicates.

C. Two-layer CRISPRa/i cascade output dependence on promoter and sgRNA. **Top:** schematic of aTc-inducible CRISPRa activating expression of sgRNA targeting mRFP1 for CRISPRi. Level-matching is achieved through engineering basal expression characteristics of the activatable promoter driving sgRNA expression as well as through modifications of sgRNA activity via 5' sgRNA truncations that reduce sgRNA spacer-target DNA complementary from 20 to 14 nucleotides. **Bottom:** CRISPRi response to increasing levels of activation of the promoter

driving sgRNA expression for three different tunings. On-target sgRNA is plotted in Red while off-target sgRNA is plotted in grey. Data are represented as mean RFP/OD_{600} ± standard deviation. Calculations for Span and Output Dynamic range (O-DR) can be found in Methods 5.2.

Figure 6: Tunable and composable dynamic CRISPRa/i circuits in *E. coli*





A. Fold-change detection behavior in I1-FFL CRISPRa/i circuits. **Top:** circuit schematic of CRISPRa/i incoherent type 1 feed forward loop (I1-FFL) implemented in *E. coli*. Here tuning of repression by node Y is achieved by changes to the transcriptional properties of the promoter at node Y or through modifications to the sgRNA expressed at node Y. aTc induction of MCP-SoxS activator serves as a titratable input to the system. **Bottom:** aTc response for 9 combinations of network topology and sgRNA tuning showing that connectivity between nodes and tuning specifies input/output relationship of complex CRISPRa/i circuits. Network topologies are indicated along the top while tuning variants are represented along the left. Greyed out arrow in CRISPRa+i indicates on-target sgRNA from an unactivated promoter. I1-FFLs must have specific component relationships to achieve fold-change detection seen here as logarithmic input/output response curves. Data are represented as the mean of measured RFP/OD_{600} ± standard deviation of three biological replicates. Data for off-target sgRNA is the same in all three plots and is collected in the J2 context (bottom left).

B. CRISPRa/i circuit expression dynamics were measured in a continuous dilution experiment. **Left:** schematic of continuous dilution experimental design and procedure. **Right:** Expression dynamics of CRISPRa, CRISPRa/i cascade, and an I-1FFL tuned for fold-change detection are compared. Data are presented as relative RFP/OD_{600} ± standard deviation of three biological replicates (Methods, 5.2). Relative RFP/OD_{600} scales data to the respective output response ranges to place them on a common scale for comparison of output expression dynamics.

STAR Methods

1. Plasmid Design and Preparation

1.1 Design

Plasmid design and sequencing analyses were performed using Benchling sequence designer.

1.2 Cloning

All PCR amplification of plasmids and fragments used Phusion DNA polymerase in GC buffer. Primers were synthesized by IDT and resuspended into nuclease-free water. All PCR reactions were treated with DpnI for longer than 1 hour and purified using Qiagen gel extraction kits. Plasmid assembly was achieved using 5X In-Fusion HD mastermix (Takara). Assembled plasmids were transformed into chemically competent NEB Turbo *E. coli* and plated onto LB-agar plates with either 100 µg/mL carbenicillin or 25 µg/mL chloramphenicol and grown overnight ~16 hours at 37 °C. Single colonies were picked from plates and grown overnight in LB shaking at 37 °C with appropriate concentrations of relevant antibiotics. Plasmids were isolated from subcultures using a DNA miniprep kit (QIAprep Spin Miniprep Kit) and Sanger sequenced (Genewiz inc.) to identify correctly assembled plasmids. Plasmids intended for use in CFS were grown in culture volumes ~20 mL to ensure adequate yields for multiple cell-free reactions. Plasmids intended for cell-free expression were further purified using a PCR purification kit (Invitrogen PureLink, Cat. K310001), eluted into nuclease-free water. Plasmid

concentrations were quantified via spectrophotometry (Nanodrop 2000c, Cat. ND-2000C).

2. Cell-Free Reactions

2.1 Cell-Free System Preparation

The cell-free system was prepared according to previously published procedures (Garamella et al., 2016; Sun et al., 2013). The cell-free system used for an experiment was thawed on ice and pooled into a 1.5 ml eppendorf tube, vortexed, and spun-down using a mini benchtop centrifuge to ensure homogeneity across samples.

2.2 Cell-Free Gene Expression Reaction

Cell-free gene expression reactions were assembled on ice from the CFS and purified DNA. A master mix with common plasmids across reactions was prepared, and 1.5 µL per reaction allocated into PCR tubes. Plasmids which were varied across reactions were added in the remaining 1 µL. The CFS was pipette mixed and added to each PCR tube in 7.5 µL for a final volume of 10 µL. PCR tubes were vortexed, spun-down using a mini benchtop centrifuge, and placed on ice. Triplicates of 2.5 µL for each reaction were pipetted into individual wells of a 96-well V-bottom plate (Costar, Cat. 3363). The plate was sealed (Costar, Cat. 3080) and analyzed on a BioTek Synergy HTX plate reader at 29 °C. sfGFP fluorescence (ex. 485 nm, em. 528 nm) and mRFP1 fluorescence (ex. 540 nm, em. 600 nm) of cell free reactions were measured every 10 min from the bottom of the plate. All reactions were run in batch mode.

3. *E. coli* Experiments

Circuits were assembled in *E. coli* through transformation of plasmid pairs. In all *E. coli* experiments CRISPRa system components (dCas9, MCP-SoxS, input sc/sgRNAs) are located on a p15A ori plasmid while sc/sgRNAs forming the second layer of a circuit and reporter construct are located on a pSC101** ori plasmid. The p15A plasmid used in constitutive CRISPRa experiments was pCK085. aTc-inducible CRISPRa experiments use pJF182 in which pTet controls expression of TetR and MCP-SoxS. Plate reader measurements were conducted using a BioTek Synergy HTX with a black flat bottom plate (Ref# 3631) using 100 µL of culture.

3.1 Constitutive CRISPRa experiments:

Endpoint CRISPRa experiments are conducted using constitutive expression of all CRISPRa components (pCK085). Circuits were assembled by transformation of CRISPRa and reporter plasmids into chemically competent MG1655 E. coli. Transformed E. coli were outgrown for 1 hour shaking at 37 °C and plated onto LB-agar with carbenicillin and chloramphenicol. Plates were grown overnight at 37 °C. Experiments were conducted by picking three individual colonies into 400 µL Teknova EZ-RDM with 0.2% glucose and appropriate antibiotics in 96 well plates (round 2 ml), covering with breathable membrane (Breathe Easier cat# Z763624) and shaking overnight at 37 °C at 1200 RPM on a Heidolph titramax 1000.

3.2 aTc-inducible CRISPRa experiments

For inducible CRISPRa experiments expression of activator protein MCP-SoxS was controlled by pTet. Upon addition of aTc (anhydrotetracycline) to media, pTet becomes de-repressed which enables titratable expression of MCP-SoxS activator. As above, circuits are assembled by transformation of CRISPRa and reporter plasmids into chemically competent MG1655 *E. coli*. Transformed *E. coli* were outgrown for 1 hour shaking at 37 °C and plated onto LB-agar with carbenicillin and chloramphenicol. Plates were grown overnight at 37 °C. Experiments were conducted by picking three individual colonies into 400 uL EZ-RDM with 0.2% glucose and appropriate antibiotics in 96 well plates (2ml). Cultures are covered with breathable membrane and left shaking overnight at 37 °C at 1200 RPM on a Heidolph titramax 1000. Overnight cultures are subsequently diluted 1:40 into a fresh plate of EZ-RDM and incubated at 37 °C shaking at 1200 RPM. Before exiting exponential phase (~3 hours) cultures are diluted 1:40 into a fresh plate of EZ-RDM and supplemented with appropriate concentrations of aTc. These cultures are again covered with a breathable membrane, incubated in the dark at 37 °C shaking at 1200 RPM, and grown overnight ~18 hours. Measurements are conducted in Costar 96 well black flat bottom plates in 100 uL culture volume.

3.3 Continuous dilution *E. coli* experiments

Strains used in continuous dilution experiments were constructed as above through double transformation with appropriate selection on plates. For CRISPRa and I1-FFL experiments MG1655 was used while for CRISPRa/i cascades, MG1655 with an integrated reporter (J23119-RFP) was used (Fontana et al., 2018b). Throughout the

experiment all steps including liquid handling were conducted at 37 °C. Individual colonies were picked and grown overnight in 500 µL of LB with appropriate antibiotics shaking at 1200 RPM in a 96 well deep well plate. In the morning cultures were diluted 1:50 into RZ-RDM with 0.2% glucose and grown as before for one hour at a total volume of 300 µl. At one hour, 100 µL of culture was removed and measured while 100 µL fresh pre-warmed media was added to each culture. This process was repeated every 20 minutes until OD_{600} stabilized, usually within ~1-2 hours. Once ODs stabilized, aTc inducer was added to cultures to a final concentration of 50 nM. Every 20 minutes for the remainder of the experiment (~8 hours), 100 µL culture was removed and measured, while the culture was resupplied with 100 µL pre-warmed inducing media.

4. CFS CRISPRa/i Modeling

The model was defined using Antimony for Python 3.7 (Table S3). Parameters were generated from log-uniform distributions spanning 4 orders of magnitude based around literature values or best estimates. Models with different parameters were evaluated using Tellurium (Choi et al., 2018; Medley et al., 2018), and fit to scaled RFP production rates of experimental data by minimizing the cumulative point-wise squared error using the Nelder-Mead simplex algorithm.

5. Data Analysis

5.1 Cell-free data analysis

Production Rate:

Throughout this work, we define production rate as:

$\dot{B}^{\alpha}(t) = \frac{dB^{\alpha}}{dt} = \frac{B^{\alpha}(t+30) - B^{\alpha}(t)}{30}$

where:

B is the measured quantity: RFP or GFP

α specifies the circuit topology and relevant plasmid concentrations

Relative Production Rates:

Relative production rates of CRISPRa mediated outputs were calculated as the ratio of CRISPRa mediated production rates divided by unregulated production rates. For CRISPRa the contribution due to unregulated basal expression was subtracted from measured output levels due to CRISPRa. This was done to isolate the timing of CRISPRa mediated gene expression from the comparatively early contribution of basal expression, and to allow observation of CRISPRa mediated gene expression dynamics under conditions where basal expression of reporter constructs dominates. Throughout this work, relative production rates are abbreviated as Rel. RFP Prod. Rate, or Rel. GFP Prod. Rate, and are calculated as:

$$B\dot{}_{\Gamma}{}^{\alpha}(t)=(B\dot{}^{\alpha}(t)-B\dot{}^{\Gamma}(t))/(B\dot{}^{\Gamma}(t))$$

where:

α is a specific CRISPRa/i circuit

Γ is constitutive expression

For CRISPRi mediated relative production rates there is no subtraction of basal expression and relative production rates are provided as:

$$B\dot{}_{\Gamma}{}^{\alpha}(t)=(B\dot{}^{\alpha}(t))/(B\dot{}^{\Gamma}(t))$$

Fold change in cascade output:

In Figure 2B, fold change in cascade output was calculated as the ratio of RFP values generated by the CRISPRa cascade compared to CRISPRa in a single layer with the same concentration of scRNA expressing plasmid Y.

〖FC〗_β^α ("y")=((B^α (t=t_"max"))⁄(B^Γ (t=t_"max")))/((B^β (〖t" = " t〗_"max"))⁄(B^Γ (t=t_"max")))= (B^α (t=t_"max"))/(B^β (〖t" = " t〗_"max"))

where:

B is measured RFP

α is CRISPRa cascade, with y nM scRNA Y

β is CRISPRa, with y nM scRNA Y

tmax is the endpoint time of the cell free reaction

Predicted fold change in cascade output:

Predicted fold change in cascade output plotted in Figure 2B was generated using the fits to scRNA titration curves provided in 2A (red line). Predicted fold changes provided by CRISPRa at a given concentration of scRNA expressing plasmid y are denoted: (FC)̂_Γ^α (y). In this experiment, CRISPRa in the first layer of the cascade (2A, left) generated by x nM of scRNA at node X is expected to direct a 24-fold increase in transcription from the target. Thus, for a given concentration of scRNA expressing

plasmid y at node Y in the second layer of the cascade, predicted fold change in cascade output is calculated as the ratio of fold changes predicted by the fit in the right panel of 2A at a scRNA concentration of 24y, and 1y.

(FC)ˆ_β^α (y)" = " ((FC)ˆ_Γ^α (24y))/(〖FC〗 _Γ^α (y))

where:

is CRISPRa cascade, with y nM scRNA Y

is CRISPRa, with y nM scRNA Y

Normalized Z output:

In Figure 2C&D normalized Z output is defined as the percent expression provided by CRISPRi or a CRISPRa/i cascade compared to an unregulated, constitutive expression control.

〖nB〗 _Γ^α (y)=100·(B^α (t=t_"max"))/(B^Γ (t=t_"max"))

where:

B is measured RFP

α is CRISPRi or a CRISPRa/i cascade with y nM sgRNA plasmid at node Y

Scaled relative production rates:

In Figure 2E relative production rates were scaled by the observed production rate of the reaction at 40 mins to place curves on a common scale before the maturation of dCas9 and the onset of CRISPRa/i control.

sB˙_Γ^α (t)=(B˙_Γ^α (t))/(B˙_Γ^α (t=40))

where:

B˙_Γ^α is relative RFP production rate

is CRISPRi or CRISPRa/i cascade with y nM sgRNA plasmid at node Y

Time to 50% repression:

Time to 50% repression in Figure 2F and S4 was defined as the time at which relative production rates due to CRISPRa/i control first reached 50% of the initial 50 min value i.e., before maturation of dCas9 and the onset of CRISPRi.

t_(50%): (B˙_Γ^α (t))/(B˙_Γ^α (t=50))≥.5

where:

B˙_Γ^α is relative RFP production rate

is CRISPRi, or CRISPRa/i cascade with y nM sgRNA plasmid at node Y

Normalized relative RFP production rates:

In figure 3A relative RFP production rates are scaled by their respective endpoints to place curves on a common scale:

nB ̇_Γ^α (t)=(B ̇_Γ^α (t))/(B^α (t=t_"max"))

where:

B is measured RFP

B ̇_Γ^α is relative RFP production rate

 is CRISPRa, CRISPRa+i, or an I1-FFl

Scaled RFP Production Rate:

In Figure 3B RFP production rates were scaled by their respective maxima, placing both curves on a common scale, allowing comparison of time dynamics.

sB ̈^α (t)=(B ̈^α (t))/(max(B ̈^α (t)))

where:

B ̈(t) is RFP production rate

 is an I1-FFL with y = .1 nM or y = 1 nM sgRNA plasmid at node Y

Percent signal propagated by CRISPRa cascade:

The percent signal propagated by the CRISPRa cascade in CFS was calculated as the fold change in cascade output ± input divided by the fold change provided by CRISPRa in the input layer.

SP(y)= 100*(〚FC〛_(β_1)^(α_1) ("y "))/(〚FC〛_Γ^(β_2) ("y "))

where:

1 is CRISPRa cascade with y nM of scRNA at node Y

1 is CRISPRa with y nM of scRNA at node Y

2 is CRISPRa with x nM of scRNA at node X

Time to 2x activation:

Time to 2-fold activation was defined as the time at which relative production rates are expected to first exceed 1, i.e., when CRISPRa mediated production rates first achieve twice that of unregulated expression. Time to 2-fold activation was calculated as the mean ± standard deviation from linear fits to relative RFP production rates for three technical replicates. Linear fits were calculated over a 1 hour interval between 80 and 160 mins corresponding to the initial linear increase in relative RFP production rates provided by CRISPRa.

t_2x：B˙_Γ^α (t)≥ 1.0

where:

(B(t))˙ is relative RFP production rate

is CRISPRa with x nM scRNA at node X

Time to half maximal expression:

In figure S5B, time to half maximal expression was calculated differently for CRISPRa and CRISPRi. For CRISPRa the contribution due to unregulated basal expression was subtracted from measured RFP levels due to CRISPRa. This was done to isolate the timing of CRISPRa mediated gene expression from the comparatively early contribution of leak, and to allow observation of CRISPRa mediated gene expression dynamics under conditions where basal expression of reporter constructs dominates. The time to half max is denoted as t1/2.

For CRISPRa this was defined as the first time point at which

t_(1⁄2)： (B^α (t)-B^Γ (t))/(B^α (t=t_"max")-B^Γ (t=t_"max"))≥ .5

where:

B is measured RFP

is CRISPRa with x nM scRNA at node X

For CRISPRi there was no subtraction of basal expression thus time to half maximal expression for CRISPRi is given by the first time point at which

t_(1⁄2)： (B^α (t))/(B^α (t=t_"max"))≥ .5

where:

B is measured RFP

is CRISPRi with x nM scRNA at node X

Relative fold change:

In Figure S12, the gene expression dynamics of CRISPRa and a CRISPRa cascade are compared to visualize the speed of signal propagation in multi-layer CRISPR circuits. For these data, fold change is a function of time:

〖rFC〗_Γ^α (t,y)= (〖FC〗_Γ^α (t,y))/(〖FC〗_Γ^α (t=t_"max" ,y))

where:

B is measured RFP

is CRISPRa or CRISPRa cascade with y nM scRNA plasmid at node Y

y is the concentration of scRNA expressing plasmid in the final layer of the circuit

Normalized fluorescence

In figure S7, RFP and GFP fluorescence were normalized by the response range for each fluorescent protein to lie on a common scale between 0 and 1.

nB^=(B^ (t=t_"max")-min(B^ (t=t_"max")))/(max(B^ (t=t_"max"))-min(B^ (t=t_"max")))

where:

B^ is a vector containing average RFP or GFP fluorescence for all tested conditions (CRISPRa, CRISPRa+i, I1-FFL, orthogonal I1-FFLs)

Number of possible network topologies

The number of possible network topologies presented in Figure S1 was calculated as:

T(M,N)=(M+1)(N*(N-1))

where:

M is the number of modes of regulation

M = 1 for CRISPRi alone

M = 2 for both CRISPRa and CRISPRi together

N is the number of nodes in the network

5.2 E. coli data analysis

Throughout this work all measured RFP levels in E. coli were normalized by measured OD600. Data are plotted as the mean RFP/OD600 ± standard deviation of three biological replicates with appropriate propagation of uncertainties.

Span:

In Figure 5C span was calculated in percent as the range of RFP expression values provided by CRISPRa/i cascades (0, 200nM aTc) divided by RFP expression levels obtained in the presence of off-target sgRNA.

$S = (B^{(\alpha_1)} - B^{(\alpha_2)}) / B^{(\beta_1)}$

where:

B is RFP/OD600 measured at endpoint

α1 is CRISPRa/i cascade with 0 nM aTc induction

α2 is CRISPRa/i cascade with 200 nM aTc induction

β1 is CRISPRa/I cascade with off target sgRNA

Output dynamic range:

Output dynamic range was calculated as the ratio of measured CRISPRa/i cascade RFP outputs at 0nM aTc and 200nM aTc. The inverse quantity was used for I1-FFL output dynamic range.

O-DR= $B^{(\alpha_1)}/B^{(\alpha_2)}$

where:

B is RFP/OD600 measured at endpoint

α1 is CRISPRa/i cascade with 0 nM aTc induction

α2 is CRISPRa/i cascade with 200 nM aTc induction

Relative RFP/OD600:

Relative RFP/OD600 presented in Figure 6B was calculated from raw RFP/OD600 data by subtraction of the minimum observed RFP/OD600 value post induction, and scaling outputs by their maximum observed values to place all curves on a scale between zero and one.

$rB^{(t)}= (B^{(t)}-\min(B^{(t>t_0)}))/(\max(B^{(t>t_0)}))$

where:

B is RFP/OD600 measured at endpoint

α is CRISPRa, CRISPRa/i cascade, or I1-FFl with 50 nM aTc induction

t0 is the time at which aTc is added to the culture

Leak:

In Figure S10 leak was calculated as the percent reduction of measured CRISPRa RFP levels at 200nM aTc induction, due to unactivated sgRNA expression.

L=100·(B^-B^β)/B^

where:

B is RFP/OD600 measured at endpoint

α is CRISPRa at 200 nM aTc

β is CRISPRa+i at 200 nM aTc

Output Range Compression:

Output range compression was defined as the output range of a CRISPRa/i circuit divided by the accessible output range. Here, output range of a CRISPRa/i cascade was defined as:

OR= B^(α_1)-B^(α_2)

While output range of CRISPRa was defined as:

OR= B^(α_2)-B^(α_1)

In a CRISPRa/i cascade, the accessible output range was calculated as the measured fluorescence provided by the CRISPRa/i cascade with an off target sgRNA

A=B^(β_1)

For I1-FFLs, the accessible output range was defined as the output range of the corresponding circuit with an off-target sgRNA, at 200 nM aTc and 0 nM aTc

A=B^(β_2)- B^(β_1)

While for CRISPRa+i, the accessible output range was defined as the output range with an off-target scRNA directing CRISPRa to activate sgRNA expression, at 200 nM aTc and 0 nM aTc

A=B^(β_2)- B^(β_1)

Thus, output range compression of CRISPRa/i cascades was defined in percent as:

ORC=100·(1-OR/A)=100·(1-(B^(α_1)-B^(α_2))/B^(β_1))

While output range compression of CRISPRa+I was defined as:

ORC=100·(1-OR/A)=100·(1-(B^(α_2)-B^(α_1))/(B^(β_2)-B^(β_1)))

where:

B is RFP/OD600 measured at endpoint

1 is CRISPRa+i, at 0 nM aTc

2 is CRISPRa+i, at 200 nM aTc

1 is CRISPRa+i, with off-target sgRNA at 0 nM aTc

2 is CRISPRa+i, with off-target sgRNA at 200 nM aTc

Percent CRISPRa induction:

Percent CRISPRa induction presented in Figure S11 was calculated by dividing measured RFP values obtained from CRISPRa by the measured RFP value provided by CRISPRa at maximal, saturating levels of aTc induction (200nM).

$$\%I(x)=100\cdot B^{(\alpha_1)}/B^{(\alpha_2)}$$

where:

B is RFP/OD600 measured at endpoint

α1 is CRISPRa at x nM aTc

α2 is CRISPRa at 200 nM aTc

6. Data and Code Availability

All data, models, and scripts used to generate the figures reported in this paper are available at: https://github.com/carothersresearch/CRISPRai_Circuits_2021

References

Adamala, K.P., Martin-Alarcon, D.A., Guthrie-Honea, K.R., and Boyden, E.S. (2017). Engineering genetic circuit interactions within and between synthetic minimal cells. Nat. Chem. *9*, 431–439.

Adler, M., and Alon, U. (2018). Fold-change detection in biological systems. Curr. Opin. Syst. Biol. *8*, 81–89.

Alon, U. (2007). Network motifs: theory and experimental approaches. Nat. Rev. Genet. *8*, 450–461.

Banerjee, D., Eng, T., Lau, A.K., Sasaki, Y., Wang, B., Chen, Y., Prahl, J.-P., Singan, V.R., Herbert, R.A., Liu, Y., et al. (2020). Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. Nat. Commun. *11*, 5385.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. *41*, 7429–7437.

Bobrovskyy, M., and Vanderpool, C.K. (2013). Regulation of Bacterial Metabolism by Small RNAs Using Diverse Mechanisms. Annu. Rev. Genet. *47*, 209–232.

Brockman, I.M., and Prather, K.L.J. (2015). Dynamic metabolic engineering: New strategies for developing responsive cell factories. Biotechnol. J. *10*, 1360–1369.

Brophy, J.A.N., and Voigt, C.A. (2014). Principles of Genetic Circuit Design. Nat. Methods *11*, 508–520.

Chappell, J., Westbrook, A., Verosloff, M., and Lucks, J.B. (2017). Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. Nat. Commun. *8*, 1051.

Choi, K., Medley, J.K., König, M., Stocking, K., Smith, L., Gu, S., and Sauro, H.M. (2018). Tellurium: An extensible python-based modeling environment for systems and synthetic biology. Biosystems *171*, 74–79.

Clamons, S., and Murray, R. (2019). Modeling predicts that CRISPR-based activators, unlike CRISPR-based repressors, scale well with increasing gRNA competition and dCas9 bottlenecking. BioRxiv 719278.

Dinh, C.V., and Prather, K.L. (2020). Layered and multi-input autonomous dynamic control strategies for metabolic engineering. Curr. Opin. Biotechnol. *65*, 156–162.

Dinh, C.V., and Prather, K.L.J. (2019). Development of an autonomous and bifunctional quorum-sensing circuit for metabolic flux control in engineered *Escherichia coli*. Proc. Natl. Acad. Sci. *116*, 25562–25568.

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

Dudley, Q.M., Karim, A.S., and Jewett, M.C. (2015). Cell-free metabolic engineering: Biomanufacturing beyond the cell. Biotechnol. J. *10*, 69–82.

English, M.A., Gayet, R.V., and Collins, J.J. (2021). Designing Biological Circuits: Synthetic Biology Within the Operon Model and Beyond. Annu. Rev. Biochem. *90*, null.

Fontana, J., Voje, W.E., Zalatan, J.G., and Carothers, J.M. (2018a). Prospects for engineering dynamic CRISPR–Cas transcriptional circuits to improve bioproduction. J. Ind. Microbiol. Biotechnol. *45*, 481–490.

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

Fontana, J., Dong, C., Kiattisewee, C., Chavali, V.P., Tickman, B.I., Carothers, J.M., and Zalatan, J.G. (2020). Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. Nat. Commun. *11*, 1618.

Gander, M.W., Vrana, J.D., Voje, W.E., Carothers, J.M., and Klavins, E. (2017). Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. Nat. Commun. *8*, 15459.

Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016). The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. ACS Synth. Biol. *5*, 344–355.

Goentoro, L., Shoval, O., Kirschner, M., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol. Cell *36*, 894–899.

Huang, H.-H., Bellato, M., Qian, Y., Cárdenas, P., Pasotti, L., Magni, P., and Del Vecchio, D. (2021). dCas9 regulator to neutralize competition in CRISPRi circuits. Nat. Commun. *12*, 1692.

Jeong, D., Klocke, M., Agarwal, S., Kim, J., Choi, S., Franco, E., and Kim, J. (2019).

Cell-Free Synthetic Biology Platform for Engineering Synthetic Biological Circuits and Systems. Methods Protoc. *2*.

Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The incoherent feed-forward loop can generate non-monotonic input functions for genes. Mol. Syst. Biol. *4*.

Karim, A.S., Dudley, Q.M., and Jewett, M.C. (2016). Cell-Free Synthetic Systems for Metabolic Engineering and Biosynthetic Pathway Prototyping. In Industrial Biotechnology, (John Wiley & Sons, Ltd), pp. 125–148.

Kiattisewee, C., Dong, C., Fontana, J., Sugianto, W., Peralta-Yahya, P., Carothers, J.M., and Zalatan, J.G. (2021). Portable bacterial CRISPR transcriptional activation enables metabolic engineering in Pseudomonas putida. Metab. Eng. *66*, 283–295.

Landberg, J., Wright, N.R., Wulff, T., Herrgård, M.J., and Nielsen, A.T. (2020). CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in Escherichia coli. Biotechnol. Bioeng. *117*, 3835–3848.

Lehr, F.-X., Hanst, M., Vogel, M., Kremer, J., Göringer, H.U., Suess, B., and Koeppl, H. (2019). Cell-Free Prototyping of AND-Logic Gates Based on Heterogeneous RNA Activators. ACS Synth. Biol. *8*, 2163–2173.

Liu, Y., Wan, X., and Wang, B. (2019). Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. Nat. Commun. *10*, 1–16.

Lucks, J.B., Qi, L., Whitaker, W.R., and Arkin, A.P. (2008). Toward scalable parts families for predictable design of biological circuits. Curr. Opin. Microbiol. *11*, 567–573.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop

network motif. Proc. Natl. Acad. Sci. *100*, 11980–11985.

Marshall, R., and Noireaux, V. (2018). Synthetic Biology with an All E. coli TXTL System: Quantitative Characterization of Regulatory Elements and Gene Circuits. In Synthetic Biology: Methods and Protocols, J.C. Braman, ed. (New York, NY: Springer New York), pp. 61–93.

Marshall, R., Maxwell, C.S., Collins, S.P., Jacobsen, T., Luo, M.L., Begemann, M.B., Gray, B.N., January, E., Singer, A., He, Y., et al. (2018). Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. Mol. Cell *69*, 146-157.e3.

McDaniel, R., and Weiss, R. (2005). Advances in synthetic biology: on the path from prototypes to applications. Curr. Opin. Biotechnol. *16*, 476–483.

Medley, J.K., Choi, K., König, M., Smith, L., Gu, S., Hellerstein, J., Sealfon, S.C., and Sauro, H.M. (2018). Tellurium notebooks—An environment for reproducible dynamical modeling in systems biology. PLOS Comput. Biol. *14*, e1006220.

Moore, S.J., MacDonald, J.T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., Kylilis, N., Bell, D.J., McClymont, D.W., Jensen, K., et al. (2018). Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. Proc. Natl. Acad. Sci. U. S. A. *115*, E4340–E4349.

Nielsen, A.A., and Voigt, C.A. (2014). Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. Mol. Syst. Biol. *10*.

Nielsen, A.A.K., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A.,

Ross, D., Densmore, D., and Voigt, C.A. (2016). Genetic circuit design automation. Science *352*.

Poole, W., Pandey, A., Shur, A., Tuza, Z.A., and Murray, R.M. (2020). BioCRNpyler: Compiling Chemical Reaction Networks from Biomolecular Parts in Diverse Contexts (Synthetic Biology).

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Qian, Y., Huang, H.-H., Jiménez, J.I., and Del Vecchio, D. (2017). Resource Competition Shapes the Response of Genetic Circuits. ACS Synth. Biol. *6*, 1263–1272.

Reis, A.C., Halper, S.M., Vezeau, G.E., Cetnar, D.P., Hossain, A., Clauer, P.R., and Salis, H.M. (2019). Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. Nat. Biotechnol. *37*, 1294–1301.

Rosenfeld, N., and Alon, U. (2003). Response Delays and the Structure of Transcription Networks. J. Mol. Biol. *329*, 645–654.

Santos-Moreno, J., and Schaerli, Y. (2020). CRISPR-based gene expression control for synthetic gene circuits. Biochem. Soc. Trans. *48*, 1979–1993.

Santos-Moreno, J., Tasiudi, E., Stelling, J., and Schaerli, Y. (2020). Multistable and dynamic CRISPRi-based synthetic circuits. Nat. Commun. *11*, 2746.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat. Genet. *31*, 64–68.

Shin, J., and Noireaux, V. (2012). An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. ACS Synth. Biol. *1*, 29–41.

Stevens, J.T., and Carothers, J.M. (2015). Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. ACS Synth. Biol. *4*, 107–115.

Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., and Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J. Vis. Exp. JoVE e50762.

Takahashi, M.K., Hayes, C.A., Chappell, J., Sun, Z.Z., Murray, R.M., Noireaux, V., and Lucks, J.B. (2015). Characterizing and prototyping genetic networks with cell-free transcription–translation reactions. Methods *86*, 60–72.

Tan, S.-I., and Ng, I.-S. (2021). CRISPRi-Mediated NIMPLY Logic Gate for Fine-Tuning the Whole-Cell Sensing toward Simple Urine Glucose Detection. ACS Synth. Biol. *10*, 412–421.

Thieffry, D., Huerta, A.M., Pérez-Rueda, E., and Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. BioEssays *20*, 433–440.

Tian, J., Yang, G., Gu, Y., Sun, X., Lu, Y., and Jiang, W. (2020). Developing an endogenous quorum-sensing based CRISPRi circuit for autonomous and tunable dynamic regulation of multiple targets in Streptomyces. Nucleic Acids Res. *48*, 8188–8202.

Tian, T., Kang, J.W., Kang, A., and Lee, T.S. (2019). Redirecting Metabolic Flux via Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in Escherichia coli. ACS Synth. Biol. *8*, 391–402.

Wang, Y.-H., Wei, K.Y., and Smolke, C.D. (2013). Synthetic biology: advancing the design of diverse genetic systems. Annu. Rev. Chem. Biomol. Eng. *4*, 69–102.

Westbrook, A., Tang, X., Marshall, R., Maxwell, C.S., Chappell, J., Agrawal, D.K., Dunlop, M.J., Noireaux, V., Beisel, C.L., Lucks, J., et al. (2019). Distinct timescales of RNA regulators enable the construction of a genetic pulse generator. Biotechnol. Bioeng. *116*, 1139–1151.

Wu, Y., Chen, T., Liu, Y., Tian, R., Lv, X., Li, J., Du, G., Chen, J., Ledesma-Amaro, R., and Liu, L. (2020). Design of a programmable biosensor-CRISPRi genetic circuits for dynamic and autonomous dual-control of metabolic flux in Bacillus subtilis. Nucleic Acids Res. *48*, 996–1009.

Xiang, Y., Dalchau, N., and Wang, B. (2018). Scaling up genetic circuit design for cellular computing: advances and prospects. Nat. Comput. *17*, 833–853.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2015). Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. Cell *160*, 339–350.

Supplemental information

Multi-Layer CRISPRa/i Circuits for Dynamic Genetic Programs in Cell-Free and Bacterial Systems

Benjamin I. Tickman[+,1], Diego Alba Burbano[+,1,2], Venkata P. Chavali[1], Cholpisit Kiattisewee[1], Jason Fontana[1], Aset Khakimzhan[3], Vincent Noireaux[3], Jesse G. Zalatan[*1,4], James M. Carothers[*,1,2]

1: Molecular Engineering & Sciences Institute
and Center for Synthetic Biology
University of Washington
Seattle, WA 98195
United States

2: Department of Chemical Engineering
University of Washington
Seattle, WA 98195
United States

3: School of Physics and Astronomy
University of Minnesota

Minneapolis, MN 55455, USA

4: Department of Chemistry

University of Washington

Seattle, WA 98195

United States

+: These authors contributed equally

*: Corresponding authors

zalatan@uw.edu

206-543-1670

jcaroth@uw.edu

206-221-4902

In preparation as an article for ***Cell Systems***

Supplementary Figures

Supplementary Tables

Table S5: Parts

Table S6: Plasmid breakdown per figure

Table S7: Cell-Free plasmid concentrations

Supplementary Figures

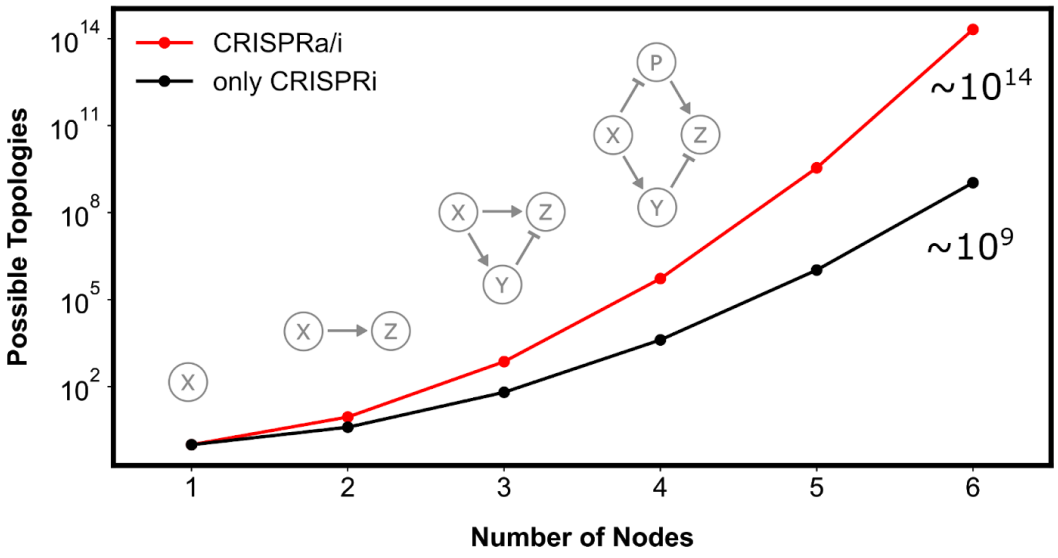


Figure S1: Number of accessible topologies with CRISPRa and CRISPRi

Number of unique circuit topologies accessible with a given number of nodes is plotted for networks composed of only CRISPRi and for networks containing both CRISPRa and CRISPRi. As the number of nodes increases CRISPRa/i control systems can generate exponentially more topologies than CRISPRi alone (Adler et al., 2017).

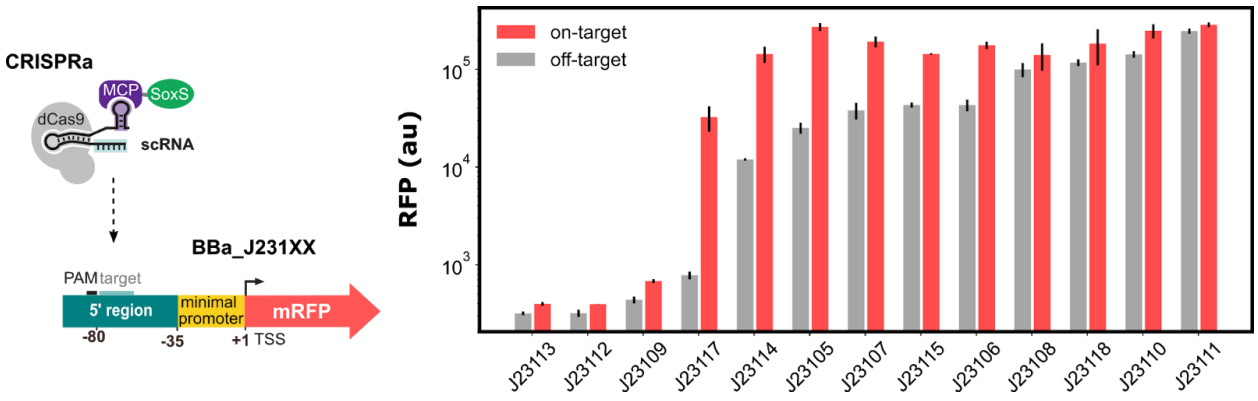


Figure S2: CRISPRa on different minimal promoters in CFS

End-point measurement of RFP values with J306 scRNA targeting a J3 reporter with different minimal promoter strengths (BBa_J231XX). J206 scRNA was used as an off-target control for each promoter strength (grey bars). Shading indicates magnitude of activated expression, and promoters are ordered by unactivated expression levels. Data are presented as the mean ± standard deviation of three technical replicates. Compare to previously published data in *E. coli* & *P. putida* (Fontana et al., 2020; Kiattisewee et al., 2021)

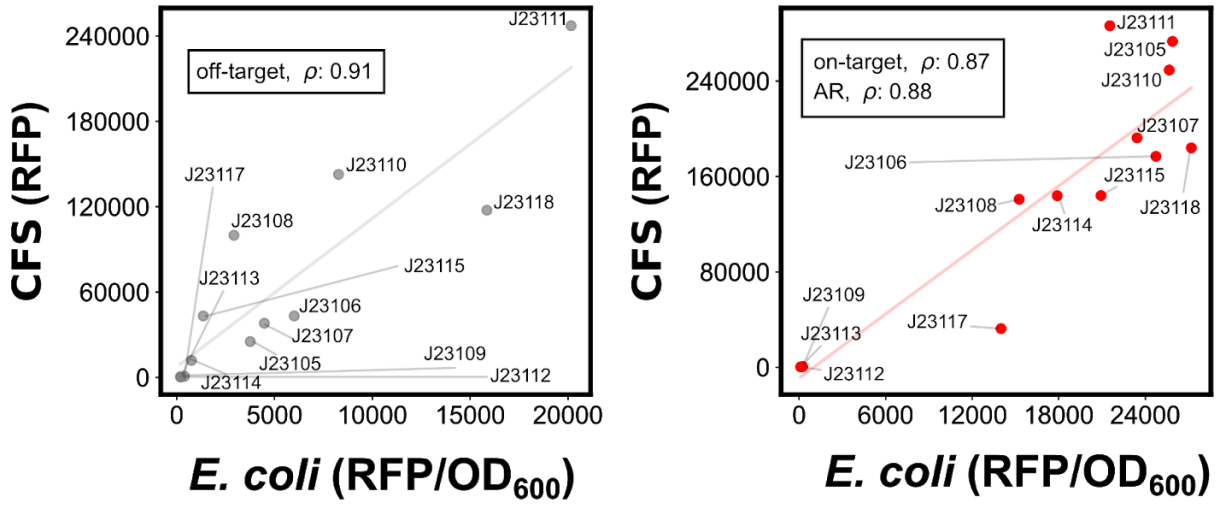


Figure S3: CRISPRa correlation between CFS and *E. coli*

Comparison of cell free system (CFS) and *E. coli* with respect to basal and activated expression levels derived from Anderson promoters (BBa_J231XX). These plots represented data from Figure S2 (CFS) and previously published data (Fontana et al., 2020). The Spearman's rank correlation coefficient (ρ) of basal and activated expression levels are 0.91 and 0.87, respectively. The activation ratio (AR) calculated from basal and activated level of each Anderson promoter is also correlated (ρ = 0.87).

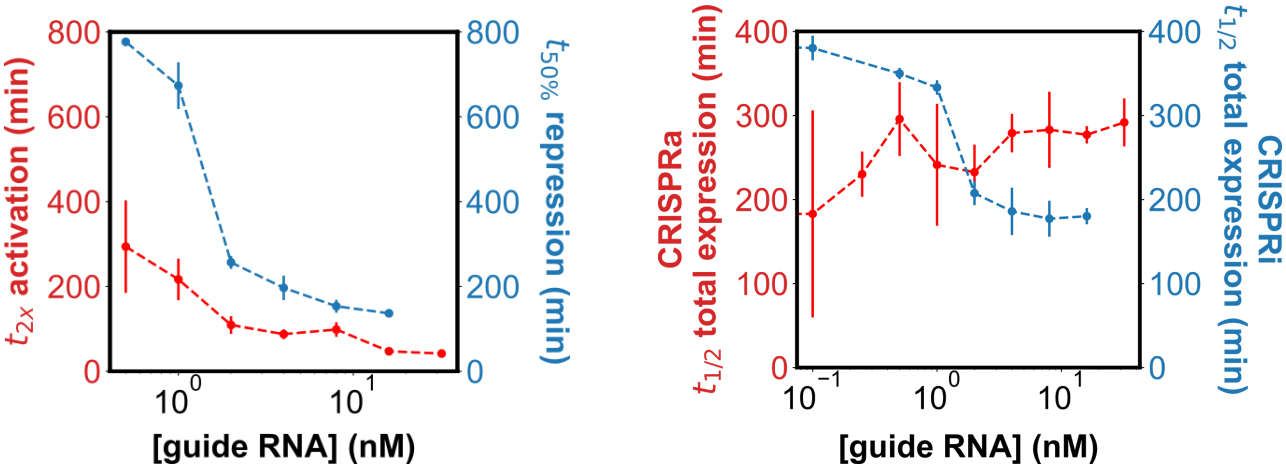


Figure S4: CRISPRa/i dynamics

Response time of CRISPRa and CRISPRi in CFS is plotted at various concentrations of plasmid expressing sc/sgRNA (left). Response time of CRISPRi is defined as time to 50% repression using relative production rates as in section 5.1. Response time of CRISPRa is defined as time to 2-fold activation as calculated in section 5.1. Data are presented as the mean +/- standard deviation of the calculated quantity for three technical replicates. Time to half maximal expression of CRISPRa and CRISPRi in CFS is plotted at various concentrations of plasmid expressing sc/sgRNA (right). Time to half maximal expression shows that sgRNA levels affect the median time of gene expression, while scRNA titrations do not. For both CRISPRa and CRISPRi time to half max is defined as the time at which the measured value of CRISPR regulated gene expression first exceeds 50% of the endpoint RFP measurement according to equations provided in section 5.1. For both CRISPRa and CRISPRi, data are presented as the average ± standard deviation of three technical replicates.

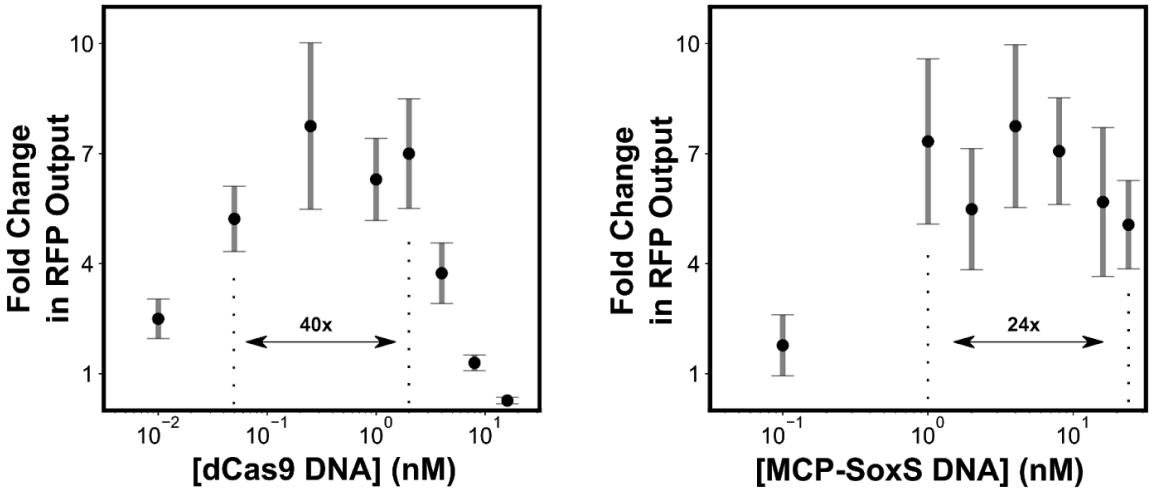


Figure S5: dCas9 and MCP-SoxS plasmid titrations

Titrations of plasmid expressing dCas9 (left) and MCP-SoxS (right) with J306 scRNA targeting the J3 promoter for CRISPRa. Data are represented as the mean ± standard deviation of three technical replicates showing fold change in output RFP levels due to CRISPRa compared to a no scRNA control measured at endpoint.

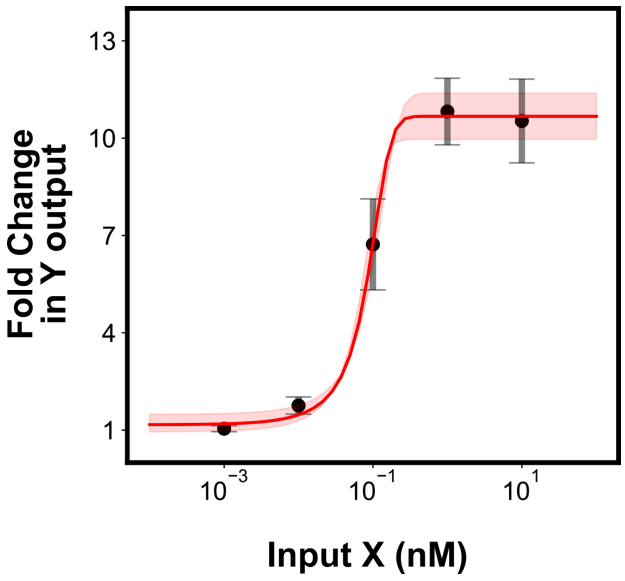


Figure S6: Titration of plasmid expressing scRNA from a strong constitutive promoter

CRISPRa response to varying expression levels of J306 scRNA activating the J3 promoter. The concentration of scRNA plasmid under the control of a strong constitutive promoter (BBa_J23119) was titrated to saturation of CRISPRa response. The output was saturated with a lower concentration of scRNA plasmid compared to that under the control of a weaker promoter (BBa_J23117), as shown in Figure 2A. Data are represented as the mean ± standard deviation of three technical replicates showing fold change in output RFP levels due to CRISPRa at various scRNA expression levels. Data is fitted to a logistic function, shown in red.

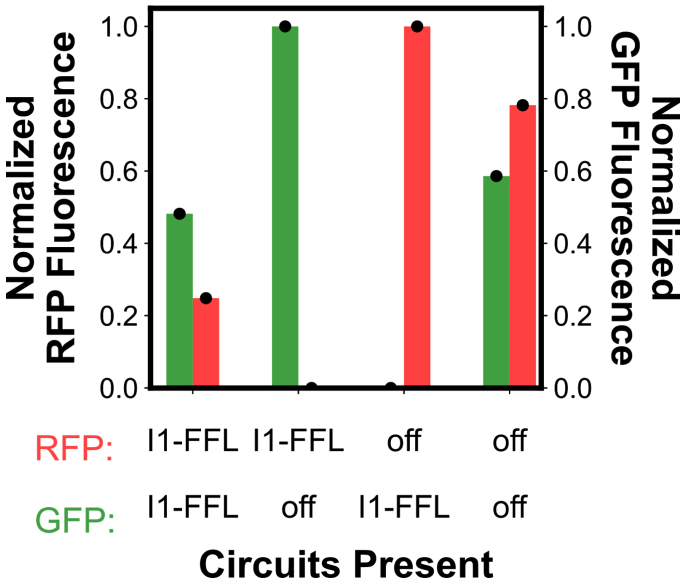


Figure S7: Orthogonal I1-FFLs

Endpoint measurements of RFP and GFP for cell-free reactions containing two orthogonal I1-FFLs with all four combinations of off/on target sgRNA. Data are normalized by the response range for each fluorescent protein to lie on a common scale between 0 and 1. Data are presented as the mean of the calculated quantity for three technical replicates with background subtraction due to autofluorescence in the GFP channel. Background subtraction is provided by the mean of three technical replicates consisting of cell-free extract mixed with pure water.

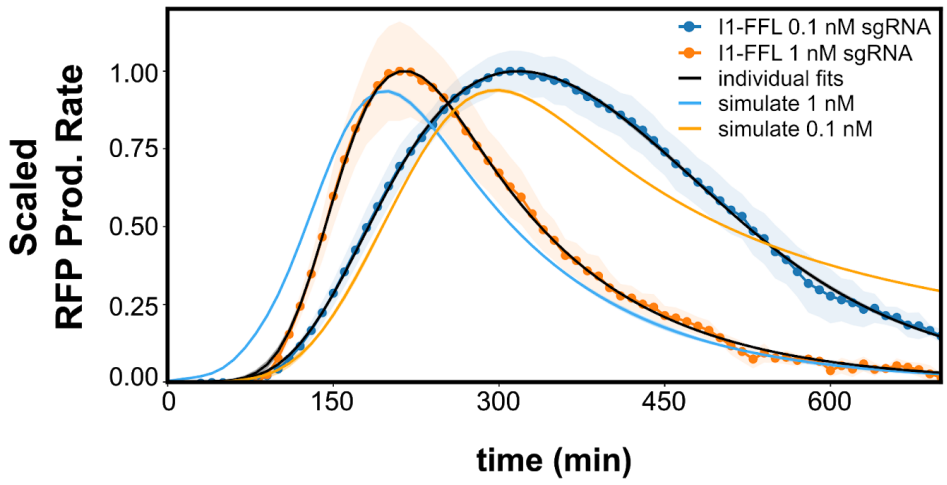


Figure S8: CFS CRSIPRa/i Model Cross Validation

Scaled RFP production rates for 2 I1-FFL plotted as a function of time with their respective fits and model predictions. Fits to each I1-FFL are shown in black and are used to predict expression dynamics for the other sgRNA concentration. Values represent the mean ± one standard deviation of three technical replicates.

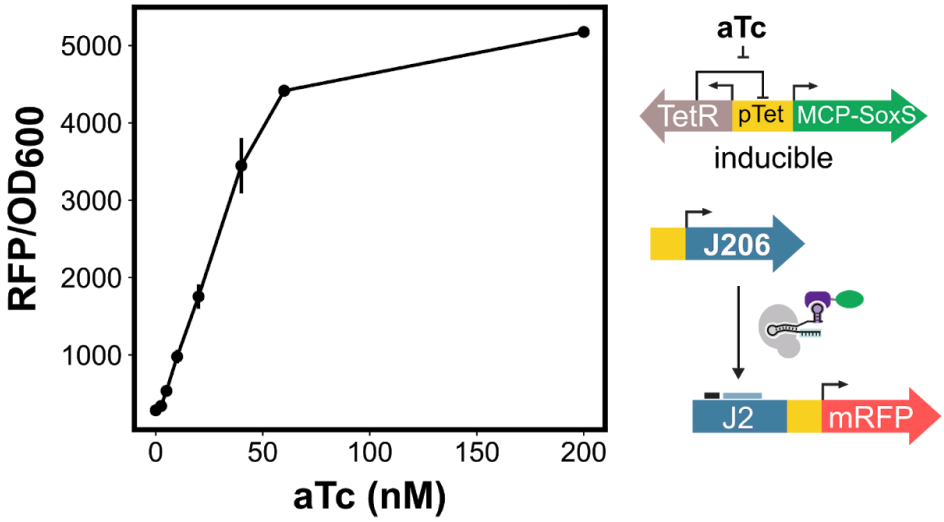


Figure S9: CRISPRa results based on J2-RFP in *E. coli*

Titration curve of inducible CRISPRa with MCP-SoxS under control of the pTet promoter and induced by aTc (0-200 nM). J206 scRNA was used for targeting the J2-RFP reporter. Compared to the J3 promoter(Figure 5A), the J2 promoter provides a similarly shaped response curve and a lower output dynamic range. These data suggest that while different promoters may have different activities, their sensitivity to MCP-SoxS induction is similar. Data are presented as mean RFP/OD_{600} ± standard deviation for three biological replicates.

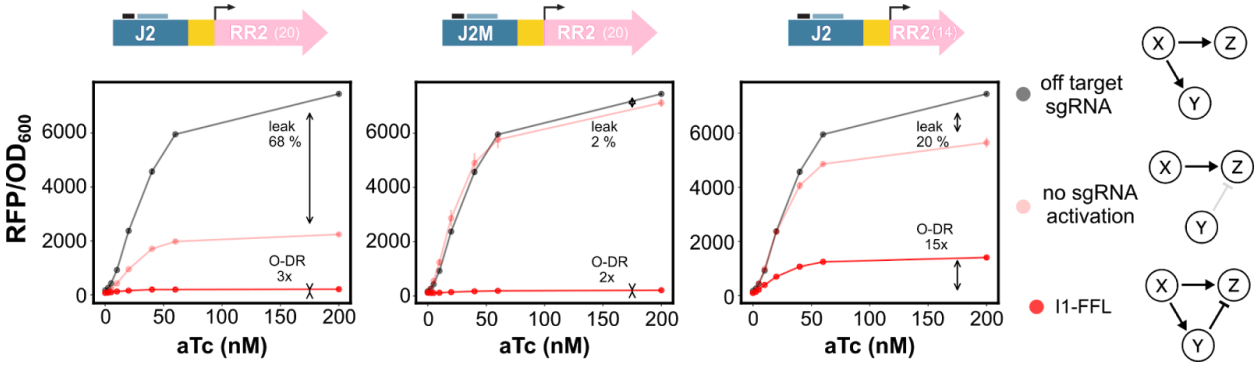


Figure S10: Tuning via promoter design and sgRNA truncations

Output RFP levels in response to aTc inputs are shown for the three different network topologies shown in Figure 6A (grey, pink, red), for three different tunings of CRISPRi levels generated by node Y as shown in Figures 5B, 6B (left, center, right). Promoter and sgRNA tuning are presented above their corresponding figure panels. Data are represented as the mean ± standard deviation of endpoint RFP/OD_{600} for three biological replicates. Data for off target sgRNA is the same in all three plots and is collected using J2 expressed off target sgRNA (left). These data are replotted in the center and right panels to provide a no repression comparison.

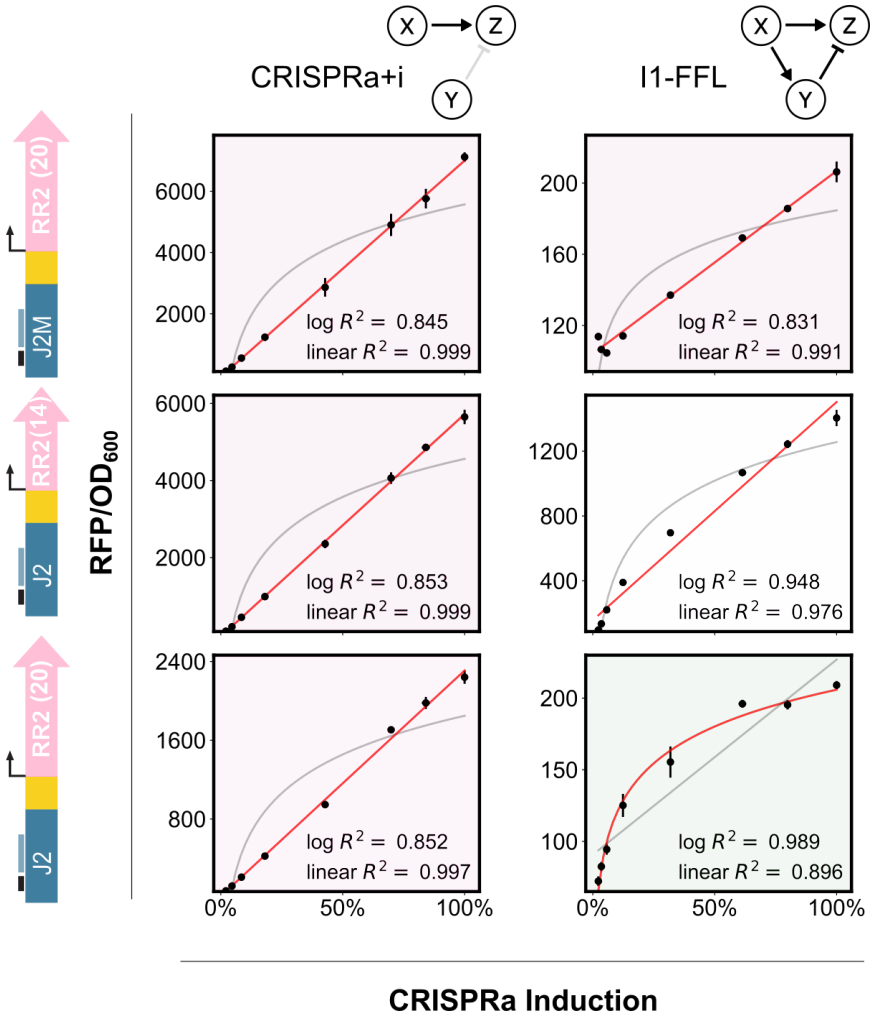


Figure S11: I1-FFL is a fold-change detector of CRISPRa inputs outside the linear range of CRISPRa response to aTc induction of MCP-SoxS

Data are replotted from figure 6A center, and right with the x-axis converted from nM aTc to %induction of the CRISPRa response, allowing observation of CRISPRa/i circuit response to increasing strength of CRISPRa outside of the linear range of pTet induction. For all tunings of CRISPRa+i, outputs respond linearly to CRISPRa (Left). For IFFLs, only specific tuning actions generate a logarithmic dependence on the input strength of CRISPRa (bottom right). Calculation for percent CRISPRa induction is provided in Methods 5.2.

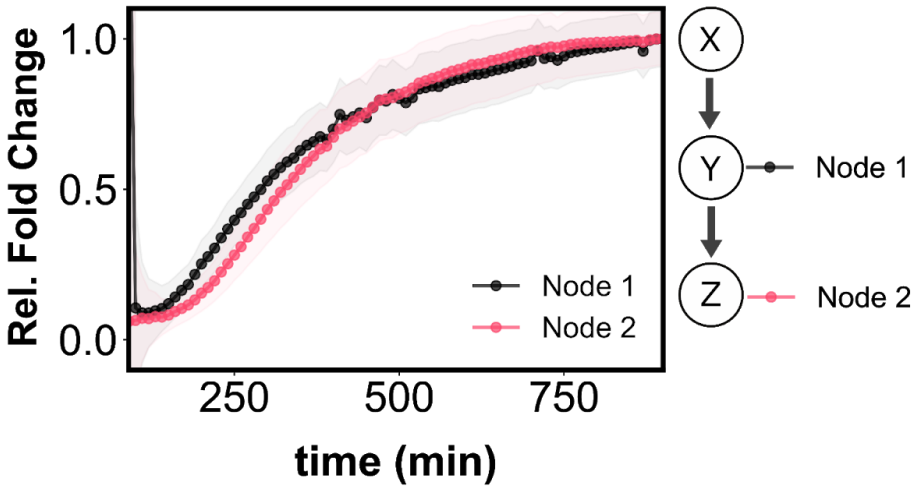


Figure S12: Speed of signal propagation CRISPRa vs. CRISPRa Cascade

Plotting of fold activation for CRISPRa and and CRISPRa cascades compared to a constitutive expression control, and scaling by their respective endpoints to lie on a common axis reveals that scRNA mediated internal layers of CRISPRa circuitry can be fast compared to the overall kinetics of CRISPRa, on the order of 20-30 minutes versus more than 6 hours for CRISPRa in a single layer (Figure 1D, right). Here, the black line corresponds to CRISPRa with 20 nM J306 scRNA plasmid targeting the J3 promoter to activate RFP. The red line corresponds to a CRISPRa cascade with 20 nM of J206 scRNA plasmid targeting 4 nM of J2 promoter expressing J306 scRNA for activation, directing CRISPRa to the J3 promoter for activation of RFP as shown in Figure 2B. Data are presented as the mean calculated quantity for three technical replicates with shading representing +/- one standard deviation. Calculation of relative fold change can be found in Methods 5.1.

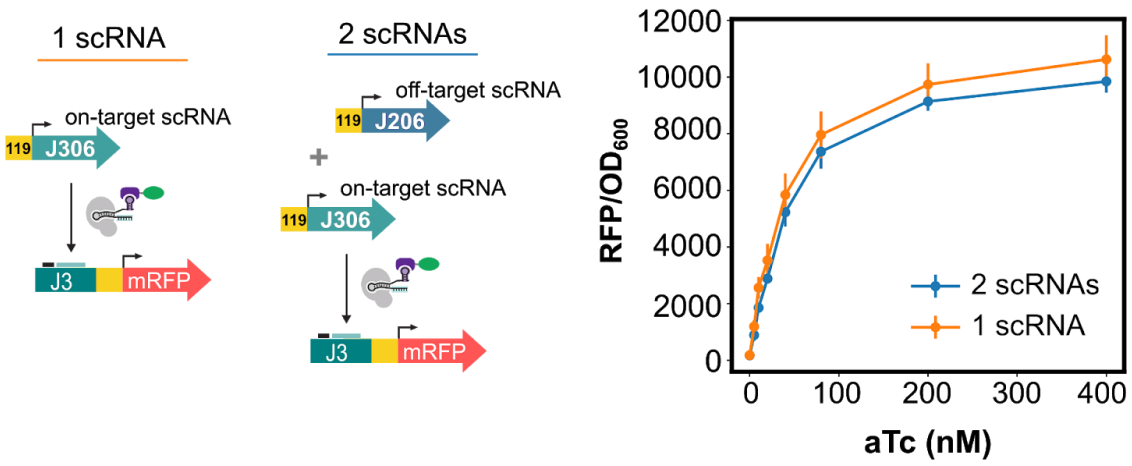


Figure S13: Expression of additional guide RNAs in *E. coli*

Titration curve of inducible CRISPRa with MCP-SoxS under control of the pTet promoter and induced by aTc (0-400 nM). In the 1 scRNA condition, J306 scRNA expressed from a strong constitutive promoter (BBa_J23119) was used for targeting the J3-RFP reporter. In the 2 scRNAs condition, both J306 and J206 scRNAs are expressed from a strong constitutive promoter (BBa_J23119). Data are presented as mean RFP/OD_{600} ± standard deviation for three biological replicates.

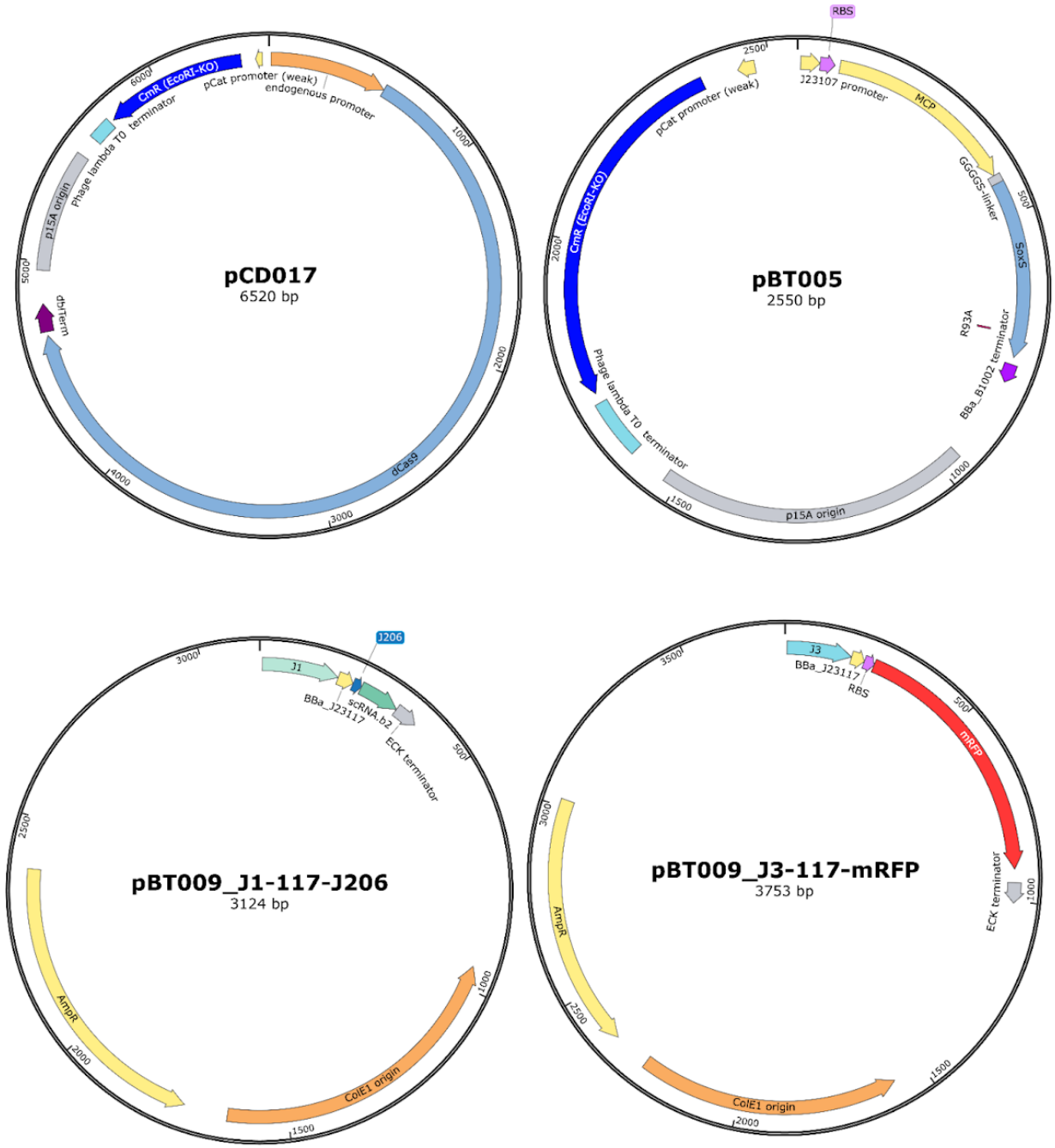
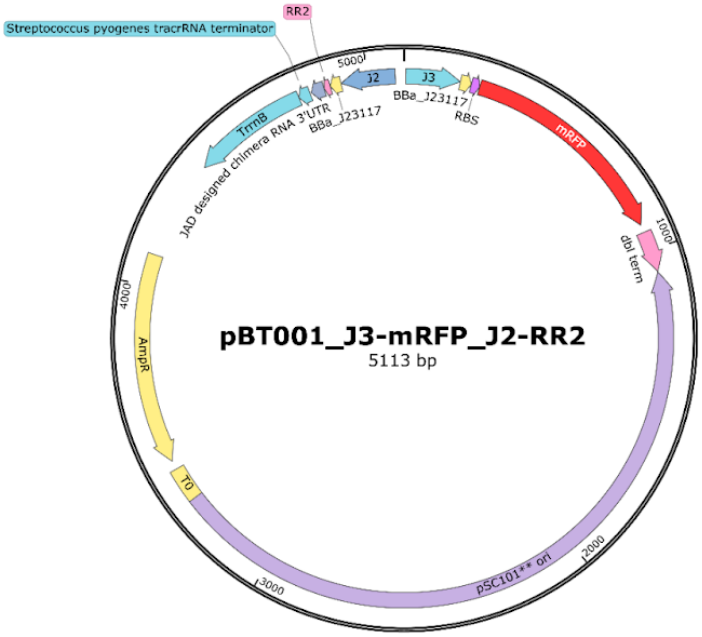


Figure S14: Representative plasmid maps for CRISPRa in CFS

Plasmids for expressing CRISPRa machinery in CFS (pCD017 for dCas9, pBT005 for MCP-SoxS, and pBT009_J1-117-J206), and reporter (pBT009_J3-117-mRFP). The promoter region driving guideRNA or fluorescent reporter as well as guideRNA identity

can be shuffled to specify individual network topologies in CFS. See Table S5 for further details.



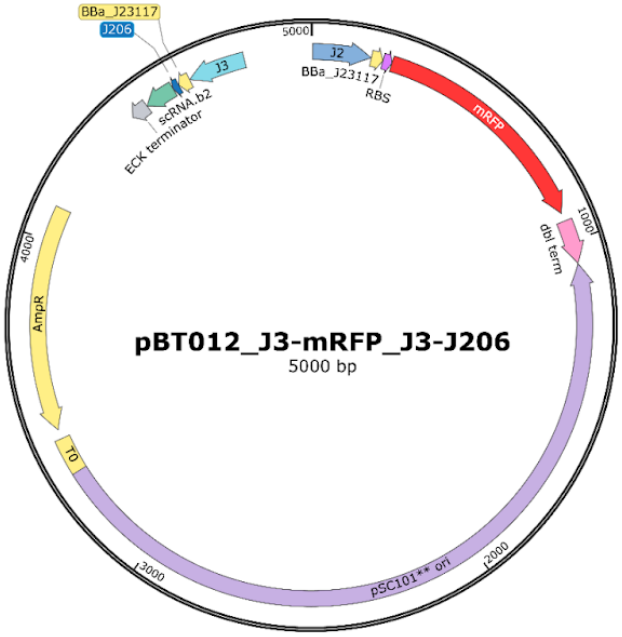


Figure S15: Representative plasmid maps for CRISPRa in *E. coli*

Plasmids for expressing CRISPRa machinery in E. coli consisted of reporter and 2nd layer sgRNA/scRNA cassette. The promoter region driving guideRNA or fluorescent reporter as well as guideRNA identity can be shuffled to specify individual network topologies. See Table S5 for further details.

Supplementary Tables

Table S1: CRISPR-mediated transcriptional circuit complexity demonstrated in this work

Name	Figure	gRNAs	Nodes	Connections	Host	Tuning variants
CRISPRa	1E, 2A_left, 2A_right, 2C_left, 3A (blue), S2, S3, S4, S5,S6,	1	2	1	CFS	N/A
CRISPRi	1E, 2C_right, 2E (black), 2F (black), S5 (blue)	1	2	1	CFS	N/A
CRISPRa+i	3A Orange	2	3	2	CFS	N/A
CRISPRa Cascade	2B, S12 (red)	2	3	2	CFS	N/A
CRISPRa/i Cascade	2D	2	3	2	CFS	N/A
I1-FFL	3A (red), 3B, 3C	2	3	3	CFS	N/A
I1-FFL/I1-FFL	3A Green	4	6	6	CFS	N/A

CRISPRa	4B_left, 5B, S9, S13 (black line)	1	2	1	*E. coli*	2
CRISPRa/i Cascade	5C, 6B_middle	2	3	2	*E. coli*	3
CRISPRa Cascade	4B_right	2	3	2	*E. coli*	3
CRISPRa+a	6A_Left, S10 (black)	3	3	2	*E. coli*	1
CRISPRa+i	6A_Middle, S10 (pink), S11 (left)	2	3	2	*E. coli*	3
I1-FFL	6A_Right, 6B_Bottom, S10 (red), S11 (right)	3	3	3	*E. coli*	3

Table S2: Existing CRISPRa/i circuit in bacterial and bacterial CFS

Article	**Novelty**	**CRISPR control**	**Focus[a]**
Bikard et. al. 2013	Individual actions in single layer	Activation, Repression	T
Dong et. al. 2018	Simultaneous actions in single layer	Activation, Repression	T
Liu et. al. 2019	Positive Autoregulation, CRISPRa in *K. oxytoca*	Activation	T
Santos- 2019	Layered CRISPRi, Incoherent type 2 feed forward loop	Repression	C
Moreno et. al.	Incoherent type 1 feed forward loop	Repression	C
Westbrook et. al. 2019	Autoinduction, multiple targets	Repression	T, M
Dinh and Prather 2019	Feedback, multiple targets	Repression	M
Wu et. al. 2020	Multiple targets	Repression	M
Tian et. al. 2020	Autoregulation	Repression	C

Huang et. al. 2021	Layered CRISPRi	Repression	B
Tan and Ng 2021	CRISPRa with dCas9-AsiA	Activation	T
Ho et. al. 2020	Simultaneous multi-gene CRISPRa and CRISPRi in *P. putida*	Activation, Repression	T, M
Kiattisewee et. al. 2021	dCas12a-SoxS CRISPRa in *Paenibacillus polymyxa*	Activation	T
Schilling et. al. 2020	CRISPRa requirements in *E. coli*	Activation	T

[a]Abbreviations for **Focus** column: Biosensing (B), Circuits (C), Metabolic engineering (M), and Tool Development (T)

Table S3: CFS CRISPRa/i Model Definition

Description	**Reaction**	**Rate**
Protein production (constitutive)	dCas9_DNA -> dCas9_DNA + dCas9_OFF	ktx1 · dCas9_DNA
Protein folding and maturation	dCas9_OFF -> dCas9_ON	kmat1 · dCas9_OFF
Protein production (constitutive)	SoxS_DNA -> SoxS_DNA + SoxS_OFF	ktx1 · SoxS_DNA

Protein folding and maturation	SoxS_OFF -> SoxS_ON	kmat2 · SoxS_OFF
Protein production (constitutive)	RFP_DNA -> RFP_DNA + RFP_OFF	ktx0 · RFP_DNA
Protein folding and maturation	RFP_OFF -> RFP_ON	kmat3 · RFP_OFF
scRNA production (constitutive)	scRNA_DNA -> scRNA_DNA + scRNA	ktx2 · scRNA_DNA
scRNA degradation	scRNA -> null	kdeg · scRNA
gRNA production (constitutive)	gRNA_DNA -> gRNA_DNA + gRNA	ktx0 · gRNA_DNA
gRNA degradation	gRNA -> null	kdeg · gRNA
CRISPRa association	dCas9_ON + SoxS_ON + scRNA -> CRISPRa	ka · dCas9_ON · SoxS_ON · scRNA
CRISPRi association	dCas9_ON + gRNA -> CRISPRi	ki · dCas9_ON · gRNA
CRISPRa binding	RFP_DNA + CRISPRa -> A_RFP_DNA	caskon · RFP_DNA · CRISPRa
Protein production (CRISPRa)	A_RFP_DNA -> A_RFP_DNA + RFP_OFF	kta1 · A_RFP_DNA
CRISPRa binding	gRNA_DNA + CRISPRa -> A_gRNA_DNA	caskon · gRNA_DNA ·

		CRISPRa
gRNA production (CRISPRa)	A_gRNA_DNA -> A_gRNA_DNA + gRNA	kta2 · A_gRNA_DNA
CRISPRi binding	RFP_DNA + CRISPRi -> I_RFP_DNA	caskon · RFP_DNA · CRISPRi
CRISPRi binding	A_RFP_DNA + CRISPRi -> I_RFP_DNA	caskon · A_RFP_DNA · CRISPRi
CRISPRi binding	I_RFP_DNA + CRISPRi -> I_RFP_DNA	caskon · I_RFP_DNA · CRISPRi

Table S4: Plasmids

#	Plasmid	J-Promoter	Minimal Promoter	RBS	CDS	gRNA	terminator	res *	Ori **
1	pBT005	X	J23107	Bujard	MCP-SoxS (R93A)	X	BBa_B1002	C	A
2	pCD017	X	Sp.Cas9	Sp.Cas9	dCas9	X	dbl term	C	A
3	pBT009. J3.RFP	J3	J23117	Bujard	mRFP1	X	ECK120033736	A	E
4	pBT009. J3.GFP	J3	J23117	10Ksf	sfGFP	X	ECK120033736	A	E
5	pBT009. J2.RFP	J2	J23117	Bujard	mRFP1	X	ECK120033736	A	E
6	pBT009. J2.GFP	J2	J23117	10Ksf	sfGFP	X	ECK120033736	A	E
7	pBT009. TJ1.119.RFP	TJ1	J23119	Bujard	mRFP1	X	ECK120033736	A	E
8	pBT009.	TJ1	J23119	10Ksf	sfGFP	X	ECK120033736	A	E

	TJ1.119.GFP								
9	pJF143.J3.J23112.RFP	J3	J23112	Bujard	mRFP1	X	dbl term	A	S
10	pJF143.J3.J23113.RFP	J3	J23113	Bujard	mRFP1	X	dbl term	A	S
11	pJF143.J3.J23109.RFP	J3	J23109	Bujard	mRFP1	X	dbl term	A	S
12	pJF143.J3.J23117.RFP	J3	J23117	Bujard	mRFP1	X	dbl term	A	S
13	pJF143.J3.J23114.RFP	J3	J23114	Bujard	mRFP1	X	dbl term	A	S
14	pJF143.J3.J23115.RFP	J3	J23115	Bujard	mRFP1	X	dbl term	A	S
15	pJF143.J3.J23105.RFP	J3	J23105	Bujard	mRFP1	X	dbl term	A	S
16	pJF143.J3.J23106.RFP	J3	J23106	Bujard	mRFP1	X	dbl term	A	S
17	pJF143.J3.J23107.RFP	J3	J23107	Bujard	mRFP1	X	dbl term	A	S

18	pJF143.J3.J23108.RFP	J3	J23108	Bujard	mRFP1	X	dbl term	A	S
19	pJF143.J3.J23110.RFP	J3	J23110	Bujard	mRFP1	X	dbl term	A	S
20	pJF143.J3.J23111.RFP	J3	J23111	Bujard	mRFP1	X	dbl term	A	S
21	pJF143.J3.J23118.RFP	J3	J23118	Bujard	mRFP1	X	dbl term	A	S
22	pJF143.J3.J23119.RFP	J3	J23119	Bujard	mRFP1	X	dbl term	A	S
23	pBT009.J1.117.J206	J1	J23117	X	X	J206	ECK120033736	A	E
24	pBT009.J1.119.J206	J1	J23119	X	X	J206	ECK120033736	A	E
25	pBT009.J1.117.J306	J1	J23117	X	X	J306	ECK120033736	A	E
26	pBT009.J1.119.J306	J1	J23119	X	X	J306	ECK120033736	A	E
27	pBT009.J2.117.J306	J2	J23117	X	X	J306	ECK120033736	A	E

28	pBT009.J3.117.J206	J3	J23117	X	X	J206	ECK120033736	A	E
29	pBT009.117.J1.RR2	J1	J23117	X	X	RR2	ECK120033736	A	E
30	pBT009.117.J2.RR2	J2	J23117	X	X	RR2	ECK120033736	A	E
31	pBT009.117.J3.RR2	J3	J23117	X	X	RR2	ECK120033736	A	E
32	pBT009.117.J1.sf1	J1	J23117	X	X	sf1	ECK120033736	A	E
33	pBT009.117.J2.sf1	J2	J23117	X	X	sf1	ECK120033736	A	E
34	pBT009.117.J3.sf1	J3	J23117	X	X	sf1	ECK120033736	A	E
35	pJF182.J306	X	J23119 J23107 Sp.Cas9	X Bujard Sp.Cas9	MCP-SoxS (R93A), dCas9	J306	dbl term, BBa_B1002, pCas9	C	A
36	pJF182.J306.J206	X	J23119 J23119 J23107 Sp.Cas9	X, X, BujardSp.Cas9	MCP-SoxS (R93A), dCas9	J206 J306	dbl term, dbl term, BBa_B1002, pCas9	C	A

37	pJF182. J206	X	J23119 J23107 Sp.Cas9	X,Bujard Sp.Cas9	MCP-SoxS (R93A), dCas9	J206 J306	dbl term, dbl term, BBa_B1002, pCas9	C	A
38	pCK085. AAV	X	J23119 J23107 Sp.Cas9	X,Bujard Sp.Cas9	MCP-SoxS (R93A,S10 1A), dCas9	AAV	dbl term, BBa_B1002, pCas9	C	A
39	pCK085. 206	X	J23119 J23107 Sp.Cas9	X,Bujard Sp.Cas9	MCP-SoxS (R93A,S10 1A), dCas9	J206	dbl term, BBa_B1002, pCas9	C	A
40	pCK085. 306	X	J23119 J23107 Sp.Cas9	X,Bujard Sp.Cas9	MCP-SoxS (R93A,S10 1A), dCas9	J306	dbl term, BBa_B1002, pCas9	C	A
41	pBT012.J3. J23105-206	J3 J2	J23105 J23117	X,Bujard	J206, RFP	J206	ECK12003373 6, dbl term	A	S

42	pBT012.J3.J23117-206	J3 J2	J23117 J23117	X,Bujard	J206, RFP	J206	ECK120033736, dbl term	A	S
43	pBT012. J3M. J23117.J206	J3 M J2	J23117 J23117	X,Bujard	J206, RFP	J206	ECK120033736, dbl term	A	S
44	pBT012.J3.J23117.J306	J3 J2	J23117 J23117	X,Bujard	J306, RFP	J306	ECK120033736, dbl term	A	S
45	pBT001	J2 J3	J23117 J23117	X,Bujard	RR2, RFP	RR2	Trrnb, dbl term	A	S
46	pBT001. RR2-14	J2 J3	J23117 J23117	X,Bujard	RR2-14, RFP	RR2-14	Trrnb, dbl term	A	S
47	pBT001. n26-3	J2 M J3	J23117 J23117	X,Bujard	RR2, RFP	RR2	Trrnb, dbl term	A	S
48	pBT002	J2 J3	J23117 J23117	X,Bujard	AAV, RFP	AAV	Trrnb, dbl term	A	S
49	pBT001. 119RFP	J2 J3	J23117 J23119	X,Bujard	RR2, RFP	RR2-14	Trrnb, dbl term	A	S

50	pBT001. RR2-14. 119RFP	J2 J3	J23117 J23119	X,Bujard	RR2-14, RFP	RR2-1 4	Trrnb, dbl term	A	S
51	pBT001 .n26-3. 119RFP	J2 M J3	J23117 J23119	X,Bujard	RR2, RFP	RR2	Trrnb, dbl term	A	S
52	pBT002. 119-RFP	J2 J3	J23117 J23119	X,Bujard	AAV RFP	AAV	Trrnb, dbl term	A	S
53	pJF144.J30 6	X	J23119 J23107 Sp.Cas9	X,Bujard Sp.Cas9	MCP-SoxS (R93A), dCas9	J306	dbl term, BBa_B1002, pCas9	C	A
54	pJF144.AAV	X	J23119 J23107 Sp.Cas9	X,Bujard Sp.Cas9	MCP-SoxS (R93A), dCas9	AAV	dbl term, BBa_B1002, pCas9	C	A

*Resistance marker is denoted Res in table S4. C stands for chloramphenicol, A stands for ampicillin

Origin of replication is denoted Ori in Table S4. E stands for ColE1, A stands for p15A, and S stands for sc101

Table S5: Parts

#	part name	sequence
1	dCas9	ATGGATAAGAAATACTCAATAGGCTTAGcTATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATA AGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCT TTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAG AATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTG AAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGC TTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGC TTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATA ATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAG CTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAAT CAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATT GGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGAT ATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATC AAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATC AAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTA GAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCC ATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCG CGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTG TCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGT ACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAA GGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAA AAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGT TGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGAT AATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGG AAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTG GGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAA TCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAA AAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAA AGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTT

		ATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAG AAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCT CTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTAT GATGTCGATgcCATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATA AAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCT AAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGAT AAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTC GCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGT TTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTAT CTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATA AAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTA CTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAA ACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGC CCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAA TTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCT TATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGA TCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAA AAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGT GCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATT ATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGA TGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGT GCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATC TTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTT AGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGAC
2	MCP_SoxS_R93A	atggggcccgcttctaactttactcagttcgttctcgtcgacaatggcggaactggcgacgtgactgtcgccccaa gcaacttcgctaacgggatcgctgaatggatcagctctaactcgcgttcacaggcttacaaagtaacctgtagcgt tcgtcagagctctgcgcagaatcgcaaatacaccatcaaagtcgaggtgcctaaaggcgcctggcgttcgtactta aatatggaactaaccattccaattttcgccacgaattccgactgcgagcttattgttaaggcaatgcaaggtctcc taaaagatggaaacccgattccctcagcaatcgcagcaaactccggcatctacGGTGGCGGAGGTAGCATGTCCCA TCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGCTTAACATTGATGTAGTCGCA AAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAATGTTCCGCACGGTGACGCATCAGACGCTTGGCGATTACA TTCGCCAACGCCGCCTGTTACTGGCCGCCGTTGAGTTGCGCACCACCGAGCGTCCGATTTTTGATATCGCAATGGA CCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGCGTTTTCGCGCGGCAGTTTGATCGCACTCCCGCGGATTATCGC CACCGCCTGTAA

3	J1	GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTGACCTAT GGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCGTCCTTTGGGTTC CACCGGATACCTCCGGAC
4	J2	CGAGAACAGTATGCGCGACTCGCAAAGTCTAGCTCGCGCAGTGAAGAATCGACGTGTTTCTAATTCGACGCTACTT GAGTGAGCACTGAAAGCGCGCTGTCGCGCGACTGACCTGACGCATCCTGAGGACGTGTTCGGCTACTACACAAGTA TTAAGAGACAATGCGCTC
5	J3	AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCTGTGAAA GATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGACACAACGTCGTCT TGAAGTTGCGATTATAGA
6	J206 scRNA.b2	TAGTAGCCGAACACGTCCTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACATGAGGATCACCCATGTGCTTTTTTT
7	J306 scRNA.b2	TTGTGTCCAGAACGCTCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACATGAGGATCACCCATGTGCTTTTTTT
8	Bujard RBS	GAATTCATTAAAGAGGAGAAAGGTACC
9	sf10KRBS	TGTTTCGCGAGGGTAAGGGCTTGGTTTT
10	sfGFP	ATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACA AATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGG AAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGAT CACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAG ATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGG TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATC ACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAAC TAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGAC ACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGG ATTACACATGGCATGGATGAGCTCTACAAATAA
11	mRFP1	ATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATgcgtttcaaagttcgtatggaaggttccgttaacggtcacg agttcgaaatcgaaggtgaaggtgaaggtcgtccgtacgaaggtacccagaccgctaaactgaaagttaccaaagg tggtccgctgccgttcgcttgggacatcctgtccccgcagttccagtacggttccaaagcttacgttaaacacccg gctgacatcccggactacctgaaactgtccttcccggaaggtttcaaatgggaacgtgttatgaacttcgaagacg gtggtgttgttaccgttacccaggactcctccctgcaagacggtgagttcatctacaaagttaaactgcgtggtac caacttcccgtccgacggtccggttatgcagaaaaaaaccatgggttgggaagcttccaccgaacgtatgtacccg

		gaagacggtgctctgaaaggtgaaatcaaaatgcgtctgaaactgaaagacggtggtcactacgacgctgaagtta aaaccacctacatggctaaaaaaccggttcagctgccgggtgcttacaaaaccgacatcaaactggacatcacctc ccacaacgaagactacaccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccggtgcttaa
12	RR2 sgRNA	TGGAACCGTACTGGAACTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTT
13	sf1 sgRNA	CATCTAATTCAACAAGAATTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTT
14	hs1_AAV sgRNA	GGGGCCACTAGGGACAGGATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTT
15	hs1_AAV scRNA	GGGGCCACTAGGGACAGGATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACATGAGGATCACCCATGTGCTTTTTTT
16	BBa_J23117	ttgacagctagctcagtcctagggattgtgctagc
17	BBa_J23119	ttgacagctagctcagtcctaggtataatgctagc
18	BBa_J23112	ctgatagctagctcagtcctagggattatgctagc
19	BBa_J23113	ctgatggctagctcagtcctagggattatgctagc
20	BBa_J23109	tttacagctagctcagtcctagggactgtgctagc
21	BBa_J23114	tttatggctagctcagtcctaggtacaatgctagc
22	BBa_J23115	tttatagctagctcagcccttggtacaatgctagc
23	BBa_J23105	tttacggctagctcagtcctaggtactatgctagc
24	BBa_J23106	tttacggctagctcagtcctaggtatagtgctagc
25	BBa_J23107	tttacggctagctcagccctaggtattatgctagc
26	BBa_J23108	ctgacagctagctcagtcctaggtataatgctagc
27	BBa_J23110	tttacggctagctcagtcctaggtacaatgctagc
28	BBa_J23111	ttgacggctagctcagtcctaggtatagtgctagc
29	BBa_J23118	ttgacggctagctcagtcctaggtattgtgctagc

30	ECK120033736 terminator	aacgcatgagAAAGCCCCCGGAAGATCACCTTCCGGGGGCTTTtttattgcgc
31	dbl term	CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGC TCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA
32	TrrnB	CAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGAAGCTTGGGCCCGAACAAAAACTCatctcagaagagga tctgaatagcgccgtcgaccatcatcatcatcatcattgagtttaaacggtctccagcttggctgttttggcggat gagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcgg cagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtgggg tctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactGGGCCTTTCGT TTTATCTGTTGTTTGTCGGTGAACT
33	3'UTR mRFP1	ggatccaaactcgagtaaggatctGTGCTTTTTTT
34	3'UTR sfGFP	GTGCTTTTTTT
35	*Sp*.pCas9 (endogenous promoter of *S. pyogenes* Cas9)	TTACGAAATCATCCTGTGGAGCTTAGTAGGTTTAGCAAGATGGCAGCGCCTAAATGTAGAATGATAAAAGGATTAA GAGATTAATTTCCCTAAAAATGATAAAACAAGCGTTTTGAAAGCGCTTGTTTTTTTTGGTTTGCAGTCAGAGTAGAA TAGAAGTATCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCT ATGCTGTTTTGAATGGTTCCAACAAGATTATTTTATAACTTTTATAACAAATAATCAAGGAGAAATTCAAAGAAAT TTATCAGCCATAAAACAATACTTAATACTATAGAATGATAACAAAATAAACTACTTTTTAAAAGAATTTTGTGTTA TAATCTATTTATTATTAAGTATTGGGTAATATTTTTTGAAGAGATATTTTGAAAAAGAAAAATTAAAGCATATTAA ACTAATTTCGGAGGTCATTAAAACTATTATTGAAATCATCAAACTCATTATGGATTTAATTTAAACTTTTTATTTT AGGAGGCAAAA
36	dCas9 terminator_dbl term	TAACtcgagtaaggatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctg ttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata
37	TJ1	GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTGACCTAT GGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCGTCCTTTGGGTTT CCCTATCAGTGATAGAGA
38	B1002 terminator	GCGGCCGCcacgcaaaaaaccccgcttcggcggggttttttcgc
39	RR2-14	CGTACTGGAACTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG

		CACCGAGTCGGTGCTTTTTT
40	J2M	CCGAGAACAGTATGCGCGACTCGCAAAGTCTAGCTCGCGCAGTGAAGAATCGACGTGTTTCTAATTCGACGCTACT TGAGTGAGCACTGAAAGCGCGCTGTCGCGCGACTGACCTGACGCATCCTGAGGACGTGTTCGGCTACTACTCACGC AGGCTACTGCGAGTCCGGA
41	J3M	AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCTGTGAAA GATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGACACAACTCACGCA GGCTACTGCGAGTCCGGA
42	TetR:pTet:MCP-SoxS_R93A	tccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacg ctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttataCCTAGagcgcccggaagagagt caattcagggtggtgaatctgcagTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAAT TCAAGGCCGAATAAGAAGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCA TACTATCAGTAGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGT AAAATGCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAAAGGCTAATTGA TTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGCTCCATCGCGATGACTTAGTA AAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAAGGGCAAAAGTGAGT ATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTTTTACATGCCAATACAATGTA GGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGTTAAACCTTCGATTCCGACCTCATTAAGCAGCTCTA ATGCGCTGTTAATCACTTTACTTTTATCTAATCTAGACATCATTAATTCCTAATTTTTGTTGACACTCTATCGTTG ATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGAATTCATTAAAGAGGAGAAAGGTACCatggggccc gcttctaactttactcagttcgttctcgtcgacaatggcggaactggcgacgtgactgtcgccccaagcaacttcg ctaacgggatcgctgaatggatcagctctaactcgcgttcacaggcttacaaagtaacctgtagcgttcgtcagag ctctgcgcagaatcgcaaatacaccatcaaagtcgaggtgcctaaaggcgcctggcgttcgtacttaaatatggaa ctaaccattccaattttcgccacgaattccgactgcgagcttattgttaaggcaatgcaaggtctcctaaaagatg gaaacccgattccctcagcaatcgcagcaaactccggcatctacGGTGGCGGAGGTAGCATGTCCCATCAGAAAAT TATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGCTTAACATTGATGTAGTCGCAAAAAAATCA GGCTATTCAAAGTGGTACTTGCAACGAATGTTCCGCACGGTGACGCATCAGACGCTTGGCGATTACATTCGCCAAC GCCGCCTGTTACTGGCCGCCGTTGAGTTGCGCACCACCGAGCGTCCGATTTTTGATATCGCAATGGACCTGGGTTA TGTCTCGCAGCAGACCTTCTCCCGCGTTTTCGCGCGGCAGTTTGATCGCACTCCCAGCGATTATCGCCACCGCCTG TAAGCGGCCGCcacgcaaaaaaccccgcttcggcggggttttttcgc
43	E.coli scRNA promoter, J23119(SpeI)+f	tgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgttactagATTGACAGCTAGC TCAGTCCTAGGTATAATACTAGT

	lanking sequence	
44	*E. coli* scRNA terminator	GAAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATCAT TGAGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGA ACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAC TCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAA ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACT
44	*E. coli* constitutive J23107:MCP-SoxS_R93A,S101A	agcgcccggaagagagtcaattcagggtggtgaatctgcagtttacggctagctcagccctaggtattatgctagc GAATTCATTAAAGAGGAGAAAGGTACCatggggcccgcttctaactttactcagttcgttctcgtcgacaatggcg gaactggcgacgtgactgtcgccccaagcaacttcgctaacgggatcgctgaatggatcagctctaactcgcgttc acaggcttacaaagtaacctgtagcgttcgtcagagctctgcgcagaatcgcaaatacaccatcaaagtcgaggtg cctaaaggcgcctggcgttcgtacttaaatatggaactaaccattccaattttcgccacgaattccgactgcgagc ttattgttaaggcaatgcaaggtctcctaaaagatggaaacccgattccctcagcaatcgcagcaaactccggcat ctacGGTGGCGGAGGTAGCATGTCCCATCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGAC CAGCCGCTTAACATTGATGTAGTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAATGTTCCGCACGG TGACGCATCAGACGCTTGGCGATTACATTCGCCAACGCCGCCTGTTACTGGCCGCCGTTGAGTTGCGCACCACCGA GCGTCCGATTTTTGATATCGCAATGGACCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGCGTTTTCGCGCGGCAG TTTGATCGCACTCCCGCGGATTATCGCCACCGCCTGTAAGCGGCCGCcacgcaaaaaaccccgcttcggcggggtt ttttcgc

Table S6: Plasmid breakdown per figure

Figure	**Plasmid**	**Strain**
1D	pBT005,pCD017, pBT009.J3.117.RFP, pBT009.J1.117.J306	TXTL
1E	pCD017,pBT009.TJ1.119.GFP, pBT009.J1.117.SF1	TXTL
1F	pBT005,pCD017, pBT009.J3.117.RFP, pBT009.J1.117.J306, pBT009.TJ1.119.RFP, pBT009.J1.117.RR2	TXTL
2A	pBT005, pCD017, pBT009.J1.117.J206, pBT009.J1.117.J306, pBT009.J2.117.RFP, pBT009.J3.117.RFP	TXTL
2B	pBT005, pCD017, pBT009.J1.117.J206, pBT009.J2.117.J306, pBT009.J3.117.RFP	TXTL
2C	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J3.117.RFP, pBT009.J3.117.RR2, pBT009.TJ1.119.RFP	TXTL
2D,E,F	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J3.117.RR2, pBT009.TJ1.119.RFP	TXTL

3A_Blue	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J3.117.RFP	TXTL
3A_Orange	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J3.117.RFP, pBT009.J1.117.RR2	TXTL
3A_Red	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J3.117.RFP, pBT009.J3.117.RR2	TXTL
3A_Green	pBT005, pCD017, pBT009.J1.117.J206, pBT009.J3.117.RFP, pBT009.J3.117.RR2, pBT009.J1.117.J306, pBT009.J2.117.SF1, pBT009.J2.117.GFP	TXTL
3B,C	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J3.117.RR2, pBT009.J3.117.RFP	TXTL
4B_inducible	pJF182.J306, pJF143.J3	MG1655
4B_constitutive	pCK085.J306, pJF143.J3	MG1655
4C_left	pJF182.J206, pBT001.J2.RR2.J3.119.RFP	MG1655
4C_Center	pJF182.J206, pBT001.J2.117.RR2-14.J3.119.RFP	MG1655
4C_Right	pJF182.J206, pBT001.J2M.117.RR2.J3.119.RFP	MG1655

4D_Left_CRISPRa O/T	pCK085.J306, pBT012.J3.J306.J2.RFP	MG1655
4D_Left_CRISPRa	pCK085.J206, pBT012.J3.117.J306.J2.117.RFP	MG1655
4D_middle left_OT	pCK085.HS1, pBT012.J3.105.J206.J2.117.RFP	MG1655
4D_middle left_on	pCK085.J306, pBT012.J3.105.J206.J2.117.RFP	MG1655
4Dmiddle right_OT	pCK085.HS1, pBT012.J3.117.J206.J2.117RFP	MG1655
4Dmiddle right_on	pCK085.J306, pBT012.J3.117.J206.J2.117RFP	MG1655
4D_right_OT	pCK085.HS1, pBT012.J3M.117.J206.J2.117.RFP	MG1655
4D_right_on	pCK085.306, pBT012.J3M.117.J206.J2.117RFP	MG1655
5A_Top_Left	pJF182.J206.J306, pBT002	MG1655
5A_Middle_Left	pJF182.J206.J306, pBT002	MG1655
5A_Bottom_Left	pJF182.J206.J306, pBT002	MG1655

5A_Top_Middle	pJF182.J306, pBT001.J2M.117.RR2.J3.117.RFP	MG1655
5A_Middle_Middle	pJF182.J306, pBT001.J2.117.RR2-14.J3.117.RFP	MG1655
5A_Bottom_Middle	pJF182.J306, pBT001.J2.117.RR2.J3.117.RFP	MG1655
5A_Top_Right	pJF182.J206.J306, pBT001.J2M.117.RR2.J3.117.RFP	MG1655
5A_Middle_Right	pJF182.J206.J306, pBT001.J2.117.RR2-14.J3.117.RFP	MG1655
5A_Bottom_Right	pJF182.J206.J306, pBT001.J2.117.RR2.J3.117.RFP	MG1655
5B_Top	pJF182.J306, pJF143.J3	MG1655
5B_Middle	pJF182.J206, pBT001.J2.117.RR2.J3.117.RFP	MG1655:RFP
5B_Bottom	pJF182.J206.J306, pBT001.J2.117.RR2.J3.117.RFP	MG1655
S2	pCD017, pBT005, pBT009.J1.117.J306, pBT009.J1.117.J206,	TXTL

	pJF143.J3.J23113.RFP, pJF143.J3.J23112.RFP, pJF143.J3.J23109.RFP, pJF143.J3.J23117.RFP, pJF143.J3.J23114.RFP, pJF143.J3.J23105.RFP, pJF143.J3.J23107.RFP, pJF143.J3.J23115.RFP, pJF143.J3.J23107.RFP, pJF143.J3.J23106.RFP, pJF143.J3.J23108.RFP, pJF143.J3.J23118.RFP, pJF143.J3.J23110.RFP, pJF143.J3.J23111.RFP	
S3	pCD017, pBT005, pBT009.J1.117.J306, pBT009.J1.117.J206, pJF144.J306, pJF144.hs1AAV, pJF143.J3.J23113.RFP, pJF143.J3.J23112.RFP, pJF143.J3.J23109.RFP, pJF143.J3.J23117.RFP, pJF143.J3.J23114.RFP,	TXTL, *E. coli*

	pJF143.J3.J23105.RFP, pJF143.J3.J23107.RFP, pJF143.J3.J23115.RFP, pJF143.J3.J23107.RFP, pJF143.J3.J23106.RFP, pJF143.J3.J23108.RFP, pJF143.J3.J23118.RFP, pJF143.J3.J23110.RFP, pJF143.J3.J23111.RFP	
S4	pBT005, pCD017, pBT009.J3.117.RFP, pBT009.J1.117.J306	TXTL
S5	pBT005, pCD017, pBT009.J3.117.RFP, pBT009.J1.117.J306, pBT009.J1.117.RR2, pBT009.TJ1.119.RFP,	TXTL
S6	pBT005, pCD017, pBT009.J3.117.RFP, pBT009.J1.119.J306	TXTL
S7	pBT005, pCD017, pBT009.J1.117.J206, pBT009.J2.117.RFP, pBT009.J2.117.RR2, pBT009.J1.117.J306, pBT009.J3.117.SF1, pBT009.J3.117.GFP	TXTL
S8	pBT005, pCD017, pBT009.J3.117.RFP,	TXTL

	pBT009.J3.117.RR2, pBT009.J1.117.J306	
S9	pJF143.J2.117.RFP, pJF182.J206	*E. coli*
S10	pJF182.J306, pJF182.J206.J306, pBT002, pBT001.RR2-14, pBT001, pBT001.N26-3	*E. coli*
S11	pJF182.J306, pJF182.J206.J306, pBT002, pBT001.RR2-14, pBT001, pBT001.N26-3	*E. coli*
S12	pBT005, pCD017, pBT009.J1.117.J306, pBT009.J1.117.J206, pBT009.J2.117.J306, pBT009.J3.117.RFP, pBT009.J2.117.RFP	TXTL

Fig #	Plasmid #1	Conc #1	Plasmid #2	Conc #2	Plasmid #3	Conc #3	Plasmid #4	Conc #4	Plasmid #5	Conc #5	Plasmid #6	Conc #6	Plasmid #7	Conc #7	Plasmid #8	Conc #8
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	0						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	0.1						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	0.5						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	1						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	2						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	4						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	8						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	16						
1D, S4	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	32						
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	0								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	0.01								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	0.1								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	0.5								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	1								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	2								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	4								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	8								
1E	pBT009.TJ1.119.RFP	0.25	pCD017	4	pBT009.TJ1.119.GFP	1	pBT009.J3.117.sf1	16								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J3.117.RR2	0								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	0.01								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	0.1								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	0.5								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	1								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	2								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	4								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	8								
S4	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT007.J2.117.RR2	16								
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	0						
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	0.1						

2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	0.5							
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	1							
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	2							
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	4							
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	8							
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	16							
2A_Left, 2C_left	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J3.117.RFP	4	pBT009.J1.J306	32							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	0							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	0.1							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	0.5							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	1							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	2							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	4							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	8							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	16							
2A,Right	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J2.117.RFP	4	pBT009.J1.J206	32							
S12	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	20	pBT009.J3.J206	0							
S12	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	20	pBT009.J3.J206	4							
S12	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	20									
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	20	pBT009.J3.J206	.5							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	20	pBT009.J3.J206	1							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	20	pBT009.J3.J206	2							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	20	pBT009.J3.J206	4							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	0	pBT009.J3.J206	.5							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	0	pBT009.J3.J206	1							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	0	pBT009.J3.J206	2							
2B	pBT005	4	pCD017	2	pBT009.J2.RFP	4	pBT009.J1.J306	0	pBT009.J3.J206	4							
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	0									

2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	0.01									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	0.1									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	0.5									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	1									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	2									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	4									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	8									
2C_Right	pBT009.TJ1.119.GFP	0.25	pCD017	4	pBT009.TJ1.119.RFP	1	pBT009.J3.117.RR2	16									
2D,2E,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	0					
2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	0.1					
2D,2E,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	0.5					
2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	1					
2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	2					
2E,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	4					
2D,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	8					
2E,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J206	16	pBT009.J3.117.RR2	16					
2D,2E,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J306	16	pBT009.J3.117.RR2	0.5					
2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J306	16	pBT009.J3.117.RR2	1					
2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J306	16	pBT009.J3.117.RR2	2					
2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J306	16	pBT009.J3.117.RR2	4					
2D,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J306	16	pBT009.J3.117.RR2	8					
2E,2F	pCD017	4	pBT005	4	pBT009.TJ1.119.RFP	1	pBT009.TJ1.GFP	0.5	pBT009.J1.117.J306	16	pBT009.J3.117.RR2	16					
3A	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.25							
3A	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J1.RR2	0.25							
3A	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16									
3A	pBT005	4	pCD017	2	pBT009.J3.RFP	4											
3A, S7	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.25	pBT009.J1.J206	20	pBT009.J2.SF1	0.5	pBT009.J2.GFP_RBSmax	10	
S7	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J1.RR2	0.25	pBT009.J1.J206	20	pBT009.J2.SF1	0.5	pBT009.J2.GFP_RBSmax	10	
S7	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.25	pBT009.J1.J206	20	pBT009.J1.SF1	0.5	pBT009.J2.GFP_RBSmax	10	
S7	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J1.RR2	0.25	pBT009.J1.J206	20	pBT009.J1.SF1	0.5	pBT009.J2.GFP_RBSmax	10	
S7	pBT005	4	pCD017	2	pBT009.J1.J306	16	pBT009.J1.J206	20									
S7	pBT005	4	pCD017	2	pBT009.J1.J306	16	pBT009.J1.J206	20									

S7	pBT005	4	pCD017	2	pBT009.J1.J306	16	pBT009.J1.J206	20									
S7	pBT005	4	pCD017	2	pBT009.J1.J306	16	pBT009.J1.J206	20									
S7	pBT005	4	pCD017	2	pBT009.J1.J306	16	pBT009.J1.J206	20									
3C	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0							
3B, 3C, S8	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.1							
3C	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.25							
3C	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.5							
3C	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	0.75							
3B, 3C, S8	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	1							
3C	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	2							
3C	pBT005	4	pCD017	2	pBT009.J3.RFP	4	pBT009.J1.J306	16	pBT009.J3.RR2	4							
S2, S3	pJF143.J3.J23112	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23113	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23109	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23117	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23114	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23115	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23105	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23106	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23107	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23108	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23110	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23111	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23118	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J206	12							
S2, S3	pJF143.J3.J23112	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							
S2, S3	pJF143.J3.J23113	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							
S2, S3	pJF143.J3.J23109	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							
S2, S3	pJF143.J3.J23117	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							
S2, S3	pJF143.J3.J23114	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							
S2, S3	pJF143.J3.J23115	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							
S2, S3	pJF143.J3.J23105	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12							

S2, S3	pJF143.J3.J23106	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12						
S2, S3	pJF143.J3.J23107	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12						
S2, S3	pJF143.J3.J23108	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12						
S2, S3	pJF143.J3.J23110	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12						
S2, S3	pJF143.J3.J23111	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12						
S2, S3	pJF143.J3.J23118	3	pCD017	4	pBT005	4	pBT009.TJ1.119.GFP	0.25	pBT009.J1.119.J306	12						
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	0								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	0.01								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	0.05								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	0.25								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	1								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	2								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	4								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	8								
S5 Left	pBT005	4	pBT009.J3.RFP	4	pBT009.J1.J306	8	pCD017	16								
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	0						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	0.1						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	1						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	2						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	4						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	8						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	16						
S5 right	pBT009.TJ1.GFP	0.25	pCD017	4	pBT009.J3.RFP	4	pBT009.J1.117.J306	8	pBT005	24						
S6	pCD017	4	pBT005	4	pBT009.TJ1.GFP	0.25	pBT009.J3.RFP	4	pBT009.J1.119.J306	0						
S6	pCD017	4	pBT005	4	pBT009.TJ1.GFP	0.25	pBT009.J3.RFP	4	pBT009.J1.119.J306	0						
S6	pCD017	4	pBT005	4	pBT009.TJ1.GFP	0.25	pBT009.J3.RFP	4	pBT009.J1.119.J306	0.01						
S6	pCD017	4	pBT005	4	pBT009.TJ1.GFP	0.25	pBT009.J3.RFP	4	pBT009.J1.119.J306	0.1						
S6	pCD017	4	pBT005	4	pBT009.TJ1.GFP	0.25	pBT009.J3.RFP	4	pBT009.J1.119.J306	1						
S6	pCD017	4	pBT005	4	pBT009.TJ1.GFP	0.25	pBT009.J3.RFP	4	pBT009.J1.119.J306	10						

Supplementary References

Adler, M., Szekely, P., Mayo, A., and Alon, U. (2017). Optimal Regulatory Circuit Topologies for Fold-Change Detection. Cell Syst. *4*, 171-181.e8.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. *41*, 7429–7437.

Dinh, C.V., and Prather, K.L.J. (2019). Development of an autonomous and bifunctional quorum-sensing circuit for metabolic flux control in engineered *Escherichia coli*. Proc. Natl. Acad. Sci. *116*, 25562–25568.

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

Fontana, J., Voje, W.E., Zalatan, J.G., and Carothers, J.M. (2018). Prospects for engineering dynamic CRISPR–Cas transcriptional circuits to improve bioproduction. J. Ind. Microbiol. Biotechnol. *45*, 481–490.

Fontana, J., Dong, C., Kiattisewee, C., Chavali, V.P., Tickman, B.I., Carothers, J.M., and Zalatan, J.G. (2020). Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. Nat. Commun. *11*, 1618.

Ho, H., Fang, J.R., Cheung, J., and Wang, H.H. (2020). Programmable CRISPR-Cas transcriptional activation in bacteria. Mol. Syst. Biol. *16*.

Huang, H.-H., Bellato, M., Qian, Y., Cárdenas, P., Pasotti, L., Magni, P., and Del Vecchio, D. (2021). dCas9 regulator to neutralize competition in CRISPRi circuits. Nat. Commun. *12*, 1692.

Kiattisewee, C., Dong, C., Fontana, J., Sugianto, W., Peralta-Yahya, P., Carothers, J.M., and Zalatan, J.G. (2021). Portable bacterial CRISPR transcriptional activation enables metabolic engineering in Pseudomonas putida. Metab. Eng. *66*, 283–295.

Liu, Y., Zeng, Y., Liu, L., Zhuang, C., Fu, X., Huang, W., and Cai, Z. (2014). Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. Nat. Commun. *5*, 5393.

Santos-Moreno, J., and Schaerli, Y. (2019). Multistable and dynamic CRISPRi-based synthetic circuits. BioRxiv.

Tan, S.-I., and Ng, I.-S. (2021). CRISPRi-Mediated NIMPLY Logic Gate for Fine-Tuning the Whole-Cell Sensing toward Simple Urine Glucose Detection. ACS Synth. Biol. *10*, 412–421.

Tian, J., Yang, G., Gu, Y., Sun, X., Lu, Y., and Jiang, W. (2020). Developing an endogenous quorum-sensing based CRISPRi circuit for autonomous and tunable dynamic regulation of multiple targets in Streptomyces. Nucleic Acids Res. *48*, 8188–8202.

Westbrook, A., Tang, X., Marshall, R., Maxwell, C.S., Chappell, J., Agrawal, D.K.,

Dunlop, M.J., Noireaux, V., Beisel, C.L., Lucks, J., et al. (2019). Distinct timescales of RNA regulators enable the construction of a genetic pulse generator. Biotechnol. Bioeng. *116*, 1139–1151.

Wu, Y., Chen, T., Liu, Y., Tian, R., Lv, X., Li, J., Du, G., Chen, J., Ledesma-Amaro, R., and Liu, L. (2020). Design of a programmable biosensor-CRISPRi genetic circuits for dynamic and autonomous dual-control of metabolic flux in Bacillus subtilis. Nucleic Acids Res. *48*, 996–1009.

Engineering CRISPRa/i components for increased dynamic ranges.

These unpublished data and experiments, and analyses were collected, planned and conducted by myself, Diego Alba Burbano, Ryan Cardiff, Cholpisit Kiattisewee, and Cassandra Maranas under the supervision of James Carothers and Jesse Zalatan. The starting point for this work was generated from analysis conducted by David Sparkman Yaeger on unpublished data collected by Jason Fontana. The need for this exploration was Identified in collaboration with Diego Alba Burbano during the development of CRISPRa/i circuits in CFS and E.coli (Submitted).

Introduction

CRISPRi-based regulation of gene expression is already well established as a versatile tool for multi-gene regulation(Clamons and Murray, 2019; Gander et al., 2017; Huang et al., 2020; Kiani et al., 2014; Landberg et al., 2020; Li et al., 2020). The MCP-SoxS based CRISPRa system has emerged as the tool of choice to extend the versatility of CRISPR-based regulation to activation in prokaryotes(Dong et al., 2018; Fontana et al., 2020; Kiattisewee et al., 2021). Controllers leveraging both activation and repression provide several advantages including faster response times for layered actions in-vivo, decreased basal costs of operation, and favorable scaling of component expression requirements with increased network size as compared to those which are only capable of repression(Alon, 2007; Clamons and Murray, 2019).

Despite the demonstrated ability of CRISPRa circuits to form multi-layer expression programs (Tickman and Alba Burbano 2021), we observed degradation of signals as they were propagated through multi-layer circuitry. In this system, signal degradation arises because output dynamic ranges generated by CRISPRa nodes are smaller than the scRNA input dynamic ranges to these same CRISPRa nodes. Mathematically, the fraction of the input signal that can be recovered at the output of a cascade scales with increasing depth as A^N where A is the fraction of signal propagated between individual layers and N is the number of layers in the circuit. For circuits composed of components for which output ranges are incapable of fully spanning input ranges of downstream elements, it is readily apparent that the fraction of signal recovered at the output rapidly diminishes. Broadly, there are two ways to solve the problem of incomplete spanning of dynamic ranges; either the input dynamic range of CRISPRa/i nodes could be compressed, or the output dynamic ranges of CRISPRa/i nodes could be increased. In this work, we focus on increasing output dynamic ranges through engineering of activateable promoters. Engineering of expression characteristics of synthetic promoters targeted for

activation by CRISPRa would provide access to a wider range of CRISPRa-directed expression levels resulting in higher fidelity computation, improved robustness of CRISPR control, and increased genetic stability of metabolic pathways.

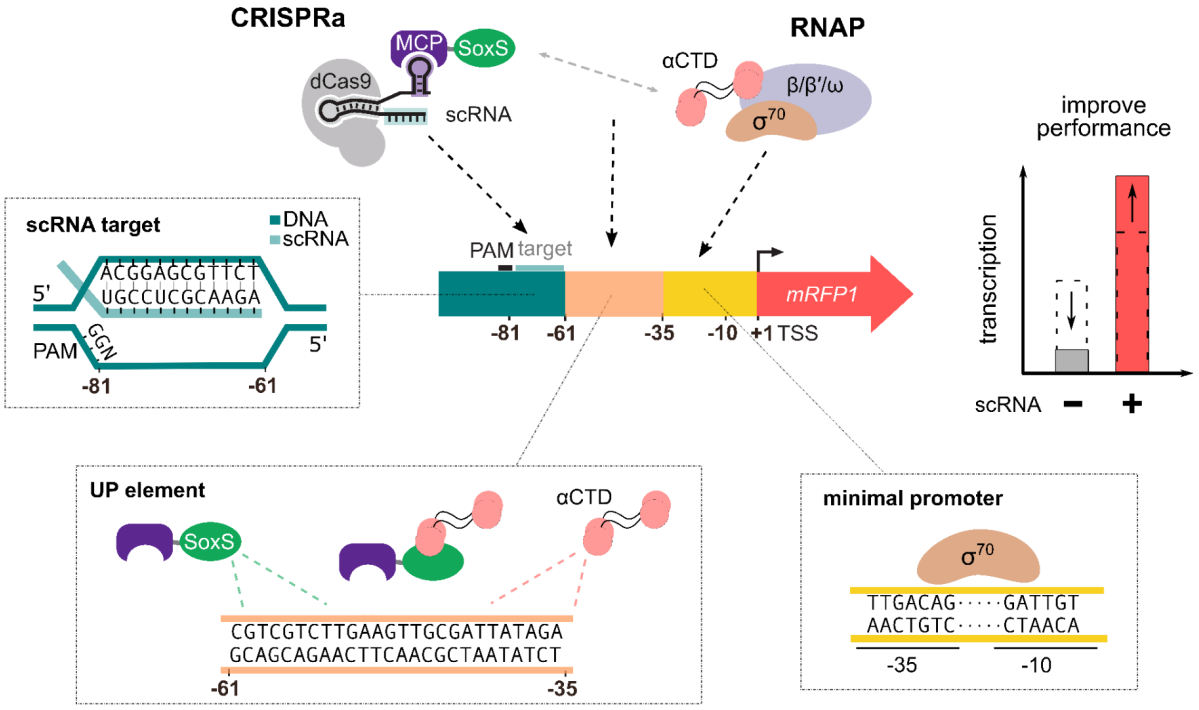
While bacterial transcriptional initiation from sigma 70 promoters is an incredibly well studied process(Brempt; Campbell et al., 2002; Djordjevic, 2011; Estrem et al., 1998; Gaal et al., 1996; Hook-Barnard and Hinton, 2007; Urtecho et al., 2019), in this work we relied on a data driven approach requiring relatively little knowledge of promoter structure. The success of this approach suggests that similar optimizations could be conducted in non-model systems, for which detailed information is unavailable. Starting from an initial unselected reporter design (Fontana et al., 2020), we systematically mutagenize minimal promoter elements, up elements, and gRNA target sequences from -3:-81 of the TSS. We characterize the effects of this mutagenesis on production of a fluorescent reporter in the presence and absence of CRISPR activation. Within each data set we observe a tradeoff between low basal expression levels and high CRISPRa-directed expression. This tradeoff establishes a pareto-optimal front composed of variants for which no other variant provides both higher activated and lower basal expression levels. Among the members of the pareto optimal front, any may be considered the best or most optimal.

Results

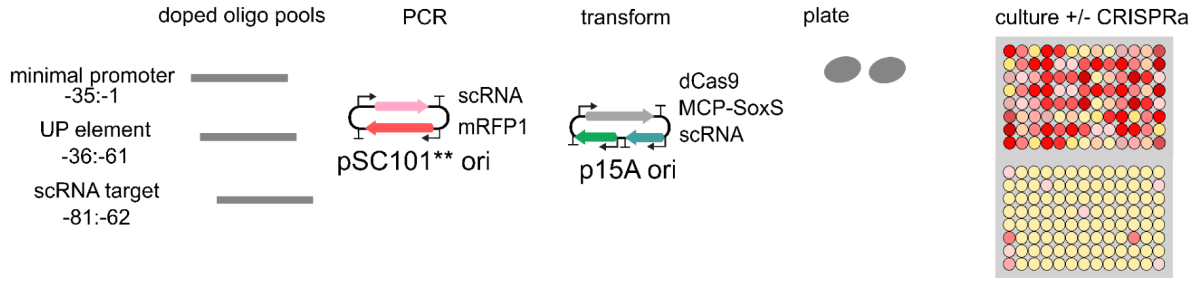
In this work, we independently mutagenize 3 regions of synthetic activateable promoters that constitute CRISPRa/i nodes. First we investigate the effect of sequence variation between -81:-62 on basal expression of CRISPRa/i nodes. Second, we investigate the effect of sequence variation within the up-elements between -61:-36 on both basal, and CRISPRa directed expression levels. Third, we investigate the effect of sequence variants in the minimal promoter from -35:-3 of the TSS. Lastly, we explore the effects of combining sequence variants from these regions to form composite parts.

Figure 1: Probing for Activatable Promoters with CRISPRa

A Anatomy of Synthetic Activatable Promoters and Relevant Interactions



B Assembly and Characterization of Libraries of Activatable Promoters in *E. coli*



A. Synthetic CRISPR-activatable promoters are divided into three distinct regions. A dCas9 binding site located -81:-62 of the TSS, an UP-element from -61:-36, and a minimal promoter from -35:-1 of the TSS. in this work, the ability of each of these regions to contribute to decreased basal expression, and increased activated expression levels is explored.

B. The effect of sequence variation within each of these regions is surveyed by library scale experiments in an inducible 2 plasmid *E. coli* system. One plasmid contains an RFP reporter, while the other plasmid contains all the components of the CRISPRa system, (dCas9, scRNA, and aTc inducible MCP-SoxS activator). Libraries are assembled by introducing sequence variants via PCR followed by transformation. Experimental strains are assembled by double transformation of reporter plasmid with CRISPR machinery. Experiments are conducted by picking colonies for overnight outgrowth into a 96 well plate, then diluting cultures 1:40 into metia with and without aTc inducer.

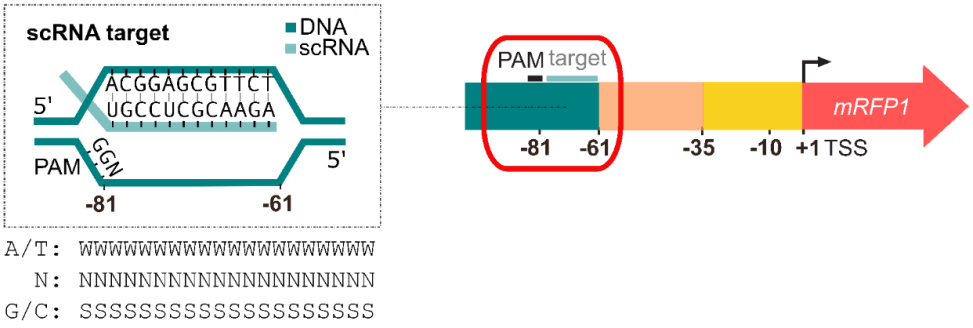
-81:-62 scRNA target sequences

Despite the fact that MCP-SoxS mediated CRISPRa experiences strict requirements with respect to distance from the TSS, the use of synthetic promoters and guideRNA targeting of dCas9 permits nearly arbitrary variation within this narrow region. This freedom allows selection of guide-RNA target sequences with desirable properties. Optimization within this region imposes three requirements. First, the chosen sequence must constitute a “good” guide target. Second, the chosen sequence should provide minimal contributions to basal expression. Third, the chosen sequence should be devoid of known transcription factor binding sites. While computational methods have been developed to screen sequences for the presence of known transcription factor binding sites and for prediction of guideRNA performance, predicting the contribution of sequence variants in this region to basal expression remains challenging (Yu et al,. 2021).

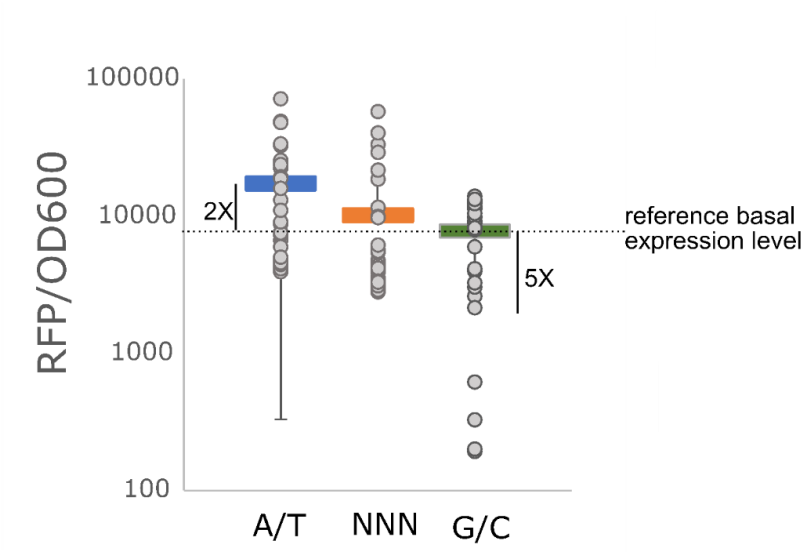
To explore the relationship between sequence identity and basal expression levels in this context, we created three libraries in which bases identities were mutagenized from -81:-62 of the TSS in the J3 promoter. The three libraries contained either random combinations of all bases (N-20), random combinations of A/T (W-20), or random combinations of G/C bases (S-20). We observed that variations in this region were mostly contained within a 10 fold range of expression levels. This variation was uneven, with long tails skewed towards higher expression levels. Generally speaking, mean fluorescence as well as the frequency of high expressing variants decreased with increasing GC content (S-20)<(N-20)<(W-20). Compared to W-20 and N-20, expression levels arising from variants in the S-20 pool were more tightly grouped. This effect is quantifiable by looking at the coefficients of variation arising from basal expression levels within each pool; .98, 1.28, and .57 for W-20, N-20, and S-20 respectively.

Figure 2: Contributions to basal expression from sequence variation -61:-82

A demarcation of mutagenized reigon



B sequence dependent effect on basal expression



A. Contribution of scRNA target site to basal expression levels was determined by measuring the distribution of fluorescence from three different libraries with a 20 nucleotide variable region between -81 and -62 of the TSS cloned into the context of the J3 promoter driving expression of RFP from a ColE1 plasmid.
B. Resulting distribution of fluorescence from each library reveals that increasing GC content in the space sequence leads to lower basal expression, and that AT rich sequences result in a high frequency of promoters with very high basal expression levels. Variants from the GC pool commonly provided ~5 fold lower basal expression than the J3 reference sequence, and the average expression level from the AT pool was 2 fold higher than the same reference variant

Table 1: basal expression properties of mutagenized libraries -81:-62 of TSS

	A/T	NNN	G/C
Mean	17365	10144	7834
Standard deviation	17034	12938	4433
Coefficient of variation	0.98097	1.2754	0.5659

-61:-36 Up-elements

Next, we investigated the effect of varying the 26 bases 3' of the scRNA target site and 5' of the minimal promoter -61:-36 of the TSS. In this region, endogenous promoters commonly contain Up-elements which interact with the alpha-CTD of RNAP providing an important contribution to basal expression levels. Various activator proteins can also bind in this region to recruit RNA polymerase and initiate transcription. SoxS belongs to such a family of activator proteins including rob and marA. These proteins activate transcription by forming protein-protein interactions with the alpha CTD, replacing the contribution of the up-element (Griffith and Wolf, 2001; Shah and Wolf, 2004). We reasoned that the dynamic range of CRISPRa/i nodes could be maximized through selection of sequences which provide minimal contributions to basal expression. Explicitly, we hypothesized that the use of "weak" up-elements would maximize the change in synthetic promoter activity provided by scRNA directed recruitment of SoxS.

Strong UP-elements are typified by AT rich homopolymeric sequences from -58:-36 of the TSS. To explore the effect of sequence variants within this region on both basal and

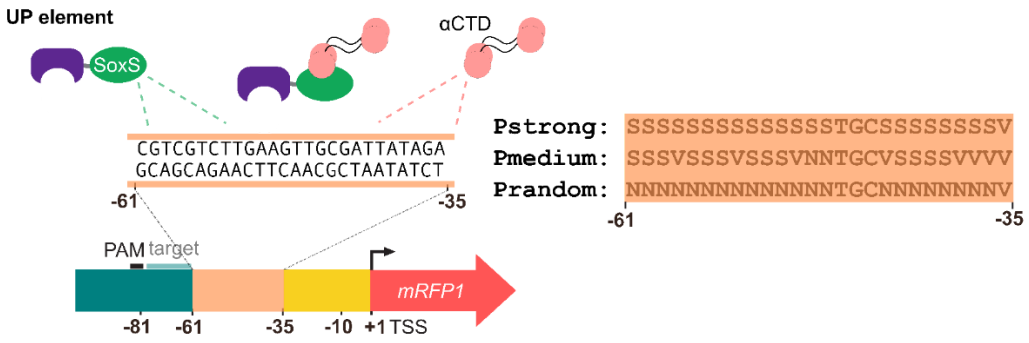
CRISPR-activated expression levels we generated three randomized libraries with increasing GC content; Prandom, Pmedium, andPstrong.

These experiments revealed a tradeoff between basal expression and activated expression levels. We observed that increasing GC content provided stark decreases to basal expression levels, and more modest decreases in activated CRISPRa/i node outputs. Importantly, while high GC content sequences provided a greater variability in activated expression level this did not preclude the existence of high activated expression levels, raising the possibility that sequence variation in this region could provide CRISPRa/i nodes with low basal expression levels without large reductions to activated expression levels.

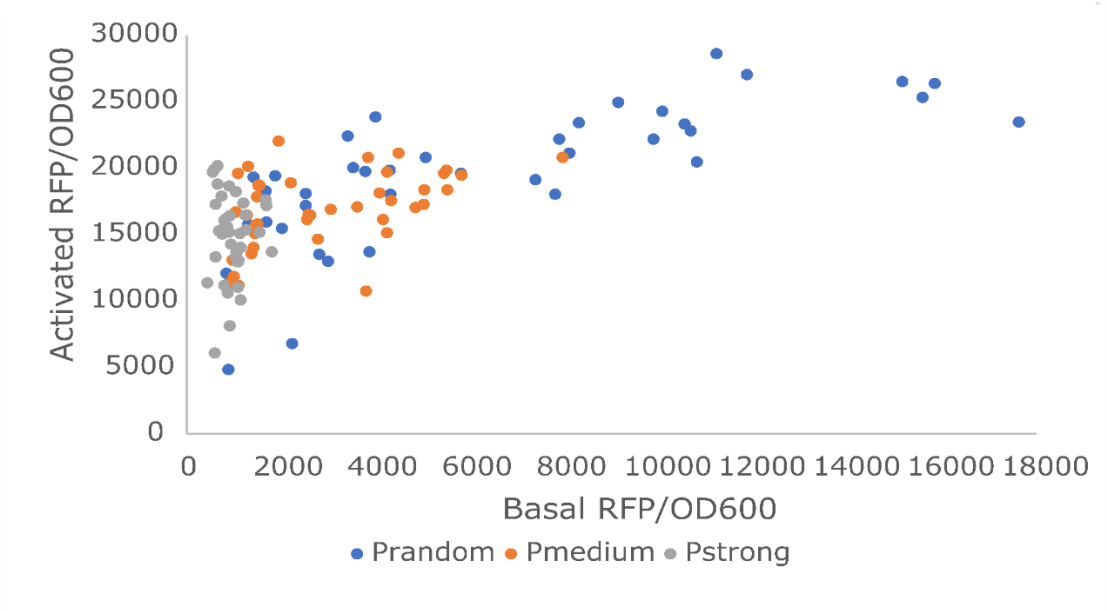
With the results of this screen in hand, we analyzed the sequence characteristics of high performing variants i.e. those with low basal expression levels and high activated expression levels. We used this information to order a new pool of randomized sequences in which individual base identities associated with high performance were fixed. From these pools we observed a much higher frequency of variants with extremely low basal expression levels, and high activated expression levels. Indeed greater than 20% of variants (38/170) achieved activation ratios of 400, constituting an increase over the previously identified highest performing variant. Furthermore, a full 10% (17/170) of tested sequences from these libraries provided activation ratios of greater than 500, and 3 variants provided activation ratios over 600. In this screen, the highest performing varint provided a 7.5 fold reduction to basal expression levels, while suffering a mere 25% reduction to activated expression level compared to an initial unselected up-element design.

Figure 3: contribution of sequence variation within the up-element to basal and CRISRPa- directed expression levels

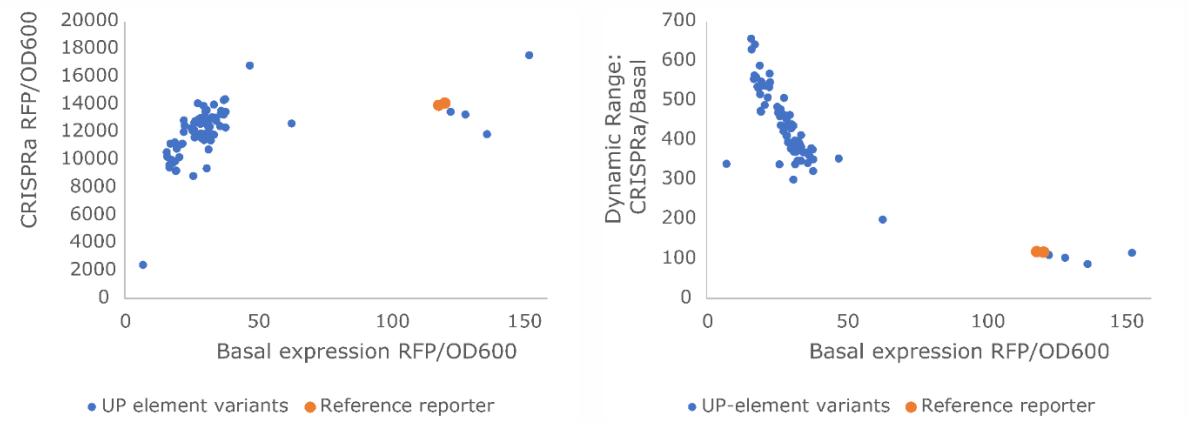
A Demarcation of mutagenized reigon



B Randomized libraries reveal a correspondence between GC content and basal expression levels



C Iterative selection and mutagenisis increases dynamic range of CRISPRa/i nodes



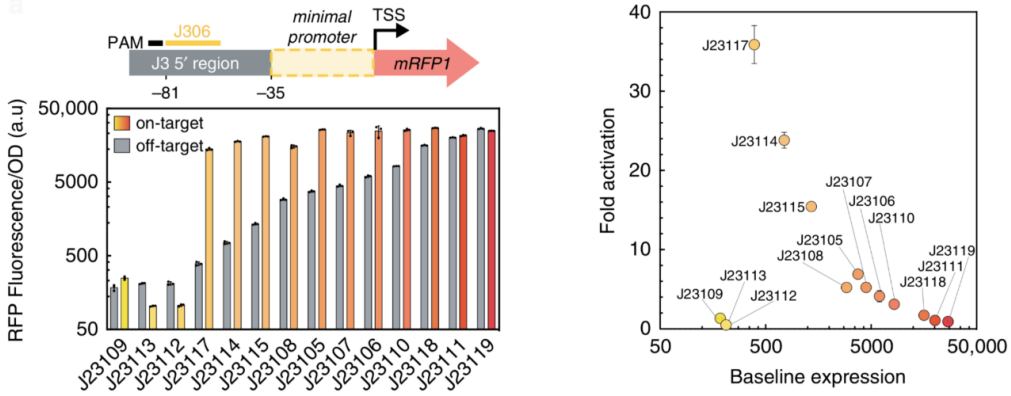
A. Contributions of sequence variations within the up element to both basal and activated expression levels were determined through the use of three different libraries cloned into the J3 sequence context driving expression of RFP
B. Randomized up element libraries reveal that increasing GC content has a strong effect on decreasing basal expression with more modest decreased to activated expression levels. Data are collected in the J3 context located on a ColE1 plasmid.
C. Sequencing variants from the experiment shown in panel B allowed the design of new pools incorporating consensus sequence variants associated with high performance i.e. low basal expression levels and high activated states. Measurement of variants from these pools reveals that library members were strongly biased towards the selection criteria. These data were collected in the J3 surrounding Context on a pSC101** plasmid. Left: variants are plotted using activated vs basal measured RFP/OD600 values. Right: comparison of dynamic range as a function of basal expression for up element variants highlights the improvement provided by sequence variation between -61:-36. In both plots, data are background subtracted to remove the contribution of autofluorescence using a strain harboring the same set of plasmids and parts, but with no RFP expression cassette.

-35:-1 Minimal promoter region

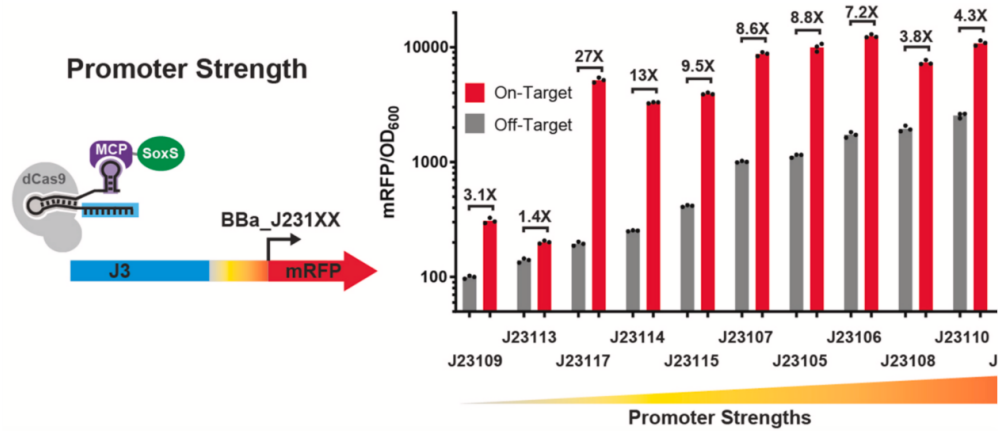
In previous work we observed a positive correlation between basal expression strength of minimal promoters and activated expression levels (Fontana et al., 2020). This relationship results in the dynamic range of CRISPRa being maximized by promoters with medium-low basal expression. This relationship has been shown to hold across hosts as diverse as p. Putida (Kiattisewee et al., 2021)**,** and e.coli CFS (Tickman & Alba et al., 2021) and appears to be a fundamental consideration governing activatability in this system.

Figure 4: Relationship between Basal and Activated expression holds across *E. coli, P. putida,* & CFS

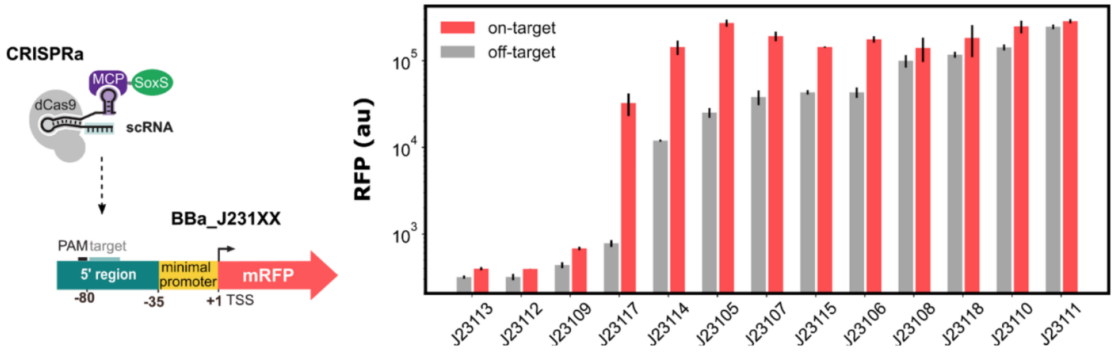
A Fontana 2020: CRISPRa on anderson promoters in *E. coli*



B Kiattisewee 2021: CRISPRa on anderson promoters in P. *putida*



C Tickman & Alba in press 2021: CRISPRa on anderson promoters in CFS



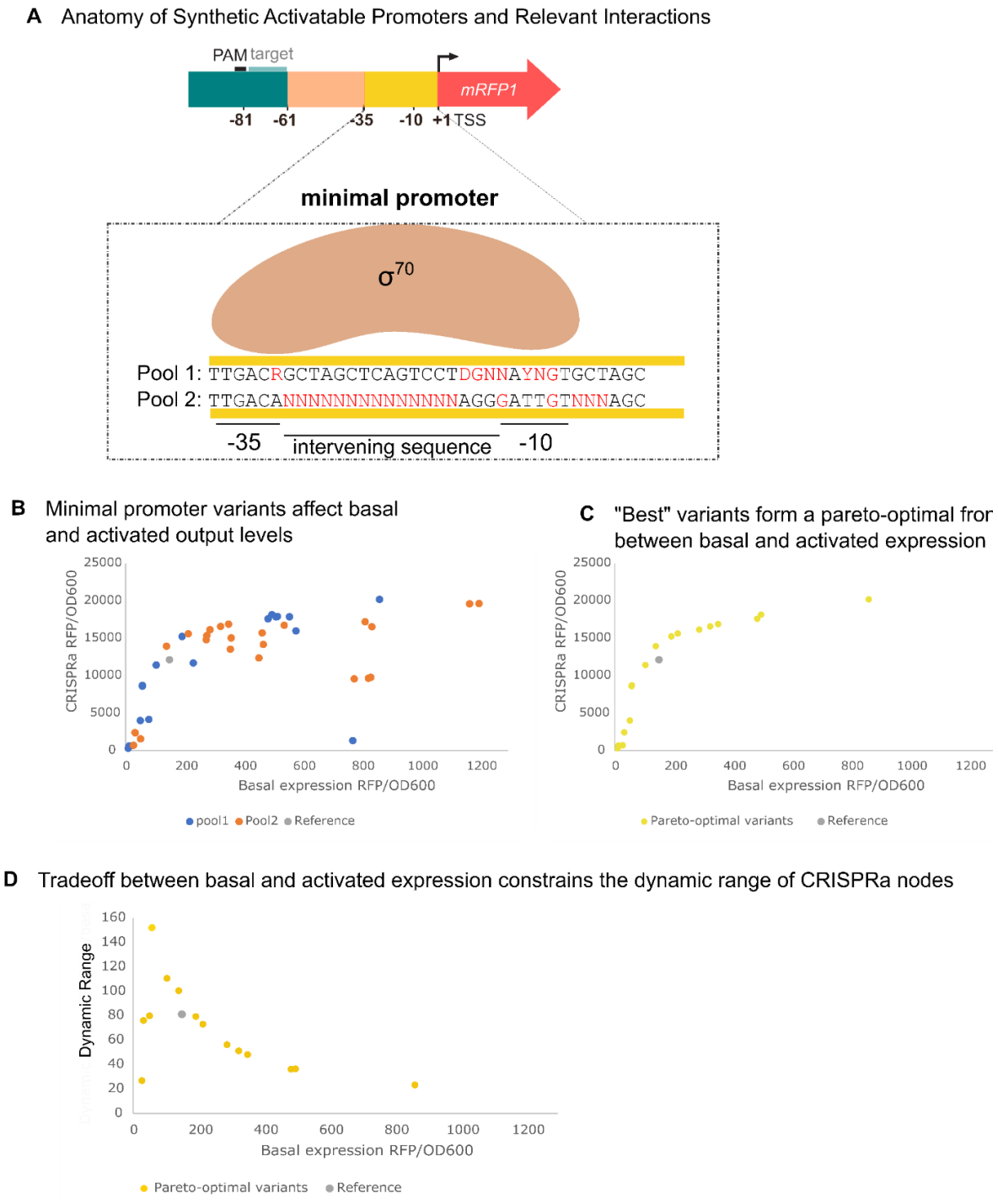
A. Constitutive CRISPRa on a library of anderson minimal promoters in the J3 context is reproduced from Fontana 2020 nat com. Data collected in *E. coli*
B. Constitutive CRISPRa on a library of anderson minimal promoters in the J3 context is reproduced from kiattisewee 2021 metabolic engineering. Data collected in *P. putida*
C. Constitutive CRISPRa on a library of anderson minimal promoters in the J3 context is reproduced from tickman and alba 2021 cell systems. Data collected in CFS.

The observed sharp dependence of CRISPRa directed dynamic range on basal expression levels, and lack of data interpolating the highest (J23117), and lowest (J23112) performing variants raised the question as to whether further improvements to reporter dynamic ranges could be realized by minimal promoter variants with lower basal expression than J23117. Starting from the best performing variant, J23117, we applied mutagenesis independently to the -10/-35 regions(-30,-15,-13,-12,-9,-8), and intervening sequence (-15:-29, -4:-6) and characterized both basal and activated expression levels. In both pools we found the same relationship between basal and CRISPRa directed expression levels, broadly consistent with the previous observations. This relationship was not absolute i.e it is not always true that a lower basal expression level will result in decreased CRISPRa directed expression. This imperfect correspondence indicates that some mutations may preferentially advantage CRISPRa directed as compared to basal expression. This observation raises the possibility that future work leveraging a mechanistic understanding of RNAP recruitment and transcriptional initiation to explore this decoupling could generate further improvements.

From these data we identify a set of variants for which no other variant has both lower basal and higher CRISPRa directed expression. This set forms a pareto-optimal front defining the best performing variants. Importantly, we found that the reference reporter which served as the starting point for these selections does not lie along this newly established optimal front. This result confirms that not only were we able to tune the strength of activateable promoters, but also that their performance could be improved by specific changes to promoter sequence. Interestingly, combining the data from these two pools, revealed that the location of the pareto optimal front remained largely unchanged despite the fact that the underlying mutations are targeted to different regions of the minimal promoter. In these data, the observed tradeoff between low basal and high activated expression levels serves to constrain the dynamic range of CRISPRa. Analogous to previous results gathered from up-element libraries, we observed

that the dynamic range of CRISPRa is highly sensitive to minimal promoter strength. Among pareto optimal variants, the top 48% of the range of observed dynamic ranges is populated by only 4 variants spanning less than 10% of the observed range of basal expression values. Importantly these same 4 variants all fell at or below the median absolute expression range of pareto optimal CRISPRa/i nodes. This observation highlights the tradeoff between basal and activated expression, of particular importance for cellular applications where absolute expression levels are constrained by available combinations of vector copy number and RBS strength in addition to the transcriptional characteristics of CRISPRa/i nodes.

Figure 5: minimal promoter variants tune output characteristics of CRISPRa/i nodes



A. Minimal promoter region is defined as the bases -35:-1 of the TSS. This region makes direct contact with the sigma 70 subunit of RNA polymerase at two 6 base regions termed -35 and -10. To assay the effect of sequence variants in this region on both activated and basal expression levels, we generated two randomized minimal promoter libraries based on BBaJ23117. In pool 1 mutagenesis was restricted to the -35 and extended -10 region, while in pool 2 mutagenesis was applied outside of the -35 and -10 sites. Minimal promoter pools were cloned into the J3 context expressing mRFP1 located on an sc101** plasmid

B. Scatter plot comparing basal and activated expression levels arising from variants within both pools as well as the un-modified J3 reference variant. Data were collected in MG1655 *E. coli* with colonies picked from plates, grown overnight in EZ media, and diluted 1:40 into EZ +/- 200nM aTc for experimental characterization. These cultures were allowed to saturate and measurements were conducted ~18 hours after dilution. These data reveal a non-linear correlation between basal and activated expression levels.

C. Selecting Pareto optimal variants i.e. variants for which no other variant has both lower basal expression and higher activated expression from both pools reveals that variants arising from both pools establish the same pareto optimal front. Data from both pools in panel B are combined and pareto optimal variants are replotted in panel C showing that the space of best solutions is unaffected by where mutagenesis is applied within the minimal promoter.

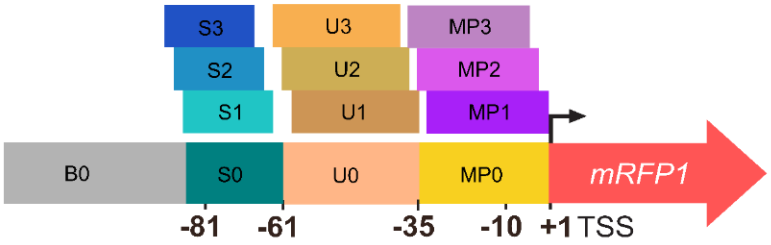
D. Data from panel C are replotted as a scatterplot comparing the fold activation provided by pareto optimal variants as a function of basal expression levels. The nonlinear relationship between basal and activated expression levels serves to restrict high dynamic range activateable promoters to a narrow range of basal expression levels.

Combinations of optimized components further improve CRISPRa/i node performance

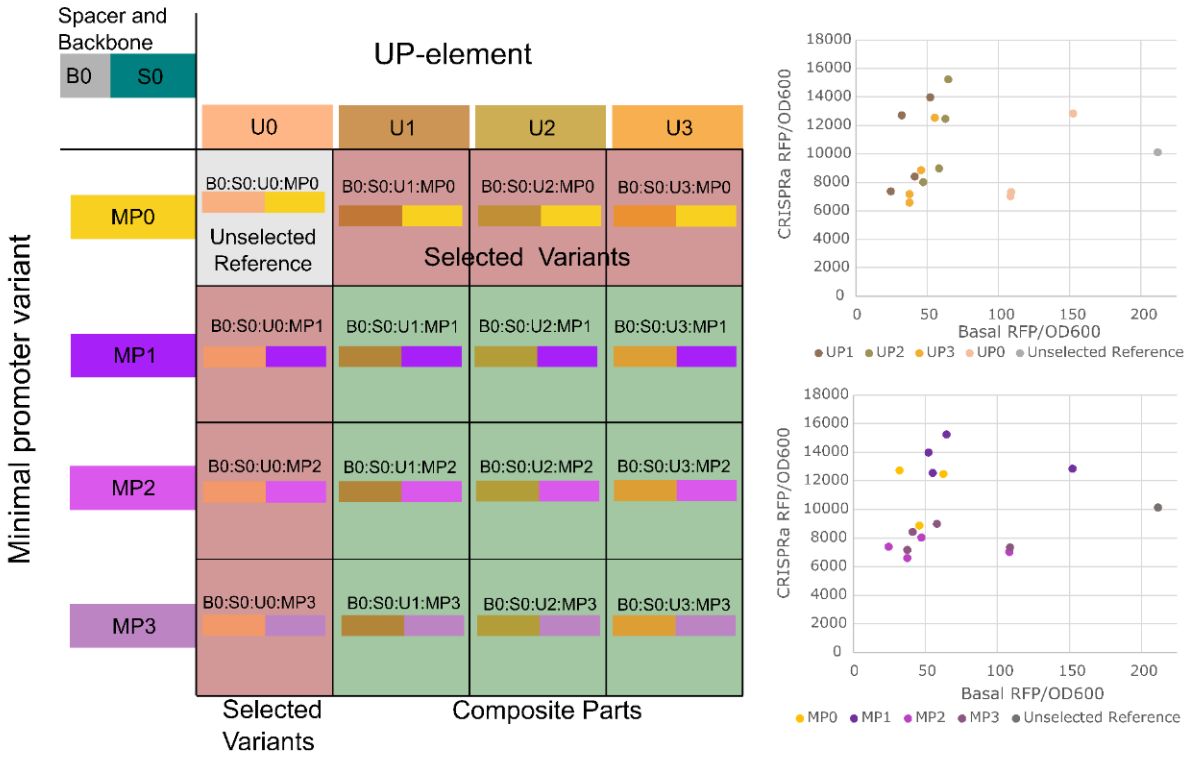
Next, we sought to determine whether CRISPRa/i node performance could be further improved by combining optimized components selected in isolation. As a first step we explored a small combinatorial library pairing UP-elements and minimal promoters. The combinatorial library was generated by pairing three minimal promoters chosen from the pareto-optimal front with three up elements, two from the pareto optimal front, and one slightly behind it (Figure 6A). As a fourth variant for both minimal promoters and up elements, we included the original context from the independent selections. Thus this new combinatorial library encompases an un-selected reference, each part in the original selected context, and all pairwise combinations (Figure 6B). Within this mini library, we observed that the coefficient of variation for basal expression within a minimal promoter context was larger than the coefficient of variation for basal expression within an up element context at .69 and .25 respectively. This means that within this combinatorial set, up-elements proved to be a stronger determinant of basal expression than minimal promoters. Interestingly, the opposite is true of activated expression levels. Here the coefficient of variation for activated expression levels is smaller within a minimal promoter context than within an up element context, at.11 and .30 respectively. Unsurprisingly, these composite parts established a new pareto optimal front. Out of the 2 up-elements, and 3 minimal promoters initially selected from their respective pareto optimal fronts, only one variant remained along the new front established by the combinatorial set. This new front is defined by four variants, three of which incorporate both selected up-elements and minimal promoters. When compared to the initial unselected control, 100% (15/15) of tested variants provided lower basal expression levels, and nearly half (7/15) provided increased activated expression levels. Together these results suggest that it may be possible to independently optimize basal and activated states through engineering of up elements and minimal promoters.

Figure 6: Composite parts can improve device performance

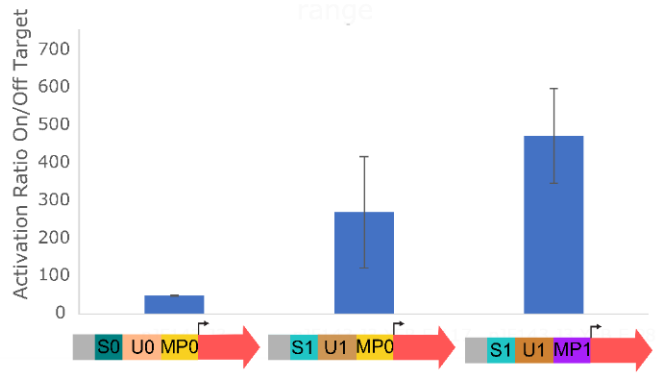
A Selected sequence variants can be combined to form composite parts



3 Performance characteristics of composite parts



C Spacer variants provide further improvement of composite parts



A. Composite parts can be formed by combining multiple sequence variants with favorable characteristics. In this work, composite parts are defined as specific combinations of spacer, up-element, and minimal promoter in the J3 upstream context.

B. Combining multiple parts results in a combinatorial space of composite parts. Here we are showing the combinatorial space established by up-elements and minimal promoter variants, with the original J3 context and 3 selected variants for each region. This establishes a 4*4 grid of variants containing the reference sequence (top left, grey), the individual variants in isolation (top and left, red), and the composite parts (bottom right, green). At right, scatter plots show the resulting basal vs activated expression. At top points are colored by up-element, while at bottom, points are colored by minimal promoter. In contrast to other experiments,these data were collected with on/off target scRNA as opposed to an inducible activator. Together these data show that composite parts comprised of minimal promoters and spacer variants can be used to improve system performance beyond that of individually selected variants

C. Composite parts incorporating spacer variants in addition to up-elements and minimal promoters can be used to further increase dynamic range of CRISPR-activatable synthetic promoters, suggesting that performance improvements arising from all three of the regions studied can be combined. Data are collected in MG1655 *E. coli* with on/off target scRNA grown to saturation in EZ media. Data is collected using three biological replicates, and background subtraction of autofluorescence is applied as previously described. Appropriate error propagation is conducted to combine the uncertainties arising from the background, basal, and activated states to arrive at the final uncertainty in fold activation.

Building on these observations, we sought to incorporate sequence variants upstream of the UP-elements (-62:-81) to further decrease basal expression levels and increase dynamic range of CRISPRa/i nodes. We observed that by combining sequences which provided low levels of basal expression in isolation termed “S1” (-62:-81) and “UP1”(-36:-61) we were able to achieve an 85% reduction to basal expression levels while suffering only a 23% reduction in CRISPRa directed expression compared to an unselected reference reporter. Addition of the best performing minimal promoter variant selected in isolation to this new context provided a further 68% reduction to basal expression levels while incurring an additional 43% reduction to CRISPRa directed expression levels. The combination of these high performing sequence variants led to a ~10 fold improvement in dynamic range and a 20 fold decrease in basal expression levels of CRISPRa/i nodes compared to the reference reporter (Figure 6C). Furthermore this composite CRISPRa/i node generated the highest dynamic range of any variant tested in this way. Together, these results establish that large improvements to device performance can be realized through routine selection and combination of high performing variants.

Despite the consistency of these results a comprehensive, predictive understanding of the relationship between CRISPRa/i node components and dynamic ranges has remained elusive. Explicitly, we have identified no simple and quantitative relationship relating individual component functions to composite behaviors such as output dynamic range. Upon consultation of the relevant literature describing transcriptional initiation and sigma 70 promoter structure in E. coli, this observation is unsurprising. Promoter activity is described as resulting from a combination of favorable interactions arising from many sources, resulting in individual component function being highly context dependent (Hook-Barnard and Hinton, 2007; Urtecho et al., 2019). Nevertheless, the fact that composite parts accounted for 75% of variants along the pareto optimal front comparing basal and activated levels established by the combinatorial library, as well as the observation that adding an scRNA target sequence providing low basal

expression levels to one of these pareto optimal variants could further improve performance, suggests that this is a tractable problem.

As an alternate approach to the aforementioned combinatorial optimization of CRISPRa/i nodes, we also pursued an iterative selection strategy to improve activateability of synthetic promoters. We hypothesized that the context dependent nature of bacterial promoters could be accounted for by simply re-selecting minimal promoter variants in the context of a high performing up-element. Here, the decision to reselect the minimal promoter as opposed to regions upstream of -35 is because we hypothesized that contributions arising from sequences upstream of the -35 are largely replaced by recruitment of activator proteins. Thus, it follows that the dynamic range of CRISPRa/i nodes could be maximized through selection of minimal promoters in the context of up-stream elements that make minimal contributions to basal expression levels.

To accomplish this, we constructed two minimal promoter libraries in the upstream context of S1 (-81:-62), and UP1 (-61:-36), both of which were previously identified to lie along pareto optimal fronts for activated and basal expression levels among sequence variants in their respective regions. In minimal promoter pool 1, we mutagenized bases -38, -36,-29:-14, -5 of the TSS. In minimal promoter pool 2 we mutagenized bases -30:-15 of the TSS. Despite the similarities between the mutagenized regions in these two pools, sequences arising from pools 1 and 2 are separated by a minimum hamming distance of 5 to ensure uniqueness of promoter variants. From each pool 96 variants were chosen at random, while a further 36 outliers were visually selected in an attempt to survey a wider range of the phenotype space.

As predicted, these iteratively selected composite components outperformed previously selected pareto-optimal variants. Compared to the unselected minimal promoter in a pareto-optimal upstream context B0:S1:UP1:MP0, we observed minimal promoter variants displaying 2.3 fold lower basal expression while providing comperabale activated levels, as well as variants providing a 35% increase in activated expression levels while leaving basal

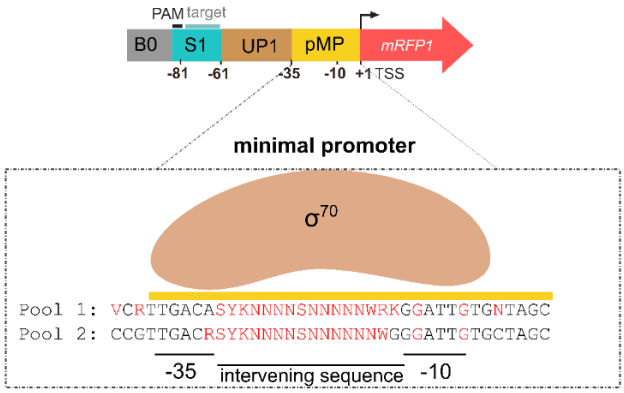
expression unchanged. This iterative selection process generated variants with basal expression levels varying across a ~10^4 fold range. At the low end of the observed basal expression range, we are unable to accurately characterize expression levels due to a combination of measurement error, and the variability associated with autofluorescence of strains. In practice, we applied a conservative background subtraction to account for autofluorescence of E.coli. Furthermore, in the following work we differentiate between analyses conducted on all data and those conducted on data for which basal expression can be quantified with reasonable accuracy (basal> ~10 RFU). From these same pools, variants with the highest basal expression reached ~33% of the expression capacity of E.coli.

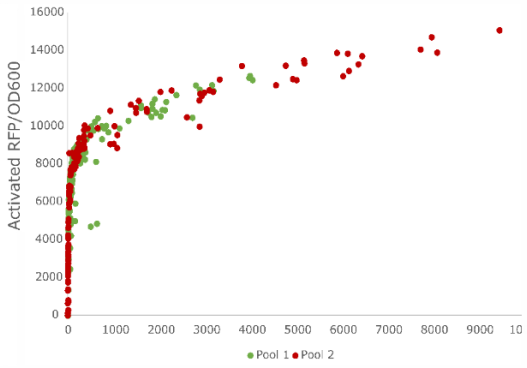
Consistent with previous observations we found that variants from both libraries established the same pareto optimal front despite differences in their underlying mutations. Within both pools, we observed a log-linear relationship between basal and activated expression levels. To confirm log-linearity in these data, we fit basal and activated expression generated by variants in pool 2 with several other models with forms including Michaelis menten, symmetrical sigmoidal (Hill equation), and asymmetrical sigmoidal. Among these models, a logarithmic fit to the data provided the highest raw R-squared, adjusted R-squared, and the lowest AIC. A model taking the form of the hill equation provided the second best fit; however, when evaluated with the AIC, this provided a delta i of >+23 when fit to all data, and a delta i of >+3.6 when fit to data with basal expression levels above 10. These delta i's provide odds ratios of roughly 10000:1 and 6:1 respectively for the likelihood of these data being described by a logarithm as opposed to a hill function. While to date no convincing mechanistic reasoning for this relationship has been presented, these analyses suggest that the data do indeed present a logarithmic relationship between basal promoter strength and CRISPRa directed expression levels. Quantifying this relationship with a logarithmic fit to the data generated from pool2 provided an equation (Activated=1482.3*ln(basal)+881.25) yielding an R-squared of .987 with basal expression levels spanning 4 orders of magnitude. Next, we

applied the same analysis to pareto optimal variants selected from orthogonal libraries in the unselected upstream context to provide a similar equation relating basal and activated levels. Despite a coarse sampling of the underlying pareto optimal front, we obtained an R-squared of .95 for a logarithmic fit comparing basal and activated expression among these pareto optimal variants. We subsequently used these equations to generate a prediction for dynamic range (activated/basal) as a function of basal expression for minimal promoter libraries in both the selected and unselected upstream context. Both the predictions and the observed data are provided in figure 7D. Within these datasets, we observe that the location of the pareto optimal front is established by sequences upstream of the minimal promoter, while the location of any pareto-optimal variant along this front is determined by the minimal promoter sequence. This analysis also provides a tangible representation of improved output characteristics, with selected upstream sequences shifting the location of the pareto optimal front providing access to higher dynamic ranges and lower basal expression from minimal promoter variants. Notably, the predicted pareto-optimal fronts for minimal promoters in the selected and unselected upstream context intersect. This intersection provides the prediction that for basal expression levels below ~55, the selected upstream context is capable of higher dynamic ranges as compared to the original unselected context, while for basal expression levels above 55, the opposite is true. This result adds another layer of nuance to our understanding of synthetic activateable promoters, providing a new avenue to enable the generation of CRISPRa/i nodes with expression characteristics that can be easily customized to achieve performance targets.

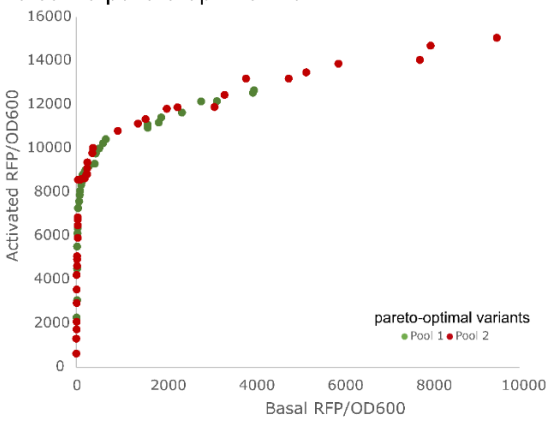
Figure 7: Composite variants define expression characteristics of CRISPRa/i nodes

A Re-selection of minimal promoters in the context of pareto optimal upstream elements

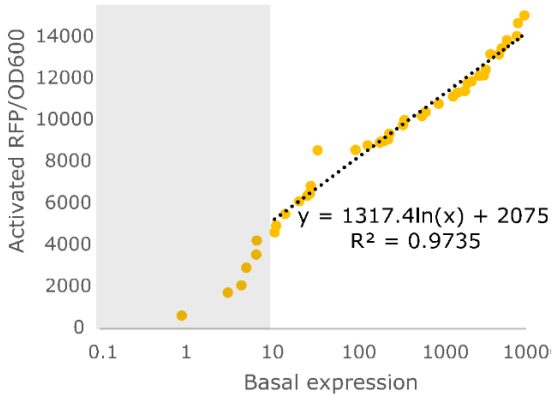




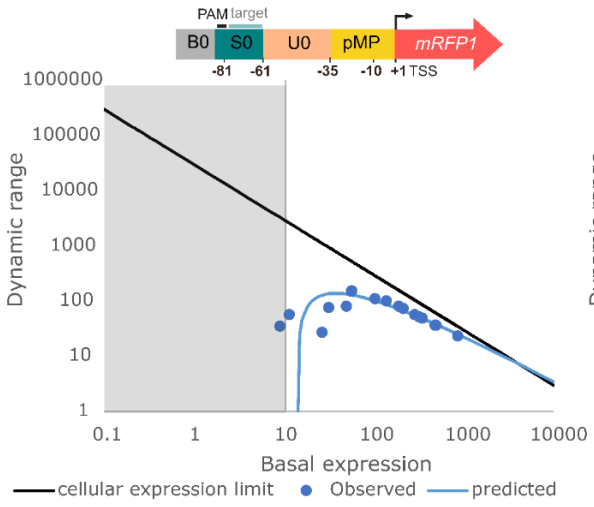
B Pareto-optimal variants from both pools establish the same pareto optimal front

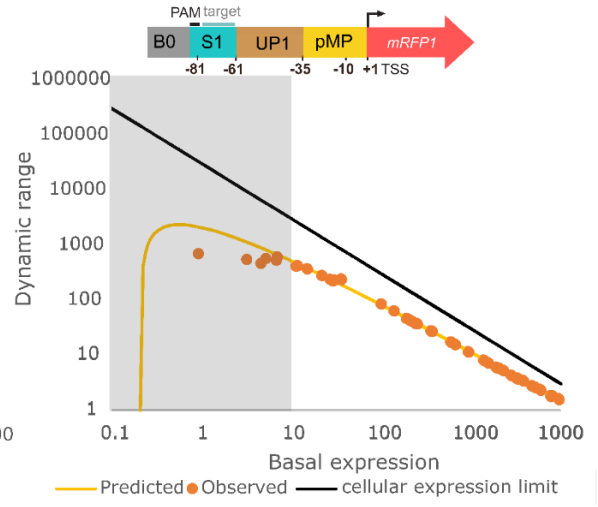


C Pareto optimal front is log linear



D Context determines position of pareto optimal front over which "best" variants can be distributed





A. Left: Two new minimal promoter libraries were designed using lessons learned from sequenced variants selected from the data presented in figure 5B. These minimal promoter libraries were cloned into a spacer/up-element context that had been selected for the ability to provide low basal expression and high activated expression when combined with the reference minimal promoter BbaJ23117. Right: scatterplot showing basal vs activated expression levels arising from both minimal promoter pools reveals the expected tradeoff between low basal expression and high activated levels previously observed in figure 5B. Data are collected in MG1655 E.coli using a aTc inducible activator system as previously described with the reporter located on a pSC101** plasmid.
B. Pareto optimal variants from the data shown in fig 7A are determined for each pool and replotted. These confirm the previous observation from figure 5C that variants from both pools establish the same pareto optimal front. Suggesting that the location of this front is not determined by the identity of the minimal promoter.
C. Pareto optimal variants from pool 2 are re-plotted in isolation on a log transformed scale to highlight the log linear relationship between basal and activated expression levels observed for minimal promoter variants. Grey region encompassing basal expression levels below ~10 denotes where the uncertainty in the autofluorescence leads to significant errors in basal expression levels and dynamic range.
D. Comparison of pareto optimal variants identified in figure 5C (Left) with pareto optimal variants obtained from pool 2 in the optimized upstream context reveals that sequences 5’ of the minimal promoter establish the location of the pareto-optimal front which defines the tradeoff between low basal and high activated expression levels. Points in the scatter plot represent observed variants. Colored lines represent the predicted dynamic range as a function of

basal expression obtained from logarithmic fits to the data. Solid black line represents the dynamic range that would arise from an activatable promoter that achieves the maximum observable expression level at a given level of basal expression, serving to bound the observable space of performance characteristics. Grey region encompassing basal expression levels below ~10 denotes where the uncertainty in the autofluorescence leads to significant errors in basal expression levels and dynamic range. In this representation, better performance is defined as variants which are further to the left and closer to the black line. Comparison of these data highlights the fact that minimal promoters selected in optimized upstream contexts can provide real improvements to the performance of activateable promoters.

Discussion

In cells, absolute expression levels from CRISPRa/i nodes are constrained by a combination of genetic copy number, as well as the strengths of the promoter and RBS. In the context of CRISPRa/i circuitry, we observed that unmodified CRISPRa nodes with a single internal layer compressed output dynamic range by 66% (Tickman & Alba Burbano et al., 2021). Under these conditions, compression of output dynamic ranges occurs entirely due to unactivated basal expression. Previously, we observed that reducing basal expression of the internal node by 3.5 fold reduced compression of output dynamic range from 66% to 31%. The 20 fold decrease to basal expression levels provided by these selected parts will further decrease compression of output dynamic ranges by internal layers of CRISPRa/i circuits enabling deeper networks capable of higher fidelity computation to be easily constructed. Application of these composite components to metabolic engineering would provide benefits not only to the ability to rapidly generate dynamic expression programs, but also to increased genetic stability and ease of pathway assembly arising from the low basal expression of engineered CRISPRa/i nodes. In this context control of CRISPRa/i programs applied through a master regulator could be used to minimize pathway and controller associated burden, growth rate defects and genetic instability before switching to production of heterologous pathway enzymes.

In CFS expression from CRISPRa/i nodes can be continuously varied by simply titrating plasmid concentrations. This ability to precisely titrate absolute expression levels greatly simplifies the level matching process as compared to cellular systems. Despite these advantages, construction of multi-layer circuitry in CFS is constrained by the reduced dynamic range of CRISPRa in CFS as compared to E. coli. Titrations of scRNA expressing plasmid in CFS revealed that CRISPRa directed output levels were sensitive across almost 3 orders of magnitude with respect to input scRNA expression levels. In CFS multi-layer circuits are

constrained by the fact that the input dynamic range of scRNA expression is larger than the output dynamic range of CRISPRa/i nodes. These engineered CRISPRa/i nodes achieve large dynamic ranges capable of spanning input dynamic ranges of downstream layers, providing high fidelity propagation of signals through layered CRISPRa/i networks. Lastly, application of these engineered CRISPRa/i components to the emerging field of cell free TX/TL based diagnostics provides not only the capacity for increasingly complex logical evaluations, but also lowers the intrinsic limits of detection by providing output stages with increased signal to noise ratios.

In this work, a simple data driven approach to selection and assembly of CRISPRa/i nodes has proven highly successful. These selected components have decreased basal expression to the limits of detection in our assay, constituting at least a 20 fold reduction compared to the previously optimal reference variant. Concomitantly, these reductions in basal expression levels have provided >10 fold increase in output dynamic range from CRISPRa/i nodes relative to the previously optimal reference variant. Intriguingly, we have observed that sequence context serves to define the shape and location of the pareto optimal front relating basal and activated expression from CRISPRa nodes. Within a given sequence context, we observed a log linear relationship between activated and basal expression spanning nearly 4 orders of magnitude, spanning greater than 50% of the range of E.coli expression capacity. Notably, this log linear relationship has also been observed from up-element libraries in the context of a defined minimal promoter sequence. This symmetry hints that these observed log-linear relationships may arise from a common underlying source that remains as yet undiscovered.

While this work has successfully provided CRISPRa/i components with increased output dynamic range, these data have revealed a tradeoff between output dynamic range and output range. While this distinction is of diminished importance in Cell free systems, functional components in cellular systems must meet target performance specifications with respect to

both of these metrics. In E.coli, currently available dynamic ranges are likely sufficient to span the input dynamic range of CRISPRa/i nodes however, the decreased range of these high dynamic range components may result in degradation of signals with increasing network depth due to incomplete spanning of input ranges for downstream layers. Fortunately, the mismatch between these input and output ranges is on the order of 2 fold. Given the large improvements in device dynamic range observed in the iteratively selected libraries, it appears likely that a similar iterative selection approach would yield variants which satisfy these specifications. Alternatively, tried and true amplification strategies such as increased genetic copy number could be used to meet absolute expression requirements in cellular CRISPRa/i circuits.

In light of this success, it is all but certain that large improvements to CRISPRa/i nodes remain undiscovered. Fortuitously, RNAP transcriptional initiation from sigma 70 promoters in E. coli has served as a model biological system, and as such a vast wealth of information exists in the published literature. Use of this mechanistic understanding to guide future selection and optimization of components can be expected to further improve CRISPRa/i node performance. As our understanding of the design rules for CRISPRa/i nodes evolves, new methods of quantifying device performance will be required. With current reporter assays, the best performing CRISPRa/i nodes could tolerate no more than a further two fold reduction in basal expression levels, and three fold increase in activated expression levels before device performance becomes unquantifiable. Development of high throughput screening approaches in cell free systems could enable quantification of device performance across a wider range of transcriptional outputs than in-vivo. However, at present the potential utility of CRISPRa/i output dynamic ranges beyond this quantification limit (~2.5*10^3) are unclear. Identification of the rules allowing facile generation of CRISPRa/i nodes with such characteristics would solidify the CRISPRa/i control system as the premiere choice for applications requiring versatile multi-gene regulation, and complex signal processing in a prokaryotic setting.

References

Alon, U. (2007). Network motifs: theory and experimental approaches. Nat. Rev. Genet. *8*, 450–461.

Brempt, M.V. Predictive design of sigma factor-specific promoters. 13.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. Mol. Cell *9*, 527–539.

Clamons, S., and Murray, R. (2019). Modeling predicts that CRISPR-based activators, unlike CRISPR-based repressors, scale well with increasing gRNA competition and dCas9 bottlenecking. BioRxiv.

Djordjevic, M. (2011). Redefining Escherichia coli σ70 Promoter Elements: −15 Motif as a Complement of the −10 Motif. J. Bacteriol. *193*, 6305–6314.

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

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. U. S. A. *95*, 9761–9766.

Fontana, J., Dong, C., Kiattisewee, C., Chavali, V.P., Tickman, B.I., Carothers, J.M., and Zalatan, J.G. (2020). Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. Nat. Commun. *11*, 1618.

Gaal, T., Ross, W., Blatter, E.E., Tang, H., Jia, X., Krishnan, V.V., Assa-Munt, N., Ebright, R.H., and Gourse, R.L. (1996). DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. Genes Dev. *10*, 16–26.

Gander, M.W., Vrana, J.D., Voje, W.E., Carothers, J.M., and Klavins, E. (2017). Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. Nat. Commun. *8*, 15459.

Griffith, K.L., and Wolf, R.E. (2001). 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.

Hook-Barnard, I.G., and Hinton, D.M. (2007). Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters. Gene Regul. Syst. Biol. *1*, 275–293.

Huang, H.-H., Bellato, M., Qian, Y., Cárdenas, P., Pasotti, L., Magni, P., and Del Vecchio, D. (2020). dCas9 regulator to neutralize competition in CRISPRi circuits (Synthetic Biology).

Kiani, S., Beal, J., Ebrahimkhani, M.R., Huh, J., Hall, R.N., Xie, Z., Li, Y., and Weiss, R. (2014). CRISPR transcriptional repression devices and layered circuits in mammalian cells. Nat. Methods *11*, 723–726.

Kiattisewee, C., Dong, C., Fontana, J., Sugianto, W., Peralta-Yahya, P., Carothers, J.M., and Zalatan, J.G. (2021). Portable bacterial CRISPR transcriptional activation enables metabolic engineering in Pseudomonas putida. Metab. Eng. *66*, 283–295.

Landberg, J., Wright, N.R., Wulff, T., Herrgård, M.J., and Nielsen, A.T. (2020). CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in Escherichia coli. Biotechnol. Bioeng. *117*, 3835–3848.

Li, S., Jendresen, C.B., Landberg, J., Pedersen, L.E., Sonnenschein, N., Jensen, S.I., and Nielsen, A.T. (2020). Genome-Wide CRISPRi-Based Identification of Targets for Decoupling Growth from Production. ACS Synth. Biol. *9*, 1030–1040.

Shah, I.M., and Wolf, R.E. (2004). 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.

Urtecho, G., Tripp, A.D., Insigne, K.D., Kim, H., and Kosuri, S. (2019). Systematic Dissection of Sequence Elements Controlling σ70 Promoters Using a Genomically Encoded Multiplexed Reporter Assay in *Escherichia coli*. Biochemistry *58*, 1539–1551.

Future directions and opportunities

Future directions in CFS

Moving forward, we seek to apply genetically encoded dynamic CRISPRai programs to enable efficient production of plant natural products in CFS. Many plant natural products of interest require the action of Class II P450s. However, Class II P450s have been avoided by engineers due to their association with membranes, and requirement for Cytochrome P450 Reductases (CPRs). In their native context these enzymes are specifically targeted for incorporation into the membranes of the endoplasmic reticulum and chloroplast through the use of N-terminal signal peptides. The activity of these enzymes has been shown to be highly sensitive to membrane phospholipid composition, with specific activities varying by up to 778 fold across differing membrane environments (Meijer et al., 1993; Collu et al., 2001). Lastly, Class 2 P450s require separate expression of their cognate reductase, the relative expression level of which must be tuned to achieve efficient electron transfer from NADPH. Together these challenges have led to a high failure rate for expression of active Class 2 P450s in heterologous hosts. Our approach is to reconstitute these Class II P450 systems in a cell free environment to enable the efficient production of plant natural products by using dynamic gene expression programs to increase yields of properly folded membrane associated catalysts, and efficient utilization of NADPH. As a testbed, we will prototype dynamic CRISPRa/i circuitry controlling expression of class II P450

systems to optimize the conversion of geraniol to 8-hydroxygeraniol, an important intermediate for the biosynthesis of many pharmaceuticals.

This work will advance the state of cell free metabolic engineering by showing that dynamic control of gene expression remains important in a cell free environment. The current state of the art involves simple overexpression of pathway enzymes followed by mixing to achieve desired stoichiometries. While this approach is convenient for prototyping many metabolic pathways, it requires enzymes that are stable, highly soluble, and are not prone to fouling. These constraints narrow the scope of metabolic pathways that are presently compatible with CFS, increase process complexity, and result in decreased titers, rates, and yields. Successful demonstration of dynamic control for metabolic engineering in CFS will enable multi-phase reactions, expand the repertoire of enzymatic activities available, and increase efficiency of substrate conversion.

Opportunities for CRISPRa/i control in CFS

Beyond applications in metabolic engineering, the ability of CRISPRa/i transcriptional circuits to provide dynamic control of gene expression, complex logical computations, and analog signal processing will be of interest to the broader cell free community. Many of the engineering challenges faced by cell free metabolic engineers are also relevant within the synthetic cell community. Here efficient allocation of resources, and coordination of gene expression across time is essential to realize the vision of self sustaining synthetic life. Overhead expression costs of current controllers employing protein based regulators scale poorly with increasing network complexity. The CRISPRa/i control system presents an opportunity for a scalable and efficient regulatory system that could one day form the backbone of transcriptional regulation in synthetic organisms.

In the context of cell free biocomputation, the CRISPRa/i control system serves as a promising option for the implementation of complex logical operations and analog signal processing due the predictable and engineerable nature of transcriptional circuits. These capabilities are expected to be of particular interest to the burgeoning field of cell free diagnostics. There is considerable interest in the use of cell free systems for diagnostics and biosensing due to their versatility and capacity for lyophilization. In diagnostic applications, inputs must be processed into transcriptional responses that can be easily measured at point of use. The potential for CRISPRa/i circuits to be stably

encoded and stored as DNAs which can be quickly and predictably assembled to execute high fidelity logical operations raises the possibility of field programmable diagnostic devices.

Opportunities for dynamic CRISPRa/i control in cellular hosts

The ability of cellular systems to convert simple feedstocks into complex biocatalysts capable of molecular recognition and transformation stands among the most awe inspiring feats of the natural world. While metabolic engineering in cellular hosts presents a more daunting challenge than cell free systems, the ability of cells to autonomously carry out the functions necessary to host engineered metabolic pathways is indispensable to the cost effective bioproduction of commodity chemicals. Over billions of years cells have evolved sophisticated systems to sense environmental conditions and internal states, integrate information, compute, and actuate to efficiently respond to their environment.

In the context of metabolic engineering, cells respond to perturbations imposed by the operation of heterologous pathways at many levels (e.g transcription, post transcriptional modifications, mRNA structural rearrangements, translation, post translational modifications, allostery, and protein conformation). Despite incredible progress, a comprehensive understanding of cellular regulatory responses to the action of heterologous pathways remains elusive. Unsurprisingly, attempts to rationally tune the operation of metabolic pathways often incite unforeseen cellular responses. While recent efforts based on flux balance analysis have successfully increased productivity across a wide range of pathways and organisms, modern engineering approaches have

increasingly relied on dynamic expression of pathway enzymes and regulation of genomic targets to further optimize production (Gupta et al., 2017; Zhang et al., 2012). Indeed it has been shown that adaptive CRISPRi circuitry employing feedback can be used to dynamically regulate the operation of metabolic pathways to increase titers of N- acetylglucosamine by 60% (Wu et al., 2020). Elsewhere it has been shown that dynamic regulatory systems which are capable of autonomously shifting between growth and production phases of a metabolic program can be used to increase T.R.Y (Dinh and Prather, 2020, 2019; Doong et al., 2018). Together, these observations demonstrate that dynamic regulatory systems can be used to improve production by matching the demands of a pathway to the available host capacity.

Concurrently, metabolic engineering efforts directed towards modifying host resource consumption have proven fruitful. In a general sense, these efforts seek to increase the fraction of feed stock that ends up in the desired product by reducing flux to biomass accumulation (Landberg et al., 2020), preventing diversion of flux out of the engineered pathway (Dinh and Prather, 2019; Tian et al., 2019; Wu et al., 2017), and to combat cellular regulatory responses opposed to pathway function (Sander et al., 2019). While such approaches have proven successful for improving titers and yields, static implementations often create marked growth rate defects, decreasing the overall rate of product formation.

The dynamic CRISPRa/i circuitry we have developed presents an exciting opportunity for the application of dynamic control to cellular metabolic engineering endeavours. The scalable nature of CRISPR-based regulation combined with the arbitrarily composable nature of CRISPRa/i nodes enables the formation of regulatory

programs capable of both feedback and feed forward architectures. These capabilities expand the accessible control space available to metabolic engineers, allowing the rapid design and assembly of complex multi-gene expression programs. Dynamic CRISPRa/i expression programs operating through the regulated expression of guideRNAs circumvent the pitfalls associated with both static and inducible control strategies by providing the ability to deliver genetically encoded, multi phase expression programs at library scale. In light of the proven ability of CRISPRa/i circuits to be rapidly assembled and tuned to form scalable and dynamic gene expression programs, their application to metabolic engineering is expected to provide access to increasingly complex and aggressive rewiring of cellular metabolism, opening a new frontier for the optimization of bioproduction.

References

Dinh, C.V., and Prather, K.L. (2020). Layered and multi-input autonomous dynamic control strategies for metabolic engineering. Curr. Opin. Biotechnol. *65*, 156–162.

Dinh, C.V., and Prather, K.L.J. (2019). Development of an autonomous and bifunctional quorum-sensing circuit for metabolic flux control in engineered *Escherichia coli*. Proc. Natl. Acad. Sci. *116*, 25562–25568.

Doong, S.J., Gupta, A., and Prather, K.L.J. (2018). Layered dynamic regulation for improving metabolic pathway productivity in Escherichia coli. Proc. Natl. Acad. Sci. *115*, 2964–2969.

Gupta, A., Brockman Reizman, I.M., Reisch, C.R., and Prather, K.L.J. (2017). Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. Nat. Biotechnol. *35*, 273–279.

Landberg, J., Wright, N.R., Wulff, T., Herrgård, M.J., and Nielsen, A.T. (2020). CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in Escherichia coli. Biotechnol. Bioeng. *117*, 3835–3848.

Sander, T., Wang, C.Y., Glatter, T., and Link, H. (2019). CRISPRi-Based Downregulation of Transcriptional Feedback Improves Growth and Metabolism of Arginine Overproducing E. coli. ACS Synth. Biol. *8*, 1983–1990.

Tian, T., Kang, J.W., Kang, A., and Lee, T.S. (2019). Redirecting Metabolic Flux via Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in Escherichia coli. ACS Synth. Biol. *8*, 391–402.

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

Wu, Y., Chen, T., Liu, Y., Tian, R., Lv, X., Li, J., Du, G., Chen, J., Ledesma-Amaro, R., and Liu, L. (2020). Design of a programmable biosensor-CRISPRi genetic circuits for dynamic and autonomous dual-control of metabolic flux in Bacillus subtilis. Nucleic Acids Res. *48*, 996–1009.

Zhang, F., Carothers, J.M., and Keasling, J.D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat. Biotechnol. *30*, 354–359.

-

Prospects for continued improvement of the CRISPRa/i system

Despite the rapid advance of CRISPRa/i component performance, in such complicated systems there will always be room for further improvements. While the utility and definition of “improvement” is at present a matter of debate, it can be expected that future applications of the CRISPRa/i system will place as of yet unimagined demands on system performance. The multi-component nature of the CRISPRa/i system provides many opportunities for the engineering of performance characteristics. Due to the popularity of CRISPR-based systems across a wide range of disciplines many groups are already engaged in engineering efforts to expand the utility of CRISPR systems broadly. Ongoing work seeks to engineer dCas9 for altered DNA binding affinities and PAM specificites, generate orthogonal sgRNA handle sequences to provide non redundant parts, engineer the affinity of dCas9 for guideRNAs, provide computational predictions for the strength of DNA binding asa function of target sequence, engineer conditionally reversible cas9 binding interactions, increase the specificity of dCas9 binding interactions, provide strategies for improved guideRNA processing and program delivery, development of control strategies to improve the robustness of multiplexed CRISPR systems, and development of anti-CRISPR systems to efficiently inactive cas9 complexes.

Within the carothers group, ongoing work seeks to expand the versatility of the CRISPRa/i system through development of alternate activator systems, improved understanding of the rules governing activation, development of higher performance CRISPRa/i nodes through promoter engineering, assimilation of responsive input stages, predictive design of high

performance guideRNAs, conditional guideRNA processing strategies, adaptation of the CRISPRa/i system to non model organisms, expansion of the characterized toolbox of CRISPRa/i components, and further determination of the rules and strategies governing efficient application of CRISPRa/i control.

In the face of such a diverse set of opportunities, provision of a concrete set of targets for future capabilities with proven applications is desirable. Continued development of the CRISPRa/i system to enable dynamic ranges greater than 10^3, absolute expression ranges spanning greater than 60% of cellular capacity, a characterized toolbox of greater than 15 nodes, arbitrary targetability with respect to sequence and position, engineering of conditional guideRNA activities, conditionally reversible CRISPR-regulation, greater than 4 independent non chemical inputs, greater than 2 orthogonal effector domains, and a predictive understanding of the determinants of the sensitivity of CRISPR regulation to guideRNA concentration would solidify the CRISPRa/i system as the tool of choice for a wide array of synthetic biology applications. While these milestones and benchmarks may seem aspirational, none of these advances rely on entirely unproven technology. More specifically, the ground work enabling all of these capabilities has already been proven independently. As such, the realization of these performance goals is simply a matter of combining and iteratively improving upon established solutions and methods. Undoubtedly, in conjunction with the larger CRISPR community, these advances and others will continue to satisfy the ever increasing demands placed on genetic control systems, enabling evermore sophisticated engineering of biological systems.

Acknowledgments

Acknowledgments are provided below with individual people placed into groups for compactness. These groups are listed in no particular order

Ron and Liz Tickman:

- For fostering a sense of curiosity in the natural world, and providing me with the belief that any observation can be understood if viewed in the right framework, and approached with sufficient rigor and determination. For providing unwavering support over my long and at times meandering path through the educational system, grades K-23. For raising 4 of the best friends I could ask for, and after a time joining them.

Jeremy, Max, Spencer, and William Tickman:

- By pursuing your own individual technical interests you have provided a unique forum for technical discussions. Each of your unique perspectives has provided me with a new lens with which to view scientific problems.

Tom Haff, Linda Sorenson:

- The importance of good educators early in scientific education can not be understated. Thankyou for helping me find my passion for science during your biology, chemistry, and physics classes at Issaquah highschool.

David Kimmelman, Stefan Stoll, Pradipsinh Rathod, Daryl Pedigo, Matt DePies, Jesse Zalatan, Paul Yager, Paul Wiggins:

- For providing excellent in-class instruction during my time at UW. Your lectures, and course materials provided rare insight and inspiration.

William Voje, Cassandra Burke, Jason Stevens, Chuhern Hwang, David Sparkman-Yaeger, Jason Fontanna, Ian Faulkner:

- For showing me the ropes of synthetic biology, introducing me to the way of thinking required of a synthetic biologist.

David Sparkman-Yaeger, Jason Fontanna, Ian Faulkner, Cholpisit Kiattisewee, Diego Alba Burbano, Widi Sugianto, Ryan Cardiff, and Ava Karanjia:

- For the immense help of sharing your ears, your minds, and your plasmids with me throughout my time in the Carothers lab I can not thank you enough. Thankyou for engaging with my work, helping to refine it by challenging my direction and ideas. Without your collective contributions this would not be possible

Jason Fontanna, Chen Dong:

- For laying the groundwork upon which my thesis is based

Diego Alba Burbano:

- For helping immensely to evolve the ideas, and representations allowing the communication of our work. For accompanying me on our journey through the development of multi-layer CRISPRa/i programs in CFS and E.coli. Your contributions have been invaluable, I can say with certainty we would not be here today without them.

James Carothers:

- For providing the right mentorship when it was needed. For sharing your vision of the future. For guiding me toward promising research directions and away from avoidable errors I am truly grateful. For providing a research environment, assembling a diverse and collaborative group that fosters the open transfer of ideas and assistance. For the ability to ask questions to which I neither know, or do not know the answer, expanding my thinking in the process. For your patience. All the work you have done to tailor your teaching process to my needs is appreciated.

Jesse Zalatan, Eric Klavins, Herbert Sauro:

- For providing open ears and a critical eye. For helping me to shape my direction through graduate school and engaging with my work. For asking questions that required me to understand my work in a larger context.

Jens Gundlach:

- For first showing me what a functional, collaborative research group looks like. For assembling a fun and hard working group of scientists. For modeling balance in academic life, fostering creativity, and for helping me realize that my passion for science could lead me to graduate studies.

Dylan Nehrenberg:

- Your strength in the face of incredible adversity has helped inspire me to be the best version of myself. Perspective in life is hard earned, the grace with which you have handled yourself, and your continued commitment to service throughout this trying time will forever serve as an example. Thankyou for leading by example.

Appendices

In the materials provided below are included for reference, and as a record of my contributions

Appendix 1: “Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements”

- Conceived of, conducted, and analyzed the experiments for quantification of CRISPRa on minimal promoters from the Anderson promoter series.
- Reviewing of the written manuscript, provision of feedback

Appendix 2: “Membrane Augmented Cell-Free Systems: A New Frontier in Biotechnology”

- Contributed to the writing of materials pertaining to control of gene expression in membrane augmented CFS.
- Provided editing assistance and feedback on the written manuscript

Appendix 3: “Complex dependence of CRISPR-Cas9 binding strength on guide RNA spacer lengths”

- Supplied genetic components allowing for interrogation of the biophysics of dCas9 target binding
- Provided technical assistance facilitating the use of the CRISPRa system
- Provided clarification regarding the theoretical basis for and limitations of a statistical mechanics based framework for interpretation of results.
- Provided editing assistance during manuscript writing

Appendix 1: Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-15454-y OPEN

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

Jason Fontana [1,5], Chen Dong [2,5], Cholpisit Kiattisewee [1], Venkata P. Chavali [3], Benjamin I. Tickman [1], James M. Carothers [1,3,4✉] & Jesse G. Zalatan [2,3,4✉]

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

[1] Molecular Engineering & Sciences Institute, University of Washington, Seattle 98195 WA, USA. [2] Department of Chemistry, University of Washington, Seattle 98195 WA, USA. [3] Department of Chemical Engineering, University of Washington, Seattle 98195 WA, USA. [4] Center for Synthetic Biology, University of Washington, Seattle 98195 WA, USA. [5] These authors contributed equally: Jason Fontana, Chen Dong. ✉email: jcaroth@uw.edu; zalatan@uw.edu

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

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

We demonstrate here that the rules for targeting CRISPRa to effective sites in *E. coli* are surprisingly stringent. In prior work, we found that CRISPRa in *E. coli* was effective at target sites located in a narrow 40 base window between 60 and 100 bases upstream of the transcriptional start site (TSS)[3]. Here, we show that multiple factors combine to make the requirements for effective sites even stricter. We demonstrate that the basal promoter strength of the target gene and the sequence composition between the target site and the minimal promoter can have marked effects on gene activation. Further, by scanning the 40 base window at single base resolution, we find sharp peaks of activity and broad regions of inactivity that occur in a periodic 10–11 base pattern, corresponding to one helical turn along the DNA target. The observation that only a few precisely positioned target sites upstream of the TSS are effective for CRISPRa poses a significant challenge, as many genes will likely lack an NGG PAM sequence at exactly the right position necessary for *Streptococcus pyogenes* dCas9. These stringent requirements may explain why CRISPRa and other tools for gene activation in bacteria have lagged far behind comparable tools in eukaryotic systems, where such strict target site requirements are absent[5].

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

Results

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

A distance metric for target sites is not effective. To determine whether we could predictably activate endogenous genes with CRISPRa, we selected three candidate genes with appropriately positioned PAM sites upstream of the TSS. Previously, we demonstrated that CRISPRa can activate heterologous promoters up to 50-fold with target sites positioned within a 40 base window between 60 and 100 bases upstream of the TSS[3]. We therefore targeted the CRISPR–Cas complex to the same window upstream of the candidate target genes. First, we targeted the *aroK-aroB* operon, which expresses enzymes involved in aromatic amino-acid biosynthesis, whose programmed overexpression could be useful for bioproduction[9]. Targeting the CRISPR–Cas complex to two sites within the optimal 40 base window resulted in no statistically significant increases in gene expression. Further, sites inside and outside of the 40 base window gave similar effects

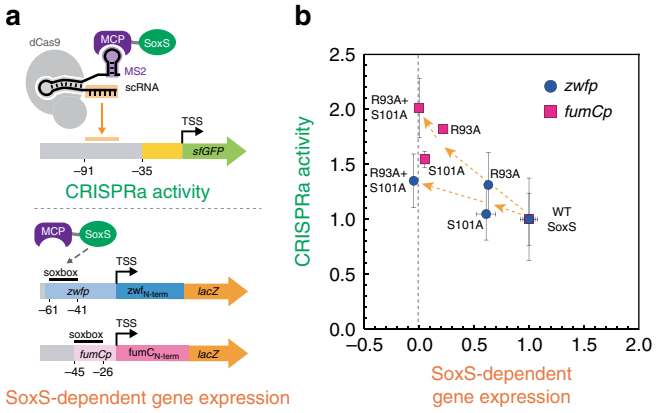


Fig. 1 A SoxS double mutant maintains CRISPRa activity and does not activate endogenous SoxS targets. a Reporter system for measuring the CRISPRa activity and endogenous SoxS-dependent gene expression of wild-type or mutant SoxS constructs. CRISPRa activity was determined in a strain harboring a genomically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

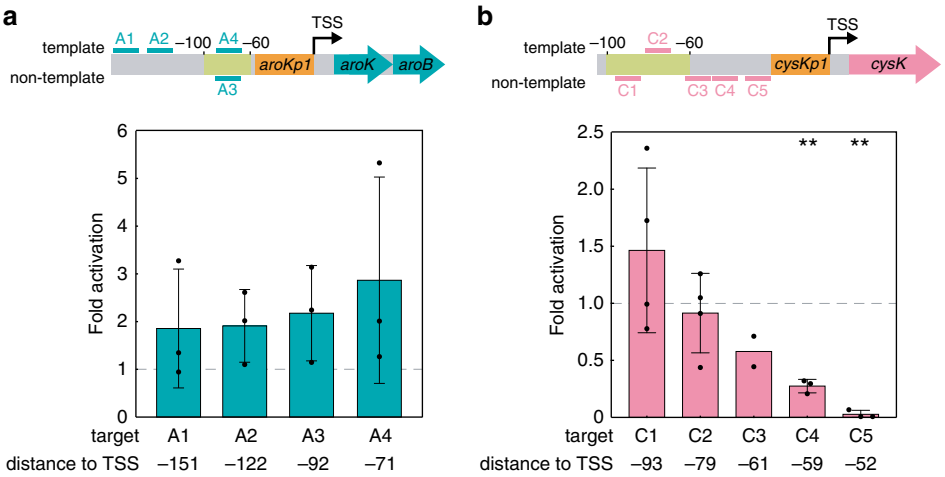


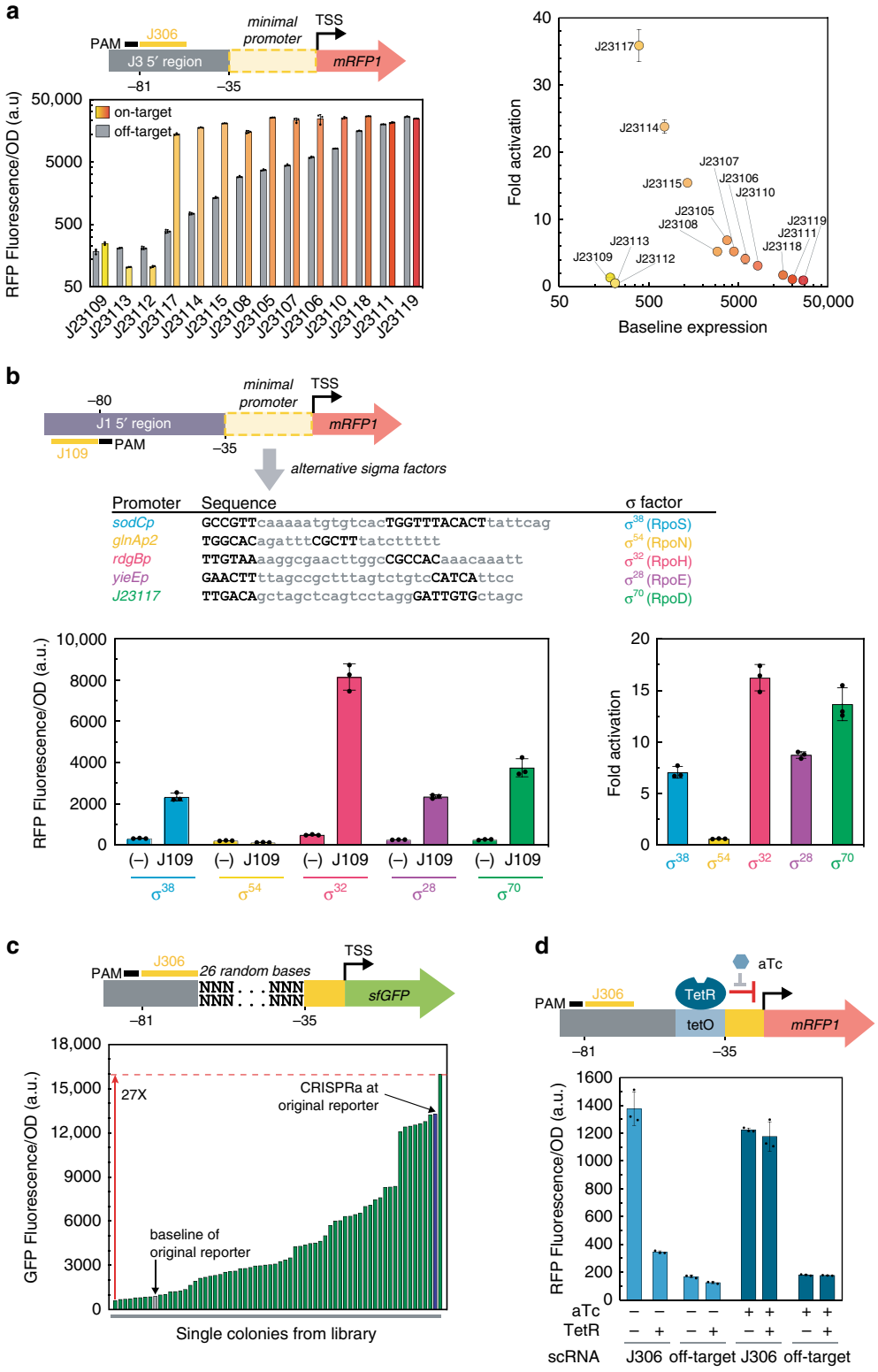
Fig. 2 A simple distance metric does not predict CRISPRa activity. a CRISPRa on the *aroK-aroB* operon. Two scRNA target sites within the 40 base window where CRISPRa is effective (−100 to −60) in heterologous reporter genes (A3–A4) and two sites further upstream (A1–A2) were chosen for the *aroKp1* promoter. **b** CRISPRa on the *cysK* gene. Three scRNA target sites within the 40 base window where CRISPRa is effective in heterologous reporter genes (C1–C3) and two sites further downstream (C4–C5) were chosen for the *cysKp2* promoter. The C4 and C5 sites resulted in repression; targeting these sites close to the core promoter may interfere with RNA polymerase binding. Gene expression was measured using RT-qPCR. Fold activation represents expression levels relative to a strain expressing an off-target scRNA (hAAVS1). In **a** and **b**, bars represent the average ± standard deviation calculated from $n = 4$ (C1, C2), $n = 3$ (A1–4, C4,C5), or $n = 2$ (C3) biologically independent samples. Some individual replicates in samples A1–4 and C1–2 appear to show activation, but a two-tailed unpaired Welch's *t* test indicates that the average differences relative to the off-target control are not statistically significant (*p* value > 0.05). Targeting CRISPRa to C4–5 resulted in repression, likely owing the short distance between the sites and the promoter. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's *t* test (**p value < 0.01). Exact *p* values: A1: 0.18, A2: 0.27, A3: 0.36, A4: 0.17, C1: 0.29, C2: 0.66, C3: 0.098, C4: 0.0022, C5: 0.00043. Source data are provided as a Source Data file.

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

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

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

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



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

CRISPRa is sensitive to intervening sequence composition. To determine whether the sequence composition between the target site and the −35 site affects CRISPRa, we constructed a promoter library with randomized sequences in this intervening region. We analyzed single colonies from this library and observed gene activation with a broad distribution over a 27-fold range (Fig. 3c). Although most variant sequences can still be activated (more than twofold) with CRISPRa, the large variation in activity was

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

unexpected because each reporter gene was driven by the same minimal promoter and contained the same scRNA target site. One possible interpretation of this result is that these randomized intervening sequences contain binding sites for endogenous transcriptional regulators; there is evidence that binding sites can emerge with relatively high frequency from random sequences[22]. These sites could potentially affect CRISPRa by directly blocking access to a scRNA target site, by blocking RNA polymerase binding, or by interfering with the ability of a CRISPRa effector protein to engage with RNA polymerase.

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

To determine whether transcription factor-binding sites appear in the library of randomized intervening sequences, we sequenced 29 variants spanning the full range of observed activation levels (Supplementary Table 6). Only five intervening sequences contained exact matches to a known consensus transcription factor-binding motif. However, all sequences contained at least one match within a single base of a known motif, and it is well established that DNA-binding proteins can recognize sites that deviate from the consesus[26]. There was no significant correlation between gene activation by CRISPRa and the number of these motifs (Spearman rank order correlation $r_s = 0.29$, $p = 0.11$, Supplementary Fig. 3A), but we note that it is not known which of these motifs actually bind endogenous transcription factors. We did find that intervening sequences that give more effective CRISPRa tend to be more GC-rich, though we do not yet understand the basis for this trend ($r_s = 0.42$, $p = 0.02$, Supplementary Fig. 3B & C). Nonetheless, these experiments indicate that the composition of the intervening sequence between the CRISPR–Cas complex and the minimal promoter is an important factor determining the level of CRISPRa.

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

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

Because we demonstrated that sequence composition can have unexpected effects on CRISPRa (Fig. 3), we tested whether the periodicity of CRISPRa was similar in different sequence contexts. We obtained comparable periodic phase dependence when different nucleotide sequences were used to shift the scRNA target site, and when the bases were inserted at a different location in the promoter (Supplementary Fig. 5A). Similar results were also obtained when we performed the base shift experiment with a reporter that had a different 5' upstream sequence (Supplementary Fig. 5B) or where the minimal BBa_J23117 promoter was replaced by endogenous *aroK* promoter (Supplementary Fig. 5C). Further, the sharp positioning dependence was observed when targeting the template or non-template strand of the reporter (Supplementary Fig. 5D). Finally, one possible confounding effect could arise if the basal expression level of the reporter gene changes when bases are inserted, which can affect the efficacy of CRISPRa (Fig. 3a). However, we observed that basal expression from the original reporter and the +5 base shifted reporter were indistinguishable (Supplementary Fig. 5E). Together, these experiments confirm that bacterial CRISPRa is sensitive to periodicity in multiple different sequence contexts.

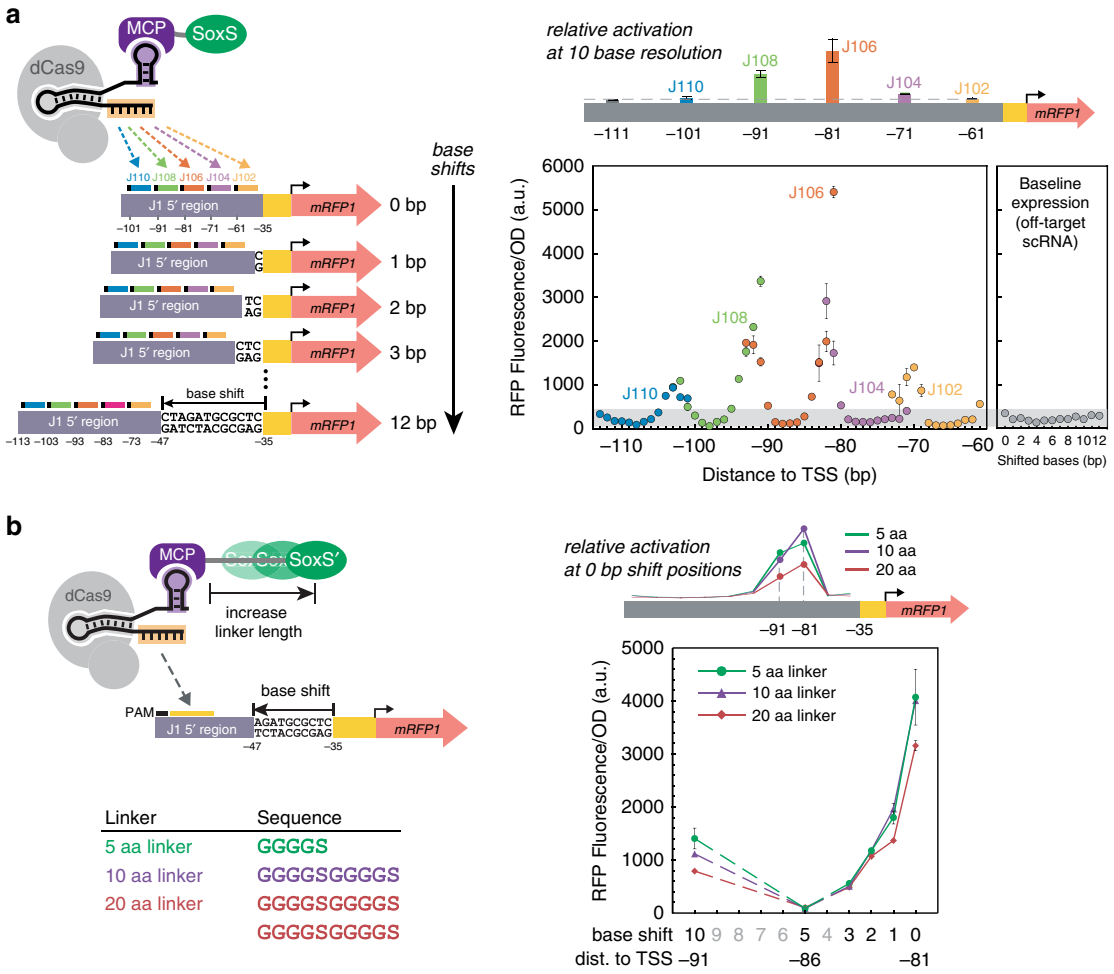


Fig. 4 CRISPRa is sensitive to the precise position of the scRNA target. a CRISPRa displays periodic positioning dependence with peak activities every 10–11 bases between −60 and −100 from the TSS. Reporter genes were constructed by inserting 0–12 bases upstream of the −35 region of the J1-J23117-mRFP1 reporter. Five scRNA sites (J102, J104, J106, J108, J110) with positions −61, −71, −81, −91, −101 from the TSS on the non-template strand of the original promoter were targeted. In this way, the complete −61 to −113 region can be covered at single base resolution. The color coding indicates data for the same target site shifted across a 12 base window. The panel on the right shows the baseline expression of reporters with shifted bases when an off-target scRNA was used (J206). The gray area represents the range of the baselines among the reporter series. For comparison, previous CRISPRa data for the J102, J104, J106, J108, J110 target positions at 10 base resolution are shown on the schematic above the plot[3]. **b** Extending the linker length between MCP and SoxS does not change the position dependence of CRISPRa. The J1-J23117-mRFP1 reporter plasmid series with base shifts were delivered together with CRISPRa components for targeting J106. The MCP-SoxS(R93A) effector contained 5aa, 10aa, and 20aa linkers. For comparison, previous CRISPRa data with 5aa, 10aa, or 20aa linker between MCP and SoxS targeting at the −81 and −91 positions are shown on the schematic above the plot[3]. Values in **a** and **b** represent the average±standard deviation calculated from $n = 3$ biologically independent samples. Source data are provided as a Source Data file.

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

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

Tuning structure to expand target site range is ineffective. If rotating the CRISPRa complex out of phase along the DNA prevents SoxS from interacting with RNA polymerase, then a

longer amino-acid linker to SoxS might allow effective CRISPRa at more scRNA sites. To test this possibility, we extended the linker between MCP and SoxS from five 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 (Fig. 4b). We obtained similar results using a linker with a different amino-acid composition (Supplementary Fig. 8A).

Another potential approach to expand the range of effective CRISPRa sites would be to change the spatial position of the MCP-SoxS protein by altering the position of the MS2 hairpin that binds MCP. We therefore tested multiple alternative scRNA designs that present the MS2 hairpin at different locations. Extending the MS2 stem by 2, 5, 10, and 20 bp resulted in progressively lower CRISPRa activity, but no change in the position of the target sites that were most effective (Supplementary Fig. 8B). Similarly, no changes were observed with alternative scRNA designs with one or two MS2 hairpins presented from different locations within the scRNA structure (Supplementary Fig. 8C).

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

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

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

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

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

We first examined the *yajG* gene, which had two plausible target sites, one of which was only compatible with dxCas9(3.7). We also included an additional site predicted to be out of phase and ineffective for CRISPRa. We observed significant, ~4–6-fold gene activation for the two sites located at the predicted peak of activity at −80/−81, and no activation at the out of phase site at −87 (Fig. 6a). The site at −81 is inaccessible to dCas9, and we only observed activation with dxCas9(3.7). We proceeded to test an additional five genes with partial success. We observed significant activation at *poxB* (~10-fold) and *uxuR* (approximately twofold) (Fig. 6b). We validated these results by performing RT-qPCR on the endogenous *yajG* and *poxB* loci. Targeting CRISPRa to these genes resulted in increases in RNA levels (Supplementary Fig. 11). Targeting CRISPRa to *araE* produced a statistically significant difference in expression, but the activation measured was modest (1.13-fold). For the remaining two candidate genes, *ansB* was modestly repressed at one of the target sites and we did not observe a statistically significant difference in expression at *ppiD*. Similarly, one of the *ldhA* sites that we targeted in initial experiments (Supplementary Fig. 1) was at a predicted optimal site at −91 and failed to give substantial activation. Thus, of seven endogenous genes tested with target sites that we predict should be effective

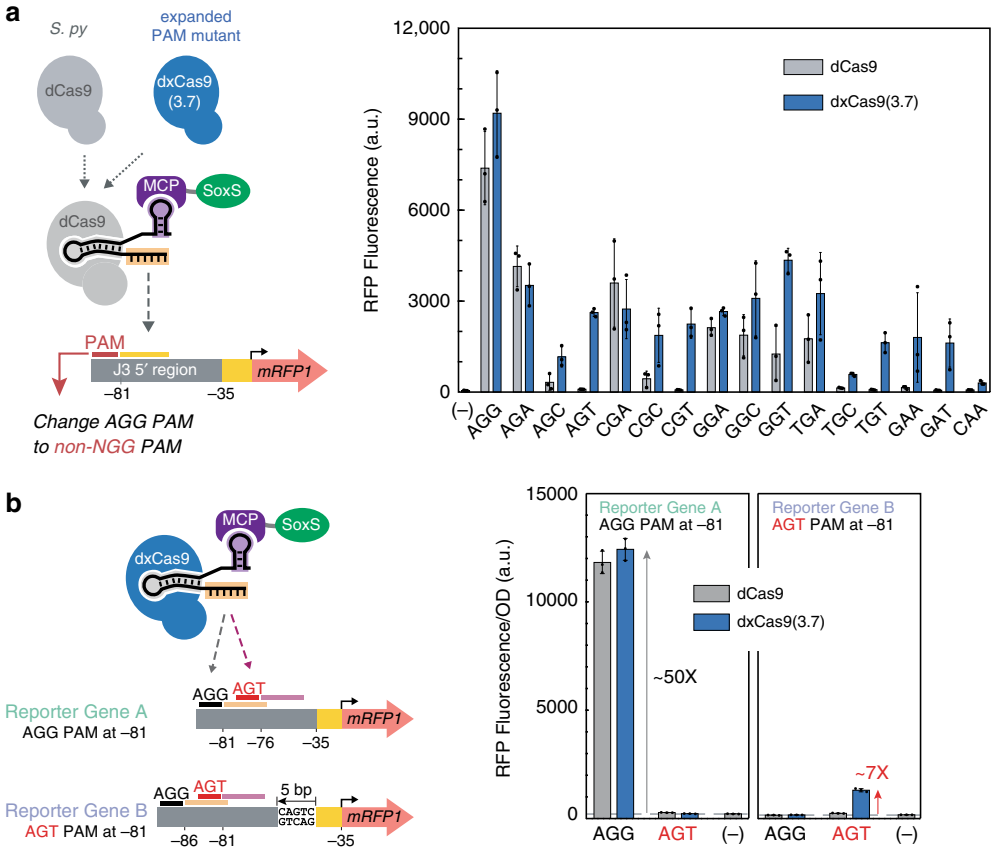


Fig. 5 dxCas9(3.7) expands the range of targetable scRNA target sites by recognizing alternative PAMs. a CRISPRa with dxCas9(3.7) displayed activity on non-NGG PAM sites with AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA sequences. CRISPRa activity with dxCas9(3.7) on non-NGG PAM sites was generally lower (6-fold to 89-fold activation relative to a control without a scRNA) compared to the AGG PAM site (188-fold activation). *Sp*-dCas9 also displayed moderate CRISPRa activity at non-NGG PAM sites with AGA, CGA, GGA, GGC, GGT, TGA sequences, consistent with published reports[6]. Reporter plasmids were constructed by replacing the AGG PAM site for the J306 target in the J3-J23117-mRFP1 reporter with alternative PAM sequences that have been previously reported to be recognized by dxCas9(3.7) in human cells[6]. The (−) sign indicates a control expressing the original reporter with the AGG PAM and the CRISPRa components with *Sp*-dCas9, the activation domain and no scRNA. **b** dxCas9 (3.7) can activate promoters that cannot be activated by *Sp*-dCas9. When the scRNA target at the optimal position (−81 to the TSS) has an AGG PAM site, both *Sp*-dCas9 and dxCas9(3.7) increased gene expression by 50-fold. When the scRNA target at the optimal position has an AGT PAM site, only dxCas9 (3.7) displayed a sevenfold increase in gene expression while *Sp*-dCas9 was inactive. The reporter gene has a target with an AGG PAM (M1) and a target with an AGT PAM (M2) upstream of a BBa_J23117 minimal promoter. In reporter gene A, the AGG target was located −81 to the TSS on the non-template strand and the AGT target was located −76 to the TSS on the non-template strand. In reporter gene B, 5 bases were inserted upstream of the −35 region, shifting the locations of the AGG target and AGT target to −86 and −81, respectively. The (−) sign indicates a negative control strain that contains the reporter plasmid and a plasmid expressing Sp-dCas9, the activation domain and an off-target scRNA (J206). Bars in **a** and **b** represent the averagerom the *E. coli* promoter±standard deviation calculated from $n = 3$ biologically independent samples. Source data are provided as a Source Data file.

(the six genes from Fig. 6b and *ldhA* from Supplementary Fig. 1), we were able to activate three genes with more than twofold increases in gene expression.

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

Discussion

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

The most unexpected property that we observed with bacterial CRISPRa was its sharp, periodic dependence on-target site position. This behavior is quite distinct from CRISPRa in eukaryotes, where a broad range of sites upstream of the TSS are effective[40], possibly because eukaryotic activators typically recruit transcription factors

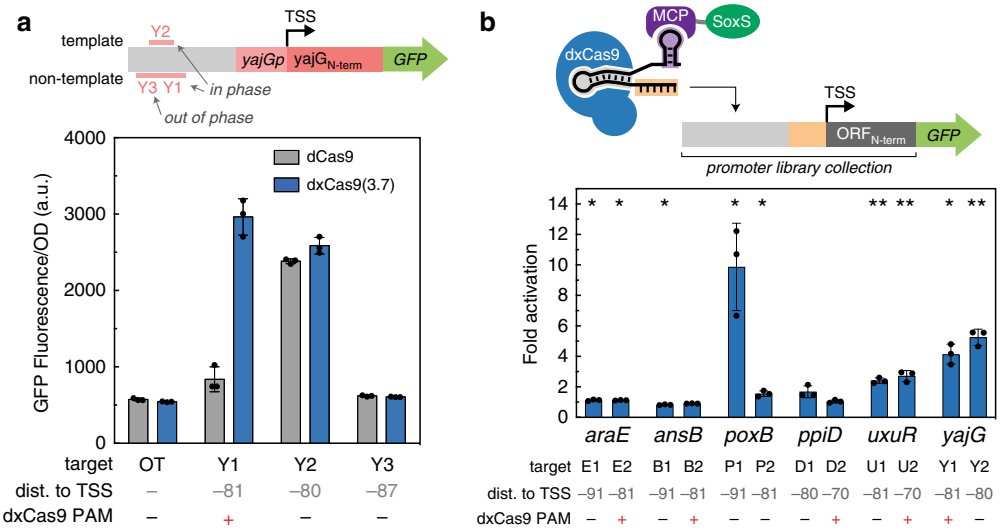


Fig. 6 Predictive rules enable endogenous activation. a CRISPRa using dCas9 and dxCas9(3.7) was targeted to a *yajG* reporter plasmid from the *E. coli* promoter collection[36]. Three scRNA target sites were selected; two sites were located at the positions where CRISPRa was most effective (Y1–2), and one was located out of phase (Y3). A negative control (OT) expressing an off-target scRNA (J306) was included. **b** CRISPRa was targeted to *yajG* and five additional promoters from the *E. coli* promoter collection (Supplementary Methods). Two scRNA sites located at the positions where CRISPRa was most effective were targeted for each gene using dxCas9(3.7). Samples are arranged by baseline expression of the target genes, in ascending order left to right. Fold activation indicates the median fluorescence of strains relative to an off-target control (J306). Values **a** and **b** represent the average±standard deviation calculated from $n = 3$ biologically independent samples. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's *t* test (**p*-value < 0.05, ***p* value < 0.01). Exact *p* values: E1: 0.036, E2: 0.024, B1: 0.031, B2: 0.141, P1: 0.033, P2: 0.021, D1: 0.088, D2: 0.585, U1: 0.0008, U2: 0.001, Y1: 0.013, Y2: 0.003. Source data are provided as a Source Data file.

and chromatin modifying machinery rather than directly recruiting RNA polymerase. There is precedent for bacterial transcriptional activators that are sensitive to target site periodicity[27–34], but the marked changes in activity with only single base shifts is surprising. Moreover, it is puzzling that we were unable to predictably alter or broaden the range of sites that are effective. Our models for how activators interact with bacterial transcription machinery may be incomplete. It will likely be productive to continue screening for activity at out-of-phase target sites using additional systematic modifications to the CRISPRa complex structure, alternative CRISPR–Cas systems, and additional candidate transcriptional activation domains.

Despite the challenges described above for identifying effective CRISPRa sites in *E. coli*, our systematic characterization provides a framework for immediate practical applications and a path for future improvements. We now have a clear understanding of the criteria needed to design synthetic promoters that can be regulated by CRISPRa, which will enable the construction of complex, tunable synthetic multi-gene circuits. To extend the scope of CRISPRa to endogenous target genes, expanded PAM variants like dxCas9(3.7)[6], or orthologous dCas9 proteins with alternate PAM specificities[41,42] will open more DNA sites for targeting, increasing the likelihood of finding a targetable site at an optimal position relative to the TSS. These strategies lay the groundwork for more widespread use of bacterial CRISPRa in basic research and practical applications including functional genomics screens, metabolic engineering, and synthetic microbial communities.

Methods

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

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

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

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

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

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The data sets generated and analyzed during the current study are available from the corresponding author upon request. The source data underlying Figs. 1b and 2–6, as well as Supplementary Figs. 1, 2, 3A–B, and 4–11 are provided as a Source Data file.

Code availability

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

Received: 20 September 2019; Accepted: 4 March 2020;
Published online: 01 April 2020

References

1. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).
2. Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
3. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat. Commun.* **9**, 2489 (2018).
4. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
5. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome gditing and beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
6. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
7. Yang Liu, B. W. & Wan, X. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nat. Commun.* **10**, 3693 (2019).
8. Keseler, I. M. et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* **45**, D543–550 (2017).
9. Rodriguez, A. et al. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Micro. Cell Fact.* **13**, 126 (2014).
10. Byrne, C. R., N., M. K., Monroe, R. S. & Ward, K. A. 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).
11. Jiang, G. R. & Sonia Nikolova, D. P. C. Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology* **147**, 2437–2446 (2001).
12. Silander, O. K. et al. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genet.* **8**, 5 (2012).
13. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
14. Alejandro Chavez, G. C. et al. Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).
15. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
16. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).
17. Shah, I. M. & R., E. W. Novel protein–protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase α subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *J. Mol. Biol.* **343**, 513–532 (2004).
18. Gort, A. S., J., A. I. & Ferber, D. M. The regulation and role of the periplasmic copper, zinc superoxide dismutase of *Escherichia coli*. *Mol. Microbiol.* **32**, 179–191 (1999).
19. Gen Nonaka, V. A. R., Blankschien, M., Herman, C. & Gross, C. A. Regulon and promoter analysis of the *E. coli* heat-shock factor, σ^{32}, reveals a multifaceted cellular response to heat stress. *Genes Dev.* **20**, 1776–1789 (2006).
20. Rhodius, V. A., Won Chul Suh, C. A. G., Nonaka, G. & West, J. Conserved and variable functions of the σ^{E} stress response in related genomes. *PLoS Biol.* **4**, e2 (2006).
21. Zhe-Xian, T., Y., W., Quan-Sheng, L., Buck, M. & Kolb, A. 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).
22. Yona, A. H., G., J. & Alm, E. J. Random sequences rapidly evolve into de novo promoters. *Nat. Commun.* **9**, 1530 (2018).
23. Lee, T. et al. BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *J. Biol. Eng.* **5**, 12 (2011).
24. Mendoza-Vargas, A. et al. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* **4**, e7526 (2009).
25. Georgina Lloyd, S. B. & Paolo, L. Activation and repression of transcription initiation in bacteria. *Essays Biochem.* **37**, 17–31 (2001).
26. Schneider, T. D. 70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence. *Nucleic Acids Res.* **38**, 5995–6006 (2010).
27. 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).
28. Zhou, Y., W., Y.-P., Kolb, A. & Busby, S. J. W. Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Res.* **42**, 9209–9216 (2014).
29. Johannes Müller, B. M.-H., Barker, A. & Oehler, S. Dimeric *lac* repressors exhibit phase-dependent co-operativity. *J. Mol. Biol.* **284**, 851–857 (1998).
30. Straney, D. C., Straney, S. B. & Crothers, D. M. Synergy between *Escherichia coli* CAP protein and RNA polymerase in the *lac* promoter open complex. *J. Mol. Biol.* **206**, 41–57 (1989).
31. Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. Stringent spacing requirements for transcription activation by CRP. *Cell* **62**, 733–743 (1990).
32. Ushida, C. & Aiba, H. Helical phase dependent action of CRP: effect of the distance between the CRP site and the −35 region on promoter activity. *Nucleic Acids Res.* **18**, 6325–6330 (1990).
33. Müller, J., Oehler, S. & Müller-Hill, B. Repression of *lac* Promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J. Mol. Biol.* **257**, 21–29 (1996).
34. Martin, R. G., Gillette, W. K., Rhee, S. & Rosner, J. L. Structural requirements for marbox function in transcriptional activation of *mar*/*sox*/*rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. *Mol. Microbiol.* **34**, 431–441 (1999).
35. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
36. Zaslaver, A. et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **3**, 623–628 (2006).
37. Kiani, S. et al. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11**, 723–726 (2014).
38. Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M. & Klavins, E. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat. Commun.* **8**, 15459 (2017).
39. Rojo, F. Repression of transcription initiation in bacteria. *J. Bacteriol.* **181**, 2987–91 (1999).
40. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).

41. Leenay, R. T. & B., C. L. Deciphering, communicating, and engineering the CRISPR PAM. *J. Mol. Biol.* **429**, 177–191 (2017).
42. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
43. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
44. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: Identifying amino acids important for DNA binding and transcription activation. *J. Mol. Biol.* **322**, 237–257 (2002).

Acknowledgements

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

Author contributions

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

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-15454-y.

Correspondence and requests for materials should be addressed to J.M.C. or J.G.Z.

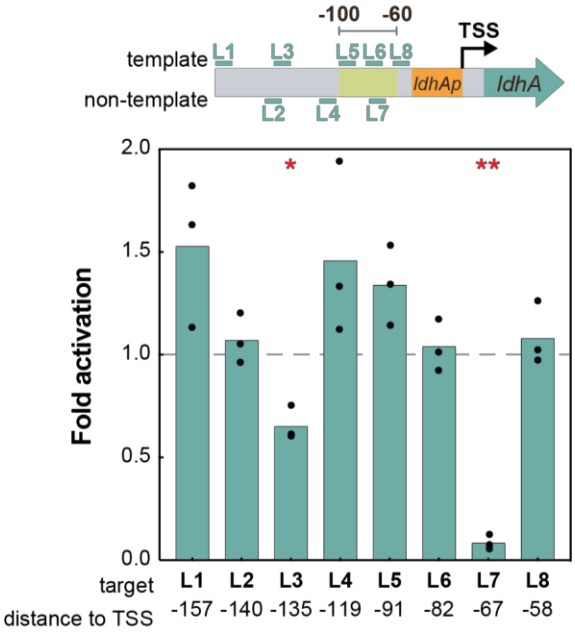
Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

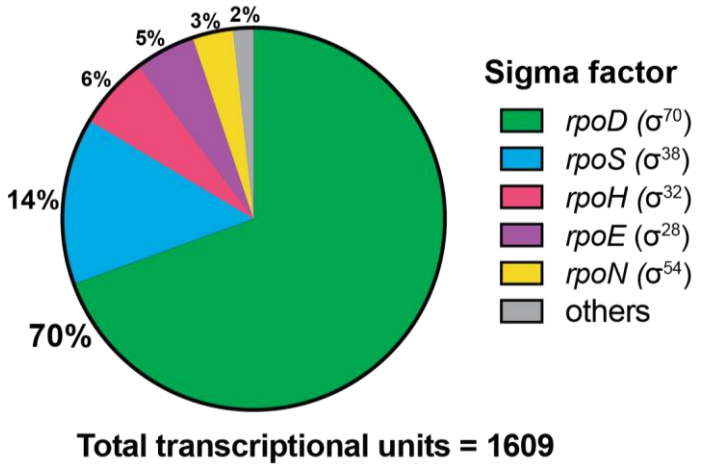
Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements

Fontana and Dong *et al.*

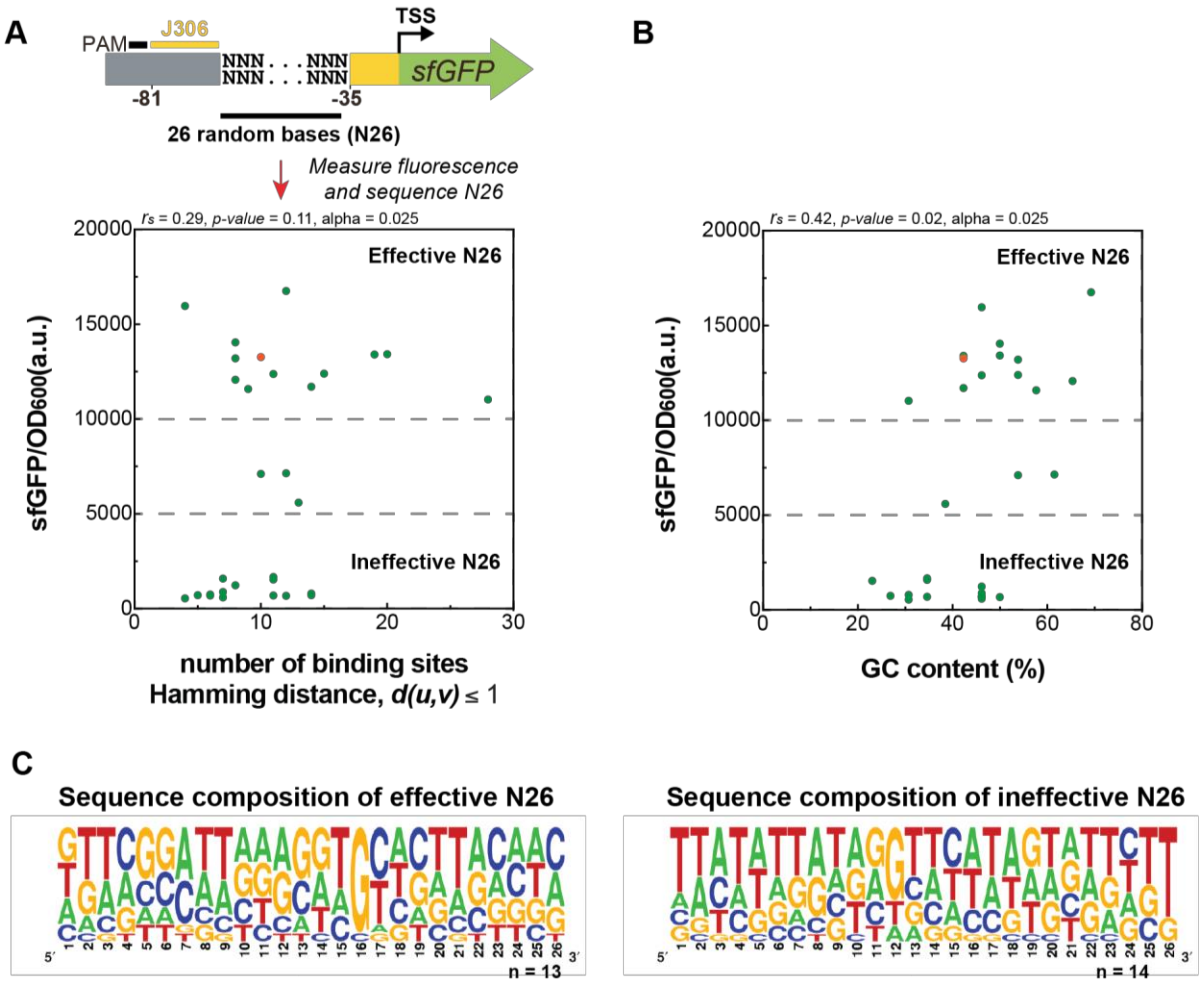


Supplementary Figure 1. CRISPRa at the endogenous gene target *ldhA* does not follow predicted trends.

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



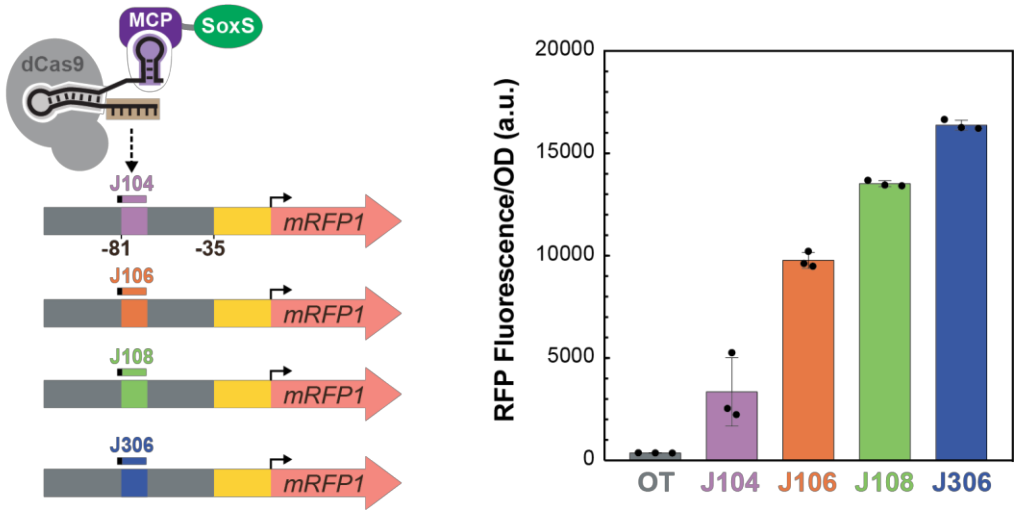
Supplementary Figure 2. Distribution of transcriptional units regulated by sigma factors. The number of *E. coli* transcriptional units regulated by sigma factors was obtained from Ecocyc[1] on the "Regulon" tab of the page relative to each sigma factor. The total number of transcriptional units represents the sum of the transcriptional units regulated by each sigma factor. Source data are provided as a Source Data file.



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

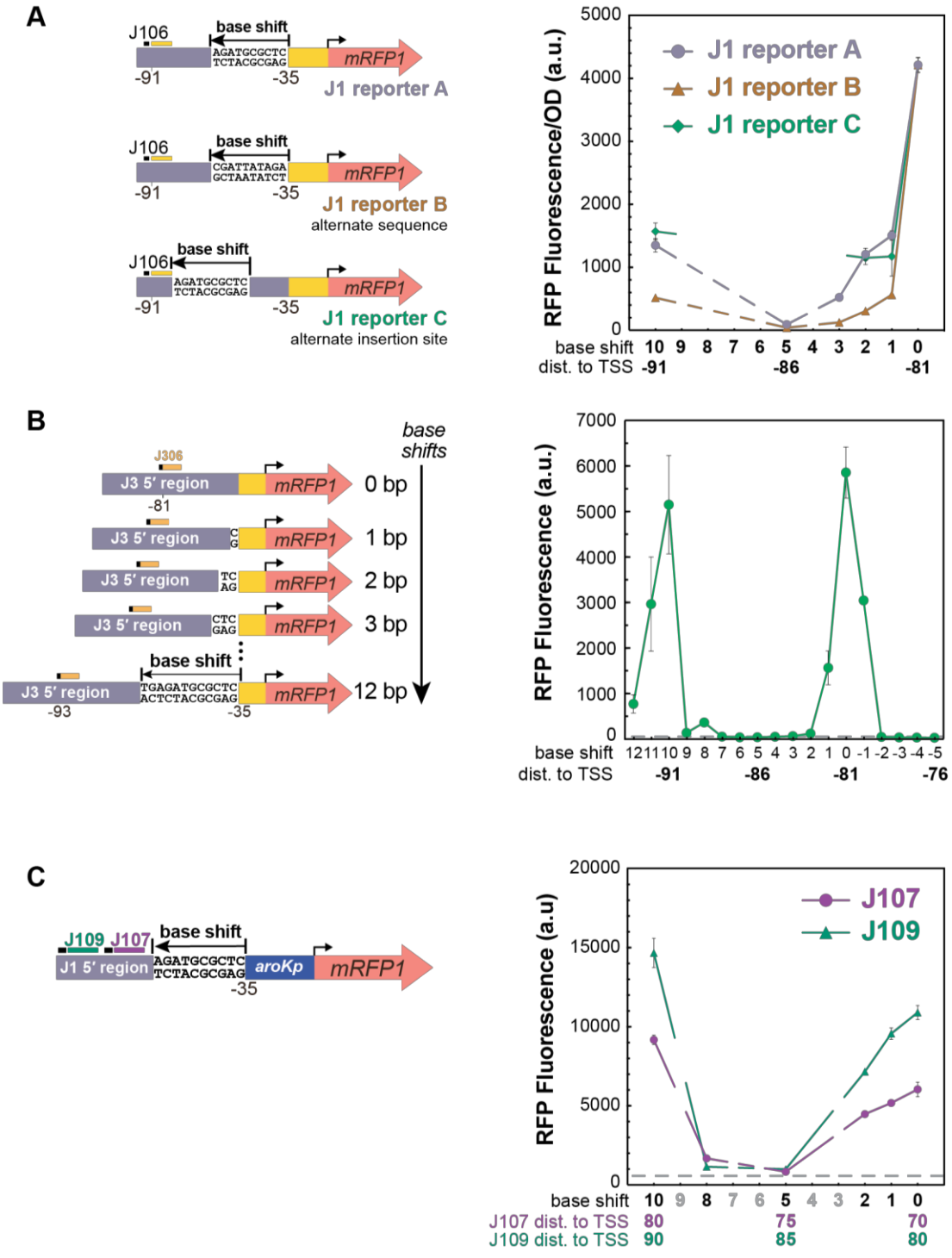
A) Plot of gene activation vs. transcription factor binding sites present in the intervening sequence between the scRNA target site and the minimal promoter. CRISPRa was targeted to a reporter library with 26 randomized bases between the scRNA target site and the -35 region on the J3-J23117-sfGFP reporter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

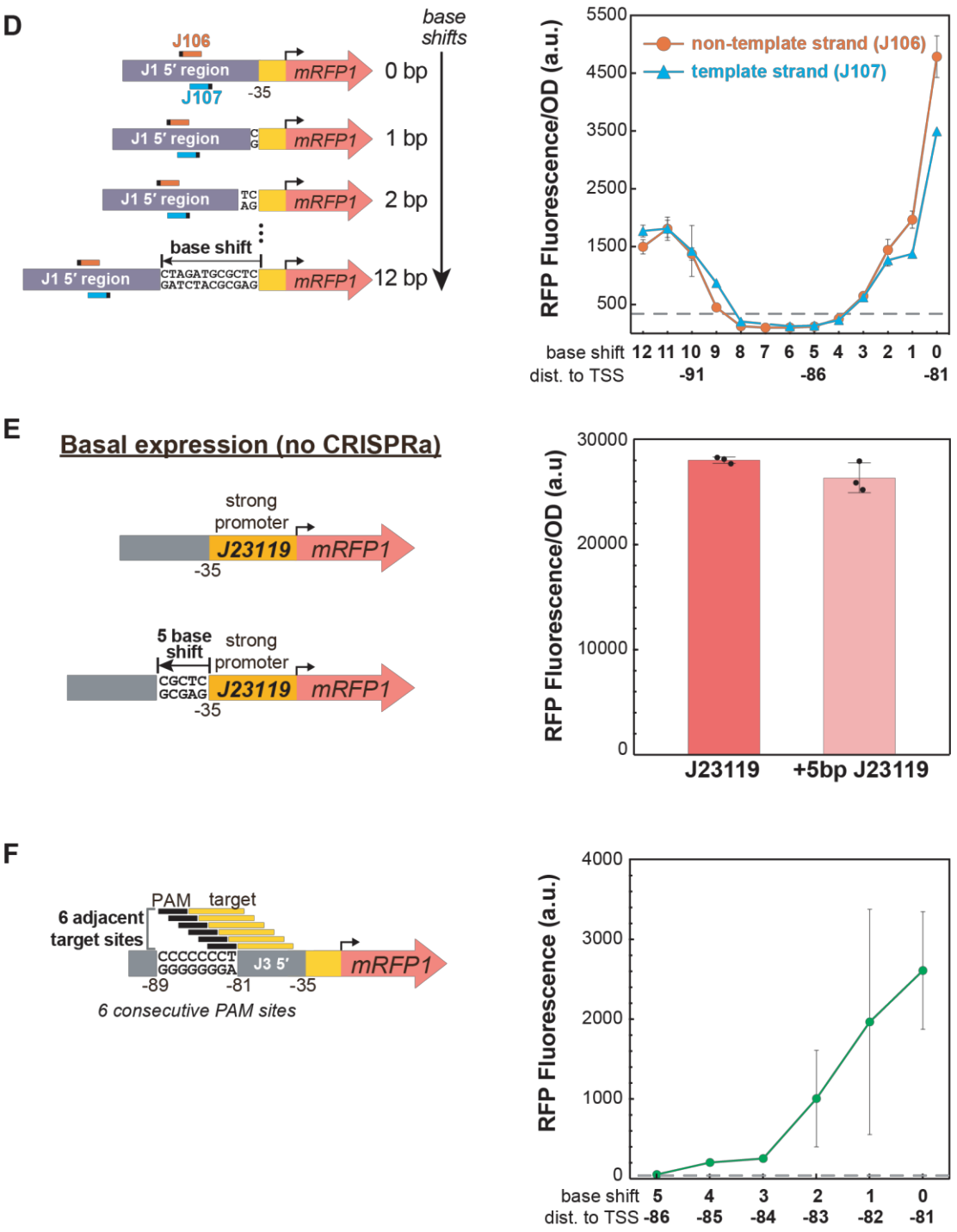
0.025. Green dots indicate the sfGFP/OD_{600} values and the number of binding sites calculated from individual colonies, and the orange dot indicates a strain where CRISPRa is targeting the original J3-J23117-sfGFP promoter. **B)** Plot of gene activation vs. GC content in the intervening sequence between the scRNA target site and the minimal promoter. r_s indicates the Spearman rank order correlation coefficient between sfGFP/OD_{600} and GC content, and its associated two-tailed *p-value* is relative to the null hypothesis of no correlation between sfGFP/OD_{600} and GC content, with a Bonferroni-corrected α = 0.025. In panels A and B, green dots indicate the sfGFP/OD_{600} values and GC content calculated from n = 1 biologically independent samples, and the orange dot indicates a strain where CRISPRa targets the original J3-J23117-sfGFP promoter. **C)** Logo plots (https://weblogo.berkeley.edu/) showing the base composition of N26 sequences effective for CRISPRa (sfGFP/OD_{600} > 10000 a.u., 13 sequences) and N26 sequences ineffective for CRISPRa (sfGFP/OD_{600} < 5000 a.u., 14 sequences). Source data of Supplementary Figure 3A and 3B are provided as a Source Data file.



Supplementary Figure 4. CRISPRa activity depends on the target sequence on the scRNA.

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


A
J106
base shift
AGATGCGCTC
TCTACGCGAG
mRFP1
-91
-35
J1 reporter A
J106
base shift
CGATTATAGA
GCTAATATCT
mRFP1
-91
-35
J1 reporter B
alternate sequence
J106
base shift
AGATGCGCTC
TCTACGCGAG
mRFP1
-91
-35
J1 reporter C
alternate insertion site
RFP Fluorescence/OD (a.u.)
5000
4000
3000
2000
1000
0
J1 reporter A
J1 reporter B
J1 reporter C
base shift 10 9 8 7 6 5 4 3 2 1 0
dist. to TSS -91 -86 -81
B
base shifts
J306
J3 5′ region
mRFP1
0 bp
-81
J3 5′ region
C
G
mRFP1
1 bp
J3 5′ region
TC
AG
mRFP1
2 bp
J3 5′ region
CTC
GAG
mRFP1
3 bp
base shift
J3 5′ region
TGAGATGCGCTC
ACTCTACGCGAG
mRFP1
12 bp
-93
-35
RFP Fluorescence (a.u.)
7000
6000
5000
4000
3000
2000
1000
0
base shift 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5
dist. to TSS -91 -86 -81 -76
C
J109
J107
base shift
J1 5′ region
AGATGCGCTC
TCTACGCGAG
aroKp
mRFP1
-35
RFP Fluorescence (a.u)
20000
15000
10000
5000
0
J107
J109
base shift 10 9 8 7 6 5 4 3 2 1 0
J107 dist. to TSS 80 75 70
J109 dist. to TSS 90 85 80

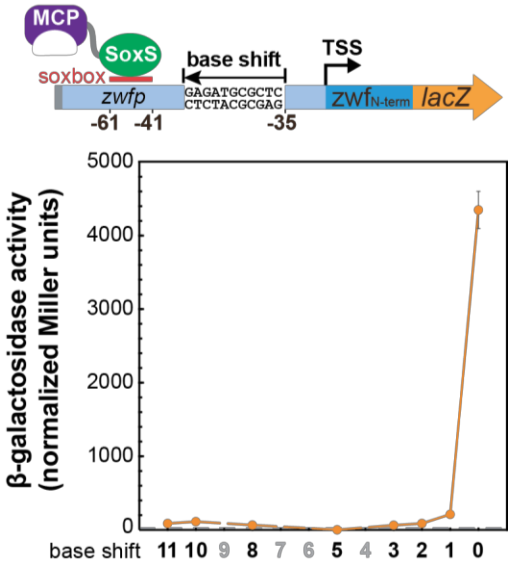


Supplementary Figure 5. The sharp positioning dependence of CRISPRa is observed across multiple promoters.

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

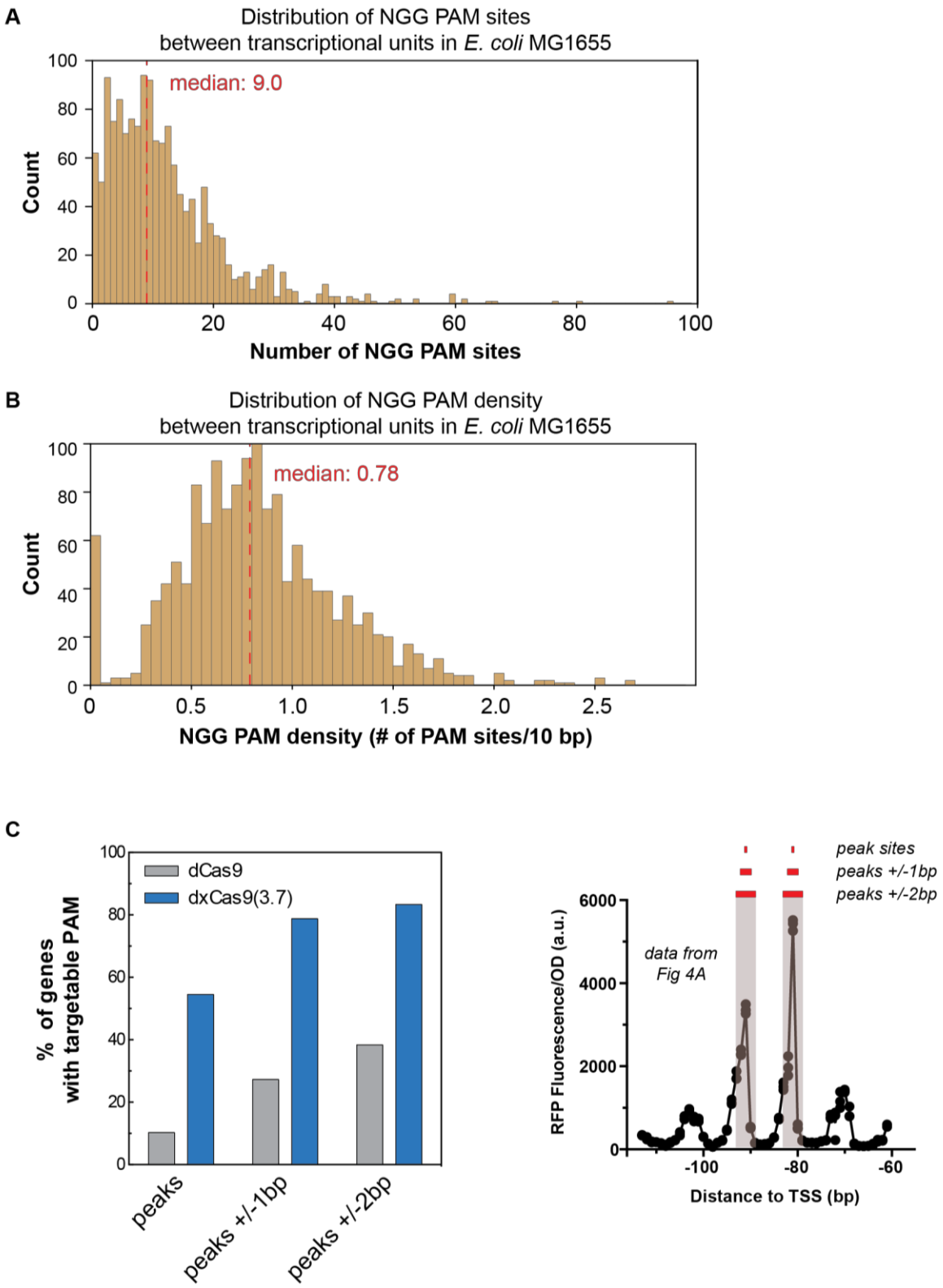
site. In the J1 reporter C, bases were inserted downstream of the J106 target site. There were modest differences between reporter A and reporter B that could indicate a contribution from sequence composition. **B)** The sharp positioning requirements of CRISPRa were observed with a different heterologous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

consecutive PAM sites on the non-template strand were introduced by placing a CCCCCCCT sequence between -89 and -81 bp to the TSS on the J3-J23117-mRFP1 reporter. Maximum gene expression was observed at the original -81 site, after which expression gradually decreased to one third of the maximum activity after moving 2 bp away (-83). Gene expression decreased further when the scRNA target was moved 3 bp and 4 bp away from the TSS (-84 and -85), and reached the baseline when moved 5 bp (-86). The grey line represents the baseline activity of the reporter strain containing an empty vector instead of the CRISPRa component plasmid. Values in panels A-F represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.



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

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

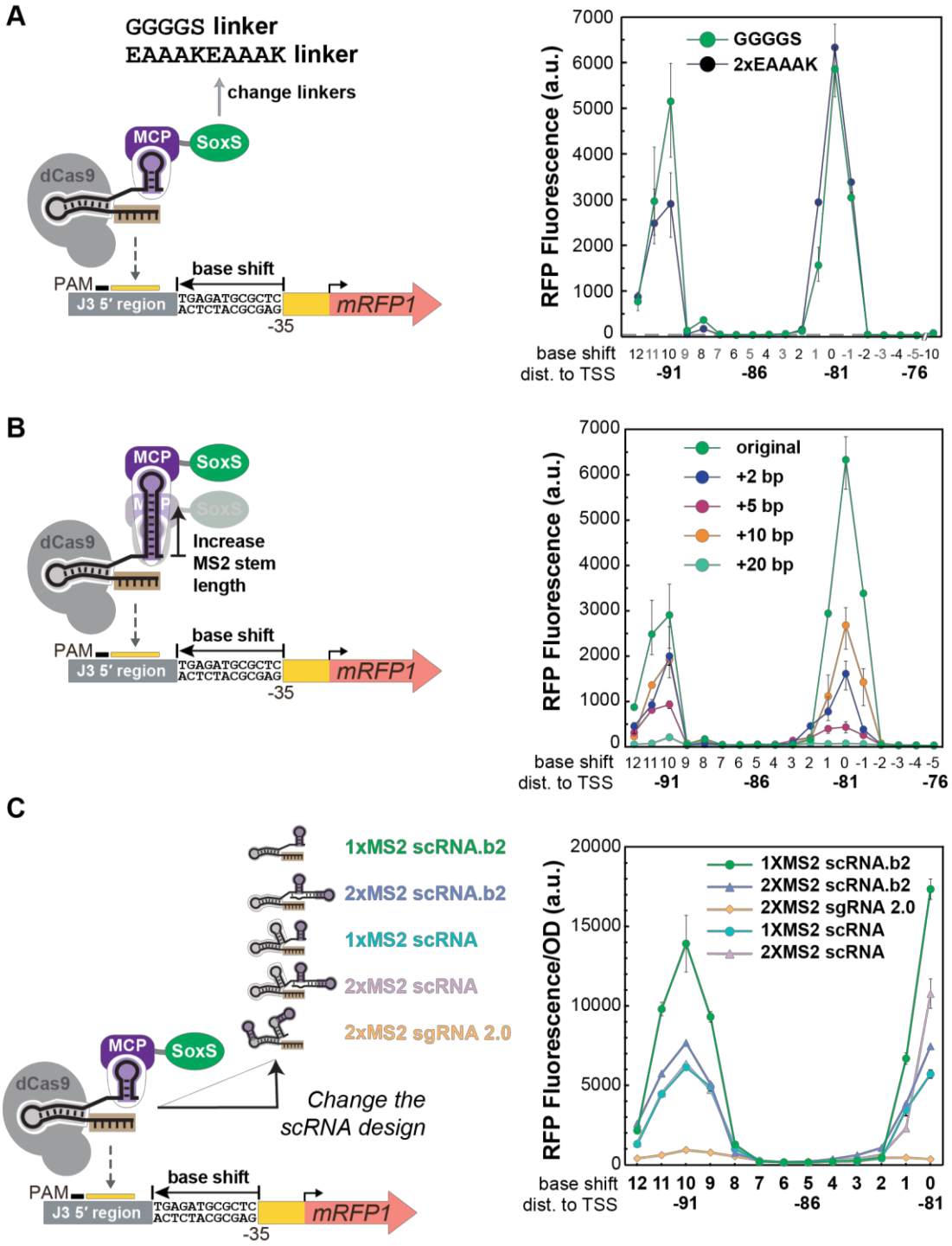


Supplementary Figure 7. Availability of PAM sites between transcriptional units in *E. coli* MG1655.

A) Distribution of the number of NGG PAM sites between transcriptional units in *E. coli*

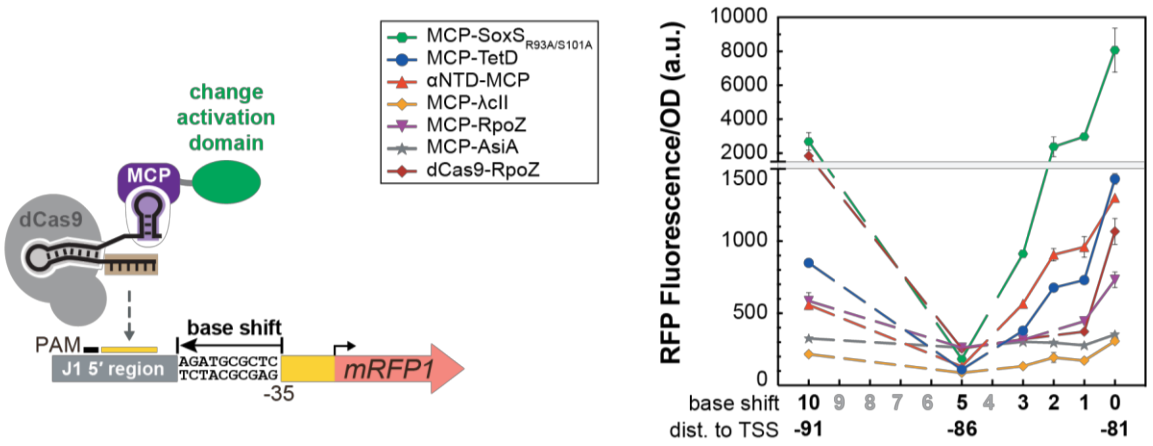
MG1655. The median number of NGG PAM sites between transcriptional units in *E. coli* MG1655 is 9.0. Methods for extracting the sequence of the DNA regions between transcriptional units in *E. coli* MG1655 and identifying available PAM sites are described in the Supplementary Methods.

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



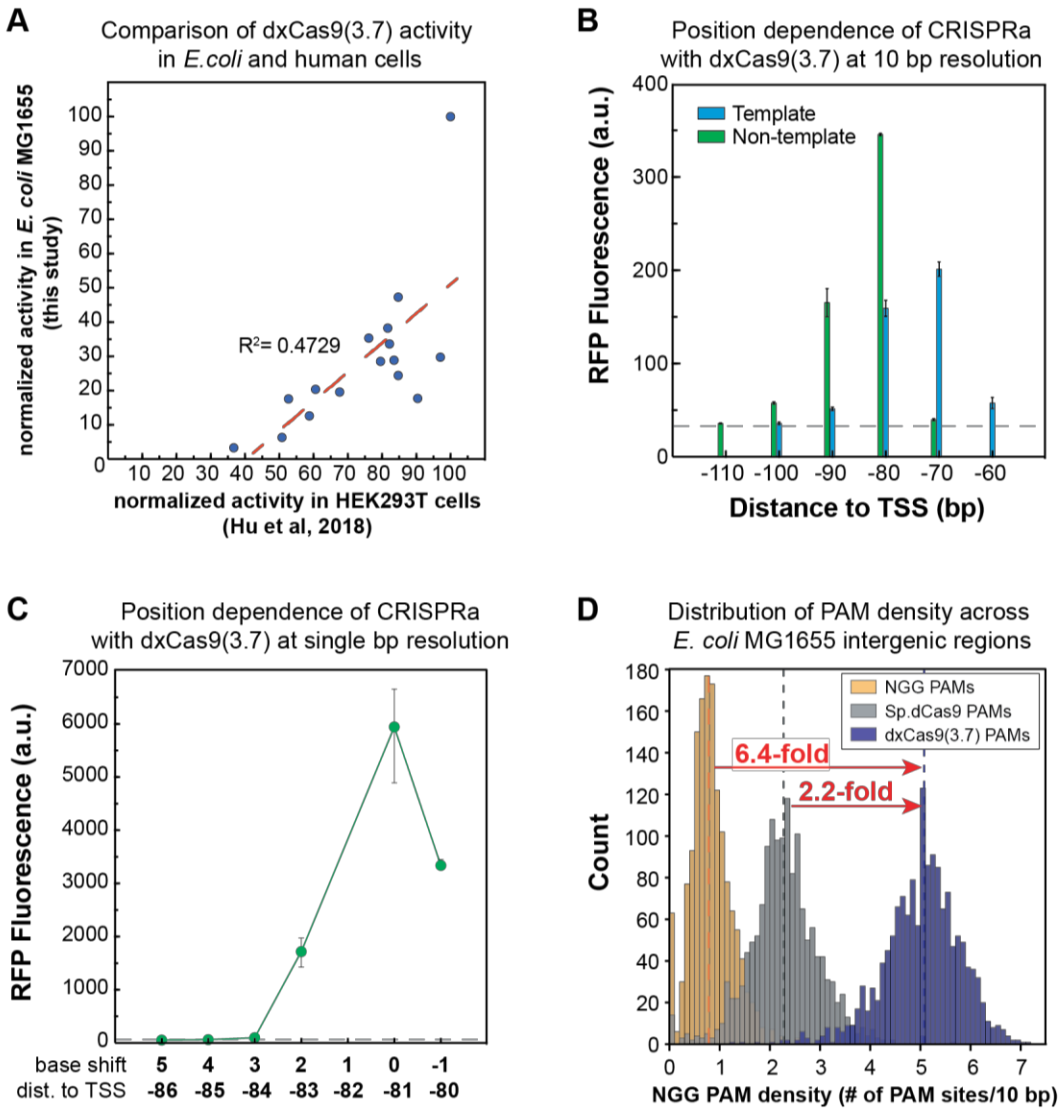
Supplementary Figure 8. Modifying the CRISPRa complex structure does relax the sharp positioning requirements of CRISPRa.

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



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

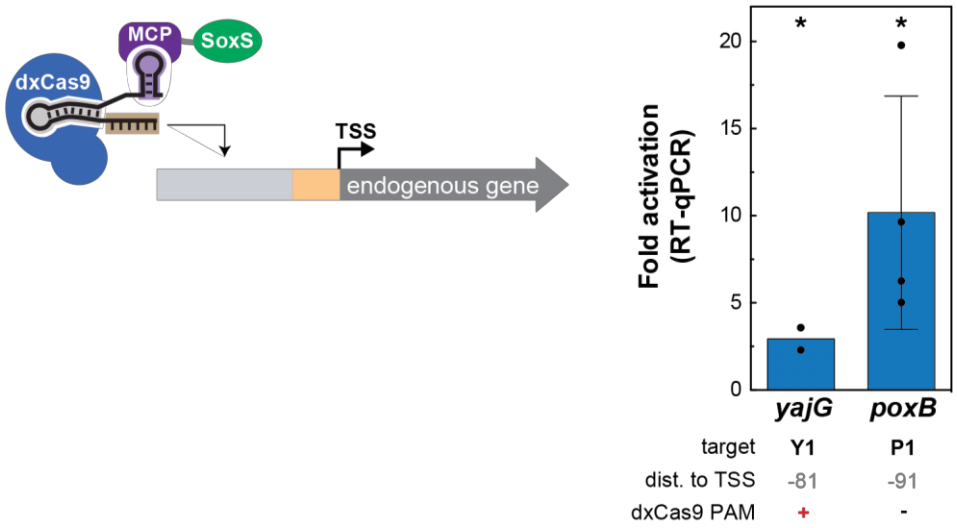
CRISPRa with MCP-TetD and αNTD-MCP activation domains[3] displayed similar positioning dependence as MCP-SoxS(R93A/S101A), where gene expression decreases as the position shifts from 0 to 3 bp, reaches the baseline at a 5 bp shift, and increases again at a 10 bp shift. CRISPRa with MCP-RpoZ[3] and dCas9-RpoZ[8]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



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

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

calculated from n = 3 biologically independent samples. The grey dotted line represents the baseline fluorescence of a strain containing the dxCas9(3.7) and MCP-SoxS(R93A/S101A) and an empty vector with no scRNAs. **C)** CRISPRa with dxCas9(3.7) displayed similar positioning dependence with single base shifts. Gene expression was significantly reduced when the scRNA target site was shifted 1-2 bp in either directions from the optimal position -81 bp to the TSS on the non-template strand. Shifting the scRNA target site 3-5 bp causes gene expression to fall to the baseline level. Reporter gene sets were constructed with 1 bp deleted or 2-5 bp inserted upstream of the -35 of the J3-J23117-mRFP1 reporter. The grey line represents the baseline activity of a strain containing the J3-J23117-mRFP1 reporter plasmid and a CRISPRa component plasmid with an off-target scRNA (hAAVS1). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. **D)** dxCas9(3.7) increases the probability of finding targetable PAM sites for CRISPRa. Histograms showing the distribution of the PAM density (defined in Supplementary Figure 7B) between transcriptional units in *E. coli* MG1655 suggest that the average likelihood of finding a dxCas9(3.7)-compatible PAM (blue) is ~6.4-times higher than finding an NGG PAM (yellow), and ~2.2-times higher than finding a *Sp*-dCas9-compatible PAM (grey). Vertical dotted lines indicate the median PAM density for each group. The median PAM density is 5.08 for dxCas9(3.7)-compatible PAMs (NGG, AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA), 2.33 for *Sp*-dCas9-compatible PAMs (NGG, AGA, CGA, GGA, GGC, GGT, TGA). and 0.78 for NGG PAMs. Methods for extracting the sequence of *E. coli* MG1655 intergenic regions, counting targetable PAM sites and calculating the PAM density are described in the Supplementary Methods. Source data are provided as a Source Data file.



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

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

Supplementary Table 1. *E. coli* strains.

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

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

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

	(3.7)	scRNA.b2	U1, U2, D1, D2, B1, B2, E1, E2, J306	SoxS(R93A/S101A), tet-inducible
Supplementary Figure 1	dCas9	1xMS2 scRNA.b2	L1-L8, hAAVS1	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 3A	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 4	dCas9	1xMS2 scRNA.b2	J104, J106, J108, J306, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5A	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5B	dCas9	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 5C	dCas9	1xMS2 scRNA.b2	J107, J109	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 5D	dCas9	1xMS2 scRNA.b2	J106, J107, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5F	dCas9	1xMS2 scRNA.b2	J306, J306+1-5	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 8A	dCas9	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A), MCP-(2xEAAAK)-SoxS(R93A/S101A)
Supplementary Figure 8B	dCas9	1xMS2 scRNA.b2, 1xMS2 scRNA.b2 + 2/5/10 bp MS2 stem extension	J306	MCP-(2xEAAAK)-SoxS(R93A/S101A)
Supplementary Figure 8C	dCas9	1xMS2 scRNA.b2, 2xMS2	J306	MCP-(5aa)-SoxS(R93A)

		scRNA.b2, 2xMS2 sgRNA2.0, 1xMS2 scRNA, 2xMS2 scRNA		
Supplementary Figure 9	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A/S101A), Alternative activators
Supplementary Figure 10B	dxCas9 (3.7)	1xMS2 scRNA.b2	J104-J113	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 10C	dxCas9 (3.7)	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 11	dxCas9 (3.7)	1xMS2 scRNA.b2	Y1, P1, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible

Supplementary Table 3. gRNA target sites.

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

			(reporter B)
M2	CGACGAGCAGAAGTGTCAGC	NT	76 (reporter A), 81 (reporter B)
aroKB_A1	GGGCAATTATTTCGTCATGA	T	151
aroKB_A2	AGATGAACGACGCGAGTTAG	T	122
aroKB_A3	TTTTACGGCTGTTTACTCAC	NT	92
aroKB_A4	TGAGTAAACAGCCGTAAAAG	T	71
cysK_C1	CCACCCCTGTTTCACACAAA	NT	93
cysK_C2	AAACCGTTTGTGTGAAACAG	T	79
cysK_C3	GACATGCAAGATGGAATAAG	NT	61
cysK_C4	ATGACATGCAAGATGGAATA	NT	59
cysK_C5	GGAAATAATGACATGCAAGA	NT	52
ldhA_L1	CAGTAATAACAGCGCGAGAA	T	157
ldhA_L2	GGATATTAACTACCCATGCT	NT	140
ldhA_L3	GCTTTATATTTACCCAGCAT	T	135
ldhA_L4	GCTTAATTTTTCGCTAAATC	NT	119
ldhA_L5	GAAAAATTAAGCATTCAATA	T	91
ldhA_L6	AGCATTCAATACGGGTATTG	T	82
ldhA_L7	AGGCGCAACCTTCAACTGAA	NT	67
ldhA_L8	ATGTTTAACCGTTCAGTTGA	T	58
yajG_Y1	TTGACGAAATAATCGCCCCT	NT	81
yajG_Y2	CATCAGTGTTTCTTTTACCA	T	80
yajG_Y3	AAATAATCGCCCCTGGTAAA	NT	87
poxB_P1	CCCGATGAAAGGAATATCAT	NT	91
poxB_P2	GGTTAAATAGCCCGATGAAA	NT	81
uxuR_U1	TGATTGACCAGTAAGTCTGT	NT	81
uxuR_U2	GATTACCCTACAGACTTACT	T	70
ppiD_D1	ACTAAGCGTTGTCCCCAGTG	T	80
ppiD_D2	GTCCCCAGTGGGGATGTGAC	T	70
ansB_B1	AGATCTACAAAGTTAGAGGC	NT	91
ansB_B2	TATATTTTGGAGATCTACAA	NT	81
araE_E1	TGCGACATGTCGTTATGTGA	NT	91
araE_E2	ATTAAATTGCTGCGACATGT	NT	81

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

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

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

Supplementary Table 4. Select *E. coli* expression plasmids[a].

Plasmid	Marker	origin	Promoter	Gene	Terminator
pCD442	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A)	*1) BBa_B0015* *2) BBa_B1002*
pCK005.1-21[b]	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J101-121 targets)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pCD564	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107*	1) dxCas9(3.7) 2) MCP-(5aa)-SoxS (R93A/S101A)	*1) BBa_B0015* *2) BBa_B1002*
pCD565	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dxCas9(3.7) 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J306 target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pCD580.-1~-5	*AmpR*	*pSC101***	*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	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*

				scRNA.b2 (J306 target)	
pJF215.x	*CmR*	*p15A*	*1) Sp.pCas9* *2) TetR-pTet* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (variable target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pJF076Sa[c]	*AmpR*	*pSC101***	*J3_BBa_J23117*	mRFP1	*BBa_B0015*
pJF143-J3[d]	*AmpR*	*pSC101***	*J3_BBa_J23117*	mRFP1	*BBa_B0015*
pJF155.1-12[c]	*AmpR*	*pSC101***	*J1_BBa_J23117 (with 1-12 bp 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. Sp.pCas9 is the endogenous Cas9 promoter from *S. pyogenes*.

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

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

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

Supplementary Table 5. Primer sequences for RT-qPCR.

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

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

Index	sfGFP/OD_{600}[a]	Sequence[b]	TF binding sites, consensus[c]	TF binding sites, $d(u,v) \leq 1$[c]
1	549.87	TATCATAGTGATGACCAGTAAACATT	0	4
2	602.70	TGCAGAAAAGGGCTCTAGTGACTTGT	0	7
3	682.05	ATCTATGCGACGTCAAACGTGATGGG	0	12
4	702.47	GCCTTTCAGTGATCCCTTGTGATCTT	0	11
5	704.82	CAATTCTAGCAAGAAATTGACTTCGT	0	6
6	710.25	TTTACGGATTGGTGCGCTATTGGGTT	0	14
7	716.04	TAATGTTGATCGCCTCTATGCGTCCT	0	5
8	746.08	TTATTATAATAGTTTTGAACGTGCCT	0	6
9	794.37	CAATATGACGTGTTGTTAATTTGGTT	1	14
10	883.74	GGTCTGTGTACTCGAACACGAGATTT	0	7
11	1230.90	TTGCTCCGCGGTTTCTCTGTAAATGT	0	8
12	1532.80	ATAAATTGCATGATTAAAGACTATTG	0	11
13	1586.70	TTCGGATTTAAGGAGATATGTTTACG	0	7
14	1679.31	TAATAGGCTAGGAATTTGAAGGGATT	0	11
15	5604.77	GCATCTACATAGTGGTACAATTGAAG	0	13
16	7110.46	GCGTCCATATATGCTGATGGTGTGGG	0	10
17	7143.40	GCGGGCTATGCCTTGGAGGGTTGTAG	0	12
18	11027.55	ATTACACACGAAGGTGTATTTACTAT	2	28
19	11589.46	TTATCTAATAGGGGCGCCCCGCTTC	0	9
20	11714.54	TTAAACAAACGAAAAGTTCGCGCCTG	1	14
21	12073.23	GATCGGGTAGGGCGCCGACAAAGGGA	0	8
22	12374.02	TGAGGGATATAGAGAGATGTTCAGCC	0	11
23	12399.33	CATCGCCGTTGCGATGCATTTTGACG	0	15
24	13207.42	GTTGTCCTTCTAGTCGCCCATGACTC	0	8
J3	**13269.13**	**CGTCGTCTTGAAGTTGCGATTATAGA**	0	10
25	13413.88	GTCGTAAATAAGTAAGTCACTCCCAC	2	19
26	13434.32	TTGAGGCCCATGCTTGTGGAAATGAC	1	20
27	14053.61	GGCAAGATGCCTCGTGCAGTAGAATA	0	8
28	15973.97	AGTCGCTGAGTAGATGTTCGTAGAGA	0	4
29	16760.50	ACACCGACTACCCCTGCTGGGCCCAG	0	12

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

sfGFP/OD_{600} of a strain expressing the J3-J23117-sfGFP promoter are included for comparison and are in bold.

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

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

Supplementary Methods

Computational analysis of PAM site availability

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

Selection of candidate endogenous genes for CRISPRa

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

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

> MCP-(5aa)-SoxS(R93A/S101A) (optimized for minimal endogenous activity, Figure 1 and subsequent)

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

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG
AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMSHQKIIQDLIAWIDE
HIDQPLNIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQ
TFSRVFARQFDRTPADYRHRL

Reporter genes

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

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

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

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

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGATTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCGAATT
CATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTT
CGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAG
GTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCC
GCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCC
TTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGG
ACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGG
TCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCT
CTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCA

CCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTC
CCACAACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAA
GGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTT
TCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCT
TTCTGCGTTTATA

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

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

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAANNNNNNNNNNNNNNNNNNNNNNNNTTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCGAATT
CATTAAAGAGGAGAAAGGTACCATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTT
GAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACG
GAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTAC
TCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGT
GCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTG
CTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGA
TGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAA
CAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAG
ACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGAC
ACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCT
GCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAGGATCCAAACTCGAGTAAGGATCTCCAGGCA
TCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCT
CTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA

Modified synthetic promoters

Promoters with different basal expression levels (Figure 3)

Annotations:

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

BBa_J23105

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtactatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23106

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtatagtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23107

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagccctaggtattatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23108

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGActgacagctagctcagtcctaggtataatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23109

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacagctagctcagtcctagggactgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23110

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtacaatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23111

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtatagtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23112

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGActgatagctagctcagtcctagggattatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23113

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGActgatggctagctcagtcctagggattatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23114

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttatggctagctcagtcctaggtacaatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23115

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttatagctagctcagcccttggtacaatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23118

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC

ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23119

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctaggtataatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Promoters regulated by alternative sigma factors (Figure 3)

Annotations:

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

J1-*sodCp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGCCGTTcaaaaatgtgtcacTGGTTTACACTtattcagGGAAT
TCATTAAAGAGGAGAAAGGTACCATG

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

J1-*glnAp2* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACTGGCACagatttCGCTTtatctttttTacggcgacGAATT
CATTAAAGAGGAGAAAGGTACCATG

J1-*rdgBp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG

TCCTTTGGGTTCCACCGGATACCTCCGGACTTGTAAaaggcgaacttggcCGCCACaaacaaattGAATT
CATTAAAGAGGAGAAAGGTACCATG

J1-*yieEp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACGAACTTttagccgctttagtctgtcCATCAttccAGAATT
CATTAAAGAGGAGAAAGGTACCATG

J3(tetO)-J23117_mRFP1 reporter (Figure 3)

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

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCACTCTATCGTTGATAGAGTttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

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

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

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

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACCTAGATGCGCTCttgacagctagctcagtcctagggattg
tgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

J3-J23117 promoter with shifted bases (Supplementary Figure 5 and subsequent)

For promoters shifted by fewer than 12bp, bases were removed from the 5' end of the inserted sequence (starting with T). For promoters with negative shifts, no sequence was inserted between the J3 upstream region and BBa_J23117, and bases were removed from the 3' end of the J3 upstream region (starting with A).

J3 upstream region, inserted sequence (12 bases), BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGATGAGATGCGCTCttgacagctagctcagtcctagggattg
tgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

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

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACCCCCCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

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

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTNNNACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Promoter to demonstrate dxCas9(3.7) versatility (Figure 5B)
J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP. The modified target site (M) is underlined. AGT PAM site.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACACTGCTGACACTTC

TGCTCGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Supplementary References

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

10. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3**, 623–628 (2006).

11. Donghyuk Kim, B. Ø. P., Jay Sung-Joong Hong, Yu Qiu, Harish Nagarajan, Joo-Hyun Seo, Byung-Kwan Cho, Shih-Feng Tsai. Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genetics* **8**, e1002867 (2012).

Appendix 2: Kruyer et al,. 2021 ACS synbio

ACS SyntheticBiology

pubs.acs.org/synthbio Review

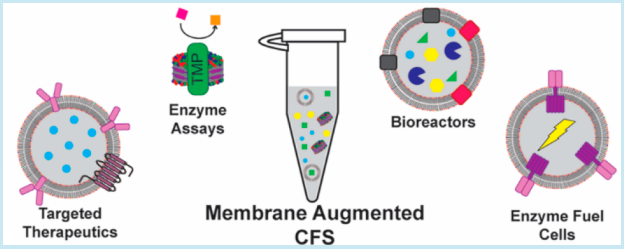
Membrane Augmented Cell-Free Systems: A New Frontier in Biotechnology

Nicholas S. Kruyer, Widianti Sugianto, Benjamin I. Tickman, Diego Alba Burbano, Vincent Noireaux, James M. Carothers, and Pamela Peralta-Yahya*

Cite This: https://doi.org/10.1021/acssynbio.0c00625 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Membrane proteins are present in a wide array of cellular processes from primary and secondary metabolite synthesis to electron transport and single carbon metabolism. A key barrier to applying membrane proteins industrially is their difficult functional production. Beyond expression, folding, and membrane insertion, membrane protein activity is influenced by the physicochemical properties of the associated membrane, making it difficult to achieve optimal membrane protein performance outside the endogenous host. In this review, we highlight recent work on production of membrane proteins in membrane augmented cell-free systems (CFSs) and applications thereof. CFSs lack membranes and can thus be augmented with user-specified, tunable, mimetic membranes to generate customized environments for production of functional membrane proteins of interest. Membrane augmented CFSs would enable the synthesis of more complex plant secondary metabolites, the growth and division of synthetic cells for drug delivery and cell therapeutic applications, as well as enable green energy applications including methane capture and artificial photosynthesis.



KEYWORDS: *cell-free systems, membrane proteins, synthetic cells, natural products, liposomes*

The functional, heterologous expression of membrane proteins is one of the missing puzzle pieces in establishing industrially relevant biological processes ranging from the production of medicinal compounds to the capture of methane (CH_4) to the bioremediation of heavy metal pollutants (Figure 1A). Plant-based medicinal compounds are synthesized *via* multienzyme cascades composed of several transmembrane cytochromes P450 (CYPs) that decorate the compounds' scaffolds.[1,2] Particulate methane monooxygenase (MMO) oxidizes CH_4 to methanol,[3−5] which enters C1 assimilation pathways in natural and synthetic methanotrophs,[3] potentially able to convert the ∼650 million tons of CO_2 equivalents produced in the U.S.[6] into high density fuels to power trucks and airplanes. Heavy metals, such as uranium from nuclear waste, can be bioremediated using MtrCAB, which facilitates the transfer of electrons from the organism to the heavy metal.[7] Toward therapeutic applications, the robust functional heterologous expression of surface receptors would support the development of drug delivery vehicles and cell therapies. For instance, G protein-coupled receptors (GPCRs) are the target of more than 30% of FDA approved drugs. Routine heterologous expression of GPCRs would facilitate the development of high-throughput screening platforms for the discovery of new drugs or the study of signaling cascades in the absence of endogenous GPCRs.[8] Access to a wider array of functional receptors would also expand cell therapies beyond detection of cell surface antigens on cancer cells to the detection of soluble, small molecule ligands around the tumor to improve targeting and reduce on-target off-tumor toxicity, *i.e.*, targeting a non-tumor tissue expressing the same antigen.[9] For example, by using GPCRs, which mediate most cellular responses to small molecules.[10] Finally, transmembrane proteins are pivotal in primary metabolism, determining the biosynthetic performance of the production host. For instance, a network of transmembrane proteins synthesizes the phospholipids needed to build the cell's membranes. The oxidative phosphorylation pathway used to produce ATP in plants, bacteria, and humans is also composed of transmembrane proteins.

Application of membrane proteins is hindered by their difficult production outside their endogenous host, with successful applications often requiring engineering of the transmembrane domain. For example, the microbial synthesis

Received: December 11, 2020

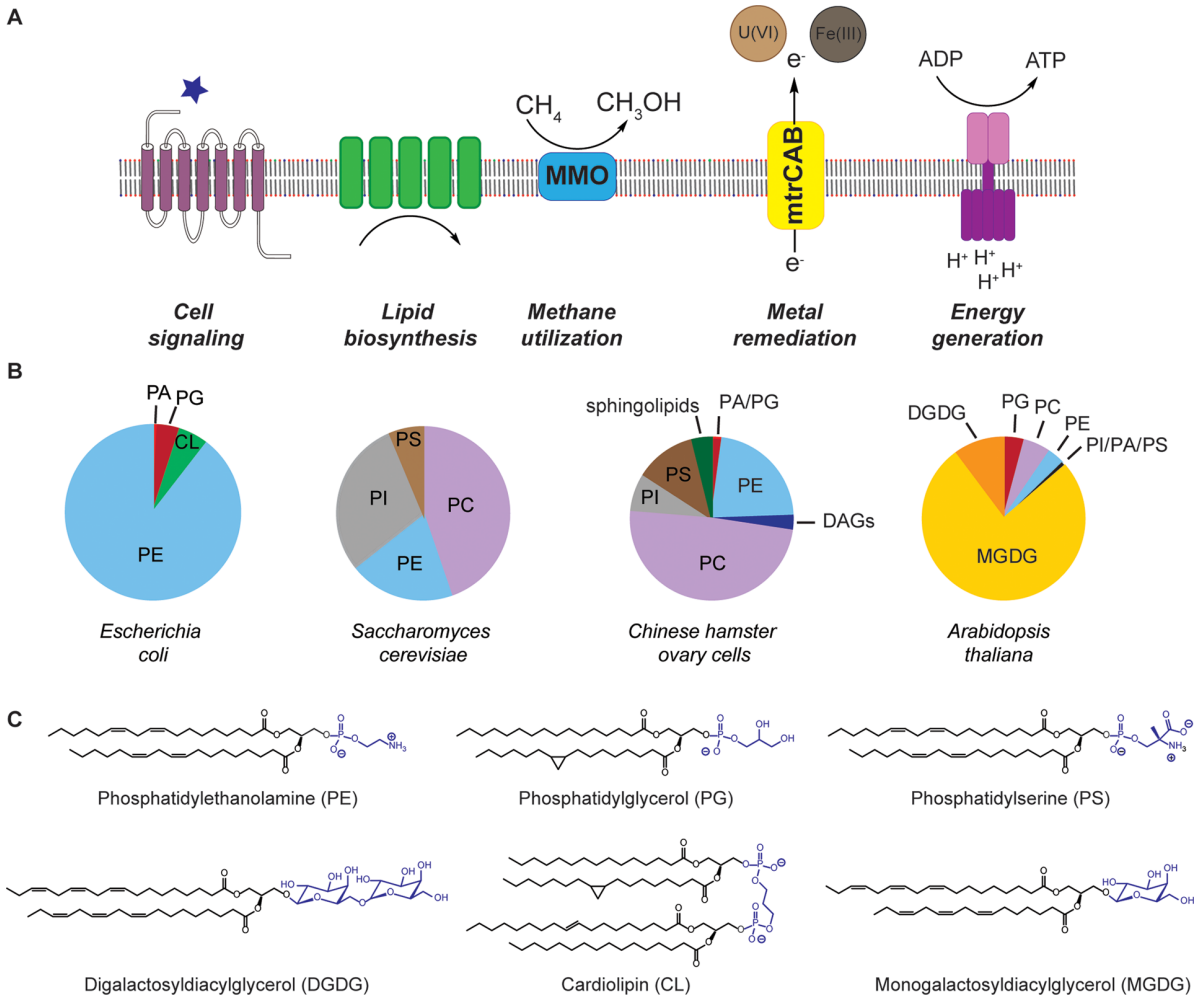


Figure 1. Biological roles of membrane proteins and cell membranes. (A) Cellular processes that occur at the cell membrane include cell signaling, lipid biosynthesis, methane utilization, metal remediation, and energy generation. (B) Membrane compositions in different organisms. PG: Phosphatidylglycerol, PA: Phosphatidic acid, PE: Phosphatidylethanolamine, CL: Cardolipin, PC: Phosphatidylcholine, PI: Phosphatidylinositol, PS: Phosphatidylserine, DAG: Diacylglycerol, MGDG: Monogalactosyldiacylglycerol, DGDG: Digalactosyldiacylglycerol. (C) Sample phospholipid structures. Phospholipid heads in blue.

of plant natural products is limited to hosts amenable to plant transmembrane CYP production, such as *Saccharomyces cerevisiae*, in spite of other hosts, such as *Escherichia coli*, having achieved higher precursor yields.[11] The development of protein chimeras and truncation of plant CYPs to generate soluble variants has been successful;[12] however, only one or two CYPs are usually engineered at a time, far from the five to ten CYPs required for many plant biosynthetic pathways.

Beyond transmembrane protein production, the physicochemical properties of the membrane, including composition, fluidity, curvature, and molecular crowding, influence the production and activity of membrane proteins.[13,14] For example, expression of *Catharanthus roseus* geraniol 10-hydroxylase in *S. cerevisiae* has a 8.3-fold lower activity than the same protein synthesized in a plant membrane, likely due to decreased enzyme stability and suboptimal reductase pairing.[15] In addition, a molecular dynamics simulation of a human CYP, CYP3A4, showed that lipid composition and electrostatics impact membrane incorporation and membrane protein orientation.[16] Indeed, the membrane compositions of mammalian, microbial, and plant cells vary vastly from one another[17−20] (Figure 1B).

As nonliving systems devoid of membranes, cell-free systems (CFSs) could be augmented with tailor-made membranes to fulfill specific membrane protein requirements and applications, thus functionally producing membrane-bound proteins that are challenging to synthesize using cells. Briefly, CFSs are composed of a cell lysate or purified cell machinery (PURE) supplemented with the nucleotides, energy sources, amino acids, cofactors, and salts necessary for transcription and translation.[21] PURE systems are often preferred for more complex protein synthesis due to reduced background and optimized conditions. For example, bacteriorhodopsin, ATP synthase, and enzymes in the lipid biosynthesis pathway have all been synthesized in PURE systems.[22] On the other hand, preparation of PURE reaction mix is low throughput and expensive. Thus, for potential scale-up applications cell lysate-based systems are required. Among cell-lysate-based CFSs, the *E. coli* based CFSs are the most commonly used platform with applications to the production of therapeutics, genetic circuit

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

engineering, construction of synthetic cells, chemical biosynthesis, and protein production.[23]

In membrane augmented CFSs, the user has complete control over composition, fluidity, and crowding of membranes, in addition to curvature and vesicle size, in the case of encapsulated CFSs. Lipids of different structure, length, saturation, and charge, a number of them commercially available, could be used to optimize the membrane composition (Figure 1C). Further, encapsulated CFSs could have different lipid compositions depending on the location of the bilayer leaflet. A membrane augmented CFS would open the doors to important bioindustrial applications, such as the development of hybrid chemical-biological processes, the use of organic solvents for *in situ* product extractions, and more efficient downstream processes for product separation. Developing genetic control systems to control both the lipid and protein composition of membrane augmented CFSs will be pivotal to achieve the high levels of membrane enzyme activity to enable these applications. For example in the bioremediation space, expression of MtrCAB in a membrane augmented CFS would enable circumvention of cell toxicity issues to facilitate applications at high contaminant concentrations.

In this review, we highlight recent advances in membrane-based CFSs and their application in the heterologous production of membrane proteins. Further, we examine the potential for genetic control systems, such as those implemented with CRISPR-based transcriptional regulation, to improve the cell-free synthesis of membrane proteins for chemical production and for the study of protein−membrane interactions. Although we are not yet at the level of on-demand membrane augmented CFS generation, the potential advantages of such systems in terms of enabling new chemistry and improving chemical bioproduction processes make it a new frontier in biotechnology.

1. PRODUCTION OF MEMBRANE PROTEINS IN CFSS

1.1. Membrane Augmented CFSs. Membrane protein production in CFSs is often limited by self-aggregation,[13] and addition of oil droplets to a PURE CFS has enabled the production of single-span transmembrane proteins, such as FasL and TRAIL used as anticancer therapeutics.[24] Oil droplets, however, are limited to the display of surface receptors, and cannot recapitulate the physicochemical properties of native lipids that support membrane protein activity.[13] Phospholipid-like additives, such as nanodiscs and liposomes, recapitulate membrane composition better (Figure 2).[25] Nanodiscs are phospholipids stabilized by membrane scaffold proteins,[26] which form a planar bilayer akin to a lipid raft. Both sides of a nanodisc are indistinguishable from one another and open to interact with the CFS environment.[13] Nanodiscs are commercially available and can be directly added to CFSs. Nanodiscs, however, may not work to produce every transmembrane protein,[13] and their composition cannot be easily changed.[25] The cost of nanodiscs (~$1/μL of CFS reaction volume), currently limits this technology to mostly research applications. Liposomes are spherical phospholipid vesicles[13] that provide a tunable environment in terms of shape, size, and composition to mimic both prokaryotic (d = 0.1−5.0 μm) and eukaryotic (d = 10−100 μm) cells. The availability of a wide array of phospholipids allows the generation of liposomes with tunable composition and size. Given liposomes' large surface area, they are the preferred scaffold for coexpression of multiple membrane proteins. Liposomes can be added to a CFS reaction or the CFS can be encapsulated within the liposomes, albeit the yield of liposomes loaded with CFS is typically small (~100 liposomes per μL of reaction mix), limiting their application. Using commercial phospholipids, liposomes have a cost of < $0.01/μL of CFS reaction, enabling high-end biotechnology applications.[27] Finally, inverted vesicles formed during the CFS cell lysis procedure have been used to produce complex membrane proteins, including the oxidative phosphorylation pathway[28] and transmembrane oligosaccharyltransferases for protein glycosylation.[29] As used, inverted vesicles require production and insertion of membrane proteins at the cellular stage, which brings the usual challenges with heterologous production of membrane proteins. The cell used for protein production controls the membrane composition and vesicle size, making it difficult to explore the extent to which these variables have an effect on enzyme activity or control these variables for the desired application.

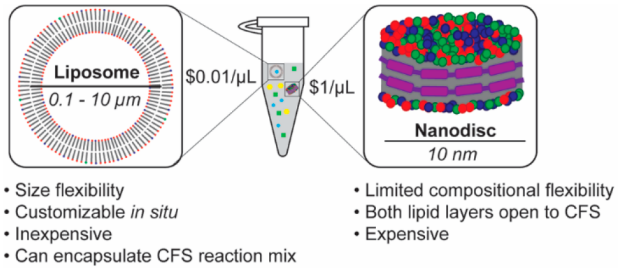


Figure 2. Membrane augmented cell-free systems. Mimetic membranes, in the form of liposomes and nanodiscs, are added directly to CFSs to aid in membrane protein production.

In situ phospholipid synthesis by CFSs would reduce the cost of membrane augmented CFSs, opening the doors to bioindustrial applications. *In situ* phospholipid synthesis has been achieved by feeding glycerol-3-phosphate and acyl-CoA. Because the lipid biosynthesis enzymes are themselves membrane-bound proteins, the CFS must be inoculated with preformed liposomes.[30] To enable feeding of fatty acids as a substrate, FadD was expressed in the CFS leading to phospholipid production and observable vesicle growth.[31] Nevertheless, low fatty acid solubility and its detrimental impact on protein stability limited this work, making the case for using glucose as the starting material in the future. Of note, as the fatty acids are located on the outside of the liposome, the phospholipids are only incorporated to the outer leaflet of the bilayer, limiting the synthesis of asymmetric membranes and potentially disturbing the incorporation of transmembrane proteins.

1.2. Chemical Composition of Membrane Augmented CFSs. Membrane proteins often rely on the phospholipids around them for activity. For example, the reaction rate, substrate affinity, and reductase coupling efficiency of the human cardiomycete epoxygenase responsible for oxidation of fatty acids and xenobiotics (CYP2J2) was shown to be impacted by the concentrations of phosphatidylcholine (PC) and phosphatidylserine (PS).[32] In another example, the bacterial structural protein MreB was polymerized when expressed in a liposome composed of PC and phosphatidylethanolamine-polyethylene glycol (PE−PEG) due to PEG's effect on membrane crowding, changing the

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Table 1. Applications of Membrane Proteins in Cell-Free Systems[a,b,c,d]

Protein Type Key
Structural/Cell Maintenance
Transporter/Porin
Energy Production
Receptor
Secondary Metabolism

Protein	Description	Application	Membrane type	Membrane composition	Protein Synthesis	CFS Method	Citation
MreB	Cell shape-determining protein	Synthetic cells	Liposome	Egg PC PE-PEG	Cell-Free	Cell lysate	[33]
				E. coli total lipid extract	Purified enzymes	N/A	[77]
HyaA	Hydrogensase-1 small chain	Protein characterization	Liposome	DOPC DGS-NTA DMPE-RhoB	Cell-Free	PURE	[78]
OmpA	Outer membrane protein A						
YfbF	Glycosyl transferase	Protocol development	Liposome	DOPC	Cell-Free	PURE	[79]
CyoE	Protoheme IX farnesyltransferase						
FtsZ	Cell division protein	Synthetic cells	Liposome	POPC POPG DOPE-RhoB	Cell-Free	PURE	[80]
FtsA							
ZipA							
GPAT	Lipid biosynthesis	Synthetic cells	Liposome	DOPC DOPE DOPG cardiolipin	Cell-Free	PURE	[30, 36]
LPAAT							
CdsA							
PgsA							
PgpABC					Purified enzymes	N/A	[31]
PssA							
Psd							
CYP2J2	Cytochrome P450	Protein characterization	Nanodisc	POPC POPS Cholesterol	Purified enzymes	N/A	[32]
CYP5A1	Cytochrome P450	Protein characterization	Nanodisc	POPC POPS POPE	Purified enzymes	N/A	[81]
CYP2B4	Cytochrome P450	Protein characterization	Nanodisc	DMPC POPC POPS	Purified enzymes	N/A	[82]
Opi3	Methyltransferase	Protein characterization	Nanodisc	DMPG DOPG DOPMME	Cell-Free	Cell lysate	[37]
OST	Oligosaccharyltransferases	Protein characterization	Nanodisc	POPC	Cell-Free	Cell lysate	[46]
		Protocol development	Inverted vesicle	N/A	Cell-Free	Cell lysate	[29]
SecYEG	Translocon	Protocol development	Liposome	Soybean lipid extract	Cell-Free	PURE	[42]
YidC	Insertase						
LepB	Signal peptidase						
MscL	Large-conductance mechanosensitive channel	Synthetic cells	Liposome	Egg PC	Cell-Free	Cell lysate	[83]
		Protein characterization	Liposome	DOPC + PEG diblock copolymer or detergent	Cell-Free	PURE	[39]
ErmE	Multidrug transporter	Protein characterization	Liposome	POPC	Cell-Free	PURE	[38]
		Protocol development	Liposome	POPC	Cell-Free	PURE	[84]
LacY	Sugar transporter	Protein characterization	Liposome	DMPC DOPC DOPE DOPG	Cell-Free	PURE	[40]
XylE							
α-hemolysin	Pore forming protein	Protein characterization	Liposome	POPC Cholesterol	Cell-Free	PURE	[85]
		Synthetic cells	Liposome	POPC Cholesterol	Cell-Free	Cell lysate	[60]
PFO	Pore protein perfringolysin O	Synthetic cells	Liposome	POPC Cholesterol	Cell-Free	Cell lysate	[59]

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Table 1. continued

Protein	Description	Application	Membrane type	Membrane composition	Protein Synthesis	CFS Method	Citation
ATP synthase	ATP synthesizing protein complex	Synthetic cells	Liposome	DOPC DOPE DOPG	Cell-Free	Cell lysate	[43]
				POPC	Purified enzymes	N/A	[68]
					Cell-Free	PURE	
				POPC POPE POPG Cholesterol	Purified enzymes	N/A	[67]
		Protocol development	Inverted vesicle	N/A	*In vivo*[1]	N/A	[28]
Bacterio-rhodopsin	Photoconverter	Synthetic cells	Liposome	POPC	Purified enzymes	N/A	[68]
					Cell-Free	PURE	
				POPC POPE POPG Cholesterol	Purified enzymes	N/A	[67]
Photo-system II	Photoconverter	Synthetic cells	Liposome	POPC POPE POPG Cholesterol	Purified enzymes	N/A	[67]
PsbS	Photosystem II subunit S	Synthetic cells	Liposome	Asolectin	Cell-Free	Cell lysate[2]	[86]
MtrCAB	Electron transfer pathway	Protein characterization	Liposome	PC	Purified enzymes	N/A	[87]
CX3CR1	Chemokine G-protein coupled receptors	Protein characterization	Liposome	POPC PE-PEG	Cell Free	PURE	[47]
CCR5			Nanodisc	POPC POPS Cholesterol			
CYP725A	Cytochrome P450	Protein characterization	Nanodisc	POPC	Purified enzymes	N/A	[88]

[a]DGS-NTA = Dioleoylglycero-[(*N*-(5-amino-1-carboxypentyl) iminodiacetic acid) succinyl] (nickel salt). DMPC = Dimyristoylphosphatidylcholine. DMPE-RhoB = Dimyristoylphosphatidylethanolamine-rhodamine B. DMPG = Dimyristoylphosphatidylglycerol. DOPC = Dioleoylphosphatidylcholine. DOPE = Dioleoylphosphatidylethanolamine. DOPE-RhoB = Dioleoylphosphatidylethanolamine-rhodamine B. DOPG = Dioleoylphosphatidylglycerol. DOPMME = Dioleoylphosphatidylmonomethylethanolamine. PC = Phosphatidylcholine. PE−PEG = Phosphatidylethanolamine-polyethylene glycol. POPC = Palmitoyl-oleoyl-phosphatidylcholine. POPE = Palmitoyl-oleoyl-phosphatidylethanolamine. POPG = Palmitoyl-oleoyl-phosphatidylglycerol. POPS = Palmitoyl-oleoyl-phosphatidylserine. [b]In the protein synthesis column: "Cell-Free" indicates that the enzyme was synthesize in the cell-free system. "Purified enzymes" indicates that the protein was synthesized *in vivo*, purified, and reconstituted in CFS. [c](1) ATP synthase produced *in vivo* and recovered in CFS after cell lysis. [d](2) Cell lysate reaction mix contains added, purified T7 polymerase.

shape of the liposome from spherical to rod-like.[33] Therefore, membrane protein activity can be altered not only by protein engineering, but also by engineering the membranes in which they are embedded.

Changing the chemical structure of the phospholipid heads or tails causes changes in the membrane's physical properties such as fluidity, thickness, and charge.[34] *In situ* changes in phospholipid composition have been achieved by using different acyltransferases to convert phosphatidic acid (PA) to either phosphatidylglycerol (PG) or PE.[30] Importantly, the PE:PG ratio was maintained to the *E. coli* membrane composition, 75% PE,[35] both genetically, by placing PE and PG production under control of different promoters and dosing the respective polymerases, and enzymatically, using PssA, which associates with PG rich membranes and catalyzes the synthesis of PE.[36] Finally, conversion of phosphatidylmonomethylethanolamine (PMME) into PC, the most prevalent phospholipid in eukaryotes, was achieved by expressing the methyltransferase Opi3 in nanodiscs.[37]

1.3. Physical Properties of Membrane Augmented CFSs. Liposome vesicle diameter, shape, curvature, and number of lamella can be optimized to create a native-like membrane environment. Optimizing the physical properties of the mimetic membrane is key to protein folding, membrane incorporation, and activity. Using the multidrug transporter ErmE as a model, it was shown that surface area-to-volume ratio in smaller vesicles improved membrane insertion. Additionally the ratio of incorporated ErmE to total synthesized ErmE was a function of liposome size, not of DNA concentration.[38] It remains to be seen if this correlation holds true for other membrane proteins, or in membranes composed of more than one type of lipid. Optimal membrane size may depend on other factors such as membrane fluidity and rigidity and may change depending on the transmembrane protein source organism. Another important membrane physical parameter is elasticity. Using diblock copolymers to increase the membrane elasticity, it was shown that decreasing the membrane area expansion modulus improved membrane folding of the mechanosensitive channel of large conductance (MscL) using the PURE system.[39] However, differing results for the model membrane protein channel rhodopsin (ChR2) indicated that this conclusion may not be generalizable to all membrane proteins and the number of transmembrane regions may influence the optimal membrane elasticity. Importantly, in

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

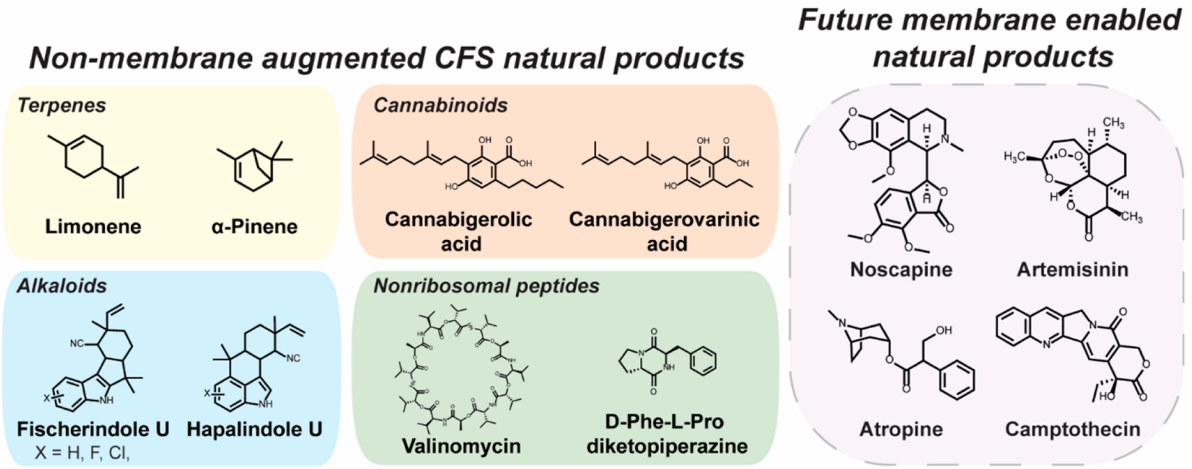


Figure 3. Production of natural products using cell-free systems. Compounds produced to date do not require membrane proteins or membrane augmented CFSs. In the dashed box, natural products that depend on the production of transmembrane proteins that could be synthesized by using a membrane augmented CFS.

addition to polymers, biosynthetic compounds, such as cholesterol, can tune the elasticity of membranes.

1.4. Membrane Protein Insertion into Membrane Augmented CFSs. Chaperones facilitate the insertion of membrane proteins in cells; however, solely membrane composition can affect the efficiency and directionality of this process. In a PURE-based CFS lacking chaperones, the 12-transmembrane proteins LacY and XylE were expressed and spontaneously incorporated into the liposome membrane. The membrane protein incorporation was lipid dependent with 2.6- and 1.5-fold increases in LacY and XylE incorporation, respectively, between the worst (100% PC) and best (100% PG) performing membrane compositions.[40] Importantly, lipid composition played a role in establishing the correct directionality for membrane insertion. While the 100% PG membrane had the highest protein incorporation, the lack of PE resulted in LacY being incorporated in an inverted membrane orientation. This result supported previous work that identified PE as a nonproteinaceous chaperone of LacY.[41] Nonspontaneous membrane protein insertion can be facilitated in a PURE system by the secYEG translocon, which successfully integrated YidC and LepB into an exogenous liposome.[42] Spontaneous cotranslational integration of the multisubunit secYEG[42] and ATP synthase[43] complexes into liposomes in CFSs suggests that many multisubunit membrane proteins will be functional in CFSs without need for additional reaction components.

In encapsulated CFSs, orientation is critical for protein function, and membrane proteins often have to asymmetrically localize in the bilayer leaflet. Using SNAP-tag modified fluorescent proteins and liposomes composed of benzylguanine-modified phospholipids, membrane asymmetry was achieved by encapsulating a CFS expressing mCherry-SNAP within the modified liposome, and suspending the encapsulated CFS in a second CFS expressing GFP-SNAP. Fluorescence microscopy confirmed mCherry was localized to the inner membrane of the liposome while GFP localized to the outer membrane.[44] Controlling protein localization is key in applications that benefit from separating intermediates in metabolic pathways. For example, the production of CMP-*N*-acetylneuraminic acid was improved 2-fold by encapsulating the first pathway enzyme, *N*-acyl-D-glucosamine-2-epimerase, within a polymersome and attaching the rest of the pathway on the outside of the polymersome, thus reducing inhibition of *N*-acyl-D-glucosamine-2-epimerase by a late pathway intermediate.[45]

2. APPLICATIONS

Expanding the use of CFSs to include membrane proteins offers a wide range of applications from single enzyme assays to use of the membrane proteins as part of longer enzyme pathways. Table 1 offers an overview of current literature in which membrane proteins are applied in a CFS using nanodiscs, liposomes, or inverted vesicles, highlighting both *in vitro* synthesized and reconstituted proteins from multiple protein classes.

2.1. Enzyme Assays. Transmembrane protein production in CFSs enables their study in the absence of endogenous metabolic pathways or metabolites that may confound the results. For example, oligosaccharyltransferase homologs have been synthesized in cell lysate based CFSs using nanodiscs and used to rapidly identify acceptor proteins,[46] bypassing competition from an endogenous glycosylation system. Furthermore, the chemokine GPCRs CX_3CR1 and CCR5 have been produced in CFSs using nanodiscs in a PURE system for structural (electron microscopy) and functional (surface plasmon resonance) studies.[47] As membrane augmented CFSs become more widely available, we expect other membrane protein classes to make use of this technology.

2.2. Energy Production. CFSs hold incredible promise for the production of fuels and chemicals due to their high productivity when compared to microbes, up to 815 mg/L/h in the case of mevalonate.[48] Cell-free production of butanol and hydrogen could be improved by extending the respective pathways using particulate MMO for the assimilation of CH_4 to enable use of a C1 feedstock.[49] Such a system would also take advantage of the improved separation and reduced effects of toxicity provided by using CFSs. It is worth noting that CFS production platforms are currently not cost competitive with microbial ones for biofuel production and have not been scaled to the volume or run as long as microbial cell factories.[50] The generation of electricity in microbial fuel cells would also benefit from improvements in membrane protein expression in

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

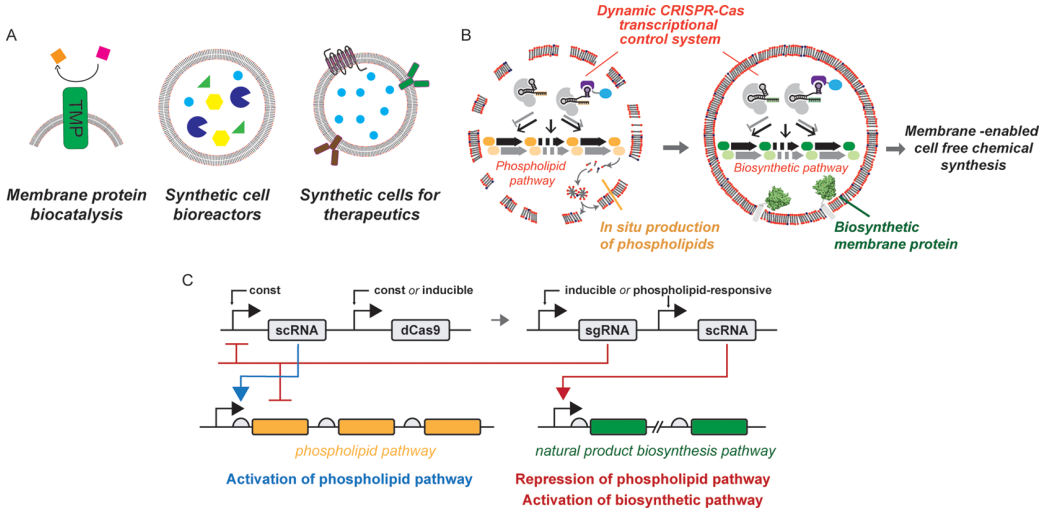


Figure 4. Applications of membrane augmented cell-free systems. (A) Potential bioindustrial applications of membrane proteins in membrane augmented cell-free systems. (B) Dynamic CRISPR-Cas control system can be implemented to stagger phospholipid and natural product biosynthesis, allowing for *in situ* liposome self-assembly prior to membrane protein production and chemical synthesis. (C) Sample two-stage control strategy for staggered phospholipid and natural product biosynthesis. First, constitutive (const) or induced expression of a modified guide RNA that enables the recruitment of a transcriptional activator (scaffold RNA, or scRNA) activates expression of the phospholipid pathway. Second, a user-supplied inducer or sufficient phospholipid concentration triggers expression of the natural product biosynthesis pathway (through targeted scRNA) as well as repression of the phospholipid pathway using a sgRNA coupled with catalytically inactive Cas9 (dCas9). Suppression of lipid biosynthesis during stage two helps conserve precious CFS resources.

CFSs by helping to overcome barriers in feedstock toxicity and limits on electric current production due to organism viability.[51] To this end, enzymatic fuel cells using purified enzymes can produce electricity from hydrogen, methanol, formate, pyruvate, and sugars without the need for membrane proteins.[52] However, these systems often have a lifetime of hours to days. Incorporation of these enzymes into a CFS would allow for enzyme and cofactor regeneration, extending the lifetime of the fuel cell. Furthermore, use of membrane augmented CFS could further improve enzyme stability and expand the range of usable fuels to include CH_4.

2.3. Chemical Biosynthesis. Despite the now large array of chemicals synthesized using CFSs,[50] none of them relies on a transmembrane protein for biosynthesis (Figure 3). Focusing on plant natural products, CFSs have been used to produce monoterpenes[53] and cannabinoids,[54] but improved expression of membrane proteins would increase the structural diversity of reachable products. Specifically, expression of transmembrane plant CYPs, such as those found in monoterpene indole alkaloid (MIA) and benzylisoquinoline alkaloid (BIA) pathways,[1] would allow CFS production of the anticancer drug Taxol, which requires 19 enzymatic steps after geranylgeranyl diphosphate with eight of those steps catalyzed by transmembrane CYPs.[55] Furthermore, synthesis of the MIA intermediate strictosidine requires four CYPs,[56] while the synthesis of the BIA intermediate noscapine requires five.[57] Today, *S. cerevisiae* is used to produce plant CYPs; however, a long doubling time (3 h) limits the rapid prototyping of CYPs and reductase partners. Highlighting the challenges associated with natural product production in *S. cerevisiae*, recent production of the tropane alkaloid scopolamine required N-terminal engineering of the membrane-bound littorine synthase from *Attropa belladonna* for functional expression in *S. cerevisiae*.[58] Finally, the membrane compositions of yeast and plants are very different (Figure 1B), likely impacting transmembrane protein functionality. Membrane augmented CFSs offer a potential upgrade on *S. cerevisiae* in terms of productivity, reduced effects of product toxicity, and flexibility in membrane composition.

2.4. Synthetic Cells. Synthetic cells, particles that mimic biological cell but have different characteristics, functions, or parts, can be generated by encapsulating CFSs in liposomes. Membrane proteins play pivotal roles in the development of synthetic cells: from growth and division *via* phospholipid biosynthesis and Z-ring proteins, to cell-to-cell communication mediated by cell surface receptors for the development of cell therapies, to the generation of NADH and ATP to extend their chemical bioproduction time.

In the context of cell-to-cell communication, a synthetic cell constitutively expressing a brain-derived neurotrophic factor was activated by addition of homoserine lactone, leading to production of the pore forming perfringolysin O, secretion of the neurotrophic factor, and differentiation in cocultured stem cells.[59] Similarly, the pore forming protein α-hemolysin has been leveraged to control uptake or secretion of doxycycline and isopropyl-β-D-thiogalactoside (IPTG), which activated downstream luciferase expression as a proof of concept reporter gene.[60]

In the context of chemical bioproduction, synthetic cells act as miniature bioreactors that, unlike cells, do not use the carbon for cell growth or maintenance, and essentially route all carbon for chemical production (Figure 4A). Thus, bioreactor synthetic cells act as immobilized enzyme catalysts, potentially enjoying easy reuse and separation, while allowing the enzyme inside the reactor to work within the simulated synthetic cell environment with limited loss of activity. Liposome volume and stability is particularly important for bioreactor synthetic cells. Decreased synthetic cell volume results in higher local substrate concentration, speeding up the reaction of low substrate affinity enzymes.[61] Nevertheless, high local product concentrations may limit yields through increased enzyme product inhibition. The stability of the liposome will determine

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

the lifetime of the reactions that take place inside it, making it crucial to understand how each of these physical properties affects liposome stability, which is known to be affected by liposome structure[62] and synthesis method.[63]

Membrane proteins often require prosthetic groups for activity. For instance, CYPs need heme, and membrane-bound glucose dehydrogenases used for NAD(P)H regeneration use pyrroloquinoline quinone (PQQ). Prosthetic groups need to be incorporated into proteins as they fold and are inserted in the membrane to achieve proper enzyme function. Often, CFSs lack the biosynthetic pathways to synthesize prosthetic groups and these compounds need to be exogenously added to the reaction. For scale-up applications, however, enriched CFSs from organisms where the prosthetic group biosynthetic pathways are expressed and/or upregulated will be needed for the functional expression of membrane proteins. Some inroads toward this goal have been made. For example, PQQ was synthesized by *Gluconobacter oxydans*-based CFS carrying the machinery to convert fed pqqA precursor to PQQ.[64] In another example, heme was biosynthesized in *E. coli* CFS *via* addition of purified 5-aminolevulinic acid synthase, and heme was successfully incorporated into P450 BM3.[65] Of note, prosthetic group biosynthesis should be carefully regulated to avoid CFS poisoning or unnecessarily diverting carbon flux from the desired product. Additionally, chaperones could be introduced to help in the incorporation of prosthetic groups, such as ferrochelatases for heme loading.[66]

The final challenge in bioreactor synthetic cells is the need for reducing power (NADH) and energy (ATP) regeneration to drive reactions for extended periods of time to reduce process cost. Cofactor regeneration in CFSs has been achieved using glyceraldehyde-3-phosphate dehydrogenase and applied to monoterpene production.[53] Production of ATP in CFSs was achieved early on using oxidative phosphorylation.[28] More recently, efforts have moved to produce ATP from light using purified ATP synthase, photosystem II, and proteorhodopsin reconstituted in liposomes,[67] or using bacteriorhodopsin to generate the proton gradient necessary to produce ATP upon light induction.[68] Importantly, all ATP production mechanisms need membrane proteins to generate an electron gradient, which requires proteins to have the correct membrane orientation. This can be partially controlled by limiting spontaneous membrane integration through modulating cholesterol and diacylglycerol concentration.[69] However, the effect of these compounds is phospholipid dependent and requires continued study on more complex lipid mixtures.[70]

3. CONTROL STRATEGIES FOR THE GENERATION OF MEMBRANE AUGMENTED CFSS

3.1. Genetic Control of Membrane Properties. CFSs have a limited amount of resources for the formation of enzymes and metabolites as well as transcription and translation machinery. For synthetic cells to become industrially relevant, the cost-effective synthesis of both phospholipids for membrane formation and membrane-bound actuating biomolecule(s) is needed. The actuating biomolecule(s) can be single enzymes for biocatalysis applications, multienzyme pathways for chemical bioproduction, or receptors for synthetic therapeutic applications. To enable these applications, a control system that can dynamically program gene expression of multiple units is needed. The control system should (1) have low overhead resource consumption, (2) be capable of turning genes both on and off to generate sequential phases of gene expression programs, (3) be tunable to precisely regulate gene expression dynamics, and (4) be scalable to allow for construction of increasingly complex gene regulatory systems.[71,72] Such a control system would enable synthesis of phospholipid-producing enzymes early in a reaction to form membranes, followed by a shift in production to membrane proteins to produce the chemicals, both with programmable stoichiometry and timing (Figure 4B,C). Control systems in liposomes will need to go beyond controlling the enzyme ratios,[36] and dynamically control on/off gene expression. This capability would allow for controlled changes in membrane properties through the course of the reaction without outside intervention.

The control system should not only regulate lipid synthesis, but also balance the expression of the membrane bound actuating biomolecule to prevent aggregation while maximizing efficiency and rate of membrane protein insertion. Most of the current understanding on membrane protein insertion relies on expression rates to prevent saturation of membrane insertion machinery.[14] In the case of liposomes, hydrophobic interactions among membrane proteins lead to self-aggregation and interactions between membrane proteins and the lipid membrane play an outsized role in protein insertion, posing additional challenges to the control system. Implementation of a gene control system capable of delivering distinct gene expression profiles would develop a better understanding of how liposomes change over time, how liposomes adsorb proteins from the CFS, and how the density of proteins already in a liposome affects how much membrane protein uptake occurs.

3.2. CRISPRa/i: A Control System for the Expression of Membrane Proteins in Membrane-Augmented CFSs. While elementary gene regulation has been implemented in CFSs,[73] efforts toward developing multigene control systems to provide precise regulation over gene expression, membrane protein insertion, and function remain at an early stage. The CRISPR-Cas system provides a powerful suite of tools for multigene transcriptional control.[71] Briefly, catalytically inactive Cas9 protein can be directed to specific DNA sequences by guide RNAs that recognize target sequences based on predictable Watson−Crick base pairing to activate (CRISPRa) or repress (CRISPRi) gene expression. Although the rules governing CRISPRa from bacterial promoters are complex,[74] a growing set of validated CRISPRa components enables the rapid construction of increasingly complex gene regulatory systems.[71] Multigene CRISPR circuits have been engineered through the regulated expression of up to seven distinct sgRNAs in the same system,[75] and CRISPRi has been shown to operate efficiently in CFSs.[76] Thus, by combining new capabilities for CRISPRa with existing tools for CRISPRi in CFSs, it should be possible to engineer multigene programs for membrane augmented CFSs operating through the regulated expression of guide RNAs. Multiguide RNA CRISPRa/i circuits could then be used to program distinct gene expression modes to enhance functional membrane protein expression. Consequently, CRISPRa/i circuits could provide an efficient mechanism for implementing dynamic multigene control, while preserving valuable CFS metabolites and cofactors.

A toolbox of pulse-generating CRISPRa/i circuits could be used to both investigate and optimize how membrane protein expression dynamics impact membrane insertion and function. Here, the network topology would specify the timing of gene

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

expression pulses or regulatory functions. Further, incorporation of input-responsive pulses of gene expression or regulation into CRISPRa/i networks would extend tunable control to semicontinuous reactions and at any specific time within a membrane CFS reaction. Ultimately, it may be possible to engineer CRISPRa/i programs as process controls for membrane augmented CFS bioindustrial applications that regulate protein expression, minimize the waste of valuable precursors and energy molecules and prevent the accumulation of destabilizing intermediates.[72]

4. FUTURE DIRECTIONS

Membrane proteins play a pivotal role in bioenergy, biomedical, and bioindustrial applications, and our ability to harness their potential hinges on their functional production outside their endogenous host. Although to some extent heterologous membrane proteins can be engineered for optimal functional heterologous production, an alternate and now more and more feasible strategy is to engineer CFSs with tailor-made augmented membranes to ensure optimal transmembrane protein activity. Although we are far from on-demand membrane augmented CFS generation, the potential advantages of such systems in terms of enabling new chemistry and improved chemical bioproduction processes make this a worthwhile endeavor. The realization that we have not only protein engineering, but also membrane engineering in our toolkit when tackling transmembrane protein challenges should help us accelerate some of these applications.

The biggest step forward in this field will be moving from using membrane augmented CFSs for protein characterization and analysis to larger scale application for chemical bioproduction, bioremediation, or synthetic cells. This will require effective scale up of membrane protein synthesis and controlled expression of phospholipid biosynthesis enzymes and biosynthetic pathway enzymes. Due to cost and tunability, liposomes appear to be the better option for scaled up membrane protein synthesis in CFSs. This said, work toward spontaneous assembly of liposomes *in situ* and without need for organic solvents would help lower process cost for liposome production. Furthermore, study and improvement on membrane protein−liposome stability will be necessary for widespread application. Metrics such as total turnover number and half-life may be useful to help quantify scalability of these systems.

Successful implementation of membrane protein synthesis in CFSs will open the door for enzymatic production of toxic chemicals in nonliving systems as well as nonliving biosensors and bioremediation tools that can be applied environmentally without risk of biological contamination, or loss of function due to environmental toxicity. Finally, more sophisticated synthetic cells aided by membrane proteins hold great promise in therapeutic applications for targeted drug therapies as well as communicating with cellular environments to make expression decisions based on external stimuli. If aided by CRISPRa/i, these decisions and logic gates can be made significantly more complex to respond to combinations of signals and give the synthetic cell temporal, on/off control over gene expression.

AUTHOR INFORMATION

Corresponding Author

Pamela Peralta-Yahya − *School of Chemical and Biomolecular Engineering and School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0002-0356-2274; Email: pperalta-yahya@chemistry.gatech.edu

Authors

Nicholas S. Kruyer − *School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Widianti Sugianto − *School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Benjamin I. Tickman − *Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, Washington 98195, United States*

Diego Alba Burbano − *Molecular Engineering & Sciences Institute and Center for Synthetic Biology and Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States*

Vincent Noireaux − *School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, United States;* orcid.org/0000-0002-5213-273X

James M. Carothers − *Molecular Engineering & Sciences Institute and Center for Synthetic Biology and Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.0c00625

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (CBET 1844152) to V.N., J.M.C., and P.P.-Y.

ABBREVIATIONS

CFS, cell-free system; CYP, cytochrome P450.

REFERENCES

(1) Cravens, A., Payne, J., and Smolke, C. D. (2019) Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun. 10*, 2142.

(2) Ehrenworth, A. M., and Peralta-Yahya, P. (2017) Accelerating the semisynthesis of alkaloid-based drugs through metabolic engineeirng. *Nat. Chem. Biol. 13*, 249−258.

(3) Clomburg, J. M., Crumbley, A. M., and Gonzalez, R. (2017) Industrial biomanufacturing: the future of chemical production. *Science 355*, No. aag0804.

(4) Ross, M. O., MacMillan, F., Wang, J., Nisthal, A., Lawton, T. J., Olafson, B. D., Mayo, S. L., Rosenzweig, A. C., and Hoffman, B. M. (2019) Particulate methane monooxygenase contains only mononuclear copper centers. *Science 364*, 566−570.

(5) Fei, Q., Guarnieri, M. T., Tao, L., Laurens, L. M. L., Dowe, N., and Pienkos, P. T. (2014) Bioconversion of natural gas to liquid fuel: opportunities and challenges. *Biotechnol. Adv. 32*, 596−614.

(6) Desai, M., and Camobreco, V. (2020) *Inventory of U.S. Greenhouse Gas Emissions and Sinks*, U.S. Environmental Protection Agency, EPA 430-R-20-002.

(7) Jiang, S., and Hur, H.-G. (2013) Effects of the anaerobic respiration of *Shewanella oneidensis* MR-1 on the stability of extracellular U(VI) nanofibers. *Microbes Environ. 28*, 312−315.

(8) Yasi, E. A., Kruyer, N. S., and Peralta-Yahya, P. (2020) Advances in G protein-coupled receptor high-throughput screening. *Curr. Opin. Biotechnol. 64*, 210−217.

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(9) Hong, M. H., Clubb, J. D., and Chen, Y. Y. (2020) Engineering CAR-T cells for next-generation cancer therapy. *Cancer Cell 38*, 473−488.
(10) Kipniss, N. H., Dingal, P. C. D. P., Abbott, T. R., Gao, Y., Wang, H., Dominguez, A. A., Labanieh, L., and Qi, L. S. (2017) *Nat. Commun. 8*, 2212.
(11) Yang, D., Park, S. Y., Park, Y. S., Eun, H., and Lee, S. Y. (2020) Metabolic engineering of *Escherichia coli* for natural product biosynthesis. *Trends Biotechnol. 38*, 745−765.
(12) Liu, X., Zhu, X., Wang, H., Liu, T., Cheng, J., and Jiang, H. (2020) Discovery and modification of cytochrome P450 for plant natural products biosynthesis. *Synth. Syst. Biotechnol. 5*, 187−199.
(13) Sachse, R., Dondapati, S. K., Fenz, S. F., Schmidt, T., and Kubick, S. (2014) Membrane protein synthesis in cell-free systems from bio-mimetic systems to bio-membranes. *FEBS Lett. 588*, 2774−2781.
(14) Guigas, G., and Weiss, M. (2016) Effects of protein crowding on membrane systems. *Biochim. Biophys. Acta, Biomembr. 1858*, 2441−2450.
(15) Collu, G., Unver, N., Peltenburg-Looman, A. M. G., van der Heijden, R., Verpoorte, R., and Memelink, J. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett. 508*, 215−220.
(16) Navrátilová, V., Paloncýová, M., Berka, K., and Otyepka, M. (2016) Effect of lipid charge on membrane immersion of cytochrome P450 3A4. *J. Phys. Chem. B 120*, 11205−11213.
(17) Furse, S., Wienk, H., Boelens, R., de Kroon, A. I. P. M., and Killian, J. A. (2015) *E. coli* MG1655 modulates its phospholipid composition through the cell cycle. *FEBS Lett. 589*, 2726−2730.
(18) Lindberg, L., Santos, A. X., Riezman, H., Olsson, L., and Bettiga, M. (2013) Lipidomic profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* reveals critical changes in lipid composition in response to acetic acid stress. *PLoS One 8*, No. e73936.
(19) Symons, J. L., Cho, K.-J., Chang, J. T., Du, G., Waxham, M. N., Hancock, J. F., Levental, I., and Levental, K. R. (2021) Lipidomic atlas of mammalian cell membranes reveals hierarchical variation induced by culture conditions, subcellular membranes, and cell lineages. *Soft Matter 17*, 288−297.
(20) Jia, Y., Tao, F., and Li, W. (2013) Lipid profiling demonstrates that suppressing *Arabidopsis* phospholipase Dδ retards ABA-promoted leaf senescence by attenuating lipid degradation. *PLoS One 8*, No. e65687.
(21) Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013) Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp. 16*, No. e50762.
(22) Kuruma, Y., and Ueda, T. (2015) The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc. 10*, 1328−1344.
(23) Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020) Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet. 21*, 151−170.
(24) Yunker, P. J., Asahara, H., Hung, K.-C., Landry, C., Arriaga, L. R., Akartuna, I., Heyman, J., Chong, S., and Weitz, D. A. (2016) One-pot system for synthesis, assembly, and display of functional single-span membrane proteins on oil-water interfaces. *Proc. Natl. Acad. Sci. U. S. A. 113*, 608−613.
(25) Henrich, E., Dötsch, V., and Bernhard, F. (2015) Screening for lipid requirements of membrane proteins by combining cell-free expression with nanodiscs. *Methods Enzymol. 556*, 351−369.
(26) Bayburt, T. H., Grinkova, Y. V., and Sligar, S. G. (2002) Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett. 2*, 853−856.
(27) Garamella, J., Garenne, D., and Noireaux, V. (2019) TXTL-based approach to synthetic cells. *Methods Enzymol. 617*, 217−239.
(28) Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., and Swartz, J. R. (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol. 4*, 220.
(29) Jaroentomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., Mrksich, M., Jewett, M. C., and DeLisa, M. P. (2018) Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun. 9*, 2686.
(30) Scott, A., Noga, M. J., de Graaf, P., Westerlaken, I., Yildirim, E., and Danelon, C. (2016) Cell-free phospholipid biosynthesis by gene-encoded enzymes reconstituted in liposomes. *PLoS One 11*, No. e0163058.
(31) Exterkate, M., Caforio, A., Stuart, M. C. A., and Driessen, A. J. M. (2018) Growing membranes *in vitro* by continuous phospholipid biosynthesis from free fatty acids. *ACS Synth. Biol. 7*, 153−165.
(32) Huff, H. C., Maroutsos, D., and Das, A. (2019) Lipid composition and macromolecular crowding effects on CYP2J2-mediated drug metabolism in nanodiscs. *Protein Sci. 28*, 928−940.
(33) Garenne, D., Libchaber, A., and Noireaux, V. (2020) Membrane molecular crowding enhances MreB polymerization to shape synthetic cells from spheres to rods. *Proc. Natl. Acad. Sci. U. S. A. 117*, 1902−1909.
(34) Renne, M. F., and de Kroon, A. I. P. M. (2018) The role of phospholipid molecular species in determining the physical properties of yeast membranes. *FEBS Lett. 592*, 1330−1345.
(35) Sohlenkamp, C., and Geiger, O. (2016) Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol. Rev. 40*, 133−159.
(36) Blanken, D., Foschepoth, D., Serrão, A. C., and Danelon, C. (2020) Genetically controlled membrane synthesis in liposomes. *Nat. Commun. 11*, 4317.
(37) Henrich, E., Löhr, F., Pawlik, G., Peetz, O., Dötsch, V., Morgner, N., de Kroon, A. I., and Bernhard, F. (2018) Lipid conversion by cell-free synthesized phospholipid methyltransferase Opi3 in defined nanodisc membranes supports an *in trans* mechanism. *Biochemistry 57*, 5780−5784.
(38) Soga, H., Fujii, S., Yomo, T., Kata, Y., Watanabe, H., and Matsuura, T. (2014) *In vitro* membrane protein synthesis inside cell-sized vesicles reveals the dependence of membrane protein integration on vesicle volume. *ACS Synth. Biol. 3*, 372−379.
(39) Jacobs, M. L., Boyd, M. A., and Kamat, N. P. (2019) Diblock copolymers enhance folding of a mechanosensitive membrane protein during cell-free expression. *Proc. Natl. Acad. Sci. U. S. A. 116*, 4031−4036.
(40) Harris, N. J., Pellowe, G. A., and Booth, P. J. (2020) Cell-free expression tools to study co-translational folding of alpha helical membrane transporters. *Sci. Rep. 10*, 9125.
(41) Nagamori, S., Vázquez-Ibar, J. L., Weinglass, A. B., and Kaback, H. R. (2003) *In vitro* synthesis of lactose permease to probe the mechanism of membrane insertion and folding. *J. Biol. Chem. 278*, 14820−14826.
(42) Matsubayashi, H., Kuruma, Y., and Ueda, T. (2014) Cell-free synthesis of secYEG translocon as the fundatmental protein transport machinery. *Origins Life Evol. Biospheres 44*, 331−334.
(43) Matthies, D., Haberstock, S., Joos, F., Dötsch, V., Vonck, J., Bernhard, F., and Meier, T. (2011) Cell-free expression and assembly of ATP synthase. *J. Mol. Biol. 413*, 593−603.
(44) Uyeda, A., Watanabe, T., Hohsaka, T., and Matsuura, T. (2018) Different protein localizations on the inner and outer leaflet of cell-sized liposomes using cell-free protein synthesis. *Synth. Biol. 3*, No. ysy007.
(45) Klermund, L., Poschenrieder, S. T., and Castiglione, K. (2017) Biocatalysis in polymersomes: improving multienzyme cascades with incompatible reaction steps by compartmentalization. *ACS Catal. 7*, 3900−3904.
(46) Schoborg, J. A., Hershewe, J. M., Stark, J. C., Kightlinger, W., Kath, J. E., Jaroentomeechai, T., Natarajan, A., DeLisa, M. P., and Jewett, M. C. (2018) A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnol. Bioeng. 115*, 739−750.
(47) Gessesse, B., Nagaike, T., Nagata, K., Shimizu, Y., and Ueda, T. (2018) G-protein coupled receptor protein synthesis on a lipid bilayer using a reconstituted cell-free protein synthesis system. *Life 8*, 54.

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(48) Dudley, Q. M., Anderson, K. C., and Jewett, M. C. (2016) Cell-free mixing of *Escherichia coli* crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. *ACS Synth. Biol. 5*, 1578−1588.
(49) Zhang, Y.-H. P. (2015) Production of biofuels and biochemicals by *in vitro* synthetic biosystems: opportunities and challenges. *Biotechnol. Adv. 33*, 1467−1483.
(50) Bowie, J. U., Sherkhanov, S., Korman, T. P., Valliere, M. A., Opgenorth, P. H., and Liu, H. (2020) Synthetic biochemistry: the bio-inspired cell-free approach to commodity chemical production. *Trends Biotechnol. 38*, 766−778.
(51) Santoro, C., Arbizzani, C., Erable, B., and Ieropoulos, I. (2017) Microbial fuel cells: from fundamentals to applications. A reivew. *J. Power Sources 356*, 225−244.
(52) Xiao, X., Xia, H.-q., Wu, R., Bai, L., Yan, L., Magner, E., Cosnier, S., Lojou, E., Zhu, Z., and Liu, A. (2019) Tackling the challenges of enymatic bio(fuel) cells. *Chem. Rev. 119*, 9509−9558.
(53) Korman, T. P., Opgenorth, P. H., and Bowie, J. U. (2017) A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun. 8*, 15526.
(54) Valliere, M. A., Korman, T. P., Woodall, N. B., Khitrov, G. A., Taylor, R. E., Baker, D., and Bowie, J. U. (2019) *Nat. Commun. 10*, 565.
(55) Jennewein, S., Wildung, M. R., Chau, M., Walker, K., and Croteau, R. (2004) Random sequencing of an induced *Taxus* cell cDNA library for identification of clones involved in Taxol biosynthesis. *Proc. Natl. Acad. Sci. U. S. A. 101*, 9149−9154.
(56) Brown, S., Clastre, M., Courdavault, V., and O'Connor, S. E. (2015) De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. U. S. A. 112*, 3205−3210.
(57) Li, Y., Li, S., Thodey, K., Trenchard, I., Cravens, A., and Smolke, C. D. (2018) Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U. S. A. 115*, E3922−E3931.
(58) Srinivasan, P., and Smolke, C. D. (2020) Biosynthesis of medicinal tropane alkaloids in yeast. *Nature 585*, 614−619.
(59) Toparlak, Ö. D., Zasso, J., Bridi, S., Serra, M. D., Macchi, P., Conti, L., Baudet, M.-L., and Mansy, S. S. (2020) Artificial cells drive neural differentiation. *Sci. Adv. 6*, No. eabb4920.
(60) Adamala, K. P., Martin-Alarcon, D. A., Guthrie-Honea, K. R., and Boyden, E. S. (2017) Engineering genetic circuit interactions within and between synthetic minimal cells. *Nat. Chem. 9*, 431−439.
(61) Tsitkov, S., and Hess, H. (2019) Design principles for a compartmentalized enzyme cascade reaction. *ACS Catal. 9*, 2432−2439.
(62) Rideau, E., Dimova, R., Schwille, P., Wurm, F. R., and Landfester, K. (2018) Liposomes and polymersomes: a comparative review towards cell mimicking. *Chem. Soc. Rev. 47*, 8572.
(63) Has, C., and Sunthar, P. (2020) A comprehensive review on recent preparation techniques of liposomes. *J. Liposome Res. 30*, 336−365.
(64) Wang, G., Zhou, Y., Ma, K., Zhang, F., Ye, J., Zhong, G., and Yang, X. (2021) Bioconversion of recombinantly produced precursor peptide pqqA into pyrroloquinoline quinone (PQQ) using a cell-free *in vitro* system. *Protein Expression Purif. 178*, 105777.
(65) Kwon, Y.-C., Oh, I.-S., Lee, N., Lee, L.-H., Yoon, Y. J., Lee, E. Y., Kim, B.-G., and Kim, D.-M. (2013) Inegrating cell-free biosyntheses of heme prosthetic group and apoenzymes for the synthesis of functional P450 monooxygenase. *Biotechnol. Bioeng. 110*, 1193−1200.
(66) Sudhamsu, J., Kabir, M., Airola, M. V., Patel, B. A., Yeh, S.-R., Rousseau, D. L., and Crane, B. R. (2010) Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produces in *E. coli*. *Protein Expression Purif. 73*, 78−82.
(67) Lee, K. Y., Park, S.-J., Lee, K. A., Kim, S.-H., Kim, H., Meroz, Y., Mahadevan, L., Jusng, K.-H., Ahn, T. K., Parker, K. K., and Shin, K. (2018) Photosynthetic artificial organelles sustain and control ATP-dependent reactions in a protocellular system. *Nat. Biotechnol. 36*, 530−535.
(68) Berhanu, S., Ueda, T., and Kuruma, Y. (2019) Artificial photosynthetic cell producing energy for protein synthesis. *Nat. Commun. 10*, 1325.
(69) Nakamura, S., Suzuki, S., Saito, H., and Nishiyama, K.-I. (2018) Cholesterol blocks spontaneous insertion of membrane proteins into liposomes of phosphatidylcholine. *J. Biochem. 163*, 313−319.
(70) Nomura, K., Yamaguchi, T., Mori, S., Fujikawa, K., Nishiyama, K.-I., Shimanouchi, T., Tanimoto, Y., Morigaki, K., and Shimamoto, K. (2019) Alteration of membrane physicochemical properties by two factors for membrane protein integration. *Biophys. J. 117*, 99−110.
(71) Fontana, J., Sparkman-Yager, D., Zalatan, J. G., and Carothers, J. M. (2020) Challenges and opportunities with CRISPR activation in bacteria for data-driven metabolic engineering. *Curr. Opin. Biotechnol. 64*, 190−198.
(72) Fontana, J., Voje, W. E., Zalatan, J. G., and Carothers, J. M. (2018) Prospects for engineering dynamic CRISPR-Cas transcriptional circuits to improve bioproduction. *J. Ind. Microbiol. Biotechnol. 45*, 481−490.
(73) Caschera, F., and Noireaux, V. (2016) Compartmentalization of an all *E. coli* cell-free expression system for the construction of a minimal cel. *Artif. Life. 22*, 185−195.
(74) Fontana, J., Dong, C., Kiattisewee, C., Chavali, V. P., Tickman, B. I., Carothers, J. M., and Zalatan, J. G. (2020) Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. *Nat. Commun. 11*, 1618.
(75) Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M., and Klavins, E. (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat. Commun. 8*, 15459.
(76) Marshall, R., Maxwell, C. S., Collins, S. P., Jacobsen, T., Luo, M. L., Begemann, M. B., Gray, B. N., January, E., Singer, A., He, Y., Beisel, C. L., and Noireaux, V. (2018) Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. *Mol. Cell 69*, 146−157.
(77) Salje, J., van den Ent, F., de Boer, P., and Löwe, J. (2011) Direct membrane binding by bacterial actin MreB. *Mol. Cell 43*, 478−487.
(78) Ando, M., Schikula, S., Sasaki, Y., and Akiyoshi, K. (2018) Proteoliposome engineering with cell-free membrane protein synthesis: control of membrane protein sorting into liposomes by chaperoning systems. *Adv. Sci. 5*, 1800524.
(79) Niwa, T., Sasaki, Y., Uemura, E., Nakamura, S., Akiyama, M., Ando, M., Shinichi, S., Mukai, S.-a., Ueda, T., Taguchi, H., and Akiyoshi, K. (2016) Comprehensive study of liposome-assisted synthesis of membrane proteins using a reconstituted cell-free translation system. *Sci. Rep. 5*, 18025.
(80) Furusato, T., Horie, F., Matsubayashi, H. T., Amikura, K., Kuruma, Y., and Ueda, T. (2018) De novo synthesis of basal bacterial cell division proteins FtsZ, FtsA, and ZipA inside giant vesicles. *ACS Synth. Biol. 7*, 953−961.
(81) Das, A., Varma, S. S., Mularczyk, C., and Meling, D. D. (2014) Functional investigations of thromboxane synthase (CYP5A1) in lipid bilayers of nanodiscs. *ChemBioChem 15*, 892−899.
(82) Ravula, T., Barnaba, C., Mahajan, M., Anantharamaiah, G. M., Im, S.-C., Waskell, L., and Ramamoorthy, A. (2017) Membrane environment drives cytochrome P450's spin transition and its interaction with cytochrome b5. *Chem. Commun. 53*, 12798−12801.
(83) Majumder, S., Garamella, J., Wang, Y.-L., DeNies, M., Noireaux, V., and Liu, A. P. (2017) Cell-sized mechanosensitive and biosensing compartment programmed with DNA. *Chem. Commun. 53*, 7349−7352.
(84) Ohta, N., Kato, Y., Watanabe, H., Mori, H., and Matsuura, T. (2016) *In vitro* membrane protein synthesis inside Sec translocon-reconstituted cell-sized liposomes. *Sci. Rep. 6*, 36466.
(85) Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y., and Yomo, T. (2013) *In vitro* evolution of α-hemolysin using a liposome display. *Proc. Natl. Acad. Sci. U. S. A. 110*, 16796−16801.

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(86) Krishnan, M., de Leeuw, T. J. J. F., and Pandit, A. (2019) Cell-free soluble expression of the membrane protein PsbS. *Protein Expression Purif. 159*, 17−20.
(87) White, G. F., Shi, Z., Shi, L., Dohnalkova, A. C., Fredrickson, J. K., Zachara, J. M., Butt, J. N., Richardson, D. J., and Clarke, T. A. (2012) Development of a proteoliposome model to probe transmembrane electron-transfer reactions. *Biochem. Soc. Trans. 40*, 1257−1260.
(88) Biggs, B. W., Rouch, J. E., Kambalyal, A., Arnold, W., Lim, C. G., Mey, M. D., O'Neil-Johnson, M., Starks, C. M., Das, A., and Ajikumar, P. K. (2016) Orthogonal assays clarify the oxidative biochemistry of Taxol P450 CYP725A4. *ACS Chem. Biol. 11*, 1445−1451.

https://doi.org/10.1021/acssynbio.0c00625
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Appendix 3: Khakimzhan et al,. 2021 Physical Biology

IOP Publishing Journal Title
Physical Biology **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

Complex dependence of CRISPR-Cas9 binding strength on guide RNA spacer lengths

Aset Khakimzhan[1], David Garenne[1], Benjamin Tickman[2], Jason Fontana[2], James Carothers[2, 3, 4] and Vincent Noireaux[1]

[1]School of Physics and Astronomy, University of Minnesota, 115 Union Street SE, Minneapolis, MN 55455, USA
[2]Molecular Engineering & Sciences Institute, University of Washington, Seattle, WA, 98195, USA. jcaroth@uw.edu.
[3]Department of Chemical Engineering, University of Washington, Seattle, WA, 98195, USA. jcaroth@uw.edu.
[4]Center for Synthetic Biology, University of Washington, Seattle, WA, 98195, USA. jcaroth@uw.edu.

E-mail: khaki005@umn.edu; noireaux@umn.edu

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

It is established that for CRISPR-Cas9 genetic guide RNAs with 17-20bp long spacer sequences are optimal for accurate target binding and cleavage. In this work we perform cell-free CRISPRa (CRISPR activation) and CRISPRi (CRISPR inhibition) experiments to demonstrate the existence of a complex dependence of CRISPR-Cas9 binding as a function of the spacer length and complementarity. Our results show that significantly truncated or mismatched spacer sequences can form stronger guide-target bonds than the conventional 18-20bp long spacers. To explain this phenomenon, we take into consideration previous structural and single-molecule CRISPR-Cas9 experiments and develop a novel thermodynamic model of CRISPR-Cas9 target recognition.

Keywords: CRISPR-Cas9; CRISPRi; CRISPRa; Cell-Free Transcription-Translation; Synthetic Biology; Thermodynamics;

1. Introduction

CRISPR-Cas9 revolutionized molecular biology by providing an easy-to-use and programmable tool for accurate genome editing [1,2]. What makes CRISPR-dCas9/Cas9 both accurate and programmable is its use of a guide RNA (gRNA) to bind to a complementary target DNA. While much has been achieved in CRISPR-dCas9/Cas9 biotechnologies [3], the design of guide RNAs still adheres mostly to empirical models and ad hoc rules [4]. One of these rules is that the length of a guide RNA spacer sequence should be between 18bp and 20bp to securely bind to the target DNA [5].

From a naïve thermodynamic perspective, the longer spacer sequences result in more RNA:DNA bonds between the Cas9/gRNA complex and the target DNA, and more RNA:DNA bonds result in a more stable bond with the target. This intuition has been confirmed by previous CRISPR-dCas9/Cas9 experiments [6–8] and used for all previous CRISPR-dCas9/Cas9 binding models. Using a cell-free transcription-translation system (TXTL), we characterize the binding efficiencies of guide RNAs with significantly truncated or mismatched spacers and demonstrate that under some conditions those guide RNAs are comparable or are better at binding to the target DNA than their conventional 20bp spacer counterparts. In a series of CRISPRi (CRISPR interference) [9] and CRISPRa (CRISPR activation) [10,11] experiments, we demonstrate that binding strength of a CRISPR-dCas9/Cas9 system does not decrease monotonically with the length/complementarity of the guide RNA, but instead dips and peaks as a function of length/complementarity.

Since our experimental results are not explained by any of the previous CRISPR-Cas9 models of binding [12,13], we developed a novel thermodynamic model based on previous experiments about the conformational changes of CRISRP-Cas9 during target interrogation. We built the

model using energies and rates obtained from structural experiments [14,15], single-molecule Forster resonance energy transfer (smFRET) experiments [16–20], magnetic tweezer experiments [20,21], atomic force microscopy experiments [22], and molecular dynamics simulations [23–26].

All of the upper mentioned experiments share a common basis: they argue that CRISPR-Cas9 target interrogation occurs in discrete steps controlled by conformational changes that are in turn coordinated with the number of formed RNA:DNA bonds. Interestingly, it is hypothesized that for purposes of increased specificity Cas9 has evolved to its cleavage active conformation to be less stable than the intermediate conformation [27]. That idea is confirmed with the magnetic bead experiment, where even though only a single PAM-distal mismatch is introduced the binding free energy increases as the system progresses into the final DNA unwinding conformation [20]. Therefore, one can imagine that a short spacer sequence will form a strong bond if it can also minimize the probability of being destabilized by conformational changes. This is the core of our model and we demonstrate how balancing the energies and the rates of conformational changes cause the emergence of strongly binding truncated and mismatched spacer sequences.

2. Methods and Materials

2.1 Materials

DNA was purchased from Integrated DNA Technologies (Coralville, IA) and Twist Biosciences (San Francisco, CA). Unless otherwise mentioned all the other reagents were purchased from Sigma Aldrich (St. Louis, MO).

2.2 DNA constructs

For silencing experiments in TXTL, the Cas9 or dCas9 enzymes were synthesized constitutively through the endogenous Sp. pCas9 promoter [11]. CRISPRi was achieved using sgRNAs (single guide RNA), expressed from the Anderson *E. coli* promoter J23119 [http://parts.igem.org/Promoters/Catalog/Anderson]. The target template for CRISPRi was the plasmid P70a-*degfp* [28,29]. The reporter protein deGFP (enhanced Green Fluorescent Protein) is a slightly truncated version of eGFP with the same fluorescent properties as eGFP. For CRISPR activation experiments in TXTL, we used the same plasmids as for CRISPRi to express Cas9 or dCas9. The scRNAs (scaffold RNA) gene was expressed from the Anderson *E. coli* promoter J23119 [http://parts.igem.org/Promoters/Catalog/Anderson]. The activator gene was expressed constitutively from the Anderson *E. coli* promoter J23107 [30]. The reporter target template for CRISPRa were the plasmids pJF143.J2, pJF143.J3, and pJF143.J4 that contain the Anderson promoter J23117 [30] upstream of the mRFP gene. The sgRNA and scRNAs were expressed using linear templates, whose degradation in TXTL was prevented by adding the Chi6 dsDNA linear template [31,32]. All the constructs have been sequenced. These plasmids are available on demand.

2.3 TXTL Target template sequences

PAM, CRISPRi target, CRISPRa targets. The target template sequences for the CRISPRi experiments:

1. sg6: GCCAGAGGTAAAATTGTCAACACGCACGGTGT
2. sg9: CCAGGGTGTCGCCCTCGAACTTCACCTCGGCGC

The target template sequences for the CRISPRa experiments:

1. J206: CTTGTGTAGTAGCCGAACACGTCCTCAGGATG
2. J306: ACGACGTTGTGTCCAGAACGCTCCGTAGGAGA
3. J406: TTTCTAGAACATCCTTTCACTTCCGGAGGCGT

2.4 TXTL reactions

The myTXTL kit (Arbor Biosciences) was used for cell-free expression. TXTL reactions are composed of an *E. coli* lysate, an amino acid mixture, an energy buffer, and the desired DNA templates. This TXTL system has been described previously [28,29]. All the batch mode TXTL reactions were incubated at 29 °C in either a bench-top incubator, for endpoint measurements, or in plate readers, for kinetic measurements. 29 °C is the optimum temperature of incubation for the myTXTL kit.

2.5. Quantitative measurements of fluorescence in TXTL reactions

Fluorescence from batch mode TXTL reactions was measured using the reporter protein deGFP (25.4 kDa, 1 mg/ml = 39.4 μM) and mRFP (25.4 kDa, 1 mg/ml = 39.4 μM). Fluorescence was measured at five-minute intervals using monochromators (deGFP Ex/Em 488/525 nm, mRFP Ex/Em 555/583) on a Biotek Synergy H1 plate reader in Costar polypropylene 96-well, V-bottom plates sealed with a mat. Endpoint reactions were measured after 16 h of incubation. To measure protein concentration (eGFP reporter), a linear calibration curve of fluorescence intensity versus eGFP concentration was generated using purified recombinant eGFP obtained from Cell Biolabs (STA-201), Inc or purified in the lab.

2.6 Estimation of CRISPRa binding fraction

Binding of CRISPR-dCas9-SoxS near the promoter site of the reporter gene activates the transcription of the reporter gene mRNA, which in turn induces the synthesis of the reporter protein observed by an increase of the fluorescence intensity of the TXTL reaction. Thus, we estimated that the fraction of bound reporters is proportional to the intensity of TXTL reaction:

$$p_{Bound} \sim I_{TXTL} \quad (1)$$

Where I_{TXTL} is the fluorescence intensity of the TXTL reaction.

2.7 Estimation of CRISPRi binding fraction

The binding of CRISPR-dCas9 to the promoter or gene interferes with the transcriptional activity of the gene,

which in turn decreases the amount of reporter protein synthesized and thus the fluorescence intensity of the TXTL reaction. By comparing the fluorescence intensities of the TXTL CRISPRi experiments using a non-targeting sgRNA (sg-NT) with the fluorescence intensity of TXTL experiments using a targeting sgRNA (sg6 or sg9), we can estimate the fraction of silenced P70a-*degfp* genes. The formula to estimate the fraction of bound genes is the following:

$$f = 1 - \frac{I_{ON}}{I_{OFF}} \quad (2)$$

And f will be referred to as the fraction coefficient, I_{ON} is the fluorescent intensity of the on-target TXTL reaction, and I_{OFF} is the fluorescent intensity of the off-target TXTL reaction. The fraction coefficient of the non-target control is always $f = 0$.

2.8 Liquid handler systematic bias correction

The liquid handler (Labcyte Echo 550) used for to perform the CRISPRi experiments systematically introduced uneven expression levels on the 96-well plate. The calculation to account for the liquid handler introduced systematic error is presented in Appendix A (Supplementary file). The uncorrected data is presented in the Appendix in Supp. Figure 1.

2.9 Computer codes

Scripts are available on demand.

3. Experimental results

3.1 CRISPR-Cas9/dCas9 in TXTL

The TXTL system used in this work enables characterizing CRISPR technologies with an excellent agreement between the observations made *in vitro* and *in vivo* [29,33]. For this work we used both CRISPRi (CRISPR interference) and CRISPRa (CRISPR activation) systems in TXTL experiments (Figure 1). While the results from CRISPRi experiments would be sufficient to make the argument, we supplement them with CRISPRa experiments to verify that the results are general and not an artifact of either method. In addition, replacing the dCas9 with a Cas9 in a CRISPRa length/complementarity experiment is used to measure the guide RNA length/complementarity needed for Cas9 cleavage. The efficiency of CRISPR systems in TXTL is estimated by measuring the expression of the CRISPRi/a targeted genes. The dCas9 necessary for CRISPR experiments was either expressed in regular TXTL reaction or was available at a fixed concentration of approximately 20-50nM in the dCas9 pre-expressed TXTL system [33]. Pre-synthesized dCas9 in the TXTL system is convenient because, first, it simplifies the experiment by maintaining the total concentration of dCas9 constant and, second, it conserves the resources (ATP, amino acids) in the reaction.

Since we are mostly interested in the binding efficiency of CRISPR-Cas9/dCas9 systems most of the experiments involve only the use of dCas9. However, as mentioned above, when we are interested in determining the necessary length/complementarity of the guide RNA needed make Cas9 cleavage active, we also performed experiments with expressing Cas9 enzymes.

In the CRISPRi experiments with pre-synthesized dCas9 we added a deGFP expressing reporter target plasmid (P70a-*degfp*) and sgRNA expressing linear DNA (sg2, sg3, sg6, sg9, sg15, and sg-NT) (Figure 1a). When a regular extract (without dCas9 pre-synthesized) was used we also added a dCas9 expressing plasmid (pCD017). For the main text we either truncated or introduced mismatches into the spacers of the sg6 sgRNA to test the effects that such mutations have on the efficiency of CRISPRi systems in TXTL. And for the supplementary section we introduced mismatches to the rest of the used targets (sg2, sg3, sg9, and sg15).

As opposed to sgRNAs used in CRISPRi, CRISPRa systems utilize a scRNAs (scaffold RNA) for target search, which in addition to the regular single guide RNA structure also contains an MS2 RNA hairpin [30,34] that can bind to an MCP [35] (MS2 coat protein). By expressing the activator protein SoxS fused to MCP we can use the binding of the CRISPRa complex to the DNA to localize the SoxS activator [36] near the promoter region. For the CRISPRa experiments with pre-synthesized dCas9 we added an mRFP expressing reporter target plasmid (pJF143.J3.117), the activator MCP-SoxS (pBT005), and the scRNA DNA (linear scRNA expressing segment of pJF144.206.x, pJF144.306x, pJF144.406x) in TXTL reactions, while for the regular TXTL system experiments we also expressed *dCas9* or *Cas9* (Figure 1b). In the CRISPRa experiments the target sequence of the guide RNA matched with a region -81 bp from the transcription start site (TSS) [34]. The efficiency and the binding fraction of CRISPRa experiments is estimated using equation (2) from the methods section.

For the mismatch experiments, the spacer-target mismatches are always introduced consecutively from the PAM distal side. If a spacer has a maximum length of 20nt, but only has 11 PAM-proximal matches, and 9 PAM distal nucleotides are not complementary to the target, then we refer to it as a 11-20mm spacer/guide RNA (Figure 1c). More generally a spacer with N PAM-proximal bonds would be referred to as an N-20mm spacer/guide RNA. For truncation experiments, we exclusively remove the PAM-distal nucleotides of the spacer. If the spacer is only 11nt long and thus can at most form 11 RNA:DNA bonds, we refer to it as 11bp spacer/guide RNA (Figure 1d). As in the case of mismatches, a truncated spacer with N PAM-proximal nucleotides would be referred to as an N bp spacer/guide RNA.

3.2 CRISPRa mismatch experiments

In mismatch experiments we changed the PAM-distal nucleotides of the fully complementary guide RNA to test the effect PAM-distal mismatches have on the binding efficiency of a CRISPRa system. For the dCas9 CRISPRa mismatch experiments we used the pre-expressed *dCas9* TXTL system. We added 2 nM of the mRFP reporter plasmid (pJF143.J3.117) and added 2.5 nM of the MCP-SoxS activator plasmid (pBT005) to the TXTL reaction. We varied the concentration of the scRNA expressing DNA (pJF144.306.xmm) between 1 nM, 2 nM, and 4 nM. We observe that the activation decreases for the 15-20 mm scRNA (Figures 2a, 2b, and 2c), thus resulting in two distinct local optimal complementarities (12-20 mm and fully complementary). We do not observe any clear dependence between the concentration of the scRNA added and the activation levels of the target plasmid. That is probably because 1 nM of scRNA DNA provided a saturating concentration of the scRNA. However, independent of the concentration, we notice that the decrease in the activation level is not monotonic with the decrease of the number of PAM-proximal matches.

To test how many matched base pairs are required for Cas9 to enter its catalytically active state we performed the same experiment, but in a regular TXTL system with 1 nM of the Cas9 plasmid (pCas9 [11]). The concentrations of scRNAs are varied between 0.25 nM, 1 nM, and 2 nM (Figures 2d, 2e, and 2f). We observed that starting from the 16-20 mm guides the activation is small in comparison to the peaks between 10-20 mm and 13-20 mm guides, which suggests that 16 RNA:DNA bonds are sufficient for the Cas9 enzymes to become catalytically active [37] and instead of activating target plasmid, the Cas9-CRISPR complexes cut the target. Along with dCas9-CRISPR and Cas9-CRISPR experiment we performed experiments in which we express Cas9 in the pre-expressed dCas9 TXTL and we observe similar results to the Cas9-CRISPR experiments (Supp. Figure 2).

3.3. CRISPRa truncation experiments

First, we performed CRISPRa truncation experiments with a pre-synthesized dCas9 TXTL. We maintained the concentration of the activator plasmid at 3 nM and the concentration of the mRFP expressing target plasmid at 2.5 nM. The concentration of the scRNA expressing linear DNA was fixed at 0.25 nM, 0.5 nM, and 1 nM (Figures 3a, 3b, and 3c). For each of the concentrations we observe drops in activation for the 11 bp and 15-16 bp spacer lengths. As we increased the concentration of the scRNA in the TXTL reaction, the relative heights of the peaks at non-standard target sequence lengths have also increased (Figure 3d). We performed similar experiments (but with less different spacer lengths) for two more scRNA spacer sequences, J206 and J406, that targeted plasmids J2 and J4 respectively (Supp. Figures 3a and 3b). The targets were also located -81 bp away from the TSS. As in the experiment with the J306 spacer and the J3 reporter, we observed that some shorter spacer lengths can activate the reporter plasmid more efficiently than the longer spacer lengths.

The second series of experiments performed with CRISPRa truncations employed the catalytically active Cas9 enzyme instead of the catalytically inactive mutant dCas9. As in the experiments described above, the concentrations of activator plasmid and mRFP reporter plasmid were maintained at 3 nM and 2.5 nM respectively. Instead of using the TXTL with pre-synthesized dCas9, we used the regular TXTL and added 2 nM of a Cas9 encoding plasmid to the reaction [11]. Instead of observing 3 peaks as with the catalytically inactive CRISPRa, we only observe 2, with expression drops for the same lengths of 11nt and 15-16nt (Figures 3e and 3f). As in the CRISPRa mismatch experiments, when the spacer length is longer than 16 bp it can form 16 canonical RNA:DNA bonds, which in turn allow the Cas9 to become catalytically active. As in the CRISPRa mismatch experiments, the spacer lengths at which we observed activation efficiency drops for the dCas9 experiments (Figure 3d) is the same length for which we observe the transition to a cleavage active Cas9 enzyme in the Cas9 experiments (Figures 3e and 3f). We performed a similar experiment (but with less different spacer lengths) for the J206 spacer and the J2 target and confirm that the effect can be observed for it as well (Supp. Figure 3c).

3.4 CRISPRi mismatch experiments

The goal of the CRISPRi mismatch experiments was to observe how decreasing the number of matching PAM-proximal base pairs affects the silencing efficiency of CRISPR-dCas9 in TXTL. For the CRISPRi mismatch experiments we used a pre-synthesized dCas9 TXTL and added mismatched sgRNA expressing DNA and a reporter target DNA plasmid (P70a-*degfp*). The concentration of the target plasmid (P70a-*degfp*) was varied between 0.33 nM, 1 nM, and 3 nM, while the concentrations of the sgRNA expressing linear DNA were varied in the range of 0.25 nM-4 nM. In the experiments in which the concentration of the reporter target DNA (P70a-*degfp*) was maintained at 1 nM we observe two target sequence match lengths for which we observe fraction coefficient drops (Figures 4a and 4b). The first drop is observed for the 12-20 mm guide, while the second drop occurs for an 18-20 mm guide. We notice that as we change the concentration of the sgRNA expressing DNA the drops can become prominent. As we increase the concentration of the target plasmid to 3 nM, we also increase the number of target sequences that need to be silenced,

therefore the fraction coefficient drop for 18-20 mm guide is more prominent (Figure 4c). When we lowered the concentration of the reporter target plasmid to 0.33 nM, we can see a decrease in the fraction coefficient at the 12-20 mm guide, which disappears as we increase the concentration of the added sgRNA expressing linear DNA (Figure 4d). As in the CRISPRa mismatch and CRISPRa truncation experiments we notice multiple optimal binding lengths: one that corresponds to the lengths at which DNA unwinding occurs [20], and the other corresponds to the length at which CRISPR-Cas9 becomes cleavage active [15]. The data in figures 4a, 4b, and 4c have been corrected as described in Appendix A, while figure 4d was not corrected. The uncorrected data is presented in Supp. Figure 1.

Since the first dip for the 12-20mm guide (Figure 4b and 4d) is not as noticeable as the dip for the 18-20mm guide (Figure 4a and 4c), we additionally performed a significance measurement comparing the 11-20mm and the 12-20mm silencing strength. To have more data for the significance experiments we performed more CRISPRi mismatch experiments (Fig 4a, 4b, 4d, and Supp. Figures 4a, 4b, 4c). We then assembled the measured fraction coefficients from Figures 4a, 4b, 4d, and Supp. Figures 4a, 4b, 4c (56 measurements for 11-20mm and 12-20mm each) and performed 10,000,000 simulations calculating the difference for randomly split groups and plotting the distribution (Supp. Figure 4d). We determined that 0.6811% of the simulations had a larger difference between the means that the experimentally measured difference, thus resulting in a significance of $p = 0.006811$.

3.5 CRISPRi truncation experiments

The goal of the CRISPRi truncation experiments was to observe how shortening the target sequence on the guide RNA affects the silencing efficiency of CRISPR-dCas9. The truncated sgRNA were designed by removing PAM-distal base pairs of the spacer sequence. The CRISPRi truncation experiments were performed with pre-synthesized dCas9 TXTL and the concentration of the reporter target plasmid was maintained (P70a-*degfp*) at 0.5 nM. The concentration of the sgRNA expressing DNA was varied between 5 pM and 8 nM and the length of the spacer was varied between 5-32 nt. As we decrease the concentration of the sgRNA expressing DNA from 150 pM to 5 pM we observe the emergence of strong binding truncations that are far from the classical 20 nt length (Figure 5a). Surprisingly, the strongest binding sgRNA truncation was the 10 nt truncation. As in the CRISPRa truncation experiments the strength of binding of a CRISPR-complex does not decrease monotonically as we truncate the target sequence. We performed the same experiment with less truncations for more sgRNA sequences (sg2, sg3, sg9, and sg15) and we observe the anomalous binding patterns for those targets as well (Supp. Figure 5).

While it takes approximately 1 h for the concentration of an expressed RNA to reach a steady state in a TXTL reaction [38], we approximate for simplicity that the concentration of sgRNAs in the TXTL reaction is constant and is proportional to the concentration of sgRNA DNA added to the reaction. The second approximation we make, is disregarding the complex formation kinetics of CRISPR-dCas9, [Cr], and simply calculating its equilibrium value using formula (B3) described in Appendix B (Supplementary file). However, after approximating the concentration of CRISPR-dCas9 complexes, we noticed that the binding of CRISPR-dCas9 for some lengths was highly sensitive. To measure the sensitivity of CRISPR-dCas9 binding we fit the concentration of CRISPR-dCas9 complexes to the fraction coefficient, f, with the Hill equation:

$$\boldsymbol{f} = \frac{[\boldsymbol{Cr}]^{\boldsymbol{n_{Hill}}}}{[\boldsymbol{Cr}]^{\boldsymbol{n_{Hill}}} + \boldsymbol{K_D}^{\boldsymbol{n_{Hill}}}} \quad (3)$$

We used equations (3) and (4) to find the apparent K_M, K_D, n^{Hill}, and α values (Figure 5b), while keeping [dC] fixed at 50 nM. The fit showed little dependency on K_M if the value of K_M was less than 2 nM (Supp. Figure 6). This result agrees with *in vitro* kinetic studies that demonstrate that K_M values are on the order of 0.1-1nM even for truncated guides [39,40]. We infer that in TXTL experiments the difference in formation of CRISPR complexes between sgRNA with different target sequence truncations is negligible, since the concentration of dCas9 and sgRNAs is significantly larger than the K_M value. The steady state amount of sgRNAs per single sgRNA DNA was found to be $\alpha = 88.4$, which is a reasonable value considering that *degfp* mRNA has an $\alpha_{degfp} \sim 30$, while the *degfp* gene is 3 times longer than an sgRNA gene [38]. The K_D values did not increase monotonically with the decrease of the length of the target sequence, showing similar conclusions to the mismatch experiments (Figure 5c). Surprisingly, different truncations of the sgRNA target sequence also exhibited noticeably different levels of modularity in the fits (Figure 5d). A previous work studying the engineering of CRISPR based circuit in *E. coli* has reported the development of a bistable switch using CRISPRi [41]. They hypothesized that the unspecific binding of dCas9 served as a sufficient condition for bistability, as opposed to possible modularity of CRISPRi. It is possible that the cooperativity observed in the TXTL CRISPRi truncation experiment is an artifact of unspecific binding as well.

4. CRISPR-Cas9 binding model

4.1. Evidence from previous experiments

Our model of CRISPR-Cas9 binding focuses on estimating the unbinding rate of CRISPR-Cas9 enzymes from the target DNA as a function of spacer length and

matching. However, estimating the off-rate of a CRISPR-Cas9 system is not trivial since CRISPR-Cas9 undergoes multiple conformational changes during target interrogation [42]. Therefore, besides simply calculating the binding energy of the formed R-loop, it is also necessary to account for the effects conformational changes have on the stability of that R-loop.

Structural studies of CRISPR-Cas9 demonstrated that both the unwinding of the target DNA and the subsequent formation of an RNA-DNA helix induce changes in the structure of the Cas9 enzyme [14,37]. The final step of those conformational changes positions the nucleating domains of Cas9 (HNH and RuvC) near the cutting locations of the target DNA strands [43]. As in the TXTL experiments, these studies show that the catalytic activity of Cas9 is present only for spacers that can form canonical PAM-distal RNA:DNA bonds. These experimental results have been confirmed and explained with MD simulations of CRISPR-Cas9 systems [23–26].

Single-molecule Forster resonance energy transfer (smFRET) experiments demonstrated the existence of intermediate conformations in which CRISPR-Cas9 has formed a stable bond with the target sequence, but the Cas9 enzyme is not in its catalytically active conformation [16–19]. This intermediate step has been hypothesized as a checkpoint conformation that improves the specificity during target recognition. For a fully matched guide RNA the dwell times in the intermediate conformations range on the scales of 0.01s to 1s until the target DNA is cut [19], while for a mismatched or truncated guide RNA the system can be stably bound in the intermediate conformation for minutes [16]. In the qualitative models of CRISPR-Cas9 target interrogation unwinding of DNA is hypothesized to act as a proofreading mechanism. Studies reporting on the probability off-target effects in stretched DNA confirm that hypothesis [44].

In agreement with the smFRET experiments, a recent magnetic tweezer study demonstrated that the target recognition process for CRISPR-Cas9 occurs in discrete steps [20]. The guide RNAs used in the work demonstrated that Cas9 unwinds the first ten PAM proximal nucleotides, forms a 10 RNA:DNA bond R-loop, and only then unwinds the rest of the target. The rate of unwinding the 10-20 base pairs (the 10 PAM-distal base pairs) varied when changing the applied torque. For PAM-distal mismatched guide RNAs the stability of the 'open' state decreased in comparison to the 'intermediate' state even though it is likely that more RNA:DNA bonds are formed in the more progressed states despite the mismatch. A similar clue was observed in the CRISPRa TXTL experiments. The binding efficiency decreased for the same truncations and matching lengths at which CRISPR-Cas9 underwent a conformational change to become catalytically active (Figure 2, Figure 3). Therefore, for our model we assume that the net binding energy offered by a single RNA:DNA bond changes depending on the conformation. This difference in the net binding energy could be a difference in the mechanical strain of each CRISPR-Cas9 conformation

4.2. Model of CRISPR-Cas9 binding

By incorporating the information from the aforementioned experiments, we construct a model of CRISPR-Cas9 binding (Figure 6a). The model proposes that the system can inhabit 3 different 1-D energy landscapes: Seed (S), Intermediate (I), and Active (A). The landscapes (S, I, and A) represent the different states/conformations of CRISPR-Cas9 during target interrogation. As in single-molecule experiments, the conformation which the system inhabits is strongly dependent on the number of RNA:DNA bonds formed. The x-axis in the figure represents the number of RNA:DNA bonds formed between the spacer and the target DNA and will be referred to as sites. A similar model relying on 1-D energy landscapes with transitions between conformations has been proposed by Eslami-Mossallam et al [45] in parallel to our work. Their model was applied to the prediction of off-target effects in CRISPR-Cas9 genome editing and outperforms previous off-target prediction models [46,47]. An important distinction between the two approaches is that their model considers conformational changes as transitions on the energy landscape, while we consider conformational changes as transitions between energy landscapes. From a physical standpoint the conformational changes should also correspond to some transition within the energy landscape of the system, therefore our model can be considered as a 2-D energy landscape: one dimension for number of RNA:DNA bonds formed and the other dimension for the current conformation of the CRISPR-Cas9 complex.

For each matching RNA:DNA bond formed the binding free energy decreases by E_C. Physically E_C corresponds to the energy acquired by replacing a matching DNA:DNA bond with a matching RNA:DNA bond, and the approximate value of E_C can be approximated from melting curve experiments [48–50]. We are also interested in the energy the system acquires when encountering an RNA:DNA mismatch/absence and we refer to the parameter as E_{MM}. For a mismatch E_{MM} corresponds to the energy acquired by replacing a matching DNA:DNA bond with a mismatched RNA:DNA bond and can be approximate with melting curve experiments [51]. For an absence E_{MM} can be modelled as a large increase in binding free energy, since there is no RNA:DNA bond to be formed and thus the system has reached its ultimate state. As in the experiments, all the mismatches are introduced serially from the PAM-distal segment of the target.

Since the binding energy of the system can decrease as more RNA:DNA bonds are formed, we need to form an assumption based on previous experiments that account for these binding energy decreases. Our model does that by assuming that the average binding free energy provided by addition of an RNA:DNA bond decreases in the more progressed conformations/landscapes of target recognition. This is shown in the schematic of the model in which the slope of the more progressed states is flatter than the slope of the preceding states. The decrease in the average contribution of RNA:DNA bonds are scaled with factors S_1 and S_2, thus making the average RNA:DNA bond S_1E_C in the intermediate state and $S_1S_2E_C$ in the active state. The energies of an average RNA:DNA mismatch/absence, E_{MM}, are also scaled for each state. The average mismatch energy is also scaled in the other conformations with S_1E_{MM} in the intermediate state and $S_1S_2E_{MM}$ in the active state (Figure 6a). The seed and the intermediate states have maximal lengths L_S and L_I, respectively. These lengths determine the maximum number of RNA:DNA bonds that can form for each CRISPR-Cas9 conformation. Both lengths are measured from the start of the sequence and experimentally have been observed to be 9-11bp for L_S [20] and 16-19bp for L_I [37]. The total length of guide RNA would determine the ultimate site of the active conformation and for the model we used a total length of 22bp. However, the energy scaling assumption requires that a CRISPR-dCas9/Cas9 system can unbind from the target DNA from both the intermediate and active conformations without having to revert to the seed conformation. As of now there is no experimental evidence that conclusively supports or denies this possibility and thus this assumption is yet to be tested.

For this model we assume that the timescale of RNA:DNA bond formation [52] are significantly smaller than the timescales of CRISPR-Cas9 target search, the timescales of target DNA unwinding [53], and the timescale of Cas9 conformational changes. This approximation is only performed to simplify the model and avoid additional rate parameters for both the formation and dissociation of both matched and mismatched RNA:DNA bonds. However, with this assumption we can approximate that the distribution of R-loop sizes is near equilibrium and thus the probability of the system being in site n in landscape X is:

$$\boldsymbol{p_{n,X} = \frac{exp(-\beta E_{n,X})}{z_x}} \quad (4)$$

where z_x is the partition function of the energy landscape X and $E_{n,X}$ is the binding energy at site n [54]. The partition function does not account for degeneracies of sites in the energy landscape. The partition function is therefore calculated by adding up the exponentials of the energies of each site in the respective landscape, which 0 RNA:DNA (site 0) bonds representing the 0 k_bT state.

In the experiments, when zero RNA:DNA bonds are formed between the target and the guide RNA, it is possible that the PAM is still bound. In our model we account for PAM binding by estimating a k_{OFF}. Unbinding of CRISPR-Cas9 from the target DNA in our model can occur in all landscapes (S, I, A) and is proportional to the probability of occupying site 0 for the respective landscape. In a super resolution microscopy experiment, it has been observed that the average dwell time for a CRISPR-Cas9 with a mismatched spacer is on the order of 10-100 ms [55,56]. We use that value and approximate the dwell time of a system at site 0 (0 RNA:DNA bonds) to be the same as the dwell time of unbinding for a spacer with noncomplementary PAM-proximal nucleotides. Therefore, the off-rates k_{XF} from any of landscapes obeys the following equation:

$$\boldsymbol{k_{XF} = k_{OFF} p_{0,X}} \quad (5)$$

That is because if we consider the landscape of a fully mismatched spacer, site 0 is the most probable site and therefore the unbinding rate of a fully mismatched guide gives a good approximation of a base k_{OFF}.

An important assumption in our model is that the system undergoes conformation changes which effect a transition to a weaker binding energy landscape. The forward transitions between landscapes (S to I, I to A) can only occur from the final site of a given landscape (when $n = L_S$ for S, and $n = L_I$ for I). This approximation is justified for seed to intermediate since experimentally it has been observed that the rate of unwinding of target DNA is slower with PAM proximal matches, which inhibit the ability to enter the partial R-loop state [20]. And for the transition I to A, as pointed out previously, PAM-distal mismatches prevent the Cas9 enzyme to enter its catalytically active states, which hints that the binding of a PAM-distal mismatch is necessary for the transition.

For a fully complementary spacer sequence the dwell time for both transitions were measured to be on the scale of 1 s [19,27]. The rates of the transitions k_{SI} and k_{IA} are then approximately 1 s^{-1}. These values of the rates were used as the base rates to calculate the speed of transitions for non-complimentary spacer sequences. Forward transitions in the model can only happen from the most progressed site, so to find rates for non-complementary targets we can adjust the rates by comparing the distributions of the landscapes. For example, the transition rate of for a non-complementary spacer k^{nc}_{SI} from seed (S) to intermediate (I) is then:

$$\boldsymbol{k^{nc}_{SI} = k_{SI} \frac{p^{nc}_{L_S,S}}{p_{L_S,S}}} \quad (6)$$

where k_{SI} is the experimental smFRET rate for a fully complementary spacer, $p^{nc}_{Ls,S}$ is the probability of being in the final site of the seed (S) landscape for a non-complementary spacer, and L_S is the length of the seed (S)

landscape. The same principle can be applied to estimate the transition rate from the intermediate (I) landscape to the active (A) landscape. The total time between binding and cleavage has been measured to be on the order of 10 s using atomic force microscopy measurements [22]

The equilibrium constants of conformational changes, K_{SI} and K_{IA}, are the exponents of their respective Gibbs free energies changes. Therefore, the difference in the Gibbs free energies ΔG_{SI} and ΔG_{IA} is what determines the reverse rates k_{IS} and k_{AI}. The default values for both parameters were set to ΔG_{SI} = -3 k_bT and ΔG_{IA} = -3 k_bT, which are estimations based on previous single molecule experiments [20] and molecular dynamics simulations [23]. For our model we assume that the reverse transition rates (k_{IS} and k_{AI}) are independent of the complementarity of the spacer to the target. We made this approximation because experimentally we only have values for the reverse transition rates for the complementary spacer CRISPR-Cas9 binding. It is likely that the reverse transition rate will increase for a significantly altered spacer sequence, but it also cannot be arbitrarily large because there must be some attempt rate that serves as an upper-bound. Since we do not know this attempt rate and do not know how the system approaches the attempt rate with increased non-complementarity, we compromise by having a simpler model with landscape independent reversal rates.

The on-rate k_{FS} is inversely proportional to the average time it takes for a single CRISPR-Cas9 to form a PAM-bond at the target site. Estimating the on-rate can be a complex calculation with significant dependence on target DNA accessibility, diffusion, and enzyme structure. In our model we assume that the on-rate is independent of the spacer sequence. This assumption is justified because searching for the correct PAM sequence and forming a PAM bond is a random process that does not rely on information regarding the target sequence. We avoided a first-principles calculation of the on-rate and instead approximated the on-rate k_{FS} from the kinetics of TXTL experiment with an sgRNA that has a fully complementary spacer sequence (Supp. Figure 7a). In a TXTL experiment with pre-synthesized dCas9, nearly complete silencing of the target DNA occurs under 1 h (Supp. Figure 7b). Considering that for the sg6-10 nt guide the binding can be considered nearly irreversible and the concentration of the prepressed dCas9 in the TXTL reaction is approximately 20-50 nM [33], we estimate that the on-rate k_{FS} should be on the order of 0.001 s^{-1} nM^{-1}.

With all the rates and processes defined we can construct the system of differential equations that model the dynamics of CRISPR-Cas9 binding (Figure 6b). The equations describing the system are the following:

$$\frac{df}{dt} = -k_{FS}[Cr][T] + k_{SF}s + k_{IF}i + k_{AF}a \quad (7)$$

$$\frac{ds}{dt} = k_{FS}[Cr][T] + k_{IS}i - (k_{SF} + k_{SI})s \quad (8)$$

$$\frac{di}{dt} = k_{SI}s + k_{AI}a - (k_{IS} + k_{IA} + k_{IF})i \quad (9)$$

$$\frac{da}{dt} = k_{IA}i - (k_{AI} + k_{AF})a \quad (10)$$

where [Cr] is the concentration of the CRISPR-Cas9 complexes from equation (Appendix B.3) and *[T]* is the concentration of the target DNA. To calculate what ratio of targets are bound we set the system of ODEs to a steady state and find that the steady state concentration of all bound targets, which is the sum of all occupants of states S, I, and A. In steady state the ratio of bound targets obeys the enzyme-ligand binding equation with a dissociation constant K_D:

$$K_D = \frac{k_{SF}(1 - \gamma\delta - \delta) + k_{IF}\delta + k_{AF}\gamma\delta}{k_{FS}} \quad (11)$$

$$\gamma = \frac{k_{IA}}{k_{AF} + k_{AI}} \quad (12)$$

$$\delta = \frac{k_{SI}}{k_{IF} + k_{IS} + k_{IA} - k_{AI}\gamma + k_{SI}(1 + \gamma)} \quad (13)$$

Equation (11) is the final derivation of the model and we used it to demonstrate the emergence of peaks as a function of spacer length and complementarity. A summary of all the default parameters and their sources are presented in Table 1.

Parameter	Approximate Value	Description
k_{FS}	0.001 nM^{-1} s^{-1} (TXTL kinetic Supp. Fig 7a and 7b)	CRISPR-dCas9 target DNA search rate/on-rate.
k_{OFF}	10-100 s^{-1} [55,56]	Off-rate of a CRISPR-dCas9 with PAM-proximal mismatches.
k_{SI}, k_{IA}	1 s^{-1} [19,27,53]	The rates of forward conformational changes (S to I, I to A).
E_C	-1 k_bT [48–50]	Average free energy difference between an RNA:DNA match and an DNA:DNA match.
E_{MM}	4 k_bT [51]	Average free energy difference between an RNA:DNA mismatch and a DNA:DNA match.
ΔG_{SI}, ΔG_{IA}	-3 k_bT [20,23]	The free energy difference between conformations (S and I, I and A).
S_1, S_2	0.75 (no reference)	The scaling factors to model the binding free energy increase in the more progressed conformations.
L_S	10 bp [20]	The maximal number of RNA:DNA bonds that can

		form in the seed conformation.
L_I	18 bp [37]	The maximal number of RNA:DNA bonds that can form in the intermediate conformation.
K_M	10 pM – 1 nM [39,40]	The equilibrium constant of dCas9/Cas9-gRNA complex formation.
[dC]	20-50 nM [33]	Concentration of dCas9 in a TXTL extract with pre-expressed dCas9.
[sg]	1 nM (TXTL experimental value)	Concentration of sgRNA/scRNA expressing DNA.
α	50 (CRISPRi fit, [38])	Steady state ratio of guide RNAs to guide RNA expressing DNA.
[T]	0.5 nM (TXTL experimental value)	Concentration of target DNA in the TXTL experiment.

Table 1: Default values of parameters for modelling of length/complementarity dependences in CRISPR-Cas9 TXTL experiments. The n/a for scales s_1 and s_2 means there are no measured values of the parameter in available literature or TXTL experiments.

A visual demonstration of the effect truncations or mismatches can carry on the energy landscapes are presented in the following figures: Figure 6c for a guide with a spacer shorter than L_S, Figure 6d for a guide with a spacer longer than Ls, but shorter than L_I, and Figure 6e for a guide longer than L_I.

4.3. Emergence of peaks in CRISPR-Cas9 truncation and mismatch experiments

We used equation (11) to test how varying the parameters of the energy landscapes and the transition rates affected the binding efficiency of CRISPR-Cas9 with truncated or mismatched spacers (Figure 7). Unless stated otherwise, each of the plots were performed by only perturbing one parameter from table 1, while keeping the rest of parameters the same.

We start with varying the parameter of the main assumption of our model. The assumptions states that the absolute values of RNA:DNA bonds and of RNA:DNA mismatches/absences decrease as CRISPR-Cas9 enters its more progressed conformations. In the seed conformation the parameters are E_C and E_{MM}, in the intermediate conformation the parameters are s_1E_C and s_1E_{MM}, and in the active conformation the parameters are $s_2s_1E_C$ and $s_2s_1E_{MM}$. For simplicity, we assigned s_1 and s_2 to be the same and varied their values from 0.5 to 1 (Figure 7a). As the scaling factors decrease, we notice an emergence of optimal lengths, which appear to be 8, 15, and 22 bp. The optimal lengths are caused by the trapping of the system in a weaker energy landscape. If the deepest site in landscape S is deeper than the deepest site in landscape I, and the transition rate k_{IS} is not negligible, then the landscape I will act as an unbinding sink in the system of differential equations (7-10). The same principle can be applied to the transition from the landscape I to landscape A.

In the TXTL experiments we observed that the optimal lengths are different for different guide RNAs (Figures 2-5). We hypothesize that it is caused by the sequence dependent nature of the lengths of the seed (S) and intermediate (I) landscapes. These lengths correspond to the average number of RNA:DNA bonds the CRISPR-Cas9-Target system needs to form to undergo an unwinding event (S to I) and conformational change (I to A). In this work, we set the landscape lengths to L_S = 10bp and L_I = 18bp. Unwinding of DNA and the conformational changes of Cas9 are complex processes that depend on many variables including the sequence. Because of the sequence dependence of DNA unwinding, there could be different options for landscape lengths L_S and L_I. We tested equation (11) with four different combinations of L_S and L_I (L_S = 8bp and L_I = 16bp, L_S = 9bp and L_I = 17bp, L_S = 10bp and L_I = 18bp, L_S = 11bp and L_I = 19bp) (Figure 7b). Depending on the lengths of the landscapes L_S and L_I the optimal lengths can also vary. It is important to note that our model only accounts for an average amount of RNA:DNA bonds needed to progress to the next step of target recognition.

We plotted the length/complementarity dependence for different E_C values (Figure 7c). Guides with larger differences between the average RNA:DNA bond energy and the DNA:DNA bond energy expectedly had more target DNA bound for all lengths. In our model a mismatch/absence does not actively destabilize the system, but instead makes the system less likely to occupy the final sites of landscapes. That in turn decreases the transition rate to the next landscape. However, since the transition to the next landscape destabilizes the system, it might be beneficial for shorter guides to have larger E_{MM}. We confirm this idea by varying the energetic cost of an RNA:DNA mismatch/absence (Figure 7d). For E_{MM} = 8 k_bT the unconventional optimal lengths (8-9 bp and 14-16 bp) have stronger bonds than for E_{MM} = 1 k_bT or E_{MM} = 2 k_bT. However, when the E_{MM} = 0 the ratio of bound targets monotonically increases with length. The larger E_{MM} values can be considered as modelling the absences, since the last RNA:DNA bond of a truncated spacer should represent the ultimate state of the landscape. The E_{MM} values that range between 2-6k_bT are better suited to demonstrate the effect of mismatches, since those are within the range of realistic mismatch energies [51].

While multiple single-molecule experiments and molecular simulations have been performed that confirm the

various conformational changes that CRISPR-Cas9 undergoes during target recognition, the equilibrium constants of the conformational changes vary between experiments. We tested how changing the value of equilibrium coefficients, K_{SI} and K_{IA}, affect the binding efficiency vs complementarity/length dependence (Figure 7e and 7f). We varied the free energies changes associated with the conformational changes from -6 k_bT to 6 k_bT. For most of the conditions we could clearly observe multiple optimal lengths, but for some of the conditions (ΔG_{SI} = 3 k_bT, 6 k_bT) the unconventional optimal values disappear.

Finally, it is of interest to understand how the stoichiometry of TXTL experiments assisted in the observation of the unconventional peak lengths. First, we consider how changing the concentration of the guide RNA expressing DNA affects the complementarity/length dependence (Figure 7g). As the concentration of the sgRNA DNA was increased, the multiple peaks of the complementarity/length plot became more visible. Conversely, if the concentration of CRISPR-Cas9 is too high the poorly binding lengths between optimums can saturate and reduce the visibility of peaks. Therefore, revealing binding peaks requires a balancing of concentrations.

Another important factor in the stoichiometry of CRISPR-Cas9 is the dependence on K_M, the equilibrium constant of Cas9-sgRNA/scRNA complex formation. Previous reports have demonstrated that depending on the length of a spacer K_M can vary. For truncated spacers, the K_M can increase to the scale of 1 nM [39] and that results in a lower concentration of CRISPR-dCas9/Cas9 complexes with truncated spacers. In the conducted TXTL experiments the concentrations of the free dCas9 and free sgRNA/scRNA were significantly larger than the possible equilibrium constants K_M of guides with truncated spacers, therefore effects caused by CRISPR-Cas9 complex formations were unnoticed. However, for an experiment in which the concentrations of the CRISPR components are closer to the K_M values, the sgRNAs/scRNAs with truncated spacers might appear as poorly binding guides. Therefore, the dCas9-gRNA binding kinetics can mask the peaks under conditions that either dCas9 or the gRNA or both have comparable concentrations to the K_M values. We calculated the expected Target bound ratio, when the concentration of the total dCas9 in the reaction is set to [dCas9] = 0.5nM and ranging the K_M values from 0.1nM to 2nM, and we indeed observe that if the K_M value increases for more truncated guide RNA, then the peaks can be masked.

Finally, the last parameter of the model, k_{CUT} determines the rate at which the target DNA is cut when the system is in the cleavage competent state and more than L_I bonds are formed. Therefore, to acquire the cleavage rate for a specific target, value for each target we use the following equation:

$$\boldsymbol{k_{CUT} = k_{CutNorm} * p_{n > L_I}} \quad (14)$$

Where $p_{n>Li}$ is the probability of having more than L_I RNA:DNA bonds in the active conformation, which can be calculated with equation (5), and $k_{CutNorm}$ is the approximate rate of cleavage for a fully matched target. By modelling the system as an absorbing Markov chain, we can calculate the mean time to DNA cleavage and, thus, the global cleavage rate of the system [57]. We do not know work that measures the target DNA cleavage rate once the CRISPR-Cas9 system is in the cleavage active conformation. However, based on an AFM experiment [22] and smFRET experiments [16,27] we can approximate the rate to be on the order of $k_{CutNorm}$ = 1 s^{-1} and still get a good understanding of the relation between length/complementarity and the mismatch/absence energy E_{MM}. We ranged E_{MM} from 2k_bT to 8k_bT and plotted the global cleavage rate as a function of the length/complementarity (Figure 7i) . Expectedly, we observe that the global cleavage rate is approximately the same as the on-rate k_{FS} for targets that can form at least L_I RNA:DNA matches. If E_{MM} is lower, then the system can cleave the target even with spacers that have lengths/complementarities of less than L_I.

5. Conclusions

Predicting the binding behaviour of CRISPR-Cas9 systems to target DNA sites and non-target DNA sites is an important problem for CRISPR-Cas9 applications. As of now, much of the prediction of on-target and off-target effects is done with empirical models. It is critical to develop physics-based models that account for experimental information, especially when it comes to the development of high-risk CRISPR-Cas9 tools. Since the CRISPR genome editing is still in relatively early stages, physics-based models are being developed part-by-part. There is still important to consider a wide range of timescales, levels of detail, and focus on critical elements of the target interrogation steps, before a more sophisticated evidence model is developed. This work demonstrates the importance of considering conformational changes and DNA unwinding during to model CRISPR-Cas9 target interrogation, which we hope can assist in the development of more complete CRISPR-Cas9 models in the future.

Our model also provides ideas for new potential experiments. First, an experiment that measures whether the binding affinity does decrease in more progressed CRISPR-Cas9 conformations that is paired with a measurement that tracks the mechanical strain the Cas9 enzymes causes on the target DNA can be performed to falsify the model. Also, simulations and experiments need to be performed to analyse the effect conformational changes themselves have on the stability of the system. We can imagine a situation in which the factors S_1 and S_2 are caused by CRISPR-Cas9 undergoing

frequent forward and reverse conformational changes during which the CRISPR system is in a transient unstable state. Finally, our model utilizes many parameters that were obtained for a small number of Cas9-gRNA-DNA systems in a cell-free environment. Therefore, to develop a more realistic model based on the dynamics of CRISPR-Cas9 there is a need for high throughput experiments that can track the rates of conformational changes during target interrogation both in-vitro and in-vivo.

Acknowledgements

A.K. and V.N. would like to thank Seth Thompson for plasmid amplifications and Diego Alba Burbano for his helpful manuscript comments.

Funding

This work was supported by the National Science Foundation [CBET 1844152 to V.N. and J.C.]; and the Binational Science Foundation [award 2018208 to V.N.].

Conflict of interest

The Noireaux laboratory receives research funds from Arbor Biosciences, a distributor of the myTXTL cell-free protein synthesis kit.

References

[1] Barrangou R and Doudna J A 2016 Applications of CRISPR technologies in research and beyond *Nature Biotechnology* **34** 933–41

[2] Anzalone A v., Koblan L W and Liu D R 2020 Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors *Nature Biotechnology* **38** 824–44

[3] Pickar-Oliver A and Gersbach C A 2019 The next generation of CRISPR–Cas technologies and applications *Nature Reviews Molecular Cell Biology* **20** 490–507

[4] Mohr S E, Hu Y, Ewen-Campen B, Housden B E, Viswanatha R and Perrimon N 2016 CRISPR guide RNA design for research applications *FEBS Journal* 3232–8

[5] Santos-Moreno J and Schaerli Y 2020 CRISPR-based gene expression control for synthetic gene circuits *Biochemical Society Transactions* **48** 1979–93

[6] Kiani S, Chavez A, Tuttle M, Hall R N, Chari R, Ter-Ovanesyan D, Qian J, Pruitt B W, Beal J, Vora S, Buchthal J, Kowal E J K, Ebrahimkhani M R, Collins J J, Weiss R and Church G 2015 Cas9 gRNA engineering for genome editing, activation and repression *Nature Methods* **12** 1051–4

[7] Yin H, Song C Q, Suresh S, Kwan S Y, Wu Q, Walsh S, Ding J, Bogorad R L, Zhu L J, Wolfe S A, Koteliansky V, Xue W, Langer R and Anderson D G 2018 Partial DNA-guided Cas9 enables genome editing with reduced off-target activity *Nature Chemical Biology* **14** 311–6

[8] Fu Y, Sander J D, Reyon D, Cascio V M and Joung J K 2014 Improving CRISPR-Cas nuclease specificity using truncated guide RNAs *Nature Biotechnology* **32** 279–84

[9] Qi L S, Larson M H, Gilbert L A, Doudna J A, Weissman J S, Arkin A P and Lim W A 2013 Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression *Cell* **152** 1173–83

[10] Zalatan J G, Lee M E, Almeida R, Gilbert L A, Whitehead E H, la Russa M, Tsai J C, Weissman J S, Dueber J E, Qi L S and Lim W A 2015 Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds *Cell* **160** 339–50

[11] Bikard D, Jiang W, Samai P, Hochschild A, Zhang F and Marraffini L A 2013 Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system *Nucleic Acids Research* **41** 7429–37

[12] Klein M, Eslami-Mossallam B, Arroyo D G and Depken M 2018 Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules *Cell Reports* **22** 1413–23

[13] Farasat I and Salis H M 2016 A Biophysical Model of CRISPR/Cas9 Activity for Rational

Design of Genome Editing and Gene Regulation *PLoS Computational Biology* **12**

[14] Jinek M, Jiang F, Taylor D W, Sternberg S H, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone A T, Charpentier E, Nogales E and Doudna J A 2014 Structures of Cas9 endonucleases reveal RNA-mediated conformational activation *Science* **343**

[15] Sternberg S H, Lafrance B, Kaplan M and Doudna J A 2015 Conformational control of DNA target cleavage by CRISPR-Cas9 *Nature* **527** 110–3

[16] Singh D, Sternberg S H, Fei J, Doudna J A and Ha T 2016 Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9 *Nature Communications* **7** 1–8

[17] Dagdas Y S, Chen J S, Sternberg S H, Doudna J A and Yildiz A 2017 A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9 *Science Advances* **3** 1–9

[18] Chen J S, Dagdas Y S, Kleinstiver B P, Welch M M, Sousa A A, Harrington L B, Sternberg S H, Joung J K, Yildiz A and Doudna J A 2017 Enhanced proofreading governs CRISPR-Cas9 targeting accuracy *Nature* **550** 407–10

[19] Osuka S, Isomura K, Kajimoto S, Komori T, Nishimasu H, Shima T, Nureki O and Uemura S 2018 Real-time observation of flexible domain movements in CRISPR–Cas9 *The EMBO Journal* **37** 1–11

[20] Ivanov I E, Wright A v., Cofsky J C, Palacio Aris K D, Doudna J A and Bryant Z 2020 Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling *Proceedings of the National Academy of Sciences of the United States of America* **117** 5853–60

[21] Szczelkun M D, Tikhomirova M S, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V and Seidel R 2014 Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes *Proceedings of the National Academy of Sciences of the United States of America* **111** 9798–803

[22] Shibata M, Nishimasu H, Kodera N, Hirano S, Ando T, Uchihashi T and Nureki O 2017 Real-space and real-Time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy *Nature Communications* **8** 1–9

[23] Palermo G, Miao Y, Walker R C, Jinek M and McCammon J A 2017 CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations *Proceedings of the National Academy of Sciences of the United States of America* **114** 7260–5

[24] Ricci C G, Chen J S, Miao Y, Jinek M, Doudna J A, McCammon J A and Palermo G 2019 Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics *ACS Central Science* **5** 651–62

[25] Casalino L, Nierzwicki Ł, Jinek M and Palermo G 2020 Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics *ACS Catalysis* **10** 13596–605

[26] East K W, Newton J C, Morzan U N, Narkhede Y B, Acharya A, Skeens E, Jogl G, Batista V S, Palermo G and Lisi G P 2020 Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics *Journal of the American Chemical Society* **142** 1348–58

[27] Yang M, Peng S, Sun R, Lin J, Wang N and Chen C 2018 The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET *Cell Reports* **22** 372–82

[28] Shin J and Noireaux V 2012 An E. coli cell-free expression toolbox: Application to synthetic gene circuits and artificial cells *ACS Synthetic Biology* **1** 29–41

[29] Garamella J, Marshall R, Rustad M and Noireaux V 2016 The All E. coli TX-TL Toolbox

2.0: A Platform for Cell-Free Synthetic Biology *ACS Synthetic Biology* **5** 344–55

[30] Dong C, Fontana J, Patel A, Carothers J M and Zalatan J G 2018 Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria *Nature Communications* **9**

[31] Marshall R, Maxwell C S, Collins S P, Beisel C L and Noireaux V 2017 Short DNA containing χ sites enhances DNA stability and gene expression in E. coli cell-free transcription–translation systems *Biotechnology and Bioengineering* **114** 2137–41

[32] Suzuki S, Ohta K I, Nakajima Y, Shigeto H, Abe H, Kawai A, Miura R, Kazuki Y, Oshimura M and Miki T 2020 Meganuclease-Based Artificial Transcription Factors *ACS Synthetic Biology* **9** 2851–5

[33] Marshall R, Maxwell C S, Collins S P, Jacobsen T, Luo M L, Begemann M B, Gray B N, January E, Singer A, He Y, Beisel C L and Noireaux V 2018 Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System *Molecular Cell* **69** 146-157.e3

[34] Fontana J, Dong C, Kiattisewee C, Chavali V P, Tickman B I, Carothers J M and Zalatan J G 2020 Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements *Nature Communications* **11** 1–11

[35] Johansson H E, Liljas L and Uhlenbeck O C 1997 RNA recognition by the MS2 phage coat protein *Seminars in Virology* **8** 176–85

[36] Griffith K L and Wolf R E 2004 Genetic evidence for pre-recruitment as the mechanism of transcription activation by SoxS of Escherichia coli: The dominance of DNA binding mutations of SoxS *Journal of Molecular Biology* **344** 1–10

[37] Sternberg S H, Lafrance B, Kaplan M and Doudna J A 2015 Conformational control of DNA target cleavage by CRISPR-Cas9 *Nature* **527** 110–3

[38] Marshall R and Noireaux V 2019 Quantitative modeling of transcription and translation of an all-E. coli cell-free system *Scientific Reports* **9**

[39] Wright A v., Sternberg S H, Taylor D W, Staahl B T, Bardales J A, Kornfeld J E and Doudna J A 2015 Rational design of a split-Cas9 enzyme complex *Proceedings of the National Academy of Sciences of the United States of America* **112** 2984–9

[40] Mekler V, Minakhin L, Semenova E, Kuznedelov K and Severinov K 2016 Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA *Nucleic Acids Research* **44** 2837–45

[41] Santos-Moreno J, Tasiudi E, Stelling J and Schaerli Y 2020 Multistable and dynamic CRISPRi-based synthetic circuits *Nature Communications* **11** 1–8

[42] Jiang F, Zhou K, Ma L, Gressel S and Doudna J A 2015 A Cas9-guide RNA complex preorganized for target DNA recognition *Science* **348** 1477–81

[43] Jiang F, Taylor D W, Chen J S, Kornfeld J E, Zhou K, Thompson A J, Nogales E and Doudna J A 2016 Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage *Science* **351** 867–71

[44] Newton M D, Taylor B J, Driessen R P C, Roos L, Cvetesic N, Allyjaun S, Lenhard B, Cuomo M E and Rueda D S 2019 DNA stretching induces Cas9 off-target activity *Nature Structural and Molecular Biology* **26** 185–92

[45] Eslami-Mossallam B, Klein M, Smagt C, Sanden K, Jones S, Hawkins J, Finkelstein I and Depken M 2020 A kinetic model improves off-target predictions and reveals the physical basis of Sp Cas9 fidelity *bioRxiv* 2020.05.21.108613

[46] Doench J G, Fusi N, Sullender M, Hegde M, Vaimberg E W, Donovan K F, Smith I, Tothova Z,

Wilen C, Orchard R, Virgin H W, Listgarten J and Root D E 2016 Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9 *Nature Biotechnology* **34** 184–91

[47] Zhang D, Hurst T, Duan D and Chen S J 2019 Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design *Proceedings of the National Academy of Sciences of the United States of America* **116** 8693–8

[48] Sugimoto N, Nakano S ichi, Katoh M, Matsumura A, Nakamuta H, Ohmichi T, Yoneyama M and Sasaki M 1995 Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes *Biochemistry* **34** 11211–6

[49] SantaLucia J 1998 A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics *Proceedings of the National Academy of Sciences of the United States of America* **95** 1460–5

[50] SantaLucia J, Allawi H T and Seneviratne P A 1996 Improved nearest-neighbor parameters for predicting DNA duplex stability *Biochemistry* **35** 3555–62

[51] Watkins N E, Kennelly W J, Tsay M J, Tuin A, Swenson L, Lee H R, Morosyuk S, Hicks D A and Santalucia J 2011 Thermodynamic contributions of single internal rAdA, rCdC, rGdG and rUdT mismatches in RNA/DNA duplexes *Nucleic Acids Research* **39** 1894–902

[52] Ouldridge T E, Šulc P, Romano F, Doye J P K and Louis A A 2013 DNA hybridization kinetics: Zippering, internal displacement and sequence dependence *Nucleic Acids Research* **41** 8886–95

[53] Gong S, Yu H H, Johnson K A and Taylor D W 2018 DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity *Cell Reports* **22** 359–71

[54] Kittel C, Kroemer H and Scott H L 1998 Thermal Physics, 2nd ed. *American Journal of Physics* **66** 164–7

[55] Jones D L, Leroy P, Unoson C, Fange D, Ćurić V, Lawson M J and Elf J 2017 Kinetics of dCas9 target search in Escherichia coli *Science* **357** 1420–4

[56] Martens K J A, van Beljouw S P B, van der Els S, Vink J N A, Baas S, Vogelaar G A, Brouns S J J, van Baarlen P, Kleerebezem M and Hohlbein J 2019 Visualisation of dCas9 target search in vivo using an open-microscopy framework *Nature Communications* **10** 1–11

[57] Barbour A D and Resnick S I 1993 Adventures in Stochastic Processes. *Journal of the American Statistical Association* **88** 1474

Figures

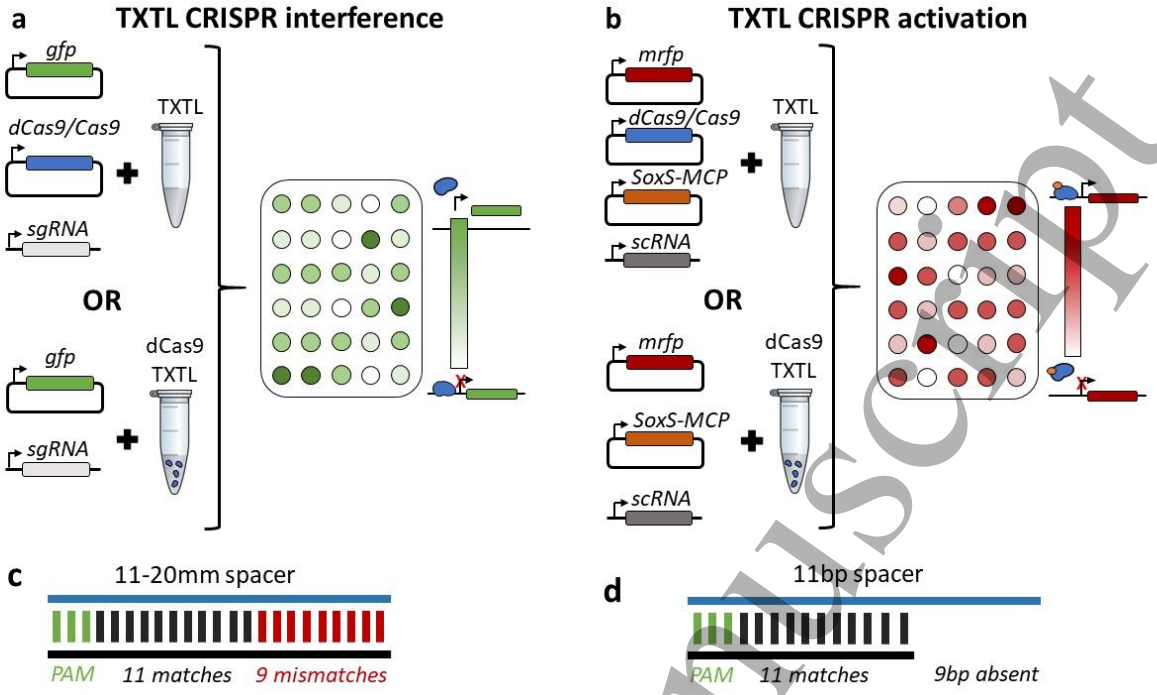


Figure 1. Schematic of the TXTL experiment. **(a)** The TXTL CRISPR interference (CRISPRi) experiment is composed of two plasmids to express the genes *degfp* and *dCas9*, and a linear DNA to express the *sgRNA* (27). The spacer sequence of the sgRNA matches with a sequence on the promoter (P70a) of the *degfp* gene (P70a-*degfp* plasmid). The TXTL reactions are incubated at 29 °C on a well plate. The measured fluorescence signals are analyzed quantitatively using either kinetics or endpoints measurements. The experiment can also be performed with a TXTL system in which dCas9 is pre-synthesized (dCas9 TXTL), thus simplifying the reaction given that the concentration of dCas9 is constant and dCas9 does not need to be dynamically synthesized. All other parts of the experiment remain the same. **(b)** The TXTL CRISPR activation (CRISPRa) is composed of three plasmids and one linear DNA part: an mRFP target plasmid, a dCas9 (pCD017) or Cas9 (pCas9) plasmid, a SoxS-MCP plasmid (pBT005), and a linear DNA to express the *scRNA* (25). The scRNA spacer matches with a sequence 70 bp upstream the ribosome binding site (RBS) of the mRFP gene. The TXTL reactions are incubated at 29 °C on a well plate. The fluorescence signals are analyzed quantitatively using either kinetics or endpoints measurements. As in CRISPRi experiments, we can also simplify the TXTL reaction by using TXTL system with pre-synthesized dCas9 (dCas9 TXTL), thus removing the need to express *dCas9* and maintaining its concentration constant. **(c)** A picture to demonstrate the naming convention of mismatched spacers. In the picture the spacer has 11 PAM-proximal matches out of 20 total, and is thus named an 11-20mm spacer. If a spacer had *N* PAM-proximal matches, then it would an *N*-20mm spacer **(d)** A picture demonstrating the naming convention for truncater spacers. If a spacer has only 11 of the PAM-proximal matches, then we refer to it as 11bp spacer. If the spacer was only *N* bp long, then it would be refered to as an *N*bp spacer.

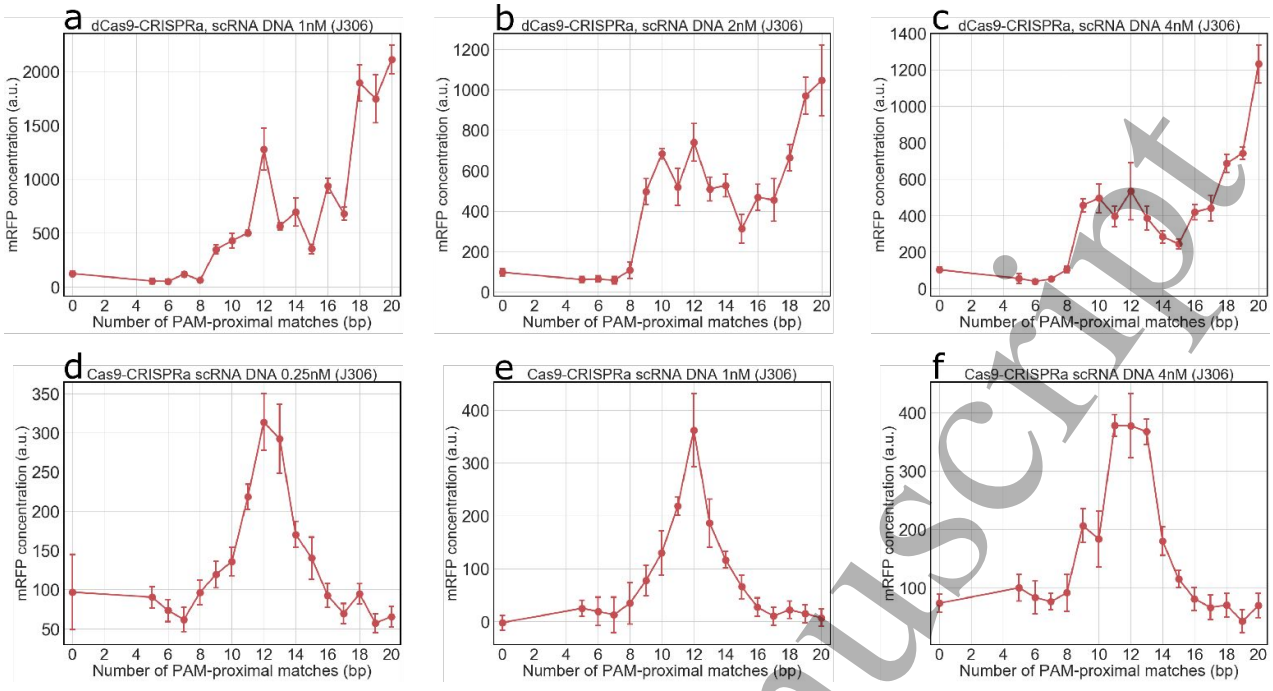


Figure 2. CRISPRa mismatch experiments in TXTL. **(a), (b), and (c)** are results of pre-synthesized dCas9 CRISPRa mismatch experiments with 1 nM, 2 nM, and 4 nM of scRNA DNA, respectively. **(d), (e), and (f)** repeat the conditions of **(a), (b), and (c)** but in addition the regular TXTL reaction has 1 nM of a Cas9 plasmid. The observed decline in the expression for the 16-20 nt match lengths is caused by catalytic activation of the CRISPR-Cas9 complex, which in turn cuts the reporter target plasmid instead of activating it.

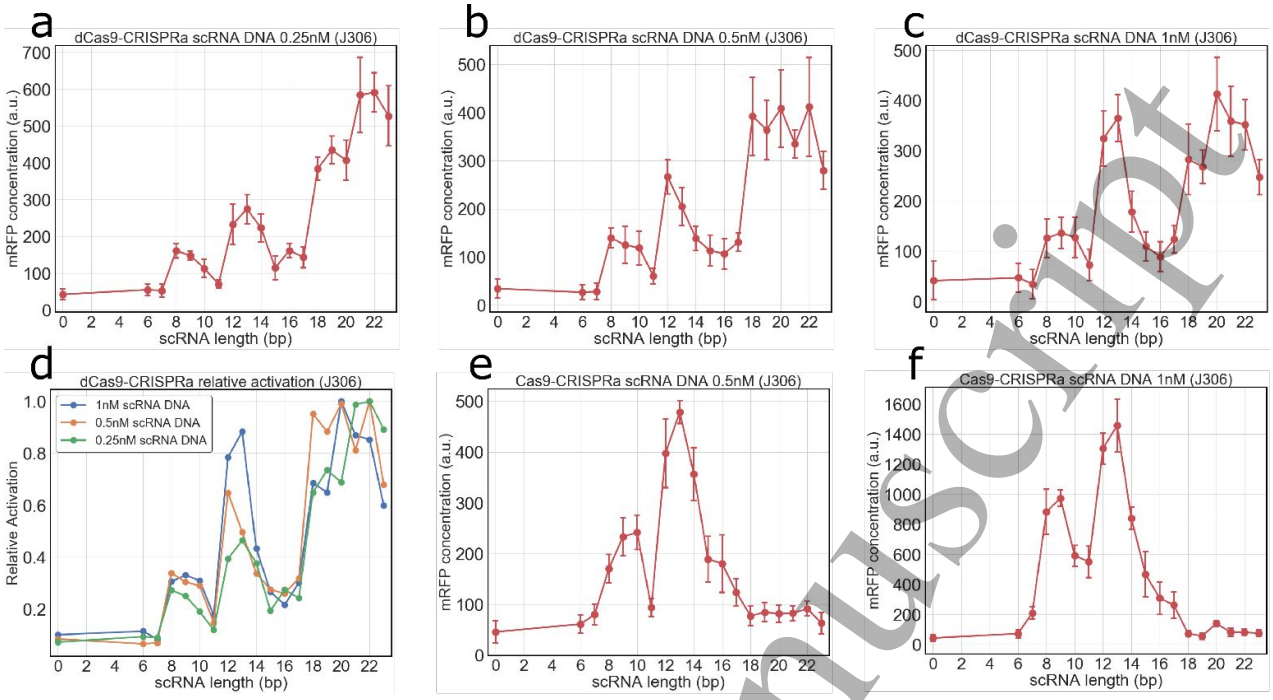


Figure 3. CRISPRa truncation experiments in TXTL. **(a), (b), and (c)** are the activation levels as a function of the scRNA spacer length (scRNA length) for 0.25 nM, 0.5 nM, and 1 nM of the scRNA DNA added to the reactions respectively. We observe that besides an optimal length of 18-20 bp, truncation lengths of 12-13 bp and 8-10 bp are also capable of activating expression. **(d)** the relative activation levels demonstrate that as we increase the concentration of the added scRNA DNA the 12-13 bp and 8-10 bp peaks also increase, indicating that with the increase of the concentration of dCas9-CRISPRa complexes we can saturate their expression to the peak values of the 18-20 bp peak. **(e) and (f)** are the activation levels in regular TXTL reaction, where instead of the dCas9 enzyme we express a Cas9 enzyme. As in **(c) and (b)** we observe the 12-13 bp and 8-10 bp peaks.

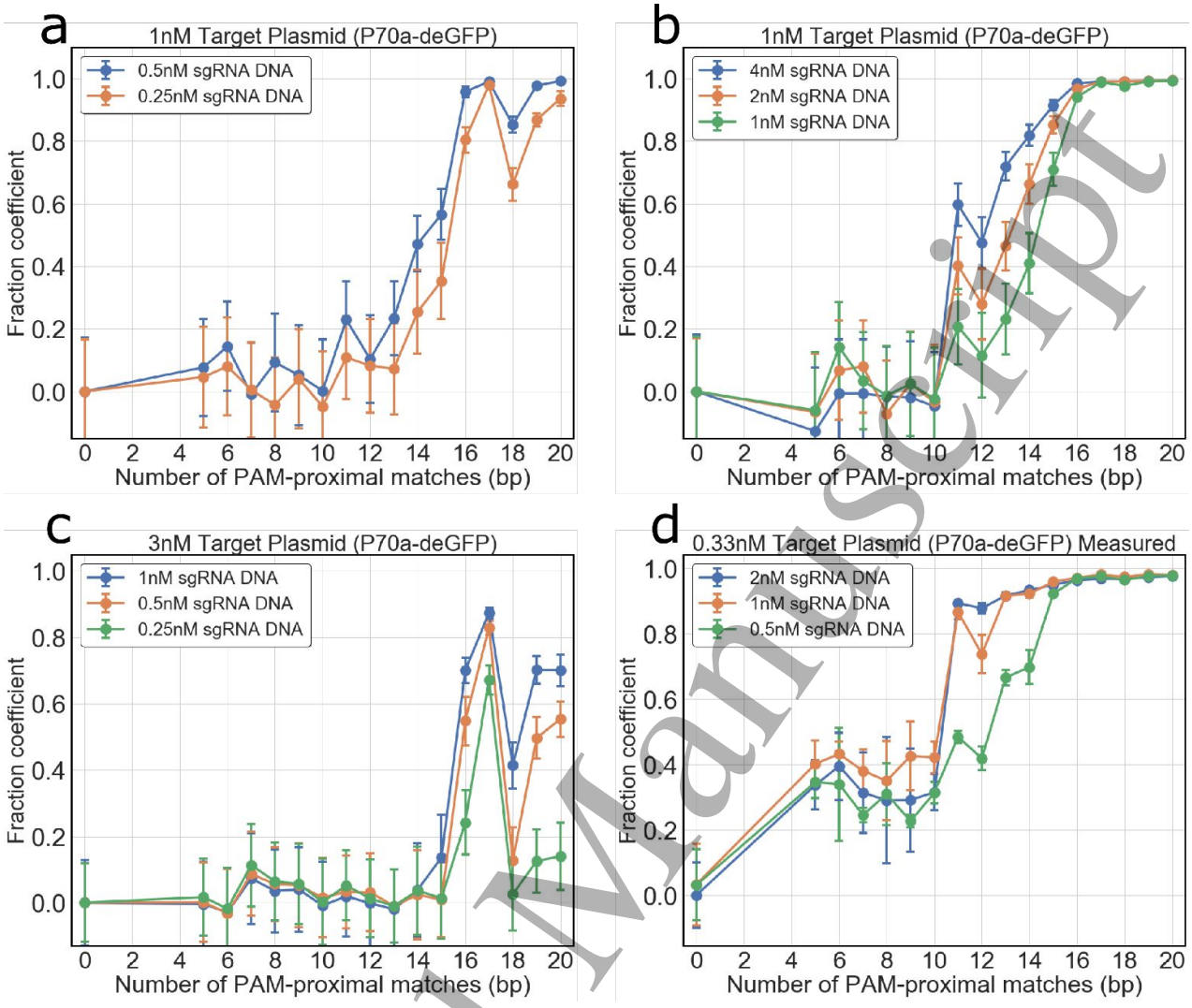


Figure 4. CRISPRi mismatch experiments in TXTL. **(a)** The concentration of the reporter plasmid P70a-*degfp* is maintained at 1 nM, while the concentration of the sgRNA DNA is varied between 0.25 nM and 0.5 nM. We observed a drop in binding when 18 nt PAM-proximal match the target. **(b)** is continuation of the **(a)**, with sgRNA DNA concentrations of 1 nM, 2 nM, and 4 nM. As we increased the concentrations the drop in the fraction coefficient at the 18 nt sgRNA is no longer visible, while the binding drop for the 12 nt PAM-proximal match sgRNA emerges. **(c)** The concentration of the target sequence plasmid is maintained at 3 nM, while the sgRNA DNA concentrations are varied between 0.25 nM, 0.5 nM, and 1 nM. As we increased the concentration of target sequence DNA from 1 nM to 3 nM, the drop at the 18 nt match becomes more prominent. **(d)** For the experiments where the concentration of the target sequence plasmid is maintained at 0.33 nM the 18 nt match drop is unnoticeable. The 12 nt match drop can be controlled by varying the concentration of the sgRNA expressing DNA added to the reaction.

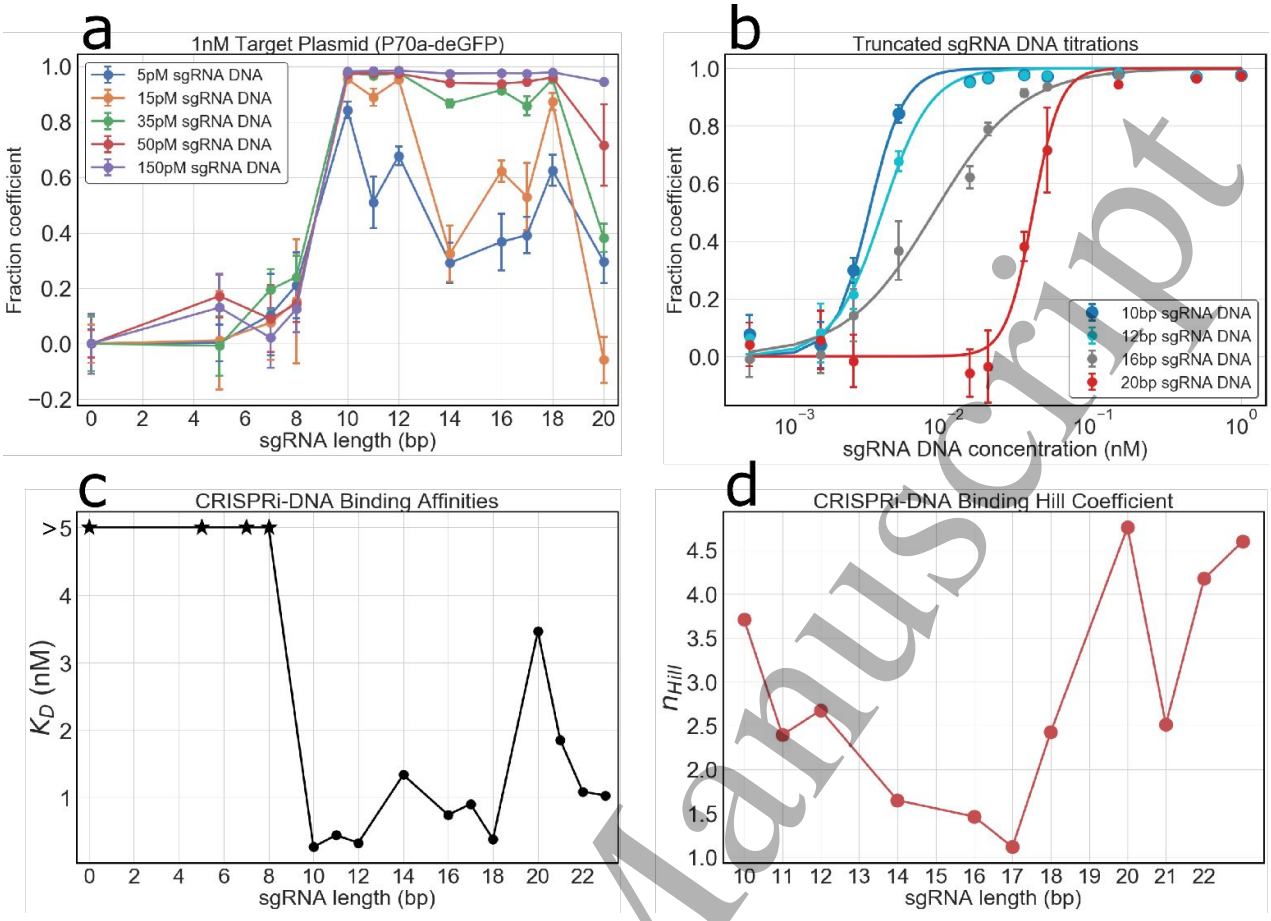


Figure 5. CRISPRi truncation experiments in TXTL. **(a)** This plot demonstrates the fraction coefficient as a function of the sgRNA spacer length (sgRNA length). The zero in the x-axis is the fraction coefficient of the off-target sgRNA, which is always f = 0 from equation [1]. **(b)** a selection of the binding curves of some sgRNA lengths. All of the curves were obtained by fitting equations [2] and [3] to the data. **(c) and (d)** respectively show the binding coefficients, K_d, and Hill coefficients, nHill, as functions of the sgRNA length. The stars indicate that the binding affinities of those truncation lengths are significantly larger than 5 nM.

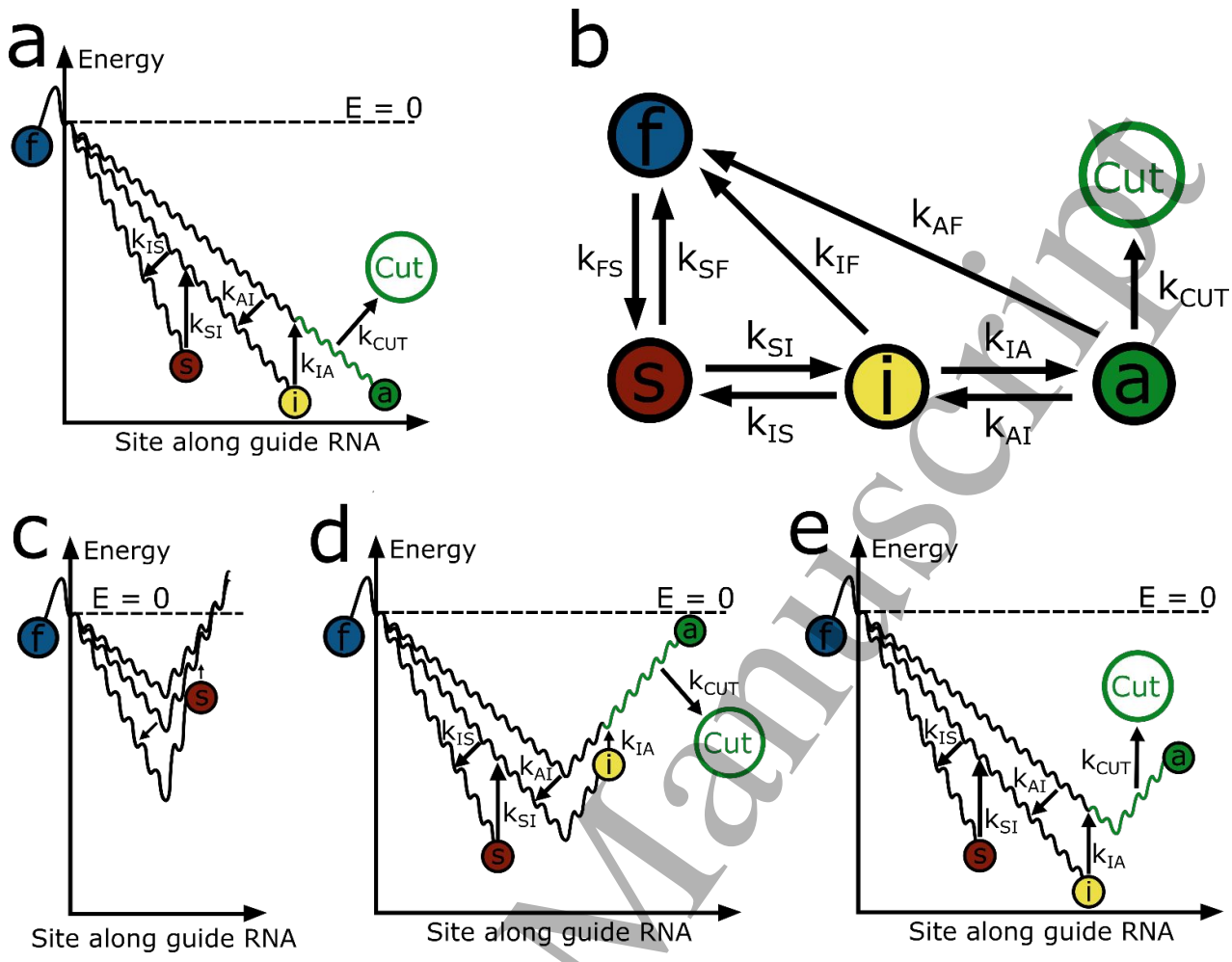


Figure 6. CRISPR-Cas9 model. **(a)** The model energy landscapes of CRISPR-Cas9 binding for a fully complentary target. The system can inhabit 3 energy landscapes of different lengths: Seed (S) - Red, Intermediate (I) - Yellow, and Active (A) - Green. The system acquires a binding energy E_C, s_1E_C, and $s_1s_2E_C$ for each formed RNA:DNA bond in landscapes S, I, and A respectively, which can be seed by the different slopes of the energy landscapes. The minimas in the landscapes correspond to sites – substates of landscapes which stand for the number of PAM-proximal RNA:DNA bonds formed. The forward transitions (k_{SI}, k_{IA}) between landscapes occur from the final site of each landscape, while the inverse transitions (k_{SI}, k_{AI}) are independent of the landscape. CRISPR-Cas9 can cut the target DNA in the active sites (painted green) of landscape A with a rate k_{CUT}. The unbinding from target DNA can occurs at a rates k_{SF}, k_{IF}, and k_{AF} for landscapes S, I, and A respectively. The unbinding rates for each landscape are caluclated by multiplying the base off-rate k_{OFF} by the probability of inhabiting site 0 (0 RNA:DNA bonds formed) in that specific landsacpe. When unbound, the CRISPR-Cas9 occupies the free state labelled as *f*. **(b)** The differential equation schematic of the CRISPR-Cas9 binding with all the rates labeled. **(c)**, **(d)**, and **(e)** are the landscapes with lengths/complementarities less than landscapes S, I, and A respectively. It can be noticed that for truncated/mismatched spacers the landscape acquires a positive absence/mismatch energy E_{MM} for each absent/mismatched RNA:DNA bond in the landscape S, and s_1E_{MM}, $s_1s_2E_{MM}$ for landscape I, A. This in turn reduced the probability of inhabiting the critical sites that control the rates of transitions and cutting, thus changing the behavior of the system.

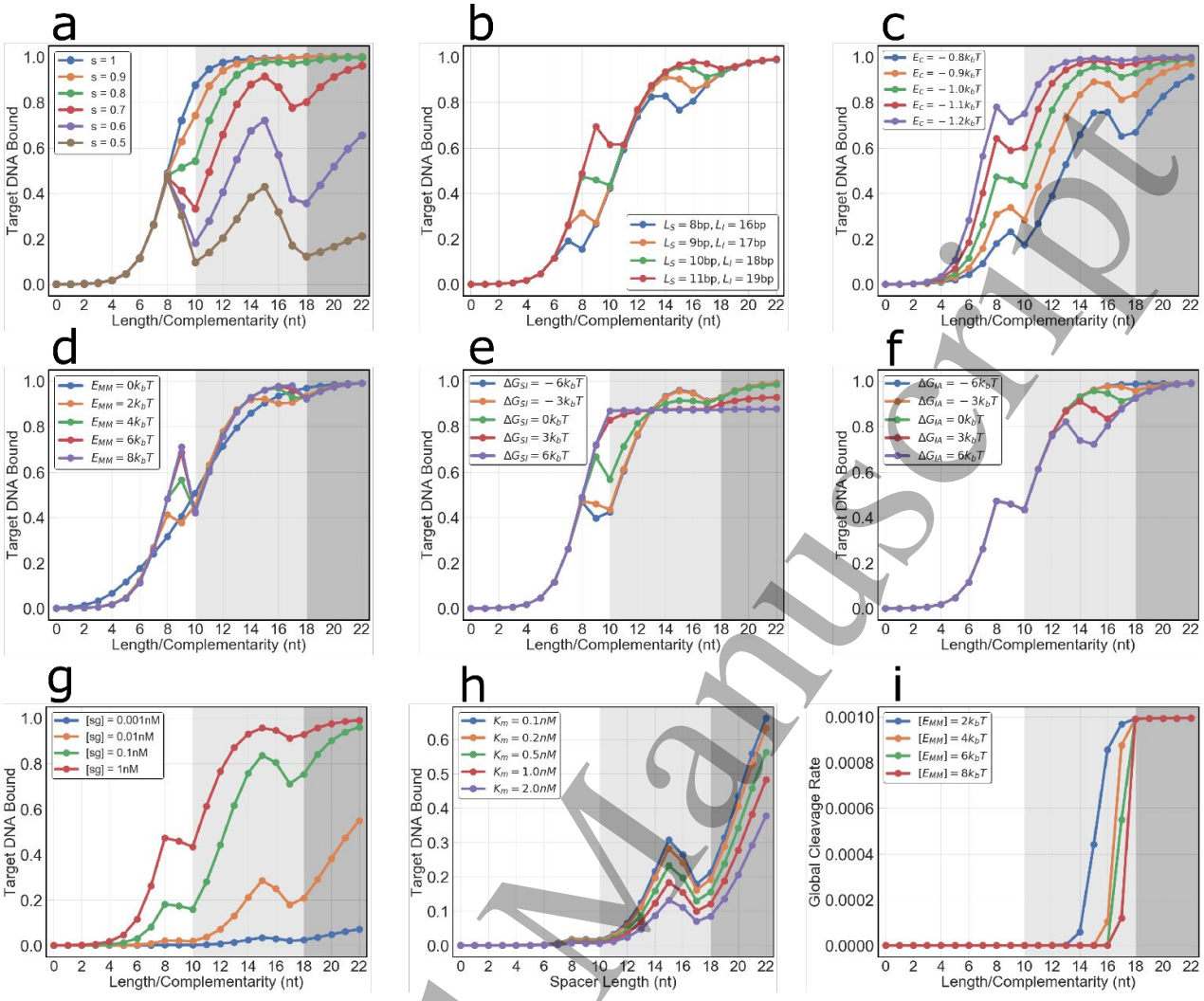


Figure 7. The emergence of multiple optimal lengths in CRISPR-Cas9 length/complementarity plots. In all plots the y-axis corresponds to the ratio of bound targets and the x-axis corresponds to the length/complementarity of the spacer sequence. The transitions to darker shaded areas keep track of the landscape lengths L_S and L_I. L_S is at the transition from white to light grey, and L_I is at the transition from light grey to dark grey. **(a)** Both energy scale factors s_1 and s_2 were set to the same value ranging from 0.5 to 1. The scale factor s = 1 means the average RNA:DNA bond E_C has the same energy in all landscapes, while s = 0.5 means an RNA:DNA bond in landscape (I) provides 0.5 E_C and an RNA:DNA bond in landscape (A) provides 0.25 E_C. It can be seen that for smaller values of s the multiple optimal lengths become more noticable. **(b)** To test how in experiment different targets could have different optimal lengths we tested the effect landscape lengths L_S and L_I have on the length/complementarity plot. The different conditions of L_S and L_I change the optimal values of and the overall profile of the length/complementarity plots. **(c)** Increasing the average RNA:DNA bond energy E_C increases the ratio of bound targets independent of the length/complemntarity. **(d)** The length/complementarity plot for multiple values of the energetic parameter E_{MM}. The values of E_{MM} were ranged between 0 k_bT to 8 k_bT. **(e)** and **(f)** are the length/complementarity plots for different values of ΔG_{SI} and ΔG_{IA}, which are the changes in the Gibbs free energy when the CRISPR-Cas9 system undergoes conformational changes during target interrogation. **(g)** The length/complementarity plot for different concentrations of the initial sgRNA/scRNA concentration added to the TXTL reaction. **(h)** The length dependence plot as under conditions when the concentration of dCas9 [dC] are thus the concentration of CRISPR-Cas9 complexes [Cr] are on the same order of as the CRISPR-Cas9 complex formation binding constant K_M. **(i)** Finally, we are interested in the global cleavage rate as a function of length/complementarity and the mismatch/absence energy E_{MM}. As expected, the target DNA is cleaved nearly immediately

upon binding if its length/complimentarity are larger or equal to L_I. The larger the E_{MM} the less likely a spacer with a length below L_I can cleave the target DNA.