Metabolic Engineering Tools for Sustainable Bioproduction in Bacterial Systems

Ryan Asher Levine Cardiff

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

Doctor of Philosophy

University of Washington

2025

Reading Committee:

James Carothers, Co-Chair
Jesse Zalatan, Co-Chair
Jorge Marchand
Lauren Rajakovich
David Shechner

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

Metabolic Engineering Tools for Sustainable Bioproduction in Bacterial Systems

Ryan Asher Levine Cardiff

Chairs of the Supervisory Committee:

James Carothers
Chemical Engineering

Jesse Zalatan
Chemistry

The chemical industry's reliance on fossil fuels drives significant carbon emissions, emphasizing the urgent need for sustainable alternatives. Microbial bioproduction presents a promising solution by enabling the synthesis of valuable chemicals from renewable feedstocks, such as CO_2 and biomass waste. However, the ability to address wide chemical markets with bioproduction is limited by insufficient tools for metabolic pathway engineering and precise gene regulation. This work describes several strategies to improve bioproduction by prototyping engineered metabolic pathways and complex gene regulatory programs. We utilize an *E. coli*-based cell-free gene expression system to engineer high-performing synthetic promoters and investigate pathways for carbon-conserving bioproduction of industrially relevant chemicals. We then discuss recent progress and opportunities at the intersection of CRISPR-based metabolic engineering and systems-level modeling approaches for improved bioproduction. Finally, we adapt the RNA-targeting CRISPR-dCas13 system as a next-generation CRISPR tool to improve metabolic regulation in bacteria.

Collectively, these works describe the development of several tools that advance metabolic engineers' ability to construct complex gene expression programs, prototype engineered carbon-conserving pathways, and precisely regulate metabolism across bacterial genomes. These tools offer new approaches to engineer microorganisms that can incorporate renewable carbon feedstocks and efficiently upcycle them into value-added chemicals, advancing the goals of sustainable bioproduction and a reduced global reliance on fossil fuels.

Table of Contents

Chapter 1: Introduction

The chemical industry relies on fossil fuels to produce nearly all manufactured goods, from medicines to materials and food products. This dependence on fossil fuels has led to considerable carbon emissions and environmental degradation in recent years [1]. In 2021, the chemicals industry was responsible for emitting approximately 3 billion metric tons of CO_2, constituting 6% of total global emissions [2,3]. This alarming contribution to greenhouse gas emissions drives the growing interest in decarbonization and the shift towards sustainable alternatives. Biology offers a solution to the chemical industry's addiction to petroleum through microbial bioproduction of chemicals used across various sectors, such as fuels, food, and materials [4–7].

Microbial bioproduction can produce valuable chemicals from renewable feedstocks through native and engineered metabolic pathways in microorganisms, offering a promising alternative to fossil fuel–dependent processes. Several products used in daily life are already produced via microbial fermentation, ranging from recombinant insulin, to ethanol, PLA plastics, and nylon [8–10]. In recent years, massive progress in the fields of metabolic engineering and synthetic biology have greatly expanded the range of accessible biochemicals that can be produced via engineered microorganisms. However, there remain several outstanding challenges for bioproduction to disrupt traditional carbon-intensive and environmentally-harmful synthesis routes.

Realizing the full potential of microbial bioproduction requires addressing several inherent challenges of rewiring cellular metabolism. Microbial bioproduction is limited by the high cost of traditional fermentation feedstocks (e.g., refined glucose) relative to the

cost of the product [11], which tends to restrict the space of compounds that can reach economic feasibility to higher-value products, such as protein biologics. To expand the potential chemical space addressable by bioproduction, strategies are urgently needed to incorporate lower cost feedstocks, such as CO_2 and other waste products, and improve bioproduction efficiency in engineered strains. However, prototyping and optimizing engineered metabolic pathways is still challenging in industrially-relevant microorganisms due to the limited tools available for metabolic engineering.

Efforts to engineer metabolic pathways for bioproduction face the universal challenge of balancing the expression of a desired pathway with the endogenous regulatory networks and metabolism of the cell [12–14]. Effective bioproduction relies on the ability to predictably up- and down- regulate endogenous and heterologous genes throughout the genome. However, regulating independent genes without perturbing larger metabolic networks often poses a significant challenge. Countless tools have been developed in bacteria to regulate arbitrary gene targets [15–17]. However, many of these tools encounter the same limitations. Repression of a gene may be deleterious to the health of the cell if it is essential for growth or contained in the same transcriptional operon as an essential gene [18]. Overexpression of a target gene may increase metabolic burden, resulting in slow growth, low production, and genetic escapees [12,13]. To make bioproduction from renewable feedstocks competitive with traditional chemical synthesis, novel methods are needed for robust control over metabolic flux. Here, we present several approaches to overcome the limitations of native metabolic networks that restrict efficient bioproduction.

First, we engineer synthetic activatable promoters with greatly improved dynamic ranges to enable the construction of complex CRISPR-based genetic circuits and regulatory networks in bacterial systems. This work represents a significant step towards building synthetic gene regulatory networks akin to those found in nature. We then investigate strategies for engineering and prototyping carbon-conserving bioproduction pathways in an *E. coli*-based cell-free gene expression system. We show that we are able to minimize competition from the native metabolism to reduce carbon loss and successfully demonstrate bioproduction of malate from CO_2 and glycine. These efforts demonstrate the utility of cell-free systems as a platform for both pathway design and carbon-efficient bioproduction. We then discuss recent progress and opportunities at the intersection of CRISPR-based metabolic engineering in microorganisms and systems-level modeling approaches for both guide RNA design and metabolism. From this review, we find that further work to develop improved models and incorporate approaches from machine learning will greatly expand the capabilities of CRISPR-based metabolic engineering. We then adapt the RNA-targeting CRISPR-dCas13 system for precise translational regulation in bacteria. First characterizing the system in *E. coli*, we find that dCas13 exhibits reduced polar effects in multi-gene operons compared to dCas9. When coupling dCas13 translational CRISPRi with dCas9 transcriptional CRISPRa, we demonstrate both a novel approach of activating single genes in multi-gene operons and improved bioproduction through multiplexed gene activation and repression. We show that tlCRISPRi is able to achieve functional effects inaccessible using dCas9, leading to improved bioproduction of a medically-relevant human milk oligosaccharide.

Chapter 2 discusses our work to develop high-performing synthetic promoters to enable the construction of complex gene regulatory networks and input-responsive genetic circuits, published in *Proceedings of the National Academy of Sciences* in 2023. In this work, we used rational promoter engineering to develop promoters with high dynamic ranges of gene expression, which allowed us to access new circuit topologies and input-responsive functions not previously demonstrated in bacterial systems. The ability to build synthetic systems with predictable functions is a crucial step in understanding the complexity of natural systems. This work represents a significant advancement in our ability to build complex gene regulation programs in bacteria.

In **Chapter 3,** we utilize cell-free systems as a platform for metabolic engineering and for prototyping of carbon-conserving pathways. This work was published in *Metabolic Engineering* in 2025. The development of novel engineered pathways in industrially relevant microbes is limited by slow engineering cycles and difficulty in independently screening and debottlenecking individual pathway steps. Additionally, cell-free systems are advantageous for metabolic engineering as they are not limited by cellular burden or regulation [19] and can utilize diverse and low-cost carbon feedstocks, such as CO_2 [20–22]. In this work, we overcome challenges associated with the native metabolism present in lysate-based cell-free systems to produce the industrial di-acid malate using glycine and 2 CO_2 equivalents through an engineered carbon-conserving pathway. These accomplishments demonstrate the capabilities of cell-free expression systems for both the prototyping of carbon-conserving pathways and the sustainable bioproduction of platform chemicals.

Chapter 4 reviews progress and prospects of systems-level modeling approaches for CRISPR-based metabolic engineering in microorganisms, published in *ACS Synthetic Biology* in 2024. CRISPR-Cas systems have enabled the development of complex metabolic engineering programs through guide RNA-directed regulation of target genes. To optimize biosynthetic pathways in microbial systems, we need improved models to inform the design and implementation of engineered metabolic pathways. Recent progress has resulted in new modeling approaches for identifying gene targets and predicting the efficacy of guide RNA targeting. This review discusses opportunities for developing improved models and incorporating approaches from machine learning to overcome current limitations and greatly expand the capabilities of CRISPR tools for metabolic engineering.

In **Chapter 5,** we demonstrate the development of the RNA-targeting CRISPR-dCas13 system as a tool for regulating metabolic pathways at the translational level, published in *Nucleic Acids Research* in 2024. The use of dCas13 as a tool for translational regulation in prokaryotes is a relatively recent development [23], however its potential is largely still unexplored. We show that the use of dCas13 allows for improved regulation of gene expression in a multi-gene operon, which represents a large portion of the genome across bacterial species. Building upon this result, we find that dCas13 and dCas9 can be used in combination for selective activation of individual genes in synthetic multi-gene operons. Finally, we apply simultaneous gene activation using dCas9 and repression with dCas13 to improve biosynthesis of a medically-relevant human milk oligosaccharide by downregulating genes that were previously inaccessible using dCas9 alone. These results suggest translational gene regulation using dCas13

may overcome transcriptional regulation effects, thereby improving our ability to precisely engineer bacteria for biotechnology and conduct systematic genetic screens.

References

(1) Ahmed, F.; Ali, I.; Kousar, S.; Ahmed, S. The Environmental Impact of Industrialization and Foreign Direct Investment: Empirical Evidence from Asia-Pacific Region. *Environ. Sci. Pollut. Res. Int.* 2022, *29* (20), 29778–29792. https://doi.org/10.1007/s11356-021-17560-w.Abstract

(2) S&P Global, *Decarbonizing Chemicals Part One: Sectorwide Challenges Will Intensify Beyond 2030.* https://www.spglobal.com/esg/insights/featured/special-editorial/decarbonizing-chemicals-part-one-sectorwide-challenges-will-intensify-beyond-2030 (accessed 2025-03-24).

(3) *Total greenhouse gas emissions in the chemical industry (Indicator)*. https://www.eea.europa.eu/en/european-zero-pollution-dashboards/indicators/total-greenhouse-gas-emissions-in-the-chemical-industry (accessed 2025-01-15).

(4) Blake, W. J.; Swartz, J. R. Cell-Free System for Converting Methane into Fuel and Chemical Compounds. US20170159058A1, June 8, 2017. https://patents.google.com/patent/US20170159058A1/en (accessed 2024-01-23).

(5) Averesch, N.; Vince, V.; Kracke, F.; Ziesack, M.; Nangle, S.; Waymouth, R.; Criddle, C. *Towards High-Performance Polyesters from Carbon Dioxide: Novel Polyhydroxyarylates from Engineered Cupriavidus Necator*; preprint; In Review, 2023. https://doi.org/10.21203/rs.3.rs-2719603/v1.

(6) Rorrer, N. A.; Nicholson, S.; Carpenter, A.; Biddy, M. J.; Grundl, N. J.; Beckham, G. T. Combining Reclaimed PET with Bio-Based Monomers Enables Plastics Upcycling. *Joule* 2019, *3* (4), 1006–1027. https://doi.org/10.1016/j.joule.2019.01.018.

(7) Grubbe, W. S.; Rasor, B. J.; Krüger, A.; Jewett, M. C.; Karim, A. S. Cell-Free Styrene Biosynthesis at High Titers. *Metab. Eng.* 2020, *61*, 89–95. https://doi.org/10.1016/j.ymben.2020.05.009.

(8) Quianzon, C. C.; Cheikh, I. History of Insulin. *J. Community Hosp. Intern. Med. Perspect.* 2012, *2* (2), 10.3402/jchimp.v2i2.18701. https://doi.org/10.3402/jchimp.v2i2.18701.

(9) Swetha, T. A.; Ananthi, V.; Bora, A.; Sengottuvelan, N.; Ponnuchamy, K.; Muthusamy, G.; Arun, A. A Review on Biodegradable Polylactic Acid (PLA) Production from Fermentative Food Waste - Its Applications and Degradation. *Int. J. Biol. Macromol.* 2023, *234*, 123703. https://doi.org/10.1016/j.ijbiomac.2023.123703.

(10) Lee, S. Y.; Kim, H. U.; Chae, T. U.; Cho, J. S.; Kim, J. W.; Shin, J. H.; Kim, D. I.; Ko, Y.-S.; Jang, W. D.; Jang, Y.-S. A Comprehensive Metabolic Map for Production of Bio-Based Chemicals. *Nat. Catal.* 2019, *2* (1), 18–33. https://doi.org/10.1038/s41929-018-0212-4.

(11) Scown, C. D.; Baral, N. R.; Yang, M.; Vora, N.; Huntington, T. Technoeconomic Analysis for Biofuels and Bioproducts. *Curr. Opin. Biotechnol.* 2021, *67*, 58–64. https://doi.org/10.1016/j.copbio.2021.01.002.

(12) Hartline, C. J.; Schmitz, A. C.; Han, Y.; Zhang, F. Dynamic Control in Metabolic Engineering: Theories, Tools, and Applications. *Metab. Eng.* 2021, *63*, 126–140. https://doi.org/10.1016/j.ymben.2020.08.015.

(13) Weiße, A. Y.; Oyarzún, D. A.; Danos, V.; Swain, P. S. Mechanistic Links between

Cellular Trade-Offs, Gene Expression, and Growth. *Proc. Natl. Acad. Sci.* 2015, *112* (9). https://doi.org/10.1073/pnas.1416533112.
(14) Gupta, A.; Reizman, I. M. B.; Reisch, C. R.; Prather, K. L. J. Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit. *Nat. Biotechnol.* 2017, *35* (3), 273–279. https://doi.org/10.1038/nbt.3796.
(15) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 2013, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.
(16) Magistro, G.; Magistro, C.; Stief, C. G.; Schubert, S. A Simple and Highly Efficient Method for Gene Silencing in Escherichia Coli. *J. Microbiol. Methods* 2018, *154*, 25–32. https://doi.org/10.1016/j.mimet.2018.10.003.
(17) Santos-Moreno, J.; Schaerli, Y. CRISPR-Based Gene Expression Control for Synthetic Gene Circuits. *Biochem. Soc. Trans.* 2020, *48* (5), 1979–1993. https://doi.org/10.1042/BST20200020.
(18) Cui, L.; Vigouroux, A.; Rousset, F.; Varet, H.; Khanna, V.; Bikard, D. A CRISPRi Screen in E. Coli Reveals Sequence-Specific Toxicity of dCas9. *Nat. Commun.* 2018, *9* (1), 1912. https://doi.org/10.1038/s41467-018-04209-5.
(19) Dudley, Q. M.; Karim, A. S.; Jewett, M. C. Cell-Free Metabolic Engineering: Biomanufacturing beyond the Cell. *Biotechnol. J.* 2015, *10* (1), 69–82. https://doi.org/10.1002/biot.201400330.
(20) Chowdhury, S.; Westenberg, R.; Wennerholm, K.; Cardiff, R. A. L.; Beliaev, A. S.; Noireaux, V.; Carothers, J. M.; Peralta-Yahya, P. Carbon Negative Synthesis of Amino Acids Using a Cell-Free-Based Biocatalyst. *ACS Synth. Biol.* 2024. https://doi.org/10.1021/acssynbio.4c00359.
(21) Cai, T.; Sun, H.; Qiao, J.; Zhu, L.; Zhang, F.; Zhang, J.; Tang, Z.; Wei, X.; Yang, J.; Yuan, Q.; Wang, W.; Yang, X.; Chu, H.; Wang, Q.; You, C.; Ma, H.; Sun, Y.; Li, Y.; Li, C.; Jiang, H.; Wang, Q.; Ma, Y. Cell-Free Chemoenzymatic Starch Synthesis from Carbon Dioxide. *Science* 2021, *373* (6562), 1523–1527. https://doi.org/10.1126/science.abh4049.
(22) Schwander, T.; Schada von Borzyskowski, L.; Burgener, S.; Cortina, N. S.; Erb, T. J. A Synthetic Pathway for the Fixation of Carbon Dioxide in Vitro. *Science* 2016, *354* (6314), 900–904. https://doi.org/10.1126/science.aah5237.
(23) Otoupal, P. B.; Cress, B. F.; Doudna, J. A.; Schoeniger, J. S. CRISPR-RNAa: Targeted Activation of Translation Using dCas13 Fusions to Translation Initiation Factors. *Nucleic Acids Res.* 2022, gkac680. https://doi.org/10.1093/nar/gkac680.

Chapter 2: Engineering activatable promoters for scalable and multi-input CRISPRa/i circuits

Ryan Cardiff [+,1], Diego Alba Burbano[+,1,2], Benjamin I. Tickman[1,2], Cholpisit Kiattisewee[1], Cassandra Maranas[1], Jesse G. Zalatan*[1,3], James M. Carothers*[,1,2]

1: Molecular Engineering & Sciences Institute and Center for Synthetic Biology
University of Washington

2: Department of Chemical Engineering
University of Washington

3: Department of Chemistry
University of Washington

+: These authors contributed equally

Published as a research article in Proceedings of the National Academy of Sciences (PNAS) on July 18th, 2023. DOI: 10.1073/pnas.2220358120

Abstract

Dynamic, multi-input gene regulatory networks are ubiquitous in nature. Multi-layer CRISPR-based genetic circuits hold great promise for building gene regulatory networks akin to those found in naturally-occurring biological systems. We develop an approach for creating high-performing activatable promoters that can be assembled into deep, wide, and multi-input CRISPR-activation and -interference (CRISPRa/i) gene regulatory networks. By integrating sequence-based design and *in-vivo* high-throughput screening, we design activatable promoters that achieve up to 1000-fold dynamic range in an *E. coli*-based cell-free system. These new components enable CRISPRa gene regulatory networks that are six layers deep and four branches wide. We show the generalizability of the promoter engineering workflow by improving the dynamic range of the light-dependent EL222 optogenetic system from 6-fold to 34-fold. Additionally, high dynamic range promoters enable conditional CRISPRa systems mediated by small molecules and protein-protein interactions. We apply these tools to build input-responsive CRISPRa/i gene regulatory networks, including feedback loops, logic gates, multi-layer cascades, and dynamic pulse modulators. Our work provides a generalizable approach for the design of high dynamic range activatable promoters and enables new classes of gene regulatory functions in cell-free systems.

Introduction

Natural biological systems employ complex gene regulatory networks (GRNs) to sense diverse environmental cues and respond to them through the coordinated expression of multiple genes (Ma et al., 2004; Cai et al., 2008; Benzinger et al., 2022). Cell-free systems (CFS) have emerged as an attractive chassis for building synthetic biological systems as they allow for rapid prototyping of genetic parts and circuits (Jung et al., 2020; Karim et al., 2020; Lehr et al., 2019; Moore et al., 2018; Swank & Maerkl, 2021). To build increasingly complex cell-free systems that can sense and respond to diverse inputs, new approaches for increasing the capabilities of synthetic GRNs are needed (Bassalo et al., 2016; Brophy & Voigt, 2014; Swank et al., 2019; Xia et al., 2019). The ability to construct complex GRNs will enable investigating the limits of biological circuit design (Alon, 2007; Brophy & Voigt, 2014), building multiplexed molecular biosensors (Jung et al., 2020b, 2022; Nguyen et al., 2021), deploying on-demand bioproduction platforms (Pardee et al., 2016; Stark et al., 2021), and the bottom-up construction of synthetic cells (Adamala et al., 2017; Aufinger et al., 2022; Garamella et al., 2019; Karzbrun et al., 2014).

CRISPR-Cas transcriptional regulation has proven a promising framework for building sophisticated genetic circuits across a variety of biological systems (Kim et al., 2019; Landberg et al., 2020; Reis et al., 2019; Santos-Moreno & Schaerli, 2020; Westbrook et al., 2019). Transcriptional units containing target sequences for CRISPR activation (CRISPRa) and/or CRISPR interference (CRISPRi), termed CRISPRa/i nodes, can be assembled into circuits with network topologies specified by a set of guide RNAs (gRNAs). Experimental and theoretical analysis indicates that the

CRISPRa/i system is well suited to design deep and wide control circuits, defined as circuits containing two or more internal nodes connected in series or parallel through orthogonal gRNAs (Clamons & Murray, 2019; Nielsen & Keasling, 2016; Tickman et al., 2021). Large CRISPRi GRNs with up to 7 sgRNAs have been constructed in yeast by implementing low leak promoters and high dynamic range repressors (Gander et al., 2017). In *E. coli*-based CFS, CRISPRi-based genetic control is well established (Marshall & Noireaux, 2019; Westbrook et al., 2019), and CRISPRa has recently been incorporated (Tickman et al., 2021), greatly expanding the circuit design space. Integration of CRISPRa and CRISPRi has enabled two-layer activation and activation-repression cascades, as well as two orthogonal I1-FFLs in the same CFS. However, CRISPRa circuitry is limited by a lack of promoter-gRNA pairs that can be interconnected with minimal signal degradation (Tickman et al., 2021). Hence, a generalizable approach for engineering activatable promoters with low basal and high activated expression levels would significantly improve CRISPRa and enable the assembly of complex, input-responsive synthetic GRNs.

Promoter engineering efforts have traditionally focused on designing constitutive and inducible promoters with predictable expression characteristics (Alper et al., 2005; Kelly et al., 2009). Tuning the strength of constitutive promoters involves designing promoter sequences that maintain either weak or strong RNA polymerase (RNAP) recruitment to the promoter (Brewster et al., 2012; Fleur et al., 2021). Inducible promoters contain recognition sites for transcriptional activators or repressors that modulate transcriptional levels upon binding (D. J. Lee et al., 2012; T. C. Yu et al., 2021). Efforts to develop inducible transcription systems with high dynamic ranges have

relied on engineering de-repression based systems such as the Lac or Tet promoters (Cress et al., 2016; Groseclose et al., 2020; Liu et al., 2019; Roney et al., 2016; T. C. Yu et al., 2021), largely due to the difficulty of rationally designing activatable promoters (Fontana et al., 2020; Chen et al., 2018). For effective activation, RNAP recruitment to the promoter should be weak in the absence of an activator, however transcription initiation should be strong upon activator-mediated RNAP recruitment (Fontana, Dong, et al., 2020; D. J. Lee et al., 2012). Hence, higher dynamic ranges with CRISPRa and other transcriptional activation systems can be achieved through promoter sequence engineering to tune RNAP interactions.

We develop an approach integrating sequence-based design and *in-vivo* high-throughput screening to generate an expandable set of high-performing promoters that exhibit both low basal and high activated expression levels. Through a sequential selection approach, we design activatable promoters with up to 1000-fold dynamic range, constituting a 33-fold improvement from previous synthetic promoters when ported to CFS (Fontana, Dong, et al., 2020; Tickman et al., 2021). These promoters enable network topologies not previously accessible, including a six-layer deep cascade and a four-branch parallel circuit. Additionally, by engineering activatable promoters, we successfully incorporate different inputs into CRISPRa/i circuits through the blue light-responsive EL222 transcriptional activator as well as protein-protein interaction (PPI)-dependent CRISPRa assembly, including SYNZIP and the small-molecule responsive abscisic acid (ABA) and gibberellic acid (GA) systems. Through the engineering of high-performing CRISPRa/i nodes as well as input signal processing modules, we show that CRISPRa/i circuits built with these components can be tuned to

achieve a broad range of functions, including deep multi-layer and wide multi-branch activation cascades, positive feedback loops, AND logic gates, and dynamic two-input pulse modulators. Overall, this work describes a new workflow for engineering activatable promoters and provides a toolbox of versatile components with immediate utility for implementing CRISPRa/i circuits. Together, these developments dramatically expand the ability to assemble large, multi-input GRNs in CFS.

Results

We first sought to characterize the impact of RNAP recruitment on both basal and activated expression levels of synthetic activatable promoters. RNAP recruitment is dependent on the affinity of the RNAP sigma subunit (σ) for the −10 and −35 hexamers of the minimal promoter, as well as the affinity of the RNAP α-CTD subunits for the UP-element sequence upstream of the minimal promoter (Ross et al., 1998; Shultzaberger et al., 2007) (Figure 1A). We systematically designed libraries of these discrete promoter regions and screened them in *E. coli*. The libraries were co-transformed with an aTc-inducible CRISPRa plasmid to enable parallel screening of basal and activated expression levels (Figure 1B, Methods 2.2) (Tickman et al., 2021). In the absence of aTc, RNAP recruitment is determined by the promoter basal strength. Upon aTc induction, the MCP-SoxS activator is expressed and localized to the promoter via CRISPRa complexes with scaffold RNAs (scRNAs) containing an MS2 hairpin. MCP-SoxS then recruits RNAP to the promoter through α-CTD interactions, activating transcription (Dong et al., 2018) (Figure 1A). This approach allows us to characterize the impact of individual promoter regions on basal and activated expression simultaneously, and combine variants with low basal and high activated expression to construct high-performing activatable promoters.

Functional interrogation of promoter regions with CRISPRa

Impact of Minimal Promoter Region on Activability

Previous work has demonstrated the importance of the minimal promoter region in determining promoter basal and activated expression levels (Fontana, Dong, et al.,

2020; Hook-Barnard & Hinton, 2007; D. J. Lee et al., 2012; Paget & Helmann, 2003). We designed two minimal promoter libraries mutagenizing the -10/-35 hexamers and the intervening sequence of the previously identified best-performing minimal promoter (BBa_J23117) within the J3 synthetic CRISPRa promoter (Figure 1C, Methods 3.1) (Fontana, Dong, et al., 2020). These libraries were co-transformed with the aTc-inducible CRISPRa plasmid expressing the J306 scRNA that targets the cognate J3 promoter. Both libraries yielded high promoter diversity with basal and activated expression levels ranging from that of a no-reporter control to a strong constitutive promoter (BBa_J23119) (Figure 1D). The set of minimal promoter variants that maintain both low basal and high activated expression levels can be conceptualized as a Pareto optimal front. In multi-objective optimization, a Pareto front defines the best-performing solutions for which no further improvements in either objective can be achieved without compromising the other (Censor, 1977). Three variants from this Pareto front exhibited up to 32% lower basal and 4% higher activated expression levels compared to the original BBa_J23117, indicating the original minimal promoter was not a part of the Pareto front. This finding suggests that promoter mutagenesis can yield improved activatable promoters beyond previous screening methods based on promoter basal strengths alone (Fontana, Dong, et al., 2020). By mutagenizing the minimal promoter of CRISPRa promoters, we generated sequences with >100-fold dynamic range in CRISPRa-mediated gene expression.

Impact of UP-Element Region on Activatibility

RNAP promoter recognition is enhanced by the AT-rich UP-elements upstream of the minimal promoter, which anchor the α-subunits of RNAP (Estrem et al., 1998; D. J.

Lee et al., 2012; Ross et al., 1998). For effective CRISPRa, RNAP should only be recruited to the promoter in the presence of an on-target scRNA. Hence, for transcriptional activation with SoxS, improvements in dynamic range could be achieved by minimizing RNAP-UP-element interactions and lowering basal expression levels. We designed five UP-element libraries mutagenizing the AT-rich *E. coli* consensus sequence with increasing GC-content (Figure 2A, Methods 3.2). As expected, the consensus UP-element and the AT-rich library had the highest basal expression levels (Figure 2B, left). On average, these libraries showed only 3-fold activation, as compared to 37- to 44-fold activation for the more GC-rich libraries (Figure S1). Qualitatively, we observed a monotonic decrease in basal expression levels and no impact on maximum activated levels with increased GC-content (Figure S1). We identified the optimal variants from each library and found a shift in the Pareto front towards lower basal and higher activated expression levels with increasing UP-element GC-content (Figure 2B, right). Specifically, the median basal and median activated expression levels of the GC-rich optimal variants were 59.7-fold lower and 1.7-fold higher than that of the AT-rich optimal variants. The original J3 synthetic promoter sat in between fronts consisting of high and low GC-content variants. By mutagenizing the UP-element, we generated promoter variants with >350-fold dynamic range in CRISPRa-mediated gene expression.

Impact of the scRNA target site region on activatability

Transcriptional activators bind upstream of the minimal promoter region to recruit RNAP to the transcription start site (TSS) (D. J. Lee et al., 2012). For CRISPRa, the optimal scRNA target site location for SoxS-mediated activation is -81 bp upstream of

the TSS (Fontana, Dong, et al., 2020). Changing the scRNA target sequence enables rapid generation of orthogonal CRISPRa promoters (Fontana, Sparkman-Yager, et al., 2020; Lian et al., 2017). Due to the proximity to the UP-element region, we reasoned that the sequence composition of the scRNA target site may have an impact on basal expression levels. We designed three scRNA target site libraries composed of varying GC-content (Methods 3.3) and measured the basal expression of each library. We found that GC-rich libraries had 4.3-fold lower median basal expression compared to AT-rich libraries (Figure 2C). Additionally, the spread of the basal expression decreased monotonically with increasing GC-content of the scRNA target site sequence (Figure 2C). Together, these results indicate GC-rich scRNA target site sequences lead to low basal expression CRISPRa promoters. To validate the CRISPRa activity at these low basal expression scRNA target sites, we then selected 10 GC-rich variants and constructed the corresponding scRNAs. All variants produced a higher fold-activation than the original J306 scRNA (Figure S2), with 3.5-fold average increase in fold-activation.

Combining promoter regions to engineer high-performing CRISPRa promoters

Engineering activatable promoters by combining optimized promoter regions

We proceeded to test if the highest performing variants from the UP-element and minimal promoter screens could be combined to yield activatable promoters with improved performance. We selected three high-performing variants from both the UP-element and minimal promoter screens, as well as the starting J3 UP-element and BBa_J23117 minimal promoter, and constructed a combinatorial set of 16 promoters.

Surprisingly, we found that promoter regions that gave the largest improvements in the original context did not necessarily give the largest improvements when tested in different contexts (Figure S3). For instance, an UP-element that gave 400-fold activation when tested with the BBa_J23117 minimal promoter only gave 270-fold activation when tested with a different high-performing minimal promoter, largely due to a decrease in activated expression (Figure S3, right). These results highlight that promoter region context effects play an important role in the design of high-performing activatable promoters.

Engineering activatable promoters through sequential promoter region screening

We tested whether promoters with improved basal and activated expression levels could be achieved by selecting minimal promoters in the context of high-performing UP-elements. We first screened the UP-element region as these libraries had a larger impact than the minimal promoter libraries on the location of the Pareto front (Figures 1D, 2B). We mutagenized the UP-element of a promoter containing the minimal promoter BBa_J23117 and a high GC-rich scRNA target. We selected three UP-element variants from the Pareto front, which had on average 90% lower basal and 12% higher activated expression than the J3 promoter (Figure 2D). We then screened minimal promoter libraries in the context of these three selected UP-element variants. We again selected three new promoter variants from the Pareto front, which had on average 83% lower basal and 56% higher activated expression than the J3 promoter (Figure 2D). Notably, with this sequential screening approach we were able to shift the Pareto front towards lower basal and higher activated expression in both the UP-element and minimal promoter screens.

Engineering deep and wide circuits with high-performing CRISPRa promoters

CRISPRa promoters expressing scRNAs for CRISPRa or sgRNAs for CRISPRi, termed CRISPRa/i nodes, can be assembled into multi-layer circuits with network topologies specified by the guide RNAs. Proper circuit function requires level-matching the input/output dynamic ranges between sequential CRISPRa/i nodes to minimize signal degradation (Gander et al., 2017; Nielsen & Keasling, 2016; Tickman et al., 2021). Promoters with lower basal and higher activated expression levels should span a higher fraction of the input dynamic range of downstream nodes, resulting in less signal degradation (Tickman et al., 2021). Therefore, we sought to evaluate how dynamic range improvements achieved at the promoter level translate into fold-activation and signal propagation improvements at the circuit level and enable construction of increasingly deep and wide CRISPRa circuits. We implement these circuits in CFS due to the ease of rapidly prototyping genetic parts and circuits (Lehr et al., 2019; Tickman et al., 2021). Given the strong correspondence between component function in bacteria and CFS (Kelwick et al., 2018; Tickman et al., 2021; Vögeli et al., 2022), we reasoned that the high-performing CRISPRa promoters designed in *E. coli* could be used to build larger and more complex GRNs in CFS.

Engineering functional CRISPRa/i nodes

We first characterized the three promoter variants selected above (Figure 2D, HP1-3) in CFS and observed up to ~1000-fold dynamic range (Figure S4). We then generated a set of orthogonal CRISPRa/i nodes to be assembled into multi-layer circuits following previously-described methods (Tickman et al., 2021). We combined the highest dynamic range activatable promoter (HP3, Table S1) with previously-screened

scRNA target sites to generate orthogonal CRISPRa/i nodes (Figure 3A). We characterized the dose-response curves for each orthogonal scRNA-promoter pair and found these new nodes gave an average activation of 890-fold (Figures 3B, S5; Table S1).

Deep CRISPRa Circuits

We investigated whether deep multi-layer cascades could be implemented using the improved CRISPRa/i nodes. We first built a two-layer CRISPRa cascade by tuning the expression levels of the input and internal CRISPRa/i nodes and achieved up to 127-fold activation (Figures S6, S7). Next, we assembled four-layer activation cascades. To compare circuit performance and dynamics in response to a scRNA input, we measured RFP expression and time to maximum expression rate (t_{max}) (Figure 3C, insert). If the input signal propagates faster than the leak from the rest of the nodes, CRISPRa-dependent expression through the network accelerates, reducing t_{max}. Therefore, a larger t_{max} between the +/- input conditions (taken as Δt_{max}) corresponds to an improved circuit function. For a cascade with equal node concentration at each layer, we found that the circuit was not input responsive (Figure 3C, middle). We then tuned node concentrations by either decreasing or increasing the concentration of each subsequent node as depth increased (denoted "*D*" and "*I*", respectively) (Figure 3C, left). Assemblies *D* and *I* both had their t_{max} significantly accelerated compared to the no input conditions (Δt_{max} of 85 min and 165 min, respectively) (Figure 3C, right). Assemblies *D* and *I* also had the lowest and highest leak and absolute expression levels (Figure 3C, middle), indicating that timing and expression level of multi-layer, input-responsive circuits can be controlled through node concentration tuning.

We then used individual scRNA-promoter dose-response curves to inform circuit assembly. This strategy achieved 3-fold higher activated expression than assembly *D* and 4.5-fold lower basal expression than assembly *I*, and a Δt_{max} of 105 min (Figure 3C). We tuned the concentrations of the first and third internal nodes of the four-layer cascade for a 8-fold improvement in fold-activation (Figure 3D). Qualitatively, higher concentrations of the third node resulted in higher activated states, while lower concentrations of the first node minimized basal expression (Figure S8). We then changed the high-performing promoter of the second internal node for a worse performing promoter with higher basal expression. We observed no difference in expression level or t_{max} with or without input scRNA, indicating the circuit was no longer input-responsive. This result underscores the importance of high-performing promoters for building deep transcriptional circuits.

We investigated how input signals are propagated through increasingly deep circuits (Figure 3E, left). We define signal propagation as the ratio of the fold-activation between the cascade output and input layer. Signal propagation was sustained above 80% until the 4th layer was added, after which it decreased rapidly (Figure 3E, right). Nevertheless, we observed measurable output differences in circuits of up to six layers. We quantify the dynamics of signal propagation in terms of signal delay, or the difference in time to reach the maximum fold-activation of the cascade between the cascade output and input layers. The two-layer cascade gave no significant difference in signal delay compared to a single-layer CRISPRa reaction. This may suggest there is a slow step in output production, such as fluorophore maturation, that masks the effect of the second layer. Beyond two layers, the signal delay showed a consistent increase

of ~50 min with subsequent additions of the third, fourth, and sixth layers (Figure 3E, right), suggesting that the response characteristics are maintained at each layer in deep circuits.

Wide CRISPRa Circuits

To identify conditions under which multiple orthogonal nodes can compose wide CRISPRa circuits, we constructed one, two, three and four parallel three-layer cascades operating in the same CFS reaction. We used a single input to activate the downstream nodes, and measured circuit performance by connecting all cascade outputs to a RFP node (Figure 3F, left). We maintained the internal node concentrations constant across parallel cascades. We observed up to 66% decrease in output fold-activation as the width of the circuit increased from one to four cascades (Figure 3F, right). This decrease came largely from higher output levels in the absence of scRNA input (Figure S9), most likely due to higher overall basal expression of internal scRNAs. We then constructed the same circuits and tuned the node concentrations proportionally to the number of parallel cascades, effectively maintaining the total node concentration constant. When constructed in this manner, we found no statistically significant difference in the fold-activation across cascades of different widths (Figure 3F, right). Hence, by tuning the concentration of orthogonal CRISPRa/i nodes, we show an arbitrary number of parallel circuits with as many as nine nodes may be regulated.

<u>Developing activatable promoters for blue-light responsive CRISPRa/i circuits</u>

Above we demonstrated that high-performing CRISPRa promoters can be generated through parallel screening of the basal and activated levels of promoter

variants. In this section, we show that the same approach can be used to generate high-performing activatable promoters for blue-light responsive gene expression using the EL222 transcriptional activator. Furthermore, by placing guide RNA expression under the control of the engineered high-performing EL222 promoter, we demonstrate light-responsive CRISPRa/i signal transduction modules and signal processing CRISPRa/i circuits.

High-performing blue-light responsive promoters

The EL222 transcriptional activator interactions with the RNAP and the DNA binding site are well-characterized, making it a suitable model system for developing optogenetic inputs for CRISPRa/i circuits (Figure 4A) (Jayaraman et al., 2016; Motta-Mena et al., 2014; Zoltowski et al., 2013). Briefly, EL222 binds an 18 bp sequence upstream of the -35 region of the *luxI* promoter and subsequently recruits RNAP through interactions with the α and σ subunits (Zoltowski et al., 2013). We mutagenized the *luxI* minimal promoter (Figure 4A, Methods 3.4), and screened variants in *E. coli* in both dark and light to select for high dynamic range (Figure 4B). Starting with a dynamic range of less than 2-fold, we observed up to 4-fold dynamic range in response to blue-light. Similar to our CRISPRa promoter screens, minimal promoters with very low (BBa_J23113) or very high (BBa_J23119) basal expressions exhibited low dynamic range in response to blue-light. We selected 4 variants with >2-fold higher dynamic range than the *luxI* minimal promoter and characterized them in CFS. These variants yielded a 34.1-fold difference in expression between light and dark, compared to just 6.2-fold for the original *luxI* minimal promoter (Figure S10). This improvement comes largely from a reduction in the basal expression from the blue-light promoter,

suggesting we successfully minimized the RNAP-minimal promoter interactions without weakening EL222-promoter interactions. More importantly, these results demonstrate this approach for engineering actionable promoters is applicable to other transcriptional activation systems.

Blue-light responsive CRISPRa/i circuits

We evaluated whether the engineered blue-light promoter transcription levels were suitable for expressing gRNAs for CRISPRa/i circuits. We titrated gRNA-expressing plasmid concentrations and compared RFP expression across dark and light conditions. For CRISPRi, the highest light-dependent change in repression was 50% (Figure S11). For CRISPRa, the highest light-dependent fold-activation was 14-fold (Figures 4C, S12). We then implemented a positive feedback loop to increase light-dependent CRISPRa output levels. In a CRISPRa/i circuit, positive feedback (PFB) can be achieved by including a downstream node that expresses a scRNA targeting an upstream node. We expected the degree of positive feedback in the system to be tunable by titrating the PFB node, with high concentrations of this node resulting in activation in the absence of blue-light. When optimally tuned, the positive feedback loop increases the light-dependent CRISPRa output levels almost 2-fold (Figure 4D). Excess PFB node led to a 7.6-fold increase in basal expression, decreasing the light-dependent activation to 1.2-fold. These results highlight that rationally designed genetic circuits built from engineered activatable promoters can be used to improve the dynamic range of input-responsive signal processing modules.

Engineered activatable promoters enable conditional CRISPRa dependent on protein-protein interactions

The versatility of protein-protein interaction (PPI)-mediated genetic regulation for coupling peptide or small-molecule binding to transcriptional outputs has long attracted interest (Gao et al., 2016; Kundert et al., 2019; Nihongaki et al., 2015; Polstein & Gersbach, 2015; Y. Yu et al., 2020). Implementation of PPI-dependent transcriptional activation has been difficult in prokaryotic systems due to strict target site requirements and low dynamic range of activatable promoters (Dong et al., 2018; Fontana, Dong, et al., 2020). Therefore, we explored if high-performing activatable promoters could permit construction of PPI-dependent conditional CRISPRa signal transduction modules.

Development of Conditional CRISPRa Systems

As an experimental testbed, we incorporated three previously characterized protein-protein heterodimerization domains into our CRISPRa system: the synthetic coiled-coil SYNZIP 5/6 pairs (Thompson et al., 2012), the abscisic acid (ABA) responsive ABI–PYL1 (Cunningham-Bryant et al., 2019; Gao et al., 2016; Liang et al., 2011), and the gibberellic acid (GA) responsive GID1–GAI (Gao et al., 2016; Miyamoto et al., 2012). We fused these heterodimerization domains to SoxS and MCP to enable conditional recruitment of SoxS to the CRISPRa complex. We generated the MCP-SZP6 and SoxS-SZP5 domains for SYNZIP-CRISPRa, the MCP-ABI and SoxS-PYL1 domains for ABA-CRISPRa, and MCP-GAI and SoxS-GID1 for GA-CRISPRa (Figure 5A). In the original J3 promoter context, we observed a 5.7-fold activation with SYNZIP-CRISPRa when SoxS was fused to the N-terminus of the SYNZIP domain, compared to 1.4-fold when fused in the opposite orientation (Figure

S13). The fold-activation of ABA-CRISPRa was also maximized when SoxS was at the N-terminus (Figure S13, right), therefore we moved forward with all SoxS N-terminus fusions. MCP-SYNZIP gave 5.7-fold activation, while SYNZIP-MCP gave only 2.8-fold (Figure S13). For ABA- and GA-CRISPRa, only C-terminus MCP fusions were tested due to the MCP-SYNZIP result and the strong precedent for using C-terminus MCP fusions in CRISPRa systems (Dong et al., 2018; Zalatan et al., 2015).

CRISPRa operates narrowly within -71:-101 bp from the TSS in a phase-dependent manner (Fontana, Dong, et al., 2020). We questioned whether the introduction of an additional protein linkage into MCP-SoxS affects the relative scRNA target site requirements (Villegas Kcam et al., 2021). We designed a CRISPRa promoter with densely packed scRNA target sites and variants with 1 bp frameshifts to allow screening with single base pair resolution between -81:-111 bp from the TSS (Figure 5B, Table S1). Surprisingly, SYNZIP-CRISPRa maintains the same preference for the -81 bp from the TSS targeting site and the same stringent 10-11 bp phase dependency seen in conventional CRISPRa (Figure 5B, S14). Having identified the optimal scRNA target site, we changed the promoter with densely packed scRNA target sites for a high-performing promoter (HP3, Table S1), and observed a 5.4-fold improvement in SYNZIP-CRISPRa fold-activation (Figure S15).

We then identified the permissible small-molecule input and protein expression levels of each conditional CRISPRa system in the context of our high-performing activatable promoters. For ABA- and GA-CRISPRa, small-molecule titrations showed that ABA-CRISPRa is responsive between 0.1-10 uM, and GA-CRISPRa is responsive between 0.1-100 uM (Figure 5C), with up to 7.9- and 9.0-fold-activation, respectively.

We screened dimer stoichiometries and expression levels by surveying a range of concentrations for both the MCP and SoxS components. SYNZIP-CRISPRa performs the best of the three systems, giving a maximal activation of 67-fold compared to a reaction with no MCP or SoxS plasmid added. Even at low CRISPRa component concentrations, SYNZIP-CRISPRa still achieves 59-fold activation (Figure 5D, left). ABA-CRISPRa gives a maximum activation of 18.6-fold (Figure 5D, middle). For GA-CRISPRa, the maximum activation of 5.9-fold was accessible in a relatively narrow range of component concentrations (Figure 5D, right). For all three conditional CRISPRa systems, higher expression of heterodimers did not necessarily improve activation. In line with the behaviors of natural scaffolds (Douglass et al., 2013; Levchenko et al., 2000), we observed a unique optimal concentration for each conditional CRISPRa system. The differences between systems may be due to the different affinities of each protein-protein interaction.

Engineering multi-input CRISPRa/i circuits

Multi-layer and multi-input circuits with conditional CRISPRa

To use conditional CRISPRa systems as inputs into more complex CRISPRa/i circuits, we began by characterizing the scRNA dose-response curve of the novel CRISPRa systems. For all three conditional systems, the amount of scRNA needed to saturate the CRISPRa response was similar to that of direct CRISPRa (Figures S5, S16). We tested the orthogonality of the small molecule systems to evaluate if they could be used together for independent gene regulation (Figure 6A). We found ABA-CRISPRa to be highly specific to its target ligand, showing no significant activation

in the presence of GA. GA-CRISPRa showed 3.1-fold cross-activation from ABA, in line with reports in eukarya (Gao et al., 2016), but maintained a 3-fold higher specificity for its target ligand, giving 10.5-fold activation from GA. These results suggest that the ABA- and GA-CRISPRa systems can be used for effective orthogonal gene regulation.

We built two types of input-responsive circuits to demonstrate the use of conditional CRISPRa for multi-input and multi-layer input processing: an AND logic gate and a CRISPRa cascade. To accomplish AND logic, we co-expressed both ABA- and GA-CRISPRa (Figure 6B, left). The addition of either ABA or GA resulted in 2-fold activation compared to the no-ligand condition. In the presence of both ligands, 4.5-fold activation was achieved, allowing us to clearly distinguish between one- and two-inputs ($p = 0.03$) and achieve AND logic. A two-layer ABA-CRISPRa cascade with both internal layers dependent on ABA gave 2.5-fold activation upon addition of ABA, showing that conditional CRISPRa can also support multi-layer information processing (Figure 6B, right).

Two-input dynamic pulse generator

Synthetic biologists aim to recreate complex, dynamic signaling networks that use multiple input-responsive regulators to tightly regulate the expression timing and magnitude of downstream targets (Benzinger et al., 2022; Cai et al., 2008; Wang et al., 2014). Towards this goal, we sought to couple blue-light CRISPRi with conditional CRISPRa into a tunable pulse generator with two-input control over the level and timing of gene expression (Figure 6C, left). Based on our system characterizations, SYNZIP-CRISPRa is well suited for integration with blue-light CRISPRi due to its high-fold activation. To simulate relative production rates as a function of CRISPRa and

CRISPRi inputs, we built upon a coarse-grained mechanistic model of CRISPRa/i regulation (Tickman et al., 2021) by introducing blue-light pulses regulating sgRNA expression (Methods 6). By simulating changes in the pulse width as CRISPRi inputs as well as changes in the scRNA concentration as CRISPRa inputs, we were able to capture changes in both the levels and timing of RFP production rates (Figure 6C, left). This simulation analysis indicates that there are regimes where blue-light CRISPRi and conditional CRISPRa can be coupled to modulate the production rate of a desired gene target.

Experimentally, we first kept the CFS in the dark for one hour to allow for EL222 expression and used varying blue-light exposure times to dynamically tune CRISPRi. To tune SYNZIP-CRISPRa, we changed the scRNA-expressing plasmid concentration. As predicted by the model, RFP production rate pulse was tunable by both the scRNA plasmid concentration and the exposure time to blue-light (Figure 6D). When compared across conditions with the same CRISPRa input, higher CRISPRi input led to 20-56% lower maximum production rates. When compared across conditions with the same CRISPRi input, higher CRISPRa input led to faster maximum production rates by 30 minutes. These results demonstrate that different CRISPRa/i-based signal transduction modules can be integrated and individually modulated to control gene expression, highlighting the potential of input-responsive CRISPRa/i GRNs for complex signal processing applications.

Discussion

Natural biological systems have evolved GRNs containing activatable promoters with wide ranges of expression characteristics that enable dynamic responses to changing environmental conditions. Engineering activatable promoters has traditionally been thought to involve a trade-off between basal and activated expression levels (Alper et al., 2005; Cress et al., 2016; Fontana, Dong, et al., 2020). In this work, we show that basal and activated expression levels can largely be decoupled to generate activatable promoters with both lower basal and higher activated expression levels than previously possible. High-throughput, sequential screening of promoter regions constructed through sequence-based design allowed us to overcome context effects and identify high-performing activatable promoters. With this approach, we successfully engineered a suite of orthogonal CRISPRa promoters that match the basal and activated expression levels of the canonical Tet inducible system and exceed those of the Lac system (Figure S17) (T. S. Lee et al., 2011).

The *E. coli* transcriptional network is governed by a hierarchical structure containing nine layers of regulation (Ma et al., 2004). Engineered activatable promoters allowed us to build multi-layer CRISPRa/i GRNs in *E. coli*-based CFS with depths and widths significantly larger than the state of the art (Garamella et al., 2016; Tickman et al., 2021), approaching the complexity of natural GRNs. Specifically, a 33-fold improvement in promoter dynamic range resulted in 80% lower signal degradation in two-layer cascades (Figures S6), and enabled deep GRNs with up to six layers of regulation. Additionally, we demonstrated wide GRNs regulating up to four parallel cascades, indicating that the CRISPRa/i framework is well suited for the design of wide

control circuits for parallel computing and multi-gene regulation. Further improvements in GRN complexity may be limited by resource constraints, such as upstream gRNAs outcompeting downstream gRNAs for dCas9 binding (Figure S7). Strategies to dynamically regulate upstream gRNA expression, such as reversing CRISPRa complex binding or implementing negative autoregulation motifs (Specht et al., 2022), could enable even larger GRNs.

Biological systems continuously monitor and process environmental signals by using signal transduction modules, such as inducible transcriptional effectors, as inputs to complex GRNs (Kaplan et al., 2008; Krishna et al., 2009). Our work provides a general framework for optimizing transcriptional activation systems at the promoter level and integrating them into CRISPRa/i GRNs. Promoter engineering of the optogenetic EL222 system enabled high light-dependent dynamic ranges, with relevant expression levels for downstream applications. Through inducible gRNA expression, we demonstrated input signal modulation through various GRN topologies, including positive feedback loops and CRISPRa/i cascades, as well as integration of different signal transduction modules into the same CRISPRa/i GRN. Given the high degree of tunability and composability of the CRISPRa/i framework (Santos-Moreno & Schaerli, 2020; Tickman et al., 2021), other GRNs with different signal processing functions, such as thresholding (Groseclose et al., 2021), amplification (Wang et al., 2014), and frequency filtering (Bashor et al., 2019), should be readily implementable. Overall, our work highlights the potential for building complex, input-responsive systems through CRISPRa/i GRNs.

Protein-protein interactions have been used widely to execute complex, input-responsive functions in eukaryotes (Beltrán et al., 2022; Foight et al., 2019; Guntas et al., 2015; Huang et al., 2020; Piraner et al., 2017). Implementing similar systems in prokaryotes has been difficult, and the development of high dynamic range promoters allowed us to successfully prototype and optimize conditional CRISPRa systems in *E. coli*-based CFS. Implementation of novel conditional CRISPRa systems may be streamlined by the fact that all systems tested here are effective when targeted -81 bp from the TSS, despite the presence of additional protein-protein interactions up to 500 amino acids in length. Additionally, conditional CRISPRa fold-activation is proportional to the strength of the protein-protein interaction (Table S2) (Carey et al., 1983; Dupeux et al., 2011; Miyamoto et al., 2012; Miyazono et al., 2009; Thompson et al., 2012; Ueguchi-Tanaka et al., 2005; Yoshida et al., 2018), informing the *a priori* selection of heterodimers for use in conditional CRISPRa. Collectively, our work suggests that other heterodimerization domains could be implemented, with minimal prototyping, as signal transduction modules for CRISPRa/i GRNs for multiplexed biosensing or screening of PPIs in CFS.

Our workflow for activatable promoter engineering enables the dynamic specification of expression levels for large networks of orthogonal gene targets. The new classes of deep, wide, and input-responsive CRISPRa/i GRNs developed here have immediate application in CFS for investigating the rules for genetic circuit design (Alon, 2007; Brophy & Voigt, 2014) and biological information processing (Adamala et al., 2017; Aufinger et al., 2022; Garamella et al., 2019; Karzbrun et al., 2014), as well as for building dynamic, multi-enzyme expression programs for self-assembling

bioproduction platforms (Grubbe et al., 2020; Kruyer et al., 2021; Pardee et al., 2016; Stark et al., 2021). Moreover, CRISPRa/i GRNs could be integrated with existing field-deployable medical diagnostics and environmental monitors to enable complex, multi-input signal processing (Jung et al., 2020a, 2022; Nguyen et al., 2021; Peruzzi et al., 2022). Moving forward, this work could serve as a stepping stone for building entirely synthetic cells and engineered living materials with GRNs that match or go beyond the complexity of natural systems.

Methods

1. Plasmid and Library Construction

1.1 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 and libraries 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. Transformed cells were 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.

For libraries, 10 uL of the outgrowth was diluted 1:20 with LB and plated onto LB-agar with carbenicillin or gentamicin to check library complexity. The remaining outgrowth was seeded into 5 mL of LB with carbenicillin or gentamicin. Transformed cells were grown overnight ~16 hours at 37 °C. Single colonies were picked from plates and grown overnight in LB with carbenicillin or gentamicin. Both single colonies and culture were sequence verified.

Plasmids were isolated from subcultures using a DNA miniprep kit (QIAprep Spin Miniprep Kit) and Sanger sequenced (Genewiz inc.) to identify correctly assembled plasmids.

2. *E. coli* Experiments

In all *E. coli* experiments CRISPRa system components (dCas9, MCP-SoxS, scRNA) are located on a p15A ori plasmid while reporter construct is located on a pSC101** ori plasmid. Plate reader measurements were conducted using a BioTek Synergy HTX with a black flat bottom plate (Ref# 3631) using 100 µL of culture. 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, covering with breathable membrane (Breathe Easier cat# Z763624) and shaking overnight at 37 °C at 1200 RPM on a Heidolph Titramax 1000. For inducible experiments, overnight cultures are subsequently diluted 1:40 into a fresh plate of EZ-RDM and supplemented with appropriate concentrations of aTc. Measurements are conducted in Costar 96 well black flat bottom plates in 100 uL culture volume.

3. Design of Promoter Region Libraries

3.1 Minimal Promoter Libraries

Of the two libraries, the first was designed by rationally mutagenizing specific bases that are known contacts of RNAP within the minimal promoter. The second library was made by randomly mutagenizing within the intervening sequence. Since the libraries yielded

similar Pareto fronts, we then combined these mutations into a third minimal promoter library, which was used in the sequential screening process.

3.2 UP-Element Libraries

We designed five UP-element libraries mutagenizing the AT-rich *E. coli* consensus sequence with increasing GC-content. We generated 5 libraries from 0% to 100% GC-content, and a library representing the *E. coli* consensus sequence (Table S3).

3.3 scRNA Target Site Libraries

We generated three scRNA target site libraries with varying compositions of GC-content (0%, 50%, and 100%) (Table S3). These libraries were used in tandem with a GC-rich UP-element.

3.4 EL222 Minimal Promoter Libraries

Starting with the native *luxI* minimal promoter, we introduced rational mutations to make it resemble a synthetic activatable promoter (J23117). We then randomly mutagenized within the -10:-35 region (Table S3).

4. Cell-Free Reactions

4.1 Cell-Free System Preparation

The cell-free system was acquired from Arbor Biosciences (myTXTL). 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. Details about plasmid preparation are provided in Methods S1.

4.2 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. For reactions containing ABA (Sigma, A4906) or GA, .1 µL of the small molecules were added alongside the plasmids. For reactions involving more than 5 plasmids, plasmids were mixed with an acoustic liquid handler robot (Echo Labcyte 525). 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. 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.

5. Optogenetic Experiments

For both *E. coli* and CFS optogenetic experiments, the cultures and reactions were prepared as described above. The incubation conditions were modified to include a blue-light illumination source (UVP Visi-Blue UV Transilluminator, 8 Watts, 460/470 nm). The samples were placed at 37 °C or 29 °C in an incubator (Thermo Forma Orbital Shaker, Model #435) with the illumination source placed atop the incubator and irradiating inwards. The distance between the illumination source and the deepwell plates is 14 cm. CFS reactions were placed inside the incubator at 29 °C at a distance

of 6 cm with the bottom of the wells facing the illumination source. In both cases, the dark conditions were kept inside a cardboard box inside the incubator. Endpoint plate reader measurements were conducted using a BioTek Synergy HTX.

6. CFS Blue-light CRISPRa/i modeling

The CFS blue-light CRISPRa/i model was expanded from the previously described CFS CRISPRa/i model (Tickman et al., 2021). The model constitutes a series of first order chemical reactions for protein and guide RNA production, CRISPR complex assembly, and DNA targeting. All model details are described in Methods S1.

7. Quantification and statistical analysis

7.1 Data analysis

Throughout this work all measured RFP levels in *E. coli* were normalized by measured OD600 with appropriate propagation of uncertainties. All metrics are described in Methods S2.

7.2 Statistics

Statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks in Figures indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.01, ∗∗∗: p-value < 0.001).

Author Information

Corresponding Authors

James M. Carothers - *Department of Chemical Engineering, Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA 98195*; Email: jcaroth@uw.edu

Jesse G. Zalatan- *Department of Chemistry, Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA 98195*; Email: zalatan@uw.edu

Authors

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

Ryan Cardiff - *Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA 98195*

Benjamin I. Tickman - *Department of Chemical Engineering, Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA 98195*

Cholpisit Kiattisewee - *Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA 98195*

Cassandra Maranas - *Molecular Engineering & Sciences Institute and Center for Synthetic Biology, University of Washington, Seattle, WA 98195*

Author Contributions

D.A.B., R.C., B.I.T., J.G.Z., and J.M.C. designed the research. D.A.B., R.C., and C.M. performed experiments. D.A.B., R.C., B.I.T., C.K., and C.M. analyzed the data. D.A.B., R.C., J.G.Z., and J.M.C. wrote the manuscript with input from all of the authors.

Acknowledgements

We thank members of the Carothers and Zalatan groups for advice, materials, and comments on the manuscript.

Competing interests

D.A.B., R.C., J.G.Z., and J.M.C are inventors on patents and/or patent applications filed by the University of Washington that describe promoter engineering, conditional CRISPRa technologies, and CRISPRa/i circuits in prokaryotic systems. J.G.Z. and J.M.C are members of the Wayfinder Biosciences scientific advisory board.

Data sharing plans

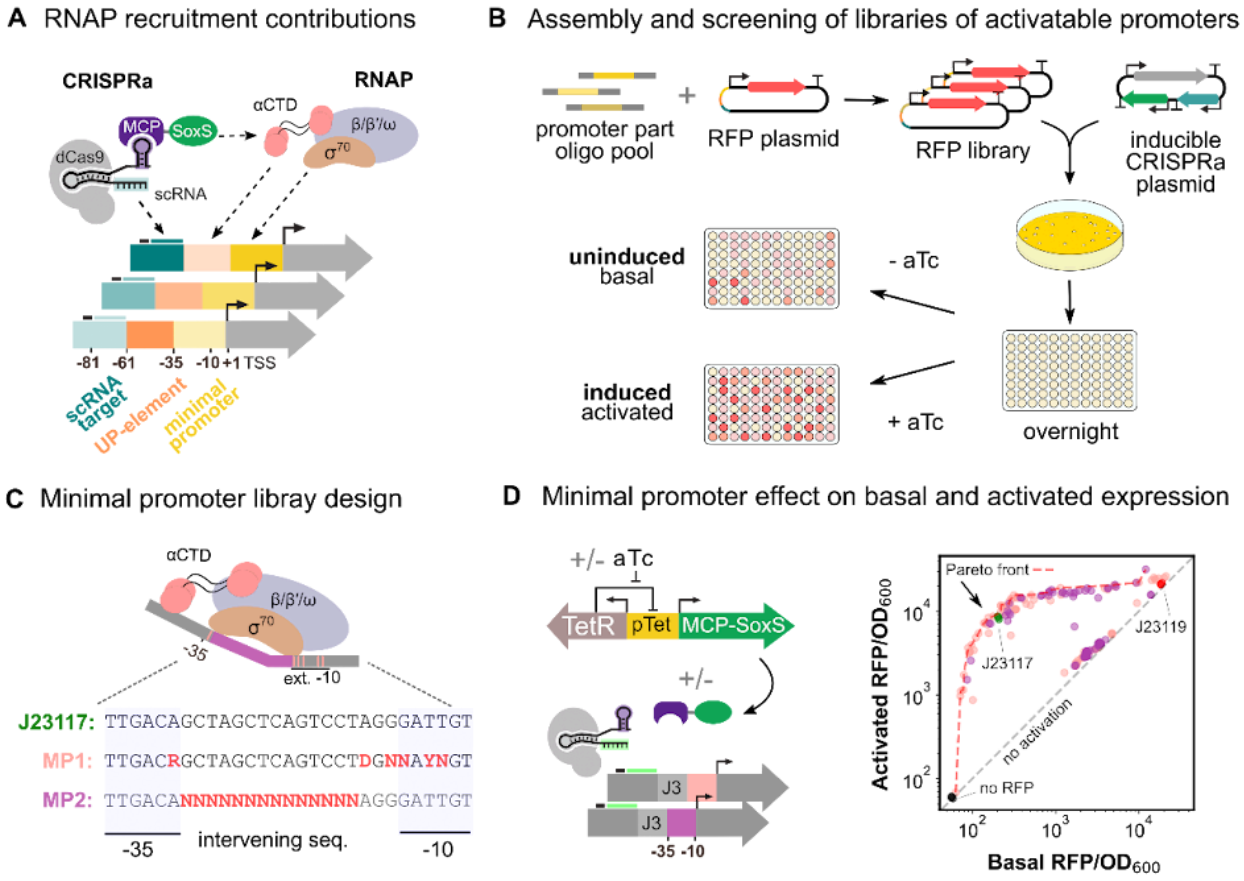
Data supporting the findings of this work are available within the paper and its supporting information files. All data and code are available on GitHub at: https://github.com/carothersresearch/CRISPRai_Circuits_2022

Funding information

This work was supported by US National Science Foundation (NSF) Award MCB 2032794 (to J.M.C. and J.G.Z.), NSF Award CBET 1844152 (to J.M.C.), NSF Award EF-1935087 (to J.M.C.), US Department of Energy (DOE) BETO DE-EE0008927 (to J.M.C. and J.G.Z.) and DOE ARPA-E DE-AR00002387-1567 (to J.M.C). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2140004 (D.A.B.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Figures

Figure 1: Functional interrogation of promoter regions with CRISPRa



A. Schematic of RNAP interactions with the CRISPRa complex and target promoter. σ^{70} affinity for the minimal promoter and α-CTD affinity for the UP-element determines RNAP recruitment to a promoter. When CRISPRa is targeted to a promoter with a complementary scRNA target site, the RNAP α-CTD domain is recruited by the SoxS transcriptional activator. RNAP-promoter and CRISPRa-promoter interactions can be modulated by modifying the DNA sequence of the different promoter regions.

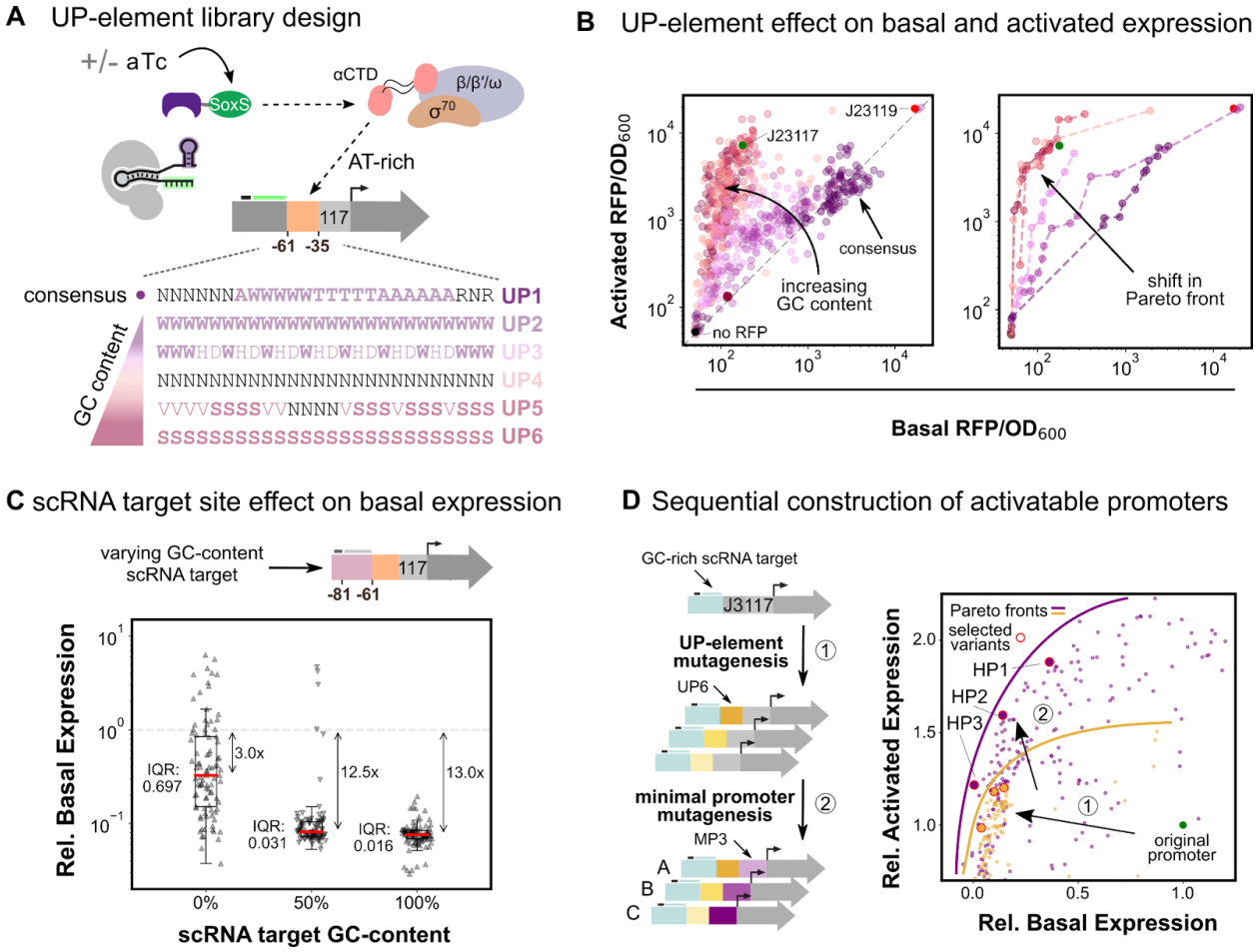
B. Workflow for the assembly and characterization of libraries of activatable promoters. A library of RFP genes with varying promoters is generated through PCR (Methods 1.1). The library is then co-transformed into *E. coli* with an aTc-inducible CRISPRa plasmid. Colonies are then seeded overnight, and subsequently diluted into media with appropriate concentrations of aTc. For each promoter variant in the libraries, basal and activated RFP levels were measured with 0 nM and 200 nM aTc, respectively (Methods 2).

C. Schematic of RNAP interaction with the minimal promoter and library design. σ^{70} recognizes specific positions in the extended -10 and -35 regions of the minimal promoter, which informed the design of the library MP1. σ^{70} binding is also influenced by the GC-content, the length, and the $^{-15}$TGn^{-1} motif of the intervening sequence, which informed the design of library MP2 (Methods 3.1).

D. Minimal promoter effect on expression levels. **Left:** Inducible CRISPRa system and minimal promoter libraries of the J3 synthetic promoter. MCP-SoxS is expressed from the aTc-inducible pTet promoter. dCas9 and J306 scRNA are constitutively expressed. **Right:** Activated and basal RFP/OD_{600} for the two minimal promoter libraries (n_{MP1} = 89, n_{MP2} = 84). Red dash line defines the Pareto front containing the best performing promoter variants (Methods S3), for which no further improvements in basal or activated levels can be achieved without compromising the other. Gray dash line defines promoter variants with equal activated and basal expression levels, indicating they are not activated by CRISPRa. The J23117 minimal promoter (green, triplicates) is included as a standard reference for CRISPRa efficiency. The J23119 minimal promoter (red, triplicates) is an example of a non-activatable promoter due

to high basal expression levels. A plasmid without RFP (black, triplicates) indicates the background fluorescence of the system.

Figure 2: Combining promoter regions to engineer high-performing CRISPRa promoters



A. Schematic of RNAP interactions with the UP-element and library design. α-CTD affinity for AT-rich UP-elements upstream of the minimal promoter helps recruit RNAP. Upon targeting with CRISPRa, UP-element RNAP recruitment contributions are largely replaced by SoxS-RNAP α-CTD interactions. UP-element libraries with increasing GC-content were designed to minimize α-CTD interactions (Methods 3.2).

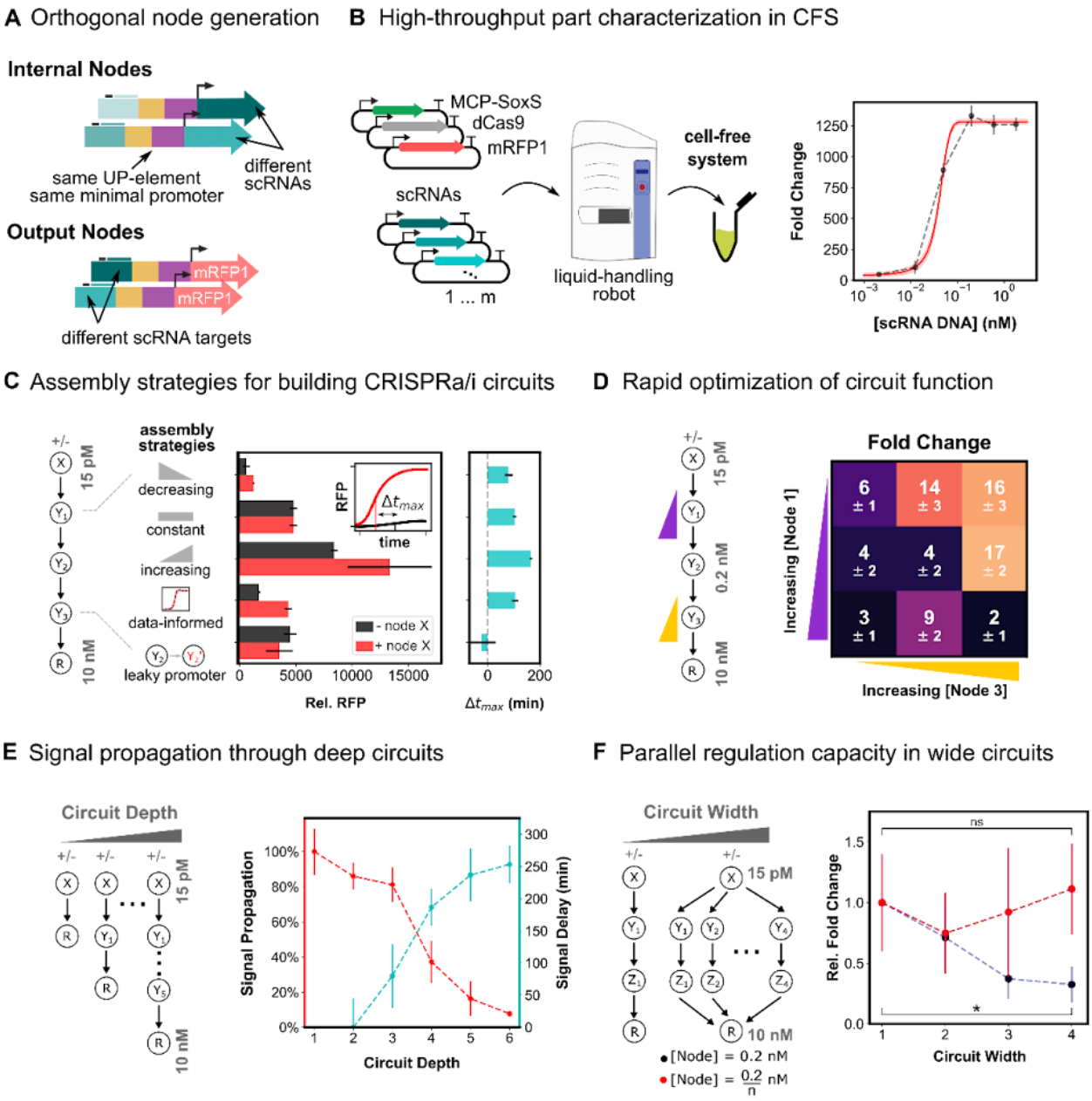
B. UP-element GC-content effect on expression levels is shown through activated and basal RFP/OD_{600} for the six UP-element libraries ($n_{UP1} = n_{UP2} = \ldots = n_{UP6} = 110$). **Left:** Increasing GC-content in the UP-element lowers the range of basal expression level, while maintaining the full range of activated expression levels. Gray dash line

defines promoter variants with equal activated and basal expression levels. **Right:** Colored dash lines define the Pareto front for each UP-element library (Methods S3). Increasing the UP-element GC-content effectively shifts the Pareto front towards lower basal expression levels.

C. scRNA target site composition effect on basal expression. Comparison of three scRNA libraries with increasing GC-content (n_{S1} = n_{S2} = n_{S3} = 93) (Methods 3.3). Basal expression levels are normalized to the standard J3 promoter basal expression level. Red lines indicate the median expression level of each distribution. The interquartile range (IQR) is calculated as the difference between the upper and lower quartiles and measures the spread of the distribution.

D. Sequential construction of activatable promoters. **Left:** Activatable promoters were constructed by sequential library mutagenesis screens starting from the J3 promoter with a GC-rich scRNA target site. Three Pareto optimal UP-elements were selected after promoter mutagenesis with a GC-rich UP-element library (1). We then mutagenized the minimal promoter of the three previously selected variants (2), and again selected three Pareto optimal variants. **Right:** Basal and activated expression levels for all mutagenesis variants normalized to the standard J3 promoter expression levels (green). Yellow points represent variants from the UP-element mutagenesis (n_{UP6} = 192) (1), while purple points represent variants from the minimal promoter mutagenesis (n_{MP3} = 279) (2). Red circles indicate selected variants from each screen, and solid lines depict the Pareto optimal fronts. Each sequential mutagenesis led to variants with both lower basal and higher activated expression levels.

Figure 3: Engineering deep and wide circuits with high-performing CRISPRa promoters



A. Schematic of orthogonal CRISPRa/i nodes for use in cell-free circuits. Internal nodes contain an orthogonal scRNA target site and express orthogonal scRNAs. Output nodes contain orthogonal scRNA target sites and express RFP. All nodes contain the same UP-element and minimal promoter (HP3).

B. High-throughput characterization of scRNA components in CFS. **Left:** Plasmids encoding each CRISPRa component are mixed using an acoustic liquid handling

robot and expressed in CFS. **Right:** scRNA-dose response curves for each node are generated by titrating the amount of scRNA plasmid from 0.5 pM to 5 nM.

C. Comparison of assembly strategies for building a four-layer CRISPRa cascade. **Left:** Internal node concentrations either decreased from 200 pM to 32 pM as depth increased, were held constant at 200 pM, or increased from 200 pM to 1.25 nM as depth increased. A fourth assembly method was tested in which internal node concentrations were 40, 200, and 170 pM, based on individual scRNA-dose response characteristics. A fifth cascade was included in which the high-performing promoter of the second internal node was replaced with the leaky J2 promoter. Input and output node concentrations were held constant across all strategies at 0 or 15 pM and 10 nM, respectively. **Center:** Cascade output RFP expression for each assembly strategy with scRNA input (red) and without (black), relative to RFP basal expression. **Right:** Change in time to maximum expression rate (Δt_{max}) for each assembly strategy (Methods S3).

D. Rapid fold change optimization of a four-layer CRISPRa cascade. **Left:** The first and third internal nodes of the cascade were varied between 40 and 160 pM, and 85 and 340 pM, respectively. The input node, second internal node, and output node were held constant at 0 or 15 pM, 0.2 nM, and 10 nM, respectively. **Right:** Fold change between with and without scRNA input for each CRISPRa cascade.

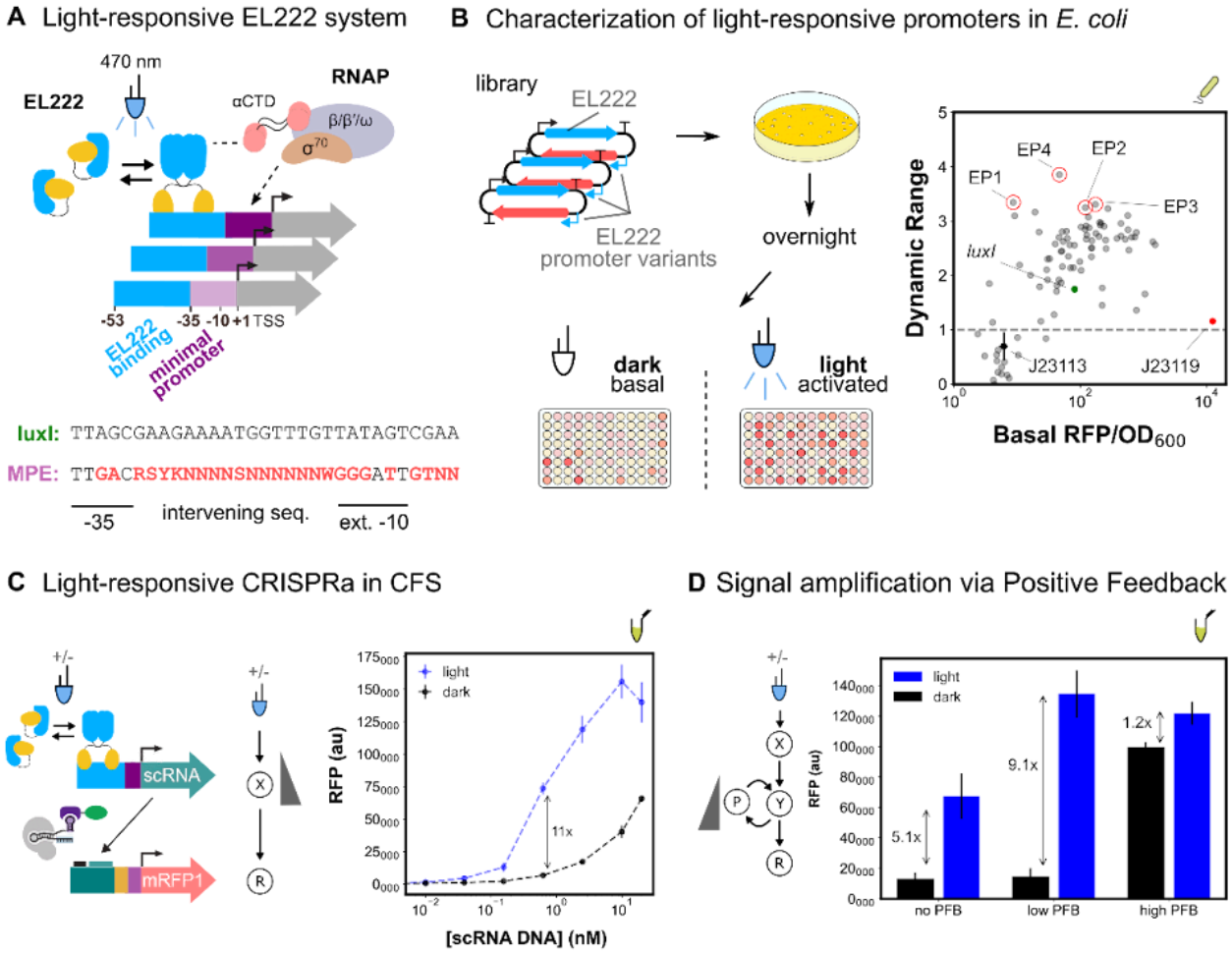
E. Signal propagation through deep CRISPRa/i circuits. **Left:** CRISPRa cascades with increasing depth. Input and output node concentrations were held constant across all cascades at 0 or 15 pM and 10 nM, respectively. All of the parallel cascade scRNA outputs are connected to the same RFP node. All node

concentrations are tabulated in Table S5. **Right:** Propagation efficiency and signal delay are shown as a function of circuit depth (Methods S3).

F. Construction of wide CRISPRa/i circuits. **Left:** CRISPRa cascades with increasing width. Input and output node concentrations were held constant across all cascades at 0 or 15 pM and 10 nM, respectively. **Right:** The concentration of each internal node was held at 0.2 nM as circuit width increased (blue), or the internal node concentration was scaled down proportionally to the width of the circuit (red), such that each internal node concentration is 0.2/n nM, where n is the number of parallel cascades. Fold-activation is given relative to a single CRISPRa cascade (Methods S3).

For all panels, values represent the mean ± standard deviation of three technical replicates.

Figure 4: Developing activatable promoters for blue-light responsive CRISPRa/i circuits



A. Schematic of EL222 light-responsive promoter system and library design. EL222 transcription factor dimerizes in response to 470 nm light and binds a specific sequence upstream of the minimal promoter. EL222 then recruits RNAP through interactions with the α-CTD domain. Minimal promoter library design is based on the original *luxI* promoter and previous minimal promoter libraries (Methods 3.4).

B. Characterization of light responsive promoters in *E. coli*. **Left:** Blue-light promoter screening (Methods 5). EL222 protein and promoter library are expressed from a single plasmid. Assembly and screening are carried out as previously described. Basal and activated expression levels are measured from cultures not exposed

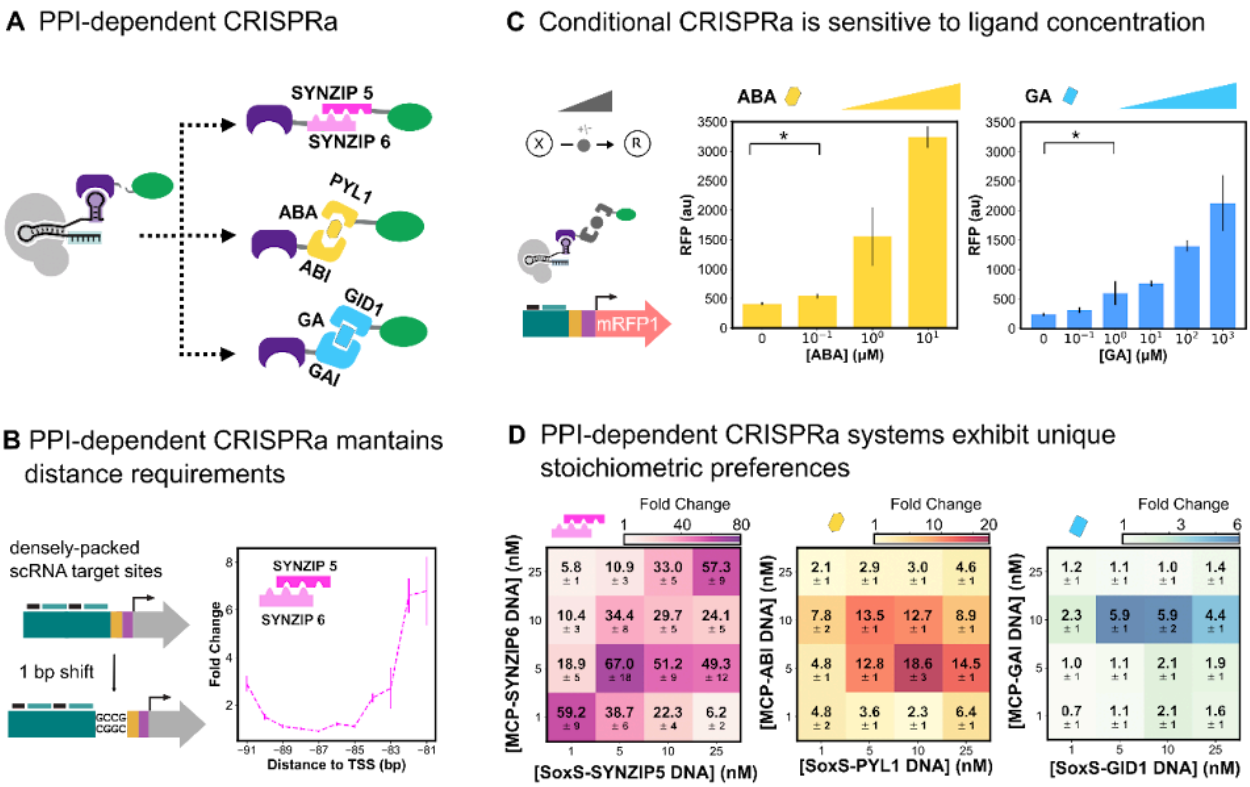
or continuously exposed to blue-light, respectively. **Right:** Basal expression and dynamic range of blue-light promoter variants (n_{MP3} = 96). Gray dash line defines promoter variants with equal activated and basal expression levels, indicating they are not activated by EL222. The J23119 minimal promoter (red) and J23113 (black) are examples of non-activatable promoters. Variants with improved performance (red circles) compared to the original *luxI* promoter (green) were selected for use in CFS.

C. Light-responsive CRISPRa in CFS. **Left:** EL222 scRNA expression from an engineered blue-light promoter and downstream CRISPRa. Reactions contain 8nM and 10 nM of EL222 and RFP plasmids respectively. **Right:** Titration of blue-light inducible scRNA plasmid concentration to maximize the fold change between blue-light dependent CRISPRa (blue) and CRISPRa due to scRNA leak in the dark (black).

D. Improvement of blue-light CRISPRa dynamic range through the construction of a positive feedback circuit. **Left:** Blue-light responsive CRISPRa cascade with positive feedback (PFB). PFB is achieved by including a downstream node that expresses a scRNA targeting an upstream node Reactions contain 15nM and 10 nM of EL222 and RFP plasmids respectively. **Right:** Blue-light dependent CRISPRa (blue) and CRISPRa due to scRNA leak in the dark (black). The amount of positive feedback was tuned by adjusting the concentration of the PFB node. “No”, “Low”, and “High” PFB concentrations correspond to 0, 3 pM, and 2 nM, respectively.

For all panels, values represent the mean ± standard deviation of three technical replicates.

Figure 5: Engineered activatable promoters enable PPI-dependent conditional CRISPRa



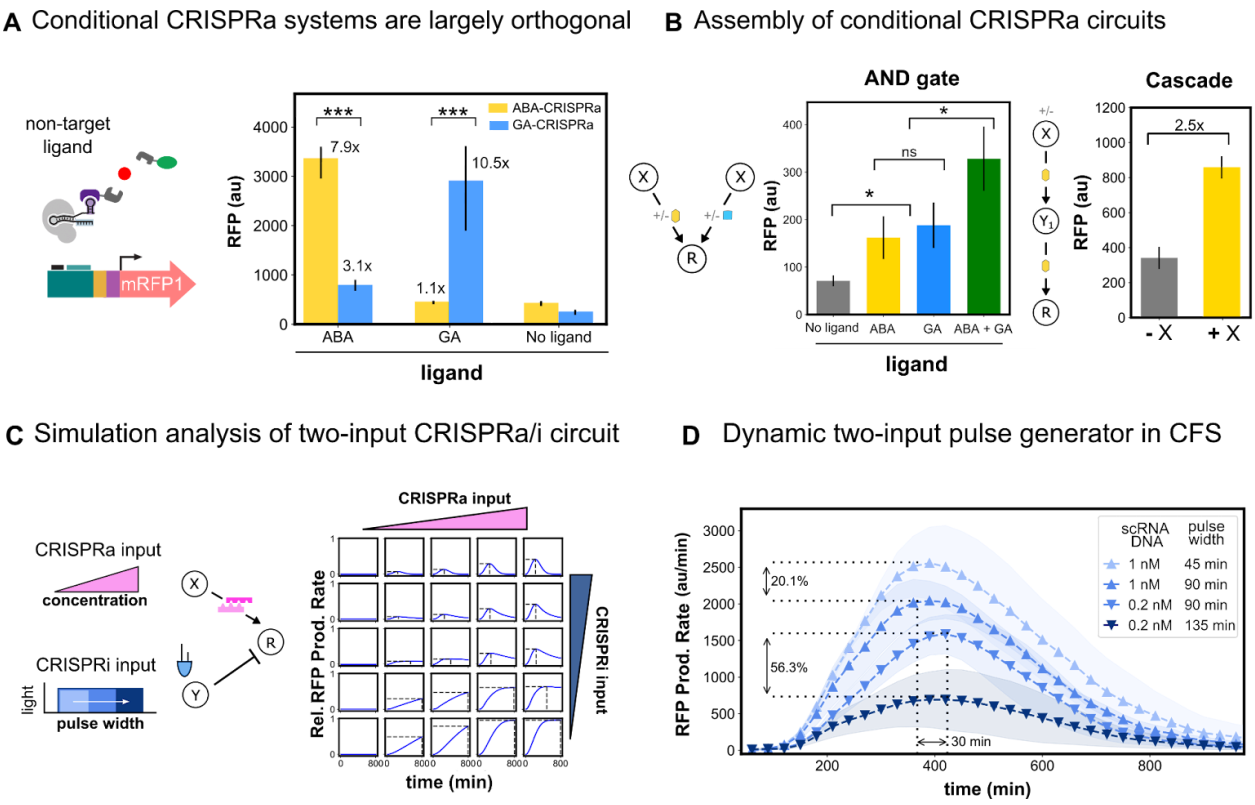
A. Schematic of different PPI-dependent CRISPRa systems. MCP-SoxS fusion is split and the two proteins are instead fused to one end of a heterodimerization domain. The heterodimerization domains used to build PPI-dependent CRISPRa systems are the SYNZIP5/SYNZIP6 pair, the abscisic acid (ABA)-responsive ABI/PYL1 domain, and the gibberellic acid (GA)-responsive GAI/GID1 domain.

B. Distance requirements of PPI-dependent CRISPRa. **Left:** Engineered promoter containing densely-packed scRNA target sites and single base pair 5' additions allows for CRISPRa targeting between -81 and -111 bp from the TSS. **Right:** Testing SYNZIP-CRISPRa between -81 and -91 bp from the TSS. SYNZIP-CRISPRa components are expressed at 5 nM. Fold change is calculated relative to an off-target scRNA for each promoter variant.

C. Tuning conditional CRISPRa response through titration of small molecule concentration. For ABA- and GA- CRISPRa, the corresponding small molecule was titrated between 0 and 10 or 0 and 10^3 μM respectively to find the optimal concentration. ABA- and GA-CRISPRa components are expressed at 10 nM.

D. Improving PPI-dependent and conditional CRISPRa response by optimizing component stoichiometries. The concentration of the plasmids expressing the MCP and SoxS components for each dimerization system were varied 1-25 nM and tested combinatorially to find the best ratio of the two heterodimers. ABA is added at 10 μM and GA is added at 10^3 μM. Fold change is given relative to a reaction with no MCP and SoxS plasmids added.

Figure 6: Engineering multi-input CRISPRa/i Circuits



A. Conditional CRISPRa response to non-cognate ligands. The orthogonality of the small molecule-responsive conditional CRISPRa systems was tested by adding either the corresponding or non-corresponding small molecule to cell-free reactions containing the components for ABA- or GA-CRISPRa. All components are added at their respective optimal screened concentrations. ABA is added at 10 μM and GA is added at 10^3 μM.

B. Assembly of conditional CRISPRa circuits. For both circuits, all components are added at their respective optimal screened concentrations. ABA is added at 10 μM and GA is added at 10^3 μM. **Left:** AND-like behavior was constructed by adding the components for both ABA- and GA- CRISPRa in a cell-free reaction. **Right:** The CRISPRa cascade was assembled by using ABA-CRISPRa to

activate expression of both the first and second node in an activation cascade. The first node was added at either 0.05 or 0 nM, and the internal and output nodes were added at 10 nM.

C. Simulation analysis of a two-input CRISPRa/i circuit using SYNZIP5/SYNZIP6 heterodimerization mediated-CRISPRa and blue-light CRISPRi (Methods 6).

D. SYNZIP-CRISPRa and blue-light CRISPRi were integrated to construct a tunable pulse generator. The amount of CRISPRa was tuned by adding either 0.2 nM or 1 nM of constitutively expressed scRNA plasmid to the CFS reaction. The sgRNA targeting RFP for CRISPRi was driven from the blue-light responsive engineered EL222 promoter. The amount of CRISPRi was tuned by adjusting the time of blue-light exposure between 45 and 135 min. RFP production rates (Methods S3) are plotted as a function of CRISPRa and CRISPRi inputs.

For all panels, values represent the mean ± standard deviation of three technical replicates.

[1] H.-W. Ma, B. Kumar, U. Ditges, F. Gunzer, J. Buer, and A.-P. Zeng, “An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs,” *Nucleic Acids Res.*, vol. 32, no. 22, pp. 6643–6649, 2004, doi: 10.1093/nar/gkh1009.
[2] L. Cai, C. K. Dalal, and M. B. Elowitz, “Frequency-modulated nuclear localization bursts coordinate gene regulation,” *Nature*, vol. 455, no. 7212, Art. no. 7212, Sep. 2008, doi: 10.1038/nature07292.
[3] D. Benzinger, S. Ovinnikov, and M. Khammash, “Synthetic gene networks recapitulate dynamic signal decoding and differential gene expression,” *Cell Syst.*, p. S2405471222000825, Mar. 2022, doi: 10.1016/j.cels.2022.02.004.
[4] J. K. Jung *et al.*, “Cell-free biosensors for rapid detection of water contaminants,” *Nat. Biotechnol.*, vol. 38, no. 12, Art. no. 12, Dec. 2020, doi: 10.1038/s41587-020-0571-7.
[5] A. S. Karim *et al.*, “In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design,” *Nat. Chem. Biol.*, vol. 16, no. 8, pp. 912–919, Aug. 2020, doi: 10.1038/s41589-020-0559-0.
[6] F.-X. Lehr *et al.*, “Cell-Free Prototyping of AND-Logic Gates Based on Heterogeneous RNA Activators,” *ACS Synth. Biol.*, vol. 8, no. 9, pp. 2163–2173, Sep. 2019, doi: 10.1021/acssynbio.9b00238.
[7] S. J. Moore *et al.*, “Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria,” *Proc. Natl. Acad. Sci.*, vol. 115, no. 19, pp. E4340–E4349, May 2018, doi: 10.1073/pnas.1715806115.
[8] Z. Swank and S. J. Maerkl, “CFPU: A Cell-Free Processing Unit for High-Throughput, Automated In Vitro Circuit Characterization in Steady-State Conditions,” *BioDesign Res.*, vol. 2021, Mar. 2021, doi: 10.34133/2021/2968181.
[9] M. C. Bassalo, R. Liu, and R. T. Gill, “Directed evolution and synthetic biology applications to microbial systems,” *Curr. Opin. Biotechnol.*, vol. 39, pp. 126–133, Jun. 2016, doi: 10.1016/j.copbio.2016.03.016.
[10] J. A. N. Brophy and C. A. Voigt, “Principles of genetic circuit design,” *Nat. Methods*, vol. 11, no. 5, Art. no. 5, May 2014, doi: 10.1038/nmeth.2926.
[11] Z. Swank, N. Laohakunakorn, and S. J. Maerkl, “Cell-free gene-regulatory network engineering with synthetic transcription factors,” *Proc. Natl. Acad. Sci.*, vol. 116, no. 13, pp. 5892–5901, Mar. 2019, doi: 10.1073/pnas.1816591116.
[12] P.-F. Xia, H. Ling, J. L. Foo, and M. W. Chang, “Synthetic genetic circuits for programmable biological functionalities,” *Biotechnol. Adv.*, vol. 37, no. 6, p. 107393, Nov. 2019, doi: 10.1016/j.biotechadv.2019.04.015.
[13] U. Alon, “Network motifs: theory and experimental approaches,” *Nat. Rev. Genet.*, vol. 8, no. 6, pp. 450–461, Jun. 2007, doi: 10.1038/nrg2102.
[14] J. K. Jung *et al.*, “Cell-free biosensors for rapid detection of water contaminants,” *Nat. Biotechnol.*, vol. 38, no. 12, pp. 1451–1459, Dec. 2020, doi: 10.1038/s41587-020-0571-7.
[15] J. K. Jung, C. M. Archuleta, K. K. Alam, and J. B. Lucks, “Programming cell-free biosensors with DNA strand displacement circuits,” *Nat. Chem. Biol.*, vol. 18, no. 4, pp. 385–393, Apr. 2022, doi: 10.1038/s41589-021-00962-9.
[16] P. Q. Nguyen *et al.*, “Wearable materials with embedded synthetic biology sensors for biomolecule detection,” *Nat. Biotechnol.*, vol. 39, no. 11, pp. 1366–1374, Nov.

2021, doi: 10.1038/s41587-021-00950-3.
[17] K. Pardee *et al.*, “Portable, On-Demand Biomolecular Manufacturing,” *Cell*, vol. 167, no. 1, pp. 248-259.e12, Sep. 2016, doi: 10.1016/j.cell.2016.09.013.
[18] J. C. Stark *et al.*, “On-demand biomanufacturing of protective conjugate vaccines,” *Sci. Adv.*, vol. 7, no. 6, p. eabe9444, Feb. 2021, doi: 10.1126/sciadv.abe9444.
[19] C. Zhou, X. Lin, Y. Lu, and J. Zhang, “Flexible on-demand cell-free protein synthesis platform based on a tube-in-tube reactor,” *React. Chem. Eng.*, vol. 5, no. 2, pp. 270–277, 2020, doi: 10.1039/C9RE00394K.
[20] K. P. Adamala, D. A. Martin-Alarcon, K. R. Guthrie-Honea, and E. S. Boyden, “Engineering genetic circuit interactions within and between synthetic minimal cells,” *Nat. Chem.*, vol. 9, no. 5, pp. 431–439, May 2017, doi: 10.1038/nchem.2644.
[21] L. Aufinger, J. Brenner, and F. C. Simmel, “Complex dynamics in a synchronized cell-free genetic clock,” *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, May 2022, doi: 10.1038/s41467-022-30478-2.
[22] J. Garamella, D. Garenne, and V. Noireaux, “TXTL-based approach to synthetic cells,” *Methods Enzymol.*, vol. 617, pp. 217–239, 2019, doi: 10.1016/bs.mie.2018.12.015.
[23] E. Karzbrun, A. M. Tayar, V. Noireaux, and R. H. Bar-Ziv, “Synthetic biology. Programmable on-chip DNA compartments as artificial cells,” *Science*, vol. 345, no. 6198, pp. 829–832, Aug. 2014, doi: 10.1126/science.1255550.
[24] D. Banerjee *et al.*, “Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale,” *Nat. Commun.*, vol. 11, no. 1, p. 5385, Dec. 2020, doi: 10.1038/s41467-020-19171-4.
[25] H. Kim, D. Bojar, and M. Fussenegger, “A CRISPR/Cas9-based central processing unit to program complex logic computation in human cells,” *Proc. Natl. Acad. Sci.*, vol. 116, no. 15, pp. 7214–7219, Apr. 2019, doi: 10.1073/pnas.1821740116.
[26] J. Landberg, N. R. Wright, T. Wulff, M. J. Herrgård, and A. T. Nielsen, “CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in Escherichia coli,” *Biotechnol. Bioeng.*, vol. 117, no. 12, pp. 3835–3848, 2020, doi: 10.1002/bit.27536.
[27] A. C. Reis *et al.*, “Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays,” *Nat. Biotechnol.*, vol. 37, no. 11, pp. 1294–1301, Nov. 2019, doi: 10.1038/s41587-019-0286-9.
[28] J. Santos-Moreno and Y. Schaerli, “CRISPR-based gene expression control for synthetic gene circuits,” *Biochem. Soc. Trans.*, vol. 48, no. 5, pp. 1979–1993, Oct. 2020, doi: 10.1042/BST20200020.
[29] A. Westbrook *et al.*, “Distinct timescales of RNA regulators enable the construction of a genetic pulse generator,” *Biotechnol. Bioeng.*, vol. 116, no. 5, pp. 1139–1151, May 2019, doi: 10.1002/bit.26918.
[30] S. Clamons and R. Murray, “Modeling predicts that CRISPR-based activators, unlike CRISPR-based repressors, scale well with increasing gRNA competition and dCas9 bottlenecking,” Synthetic Biology, preprint, Jul. 2019. doi: 10.1101/719278.
[31] J. Nielsen and J. D. Keasling, “Engineering Cellular Metabolism,” *Cell*, vol. 164, no. 6, pp. 1185–1197, Mar. 2016, doi: 10.1016/j.cell.2016.02.004.
[32] B. I. Tickman *et al.*, “Multi-layer CRISPRa/i circuits for dynamic genetic programs in cell-free and bacterial systems,” *Cell Syst.*, p. S2405471221004191, Nov. 2021,

doi: 10.1016/j.cels.2021.10.008.
[33] M. W. Gander, J. D. Vrana, W. E. Voje, J. M. Carothers, and E. Klavins, “Digital logic circuits in yeast with CRISPR-dCas9 NOR gates,” *Nat. Commun.*, vol. 8, no. 1, Art. no. 1, May 2017, doi: 10.1038/ncomms15459.
[34] R. Marshall and V. Noireaux, “Quantitative modeling of transcription and translation of an all- E. coli cell-free system,” *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Aug. 2019, doi: 10.1038/s41598-019-48468-8.
[35] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, “Tuning genetic control through promoter engineering,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 36, pp. 12678–12683, Sep. 2005, doi: 10.1073/pnas.0504604102.
[36] J. R. Kelly *et al.*, “Measuring the activity of BioBrick promoters using an in vivo reference standard,” *J. Biol. Eng.*, vol. 3, p. 4, Mar. 2009, doi: 10.1186/1754-1611-3-4.
[37] R. C. Brewster, D. L. Jones, and R. Phillips, “Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli,” *PLOS Comput. Biol.*, vol. 8, no. 12, p. e1002811, Dec. 2012, doi: 10.1371/journal.pcbi.1002811.
[38] T. L. Fleur, A. Hossain, and H. M. Salis, “Automated Model-Predictive Design of Synthetic Promoters to Control Transcriptional Profiles in Bacteria,” Synthetic Biology, preprint, Sep. 2021. doi: 10.1101/2021.09.01.458561.
[39] D. J. Lee, S. D. Minchin, and S. J. W. Busby, “Activating Transcription in Bacteria,” *Annu. Rev. Microbiol.*, vol. 66, no. 1, pp. 125–152, Oct. 2012, doi: 10.1146/annurev-micro-092611-150012.
[40] T. C. Yu *et al.*, “Multiplexed characterization of rationally designed promoter architectures deconstructs combinatorial logic for IPTG-inducible systems,” *Nat. Commun.*, vol. 12, no. 1, p. 325, Jan. 2021, doi: 10.1038/s41467-020-20094-3.
[41] B. F. Cress *et al.*, “Rapid generation of CRISPR/dCas9-regulated, orthogonally repressible hybrid T7-lac promoters for modular, tuneable control of metabolic pathway fluxes in Escherichia coli,” *Nucleic Acids Res.*, vol. 44, no. 9, pp. 4472–4485, May 2016, doi: 10.1093/nar/gkw231.
[42] T. M. Groseclose, R. E. Rondon, Z. D. Herde, C. A. Aldrete, and C. J. Wilson, “Engineered systems of inducible anti-repressors for the next generation of biological programming,” *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Sep. 2020, doi: 10.1038/s41467-020-18302-1.
[43] X. Liu *et al.*, “De novo design of programmable inducible promoters,” *Nucleic Acids Res.*, vol. 47, no. 19, pp. 10452–10463, Nov. 2019, doi: 10.1093/nar/gkz772.
[44] I. J. Roney, A. D. Rudner, J.-F. Couture, and M. Kærn, “Improvement of the reverse tetracycline transactivator by single amino acid substitutions that reduce leaky target gene expression to undetectable levels,” *Sci. Rep.*, vol. 6, p. 27697, Jun. 2016, doi: 10.1038/srep27697.
[45] J. Fontana *et al.*, “Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements,” *Nat. Commun.*, vol. 11, no. 1, p. 1618, Dec. 2020, doi: 10.1038/s41467-020-15454-y.
[46] Y. Chen *et al.*, “Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors,” *Nat. Commun.*, vol. 9, no. 1, p. 64, Dec. 2018, doi: 10.1038/s41467-017-02473-5.
[47] W. Ross, S. E. Aiyar, J. Salomon, and R. L. Gourse, “Escherichia coli promoters

with UP elements of different strengths: modular structure of bacterial promoters," *J. Bacteriol.*, vol. 180, no. 20, pp. 5375–5383, Oct. 1998, doi: 10.1128/JB.180.20.5375-5383.1998.
[48] R. K. Shultzaberger, Z. Chen, K. A. Lewis, and T. D. Schneider, "Anatomy of Escherichia coli σ 70 promoters," *Nucleic Acids Res.*, vol. 35, no. 3, pp. 771–788, Feb. 2007, doi: 10.1093/nar/gkl956.
[49] C. Dong, J. Fontana, A. Patel, J. M. Carothers, and J. G. Zalatan, "Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria," *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Jun. 2018, doi: 10.1038/s41467-018-04901-6.
[50] I. G. Hook-Barnard and D. M. Hinton, "Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters," *Gene Regul. Syst. Biol.*, vol. 1, pp. 275–293, Dec. 2007.
[51] M. S. Paget and J. D. Helmann, "The σ70family of sigma factors," *Genome Biol.*, vol. 4, no. 1, p. 203, Jan. 2003, doi: 10.1186/gb-2003-4-1-203.
[52] L. S. Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression," *Cell*, vol. 152, no. 5, pp. 1173–1183, Feb. 2013, doi: 10.1016/j.cell.2013.02.022.
[53] Y. Censor, "Pareto optimality in multiobjective problems," *Appl. Math. Optim.*, vol. 4, no. 1, pp. 41–59, Mar. 1977, doi: 10.1007/BF01442131.
[54] S. T. Estrem, T. Gaal, W. Ross, and R. L. Gourse, "Identification of an UP element consensus sequence for bacterial promoters," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 17, pp. 9761–9766, Aug. 1998.
[55] T. Gaal *et al.*, "DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture," *Genes Dev.*, vol. 10, no. 1, pp. 16–26, Jan. 1996, doi: 10.1101/gad.10.1.16.
[56] J. Fontana, D. Sparkman-Yager, J. G. Zalatan, and J. M. Carothers, "Challenges and opportunities with CRISPR activation in bacteria for data-driven metabolic engineering," *Curr. Opin. Biotechnol.*, vol. 64, pp. 190–198, Aug. 2020, doi: 10.1016/j.copbio.2020.04.005.
[57] J. Lian, M. HamediRad, S. Hu, and H. Zhao, "Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system," *Nat. Commun.*, vol. 8, no. 1, p. 1688, Nov. 2017, doi: 10.1038/s41467-017-01695-x.
[58] H. Bolouri and E. H. Davidson, "Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics," *Proc. Natl. Acad. Sci.*, vol. 100, no. 16, pp. 9371–9376, Aug. 2003, doi: 10.1073/pnas.1533293100.
[59] F. Banuett, "Signalling in the Yeasts: An Informational Cascade with Links to the Filamentous Fungi," *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 249–274, Jun. 1998, doi: 10.1128/MMBR.62.2.249-274.1998.
[60] P.-Y. Chen, Y. Qian, and D. Del Vecchio, "A Model for Resource Competition in CRISPR-Mediated Gene Repression," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec. 2018, pp. 4333–4338. doi: 10.1109/CDC.2018.8619016.
[61] D. Camsund, A. Jaramillo, and P. Lindblad, "Engineering of a Promoter Repressed by a Light-Regulated Transcription Factor in Escherichia coli," *BioDesign Res.*, vol. 2021, Sep. 2021, doi: 10.34133/2021/9857418.
[62] P. Jayaraman, K. Devarajan, T. K. Chua, H. Zhang, E. Gunawan, and C. L. Poh,

"Blue light-mediated transcriptional activation and repression of gene expression in bacteria," *Nucleic Acids Res.*, vol. 44, no. 14, pp. 6994–7005, Aug. 2016, doi: 10.1093/nar/gkw548.
[63] L. B. Motta-Mena *et al.*, "An optogenetic gene expression system with rapid activation and deactivation kinetics," *Nat. Chem. Biol.*, vol. 10, no. 3, pp. 196–202, Mar. 2014, doi: 10.1038/nchembio.1430.
[64] B. D. Zoltowski, L. B. Motta-Mena, and K. H. Gardner, "Blue light-induced dimerization of a bacterial LOV-HTH DNA-binding protein," *Biochemistry*, vol. 52, no. 38, pp. 6653–6661, Sep. 2013, doi: 10.1021/bi401040m.
[65] Y. Gao, X. Xiong, S. Wong, E. J. Charles, W. A. Lim, and L. S. Qi, "Complex transcriptional modulation with orthogonal and inducible dCas9 regulators," *Nat. Methods*, vol. 13, no. 12, pp. 1043–1049, Dec. 2016, doi: 10.1038/nmeth.4042.
[66] P. K. Jain *et al.*, "Development of Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors," *Angew. Chem. Int. Ed.*, vol. 55, no. 40, pp. 12440–12444, 2016, doi: 10.1002/anie.201606123.
[67] K. Kundert *et al.*, "Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs," *Nat. Commun.*, vol. 10, no. 1, Art. no. 1, May 2019, doi: 10.1038/s41467-019-09985-2.
[68] Y. Nihongaki, F. Kawano, T. Nakajima, and M. Sato, "Photoactivatable CRISPR-Cas9 for optogenetic genome editing," *Nat. Biotechnol.*, vol. 33, no. 7, pp. 755–760, Jul. 2015, doi: 10.1038/nbt.3245.
[69] L. R. Polstein and C. A. Gersbach, "A light-inducible CRISPR-Cas9 system for control of endogenous gene activation," *Nat. Chem. Biol.*, vol. 11, no. 3, pp. 198–200, Mar. 2015, doi: 10.1038/nchembio.1753.
[70] Z.-M. Ying, F. Wang, X. Chu, R.-Q. Yu, and J.-H. Jiang, "Activatable CRISPR Transcriptional Circuits Generate Functional RNA for mRNA Sensing and Silencing," *Angew. Chem. Int. Ed.*, vol. 59, no. 42, pp. 18599–18604, 2020, doi: 10.1002/anie.202004751.
[71] Y. Yu *et al.*, "Engineering a far-red light–activated split-Cas9 system for remote-controlled genome editing of internal organs and tumors," *Sci. Adv.*, vol. 6, no. 28, p. eabb1777, Jul. 2020, doi: 10.1126/sciadv.abb1777.
[72] M. C. Villegas Kcam, A. J. Tsong, and J. Chappell, "Rational engineering of a modular bacterial CRISPR–Cas activation platform with expanded target range," *Nucleic Acids Res.*, vol. 49, no. 8, pp. 4793–4802, May 2021, doi: 10.1093/nar/gkab211.
[73] K. E. Thompson, C. J. Bashor, W. A. Lim, and A. E. Keating, "SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains," *ACS Synth. Biol.*, vol. 1, no. 4, pp. 118–129, Apr. 2012, doi: 10.1021/sb200015u.
[74] D. Cunningham-Bryant, J. Sun, B. Fernandez, and J. G. Zalatan, "CRISPR–Cas-Mediated Chemical Control of Transcriptional Dynamics in Yeast," *ChemBioChem*, vol. 20, no. 12, pp. 1519–1523, 2019, doi: 10.1002/cbic.201800823.
[75] F.-S. Liang, W. Q. Ho, and G. R. Crabtree, "Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity," *Sci. Signal.*, vol. 4, no. 164, p. rs2, Mar. 2011, doi: 10.1126/scisignal.2001449.

[76] T. Miyamoto *et al.*, “Rapid and Orthogonal Logic Gating with a Gibberellin-induced Dimerization System,” *Nat. Chem. Biol.*, vol. 8, no. 5, pp. 465–470, Mar. 2012, doi: 10.1038/nchembio.922.
[77] J. G. Zalatan *et al.*, “Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds,” *Cell*, vol. 160, no. 1, pp. 339–350, Jan. 2015, doi: 10.1016/j.cell.2014.11.052.
[78] E. F. Jr. Douglass, C. J. Miller, G. Sparer, H. Shapiro, and D. A. Spiegel, “A Comprehensive Mathematical Model for Three-Body Binding Equilibria,” *J. Am. Chem. Soc.*, vol. 135, no. 16, pp. 6092–6099, Apr. 2013, doi: 10.1021/ja311795d.
[79] A. Levchenko, J. Bruck, and P. W. Sternberg, “Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 11, pp. 5818–5823, May 2000, doi: 10.1073/pnas.97.11.5818.
[80] B. Wang, M. Barahona, and M. Buck, “Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks,” *Nucleic Acids Res.*, vol. 42, no. 14, pp. 9484–9492, Aug. 2014, doi: 10.1093/nar/gku593.
[81] T. S. Lee *et al.*, “BglBrick vectors and datasheets: A synthetic biology platform for gene expression,” *J. Biol. Eng.*, vol. 5, no. 1, p. 12, Dec. 2011, doi: 10.1186/1754-1611-5-12.
[82] J. Garamella, R. Marshall, M. Rustad, and V. Noireaux, “The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology,” *ACS Synth. Biol.*, vol. 5, no. 4, pp. 344–355, Apr. 2016, doi: 10.1021/acssynbio.5b00296.
[83] O. Borkowski, C. Bricio, M. Murgiano, B. Rothschild-Mancinelli, G.-B. Stan, and T. Ellis, “Cell-free prediction of protein expression costs for growing cells,” *Nat. Commun.*, vol. 9, no. 1, p. 1457, Apr. 2018, doi: 10.1038/s41467-018-03970-x.
[84] D. A. Specht, L. B. Cortes, and G. Lambert, “Overcoming Leak Sensitivity in CRISPRi Circuits Using Antisense RNA Sequestration and Regulatory Feedback,” *ACS Synth. Biol.*, Aug. 2022, doi: 10.1021/acssynbio.2c00155.
[85] S. Kaplan, A. Bren, A. Zaslaver, E. Dekel, and U. Alon, “Diverse Two-Dimensional Input Functions Control Bacterial Sugar Genes,” *Mol. Cell*, vol. 29, no. 6, pp. 786–792, Mar. 2008, doi: 10.1016/j.molcel.2008.01.021.
[86] S. Krishna, L. Orosz, K. Sneppen, S. Adhya, and S. Semsey, “Relation of Intracellular Signal Levels and Promoter Activities in the gal Regulon of Escherichia coli,” *J. Mol. Biol.*, vol. 391, no. 4, pp. 671–678, Aug. 2009, doi: 10.1016/j.jmb.2009.06.043.
[87] H. R. Kempton, L. E. Goudy, K. S. Love, and L. S. Qi, “Multiple Input Sensing and Signal Integration Using a Split Cas12a System,” *Mol. Cell*, vol. 78, no. 1, pp. 184-191.e3, Apr. 2020, doi: 10.1016/j.molcel.2020.01.016.
[88] J. Beltrán *et al.*, “Rapid biosensor development using plant hormone receptors as reprogrammable scaffolds,” *Nat. Biotechnol.*, pp. 1–7, Jun. 2022, doi: 10.1038/s41587-022-01364-5.
[89] G. W. Foight *et al.*, “Multi-input chemical control of protein dimerization for programming graded cellular responses,” *Nat. Biotechnol.*, vol. 37, no. 10, Art. no. 10, Oct. 2019, doi: 10.1038/s41587-019-0242-8.
[90] G. Guntas *et al.*, “Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins,” *Proc. Natl. Acad. Sci.*, vol. 112,

no. 1, pp. 112–117, Jan. 2015, doi: 10.1073/pnas.1417910112.
[91] Z. Huang *et al.*, “Creating Red Light-Switchable Protein Dimerization Systems as Genetically Encoded Actuators with High Specificity,” *ACS Synth. Biol.*, vol. 9, no. 12, pp. 3322–3333, Dec. 2020, doi: 10.1021/acssynbio.0c00397.
[92] D. I. Piraner, M. H. Abedi, B. A. Moser, A. Lee-Gosselin, and M. G. Shapiro, “Tunable thermal bioswitches for in vivo control of microbial therapeutics,” *Nat. Chem. Biol.*, vol. 13, no. 1, pp. 75–80, Jan. 2017, doi: 10.1038/nchembio.2233.
[93] J. Carey, V. Cameron, P. L. De Haseth, and O. C. Uhlenbeck, “Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site,” *Biochemistry*, vol. 22, no. 11, pp. 2601–2610, May 1983, doi: 10.1021/bi00280a002.
[94] F. Dupeux *et al.*, “A thermodynamic switch modulates abscisic acid receptor sensitivity,” *EMBO J.*, vol. 30, no. 20, pp. 4171–4184, Oct. 2011, doi: 10.1038/emboj.2011.294.
[95] K. Miyazono *et al.*, “Structural basis of abscisic acid signalling,” *Nature*, vol. 462, no. 7273, Art. no. 7273, Dec. 2009, doi: 10.1038/nature08583.
[96] M. Ueguchi-Tanaka *et al.*, “GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin,” *Nature*, vol. 437, no. 7059, Art. no. 7059, Sep. 2005, doi: 10.1038/nature04028.
[97] H. Yoshida *et al.*, “Evolution and diversification of the plant gibberellin receptor GID1,” *Proc. Natl. Acad. Sci.*, vol. 115, no. 33, pp. E7844–E7853, Aug. 2018, doi: 10.1073/pnas.1806040115.
[98] W. S. Grubbe, B. J. Rasor, A. Krüger, M. C. Jewett, and A. S. Karim, “Cell-free styrene biosynthesis at high titers,” *Metab. Eng.*, vol. 61, pp. 89–95, Sep. 2020, doi: 10.1016/j.ymben.2020.05.009.
[99] N. S. Kruyer *et al.*, “Membrane Augmented Cell-Free Systems: A New Frontier in Biotechnology,” *ACS Synth. Biol.*, vol. 10, no. 4, pp. 670–681, Apr. 2021, doi: 10.1021/acssynbio.0c00625.
[100] J. A. Peruzzi, N. R. Galvez, and N. P. Kamat, “Engineering transmembrane signal transduction in synthetic membranes using two-component systems,” Synthetic Biology, preprint, Oct. 2022. doi: 10.1101/2022.10.30.514420.

Supplementary Figures

Supplementary Tables

Supplementary Methods

Supplementary Figures

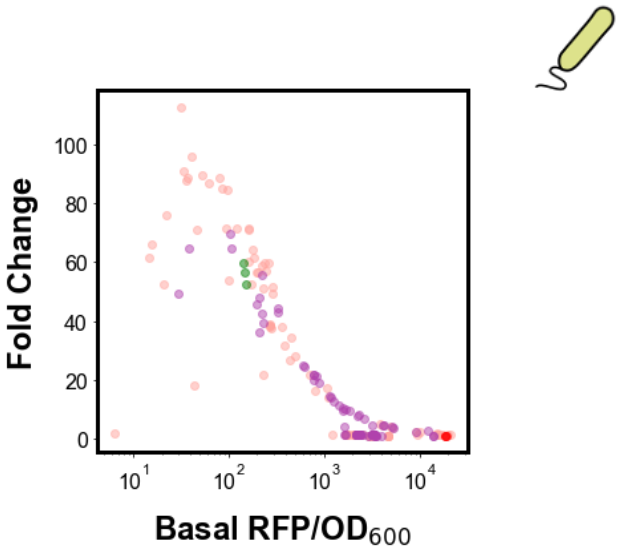


Figure S1: Minimal promoter effect on fold-change

Basal RFP/OD_{600} and fold change for the two minimal promoter libraries (n_{MP1} = 89, n_{MP2} = 84). The J23117 minimal promoter (green, triplicates) is included as a standard reference for CRISPRa efficiency. The J23119 minimal promoter (red, triplicates) is an example of a non-activatable promoter due to high basal expression levels.

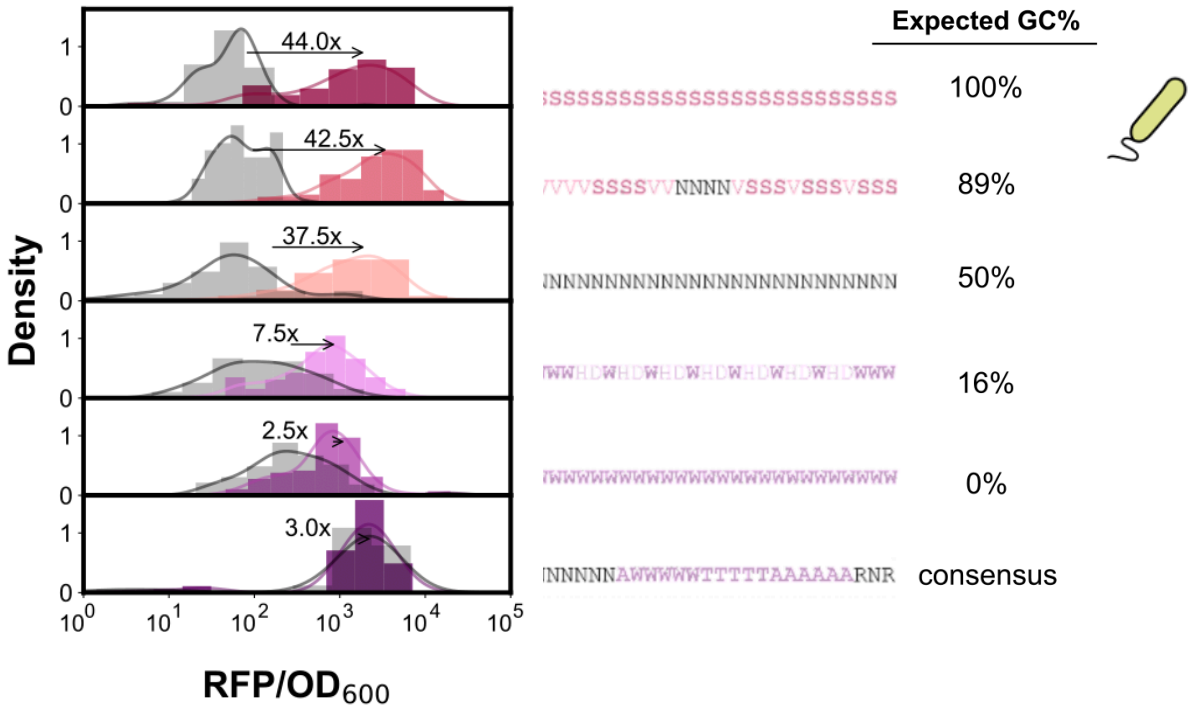


Figure S2: UP-element libraries RFP distributions

Histograms and probability density functions for basal expression (gray) and activated expression (colored) for six UP-element libraries with increasing GC-content. Median fold change is calculated as the ratio of median activation and media basal expression levels.

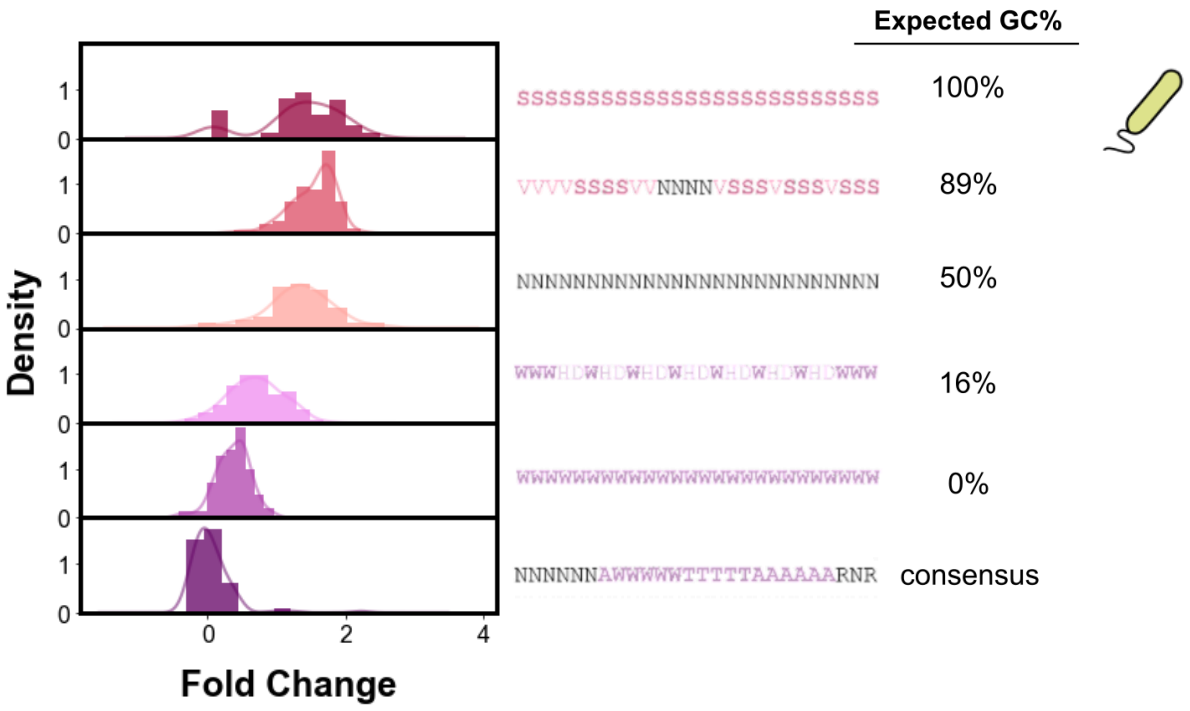


Figure S3: UP-element libraries fold-change distributions

Histograms and probability density functions for ford change for six UP-element libraries with increasing GC-content.

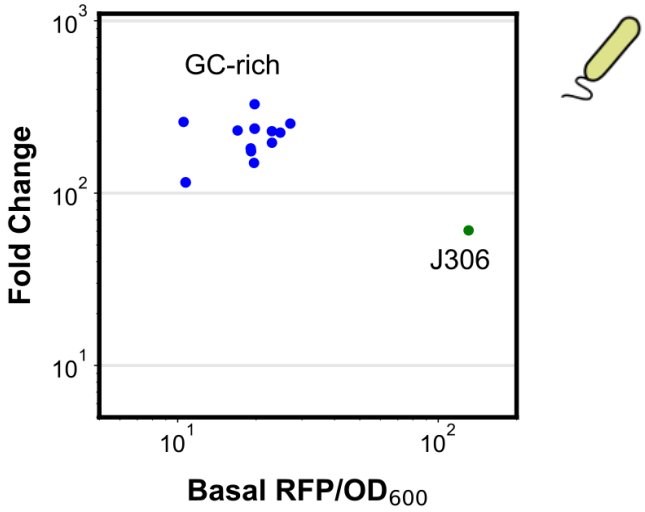


Figure S4: CRISPRa activity validation of high GC-content scRNAs

Fold change upon aTc induction and basal expression. scRNA target site sequences were initially selected based on low expression leak in *E. coli* and the corresponding scRNAs were constructed for use in CFS. Selected scRNAs were benchmarked against the standard J306 scRNA (green).

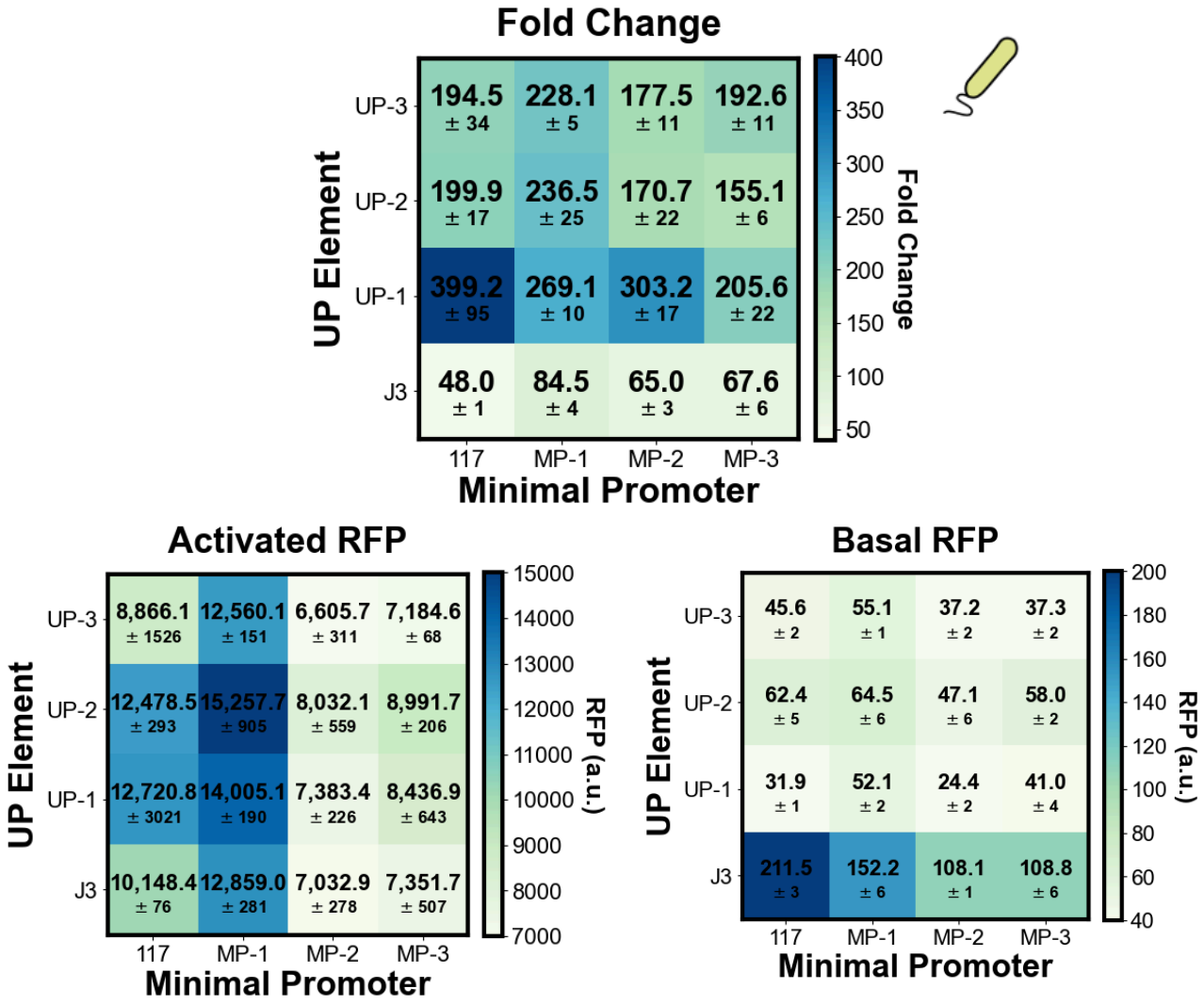


Figure S5: Combinatorial construction of activatable promoters

In addition to the J3.J23117 benchmark, three high performing variants each from the UP element and minimal promoter libraries were tested in a combinatorial manner for a total of 16 UP element/minimal promoter combinations screened in *E. coli*. We quantified the basal and activation expression of the 16 promoters with the same scRNA (bottom). Activation ratio is calculated by dividing the activated RFP expression from the inducible CRISPRa system by the basal RFP expression from each promoter (top). Values represent the mean ± standard deviation of three technical replicates.

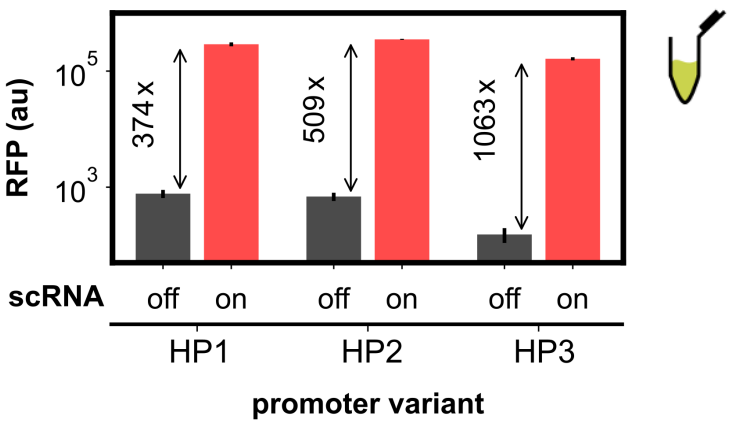


Figure S6: Activatable promoter characterization in CFS

Characterization of selected promoter variants in CFS. CRISPRa-mediated RFP expression levels (red, 0.4 nM scRNA DNA) and RFP basal expression levels (black, 0 nM scRNA DNA). Reactions contain 10 nM of RFP plasmid. Values represent the mean ± standard deviation of three technical replicates.

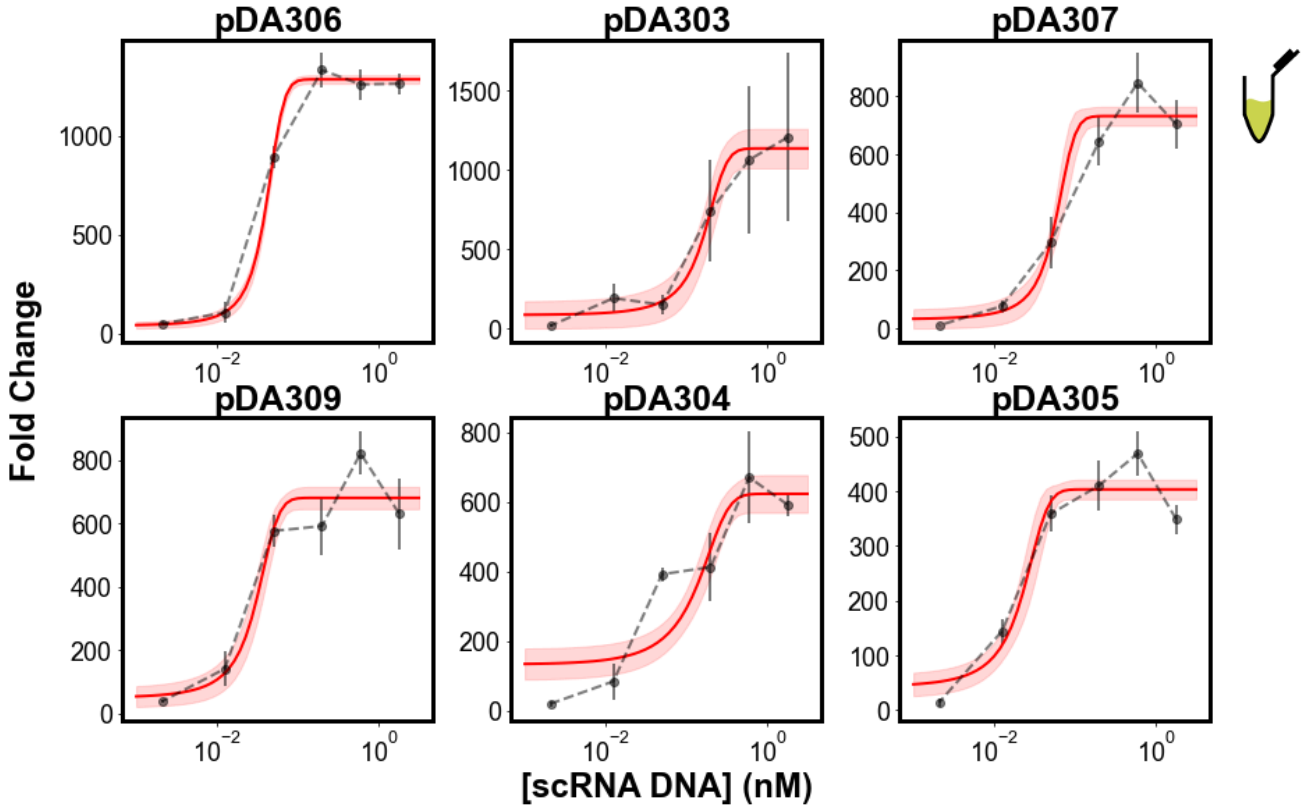


Figure S7: scRNA dose-response characterizations

scRNA dose-response curves are shown for orthogonal promoter-scRNA pairs in CFS. The scRNA-dose response curve is characterized through titrating the amount of scRNA DNA added to the CFS reaction. Reactions contain 10 nM of RFP plasmid. Red line indicates a logistic fit to the data. Values represent the mean ± standard deviation of three technical replicates.

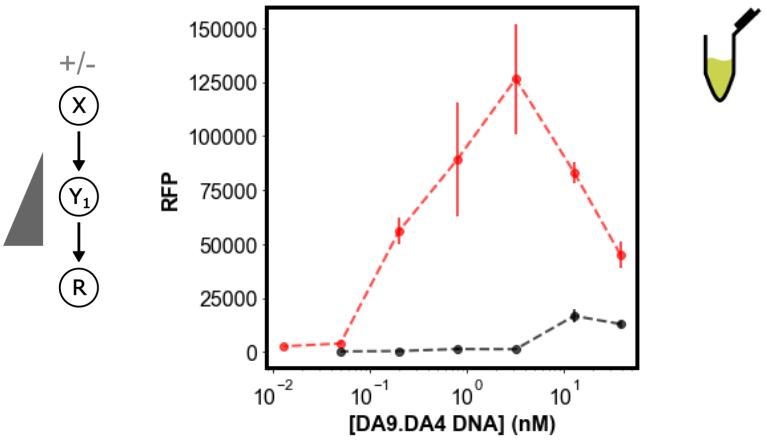


Figure S8: Titration of middle node in two-layer activation cascade

Two-layer activation cascade with high-performing components to identify the best performing internal node concentration. **Left:** Circuit schematic for measuring output RFP and fold change as a function of input scRNA. **Right:** Cascade RFP output with scRNA input (15 pM, red) and without (0 pM, black). Output node concentration is held constant at 10 nM. Values represent the mean ± standard deviation of three technical replicates.

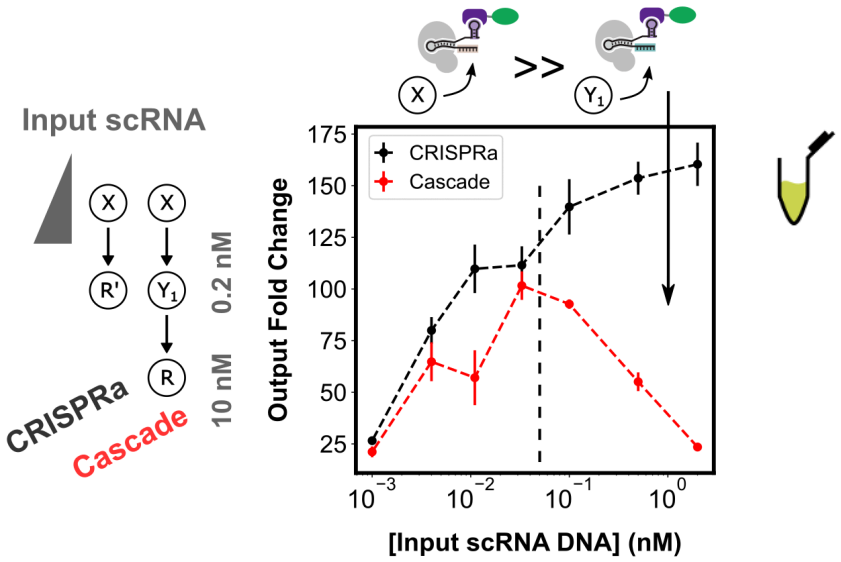


Figure S9: Signal propagation in a two-layer activation cascade

gRNA competition impact on circuit function. Left: Circuit schematic for measuring output fold change as a function of input scRNA for both CRISPRa (black) and CRISPRa cascade (red). Internal node concentration and output node concentration are held constant at 0.2 nM and 10 nM, respectively. Right: Input scRNA plasmid concentration was titrated between 1 pM and 2 nM. Black dashed line indicates saturation of CRISPRa complexes with input scRNA. Values represent the mean ± standard deviation of three technical replicates.

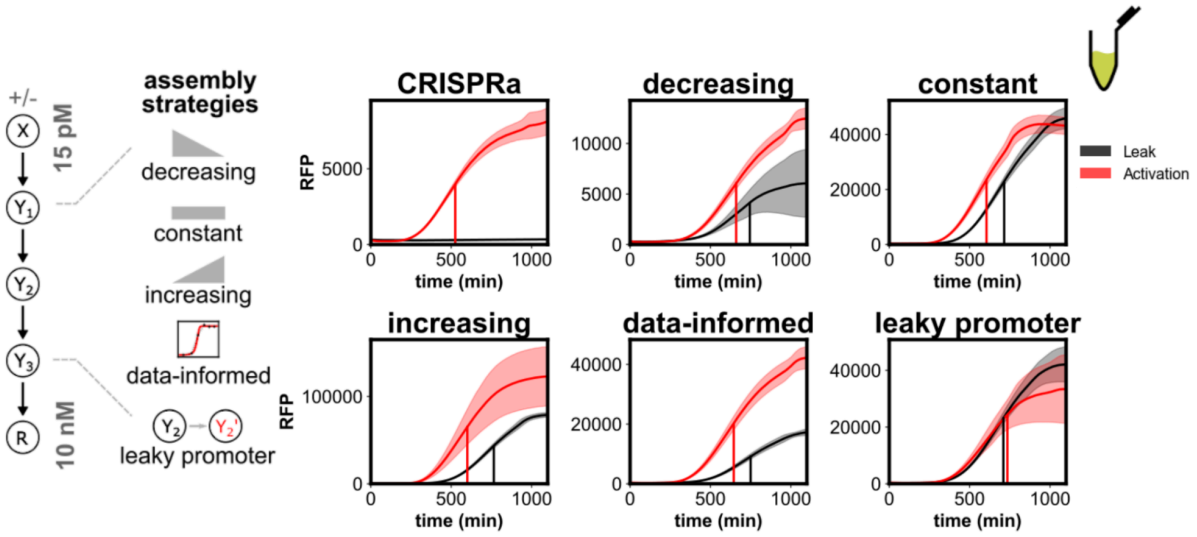


Figure S10: Time course for four-layer activation cascade assembly strategies

Comparison of the dynamics of four-layer CRISPRa cascade assemblies. **Left:** Internal node concentrations either decreased from 200 pM to 32 pM as depth increased, were held constant at 200 pM, or increased from 200 pM to 1.25 nM as depth increased. A fourth assembly method was tested in which internal node concentrations were 40, 200, and 170 pM, based on individual scRNA-dose response characteristics. A fifth cascade was included in which the high-performing promoter of the second internal node was replaced with the leaky J2 promoter. Input and output node concentrations were held constant across all strategies at 0 or 15 pM and 10 nM, respectively. **Right:** Output RFP expression for each assembly strategy with scRNA input (red) and without (black). Values represent the mean ± standard deviation of three technical replicates. Time to maximum expression rate (t_{max}) for each assembly strategy is calculated by finding the time to reach maximum RFP production rate between (Methods 7.1).

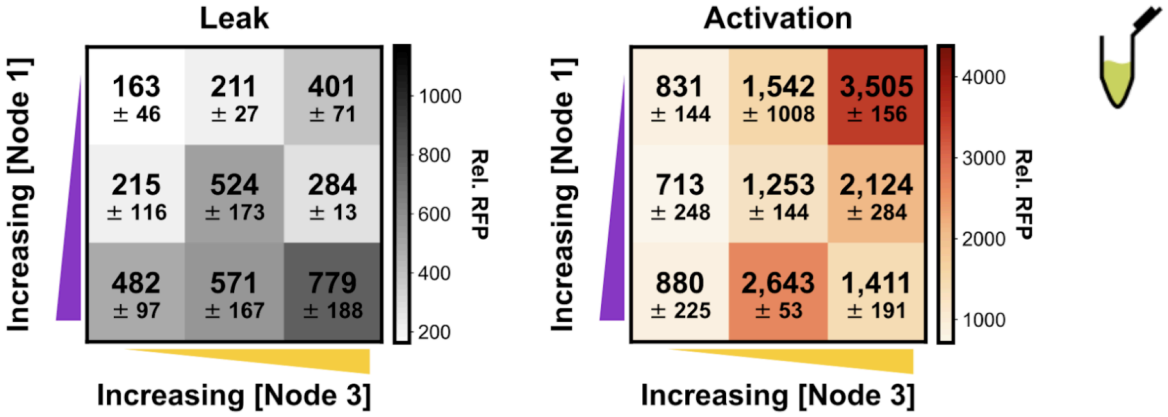


Figure S11: Four-layer activation cascade basal and activated RFP expression

Left: Basal expression levels for cascades titrating the first and third layers between 40 and 160 pM, and 85 and 340 pM, respectively. **Right:** Activated expression levels for the same cascades. The input node, second internal node, and output node were held constant at 0 or 15 pM, 0.2 nM, and 10 nM, respectively. Values are not background subtracted. Values represent the mean ± standard deviation of three technical replicates.

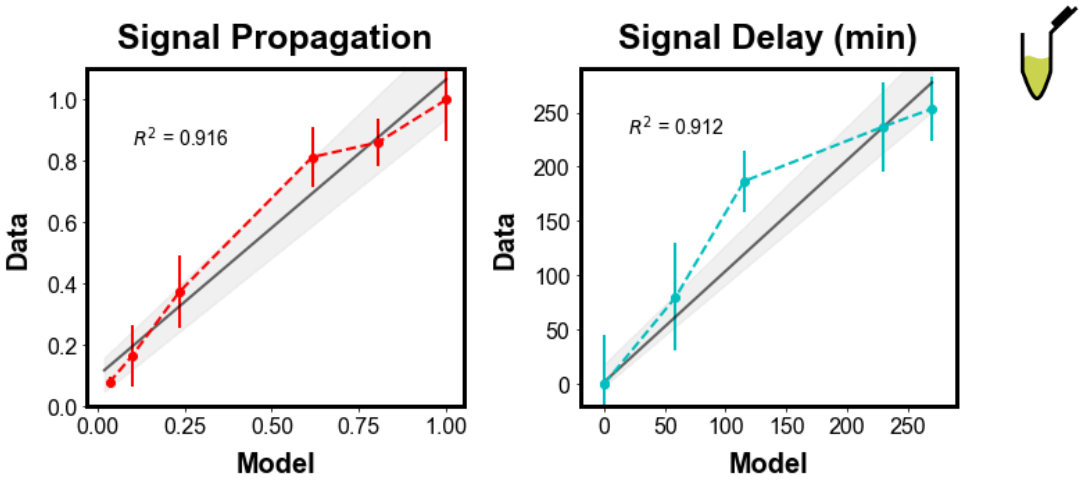


Figure S12: Signal propagation and signal delay model accuracy

Comparisons of measured and predicted signal propagation (left) and signal delay (right) for activation cascades of different depths. Signal propagation is calculated by dividing the fold-activation of the cascade output by the fold-activation from the input layer. (Methods 7.2). Signal delay is calculated as the difference between the cascade output and input layer in time to reach the maximum fold-activation (Methods 7.2). Both measures are presented as the mean ± standard deviation of three technical replicates. The fold-change from the individual promoters' dose-response curves (Figure S7) are used to iteratively predict the delay and signal propagation at the next layer (Methods S4). Black line represents the mean ± standard deviation of linear regression model measuring goodness-of-fit between model predictions and experimental data.

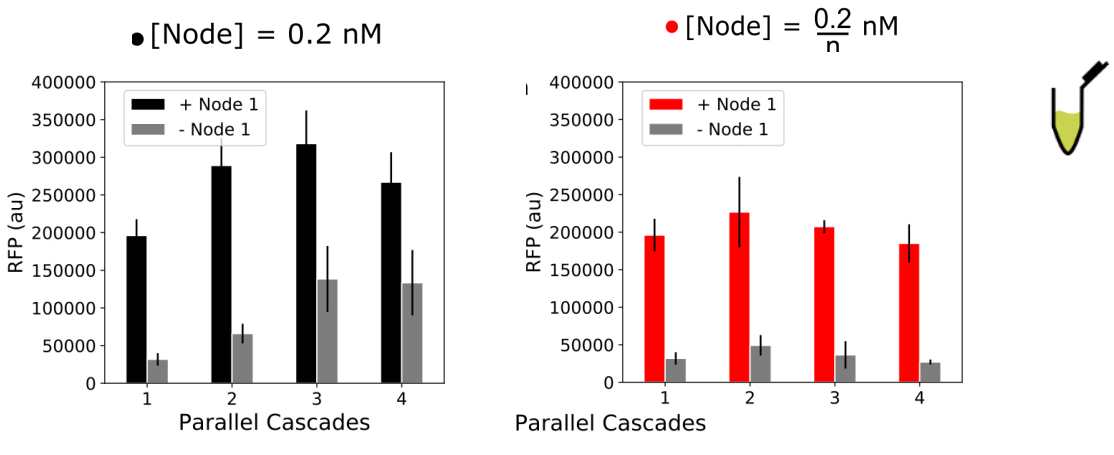


Figure S13: Wide circuit basal and activated RFP expression

Up to four parallel three-layer cascades were constructed. **Left:** The concentration of each internal node was held at 0.2 nM as circuit width increased. **Right:** The internal node concentration is scaled down proportionally to the width of the circuit, such that each internal node concentration is 0.2/n nM, where n is the number of parallel cascades.

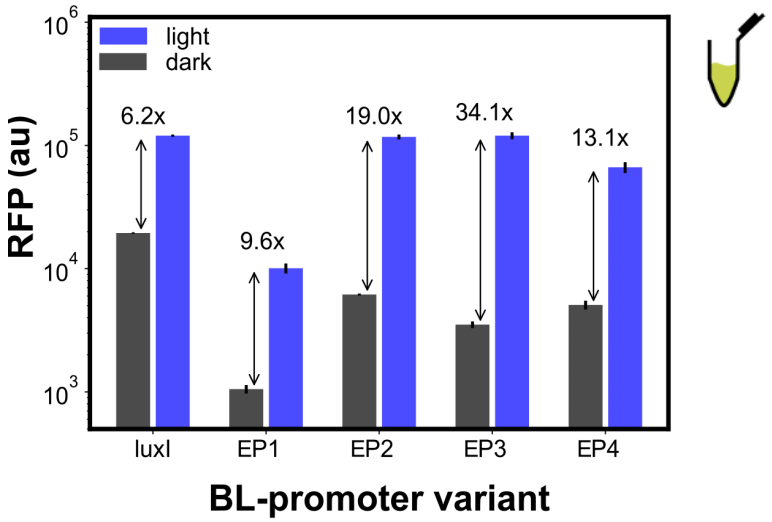


Figure S14: Blue-light promoter characterization in CFS

Characterization of selected promoter variants in CFS. Reactions contain 8 nM and 10 nM of EL222 and RFP plasmids respectively. EL222-mediated RFP expression levels (blue) and RFP basal expression levels (black). Values represent the mean ± standard deviation of three technical replicates.

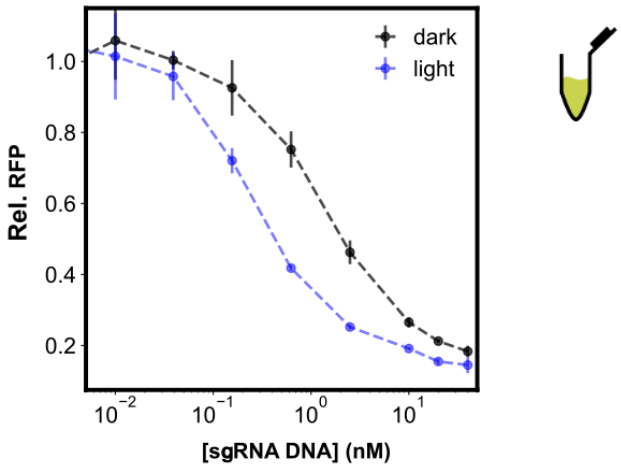


Figure S15: Blue-light CRISPRi

Titration of blue-light inducible sgRNA plasmid concentration to maximize the fold repression between blue-light dependent CRISPRi (blue) and CRISPRi due to sgRNA leak in the dark (black). Reactions contain 8 nM and 1 nM of EL222 and RFP plasmids respectively. Values represent the mean ± standard deviation of three technical replicates.

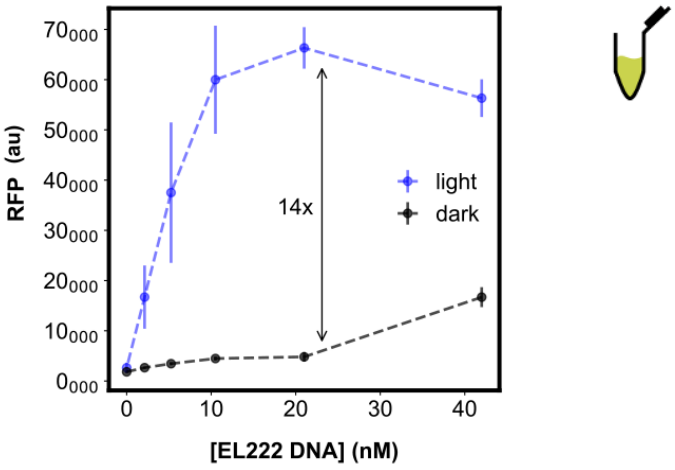


Figure S16: EL222 titration in blue-light CRISPRa

Titration of EL222 plasmid concentration to maximize the fold change between blue-light dependent CRISPRa (blue) and CRISPRa due to scRNA leak in the dark (black). Reactions contain 10 nM RFP plasmid. Values represent the mean ± standard deviation of three technical replicates.

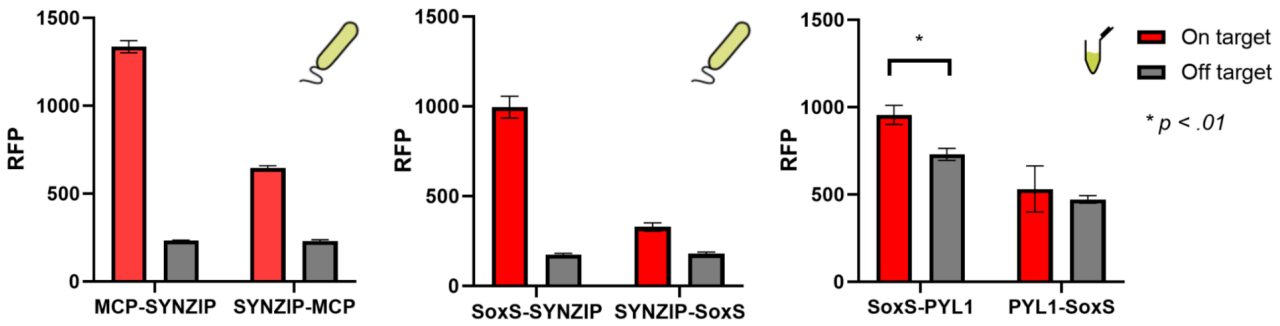


Figure S17: Fusion orientation preference for SYNZIP and ABI/PYL1

Left, Middle: MCP and SoxS fusion orientations were tested for the SYNZIP-CRISPRa system in *E. coli* using the J306 spacer at -81 bp from the TSS. The MCP test was done using SoxS-SYNZIP and the SoxS test was done using MCP-SYNZIP. **Right:** SoxS fusions were tested for the abscisic acid (ABA) CRISPRa system using MCP-ABI. The ABA constructs were tested in CFS using the R206 spacer at -81 bp from the TSS. ABA components were expressed at 5 nM. Off-target represents reactions containing a scRNA with no cognate target. Values represent the mean ± standard deviation of three technical replicates.

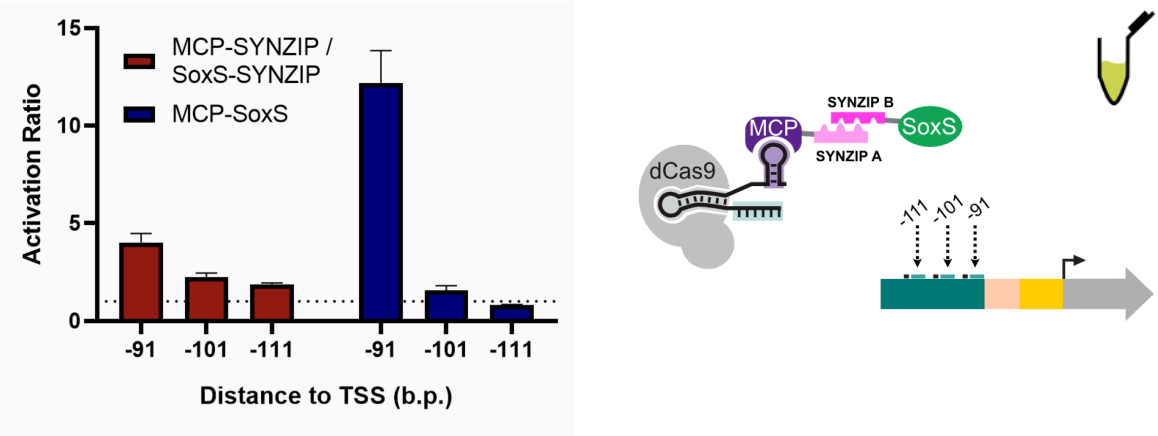


Figure S18: Dependence of SYNZIP-CRISPRa on distance to TSS

SYNZIP-CRISPRa and CRISPRa were tested at various target sites with increasing distance from the TSS in 10 bp intervals using a CRISPRa promoter with densely packed scRNA target sites. Plasmids expressing SYNZIP components are added at 5 nM each. Values represent the mean ± standard deviation of three technical replicates.

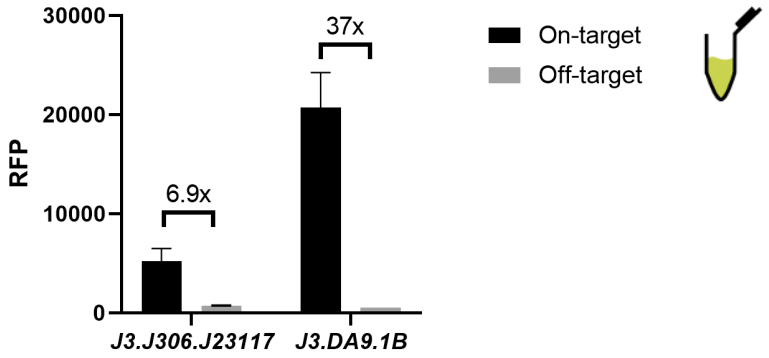


Figure S19: Improvements in SYNZIP-CRISPRa from engineered promoters

Comparison of SYNZIP-CRISPRa scRNA-dependent fold change with the previous synthetic promoter used to survey target sites and an engineered high dynamic range promoter. Off-target represents reactions containing a scRNA with no cognate target. In each reaction, the concentration of reporter DNA was 10 nM. SYNZIP components are added at 5 nM each. Reactions are background subtracted from a cell-free reaction containing no DNA. Values represent the mean ± standard deviation of three technical replicates.

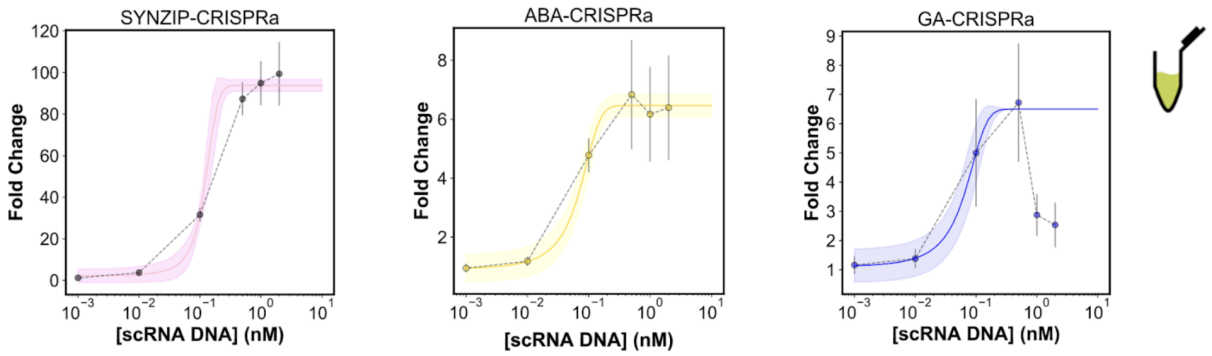


Figure S20: Conditional CRISPRa scRNA dose-response

scRNA-dose response curves were collected for conditional CRISPRa systems. scRNA concentrations were titrated between 10^{-3} and 10^{1} nM. Fold activation was calculated relative to the no scRNA condition. For all conditions, ABA is added at 10 μM and GA is added at 10^{3} μM. SYNZIP-CRISPRa components were both added at 5 nM, MCP-ABI and SoxS-PYL1 were added at 5 and 10 nM respectively, and GA-CRISPRa components were both added at 10 nM. Colored lines indicate a logistic fit to the data. Values represent the mean ± standard deviation of three technical replicates.

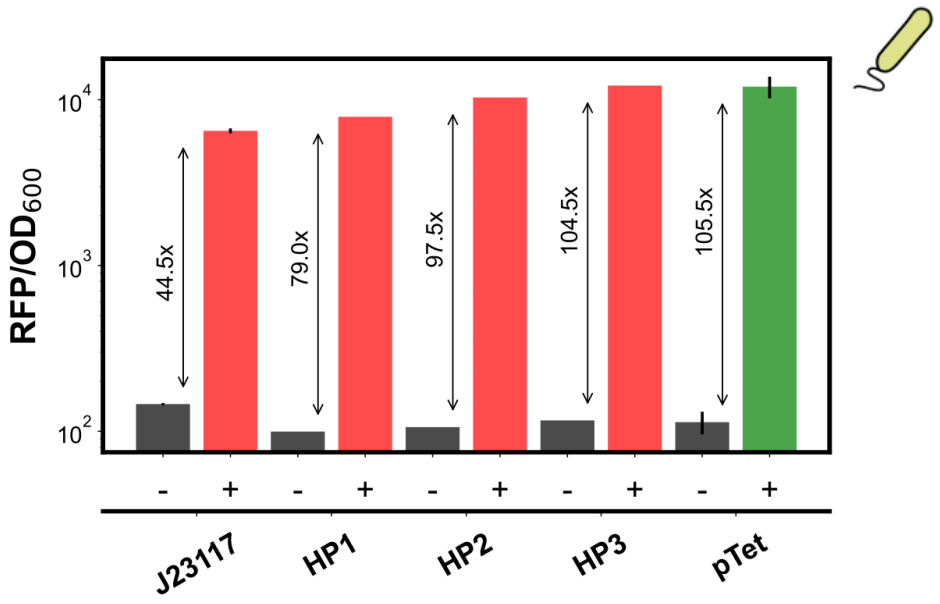


Figure S21: Comparison of CRISPRa promoters to pTet

Comparison of RFP expression levels of different CRISPRa promoters (red) to the pTet system (green). In both systems, RFP plasmid copy number and RBS remained constant. Basal expression level ("-") is measured with off-target scRNAs for the CRISPRa promoters, and 0 nM aTc for pTet. Activated expression level ("+") is measured with on-target scRNAs for the CRISPRa promoters, and 200 nM aTc for pTet. The J23117 and pTet values represent the mean ± standard deviation of three technical replicates, whereas HP1-3 values correspond to the individual variants from the sequential screen (Figure 2D).

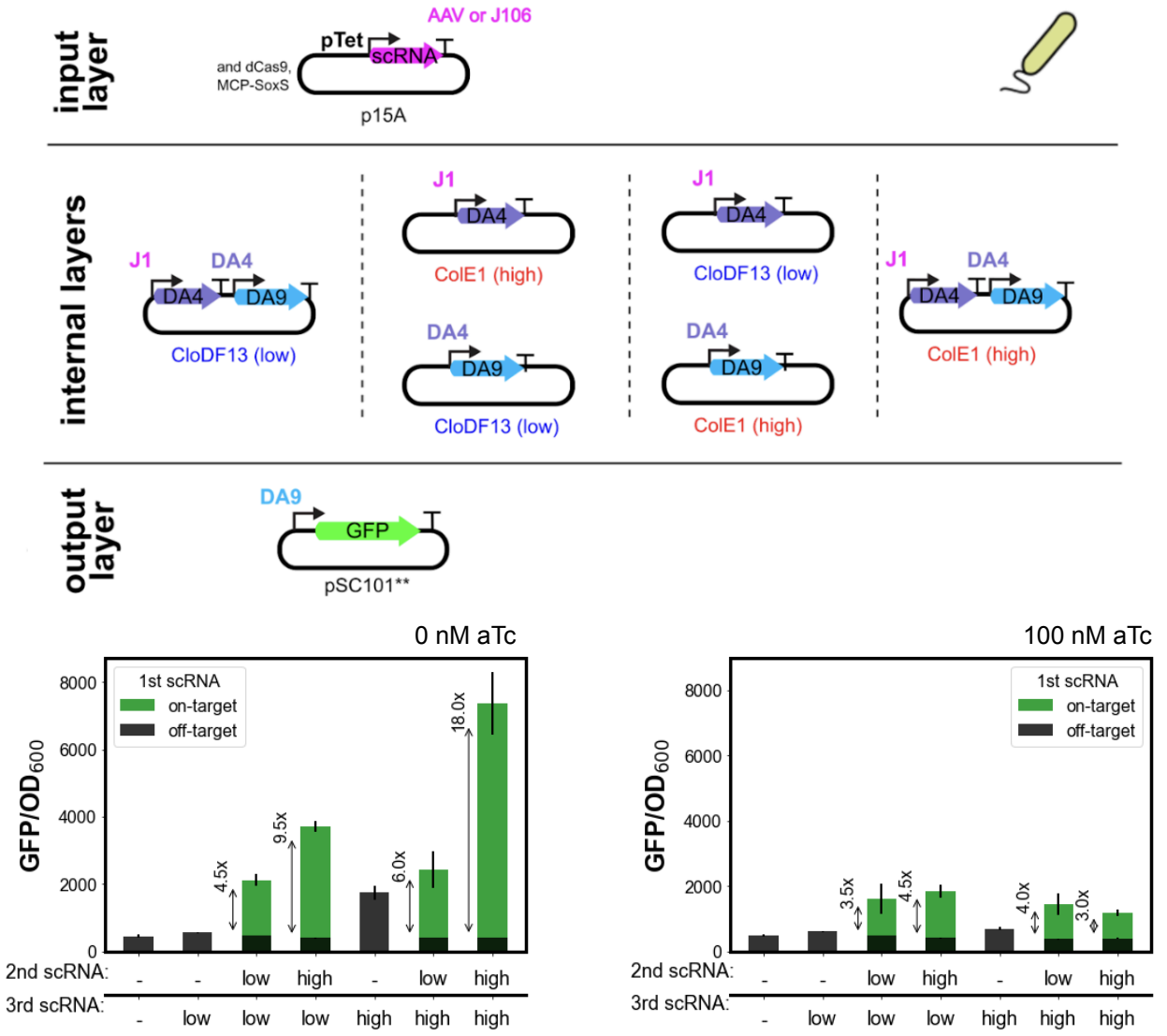


Figure S22: Three-layer activation cascades in *E. coli*

Three-layer activation cascades in *E. coli* with the input controlled by pTet and internal nodes expressed from different copy number plasmids. **Top:** Schematic of plasmids used for the different nodes. Input and output layers were kept constant across conditions, and the copy number of the plasmids encoding the two internal layers was varied between ColE1 (high copy) and CloDF13 (low copy). **Bottom, left:** GFP output for the different activation cascades at 0 nM aTc. Leak is minimized in the system when the third scRNA is expressed from a low copy number plasmid. Leak is also lowered when the second scRNA is present due to gRNA competition with the third scRNA. Expressing the second scRNA from a high copy number resulted in higher

fold-activation. **Bottom, right:** When induced with 100 nM aTc, cascade output is reduced, likely due to competition of the first scRNA with downstream scRNAs. Values represent the mean ± standard deviation of three technical replicates

Supplementary Tables

Table S1: Promoters generated in this paper

Promoter	scRNA target	UP-element	Minimal Promoter	Basal	Activated
HP1_J3.DA9.2A	AACTCTCACACGTGGCTGCA	CCGGCGGCGGCGGCTGCCGCGCGGCG	TTGACAGTTTTACGATGTGTTGGGATTGTGCTAGC	773.66 ± 123.99	289908.0 ± 22289.67
HP2_J3.DA9.1C	AACTCTCACACGTGGCTGCA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACACTTCCGGCACGAAAAGGGATTGTGCTAGC	690.33 ± 112.43	351826.66 ± 7551.20
HP3_J3.DA9.1B	AACTCTCACACGTGGCTGCA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	153.33 ± 43.81	163114.33 ± 10141.27
J3.DA2.1B	CGCCGAATGCTCTAGCGGGA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	567.33 ± 20.21	380117.0 ± 73224.73
J3.DA3.1B	CACACCTAAGTCAGGATTGT	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	595.0 ± 39.05	279047.33 ± 14991.68
J3.DA4.1B	CATATCTCTGACCTGATCGA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	263.66 ± 9.18	332795.33 ± 7755.96
J3.DA6.1B	CACATAAAAACCGCTGACTA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	370.66 ± 40.35	313809.33 ± 15988.97
J3.DA9.1B	AACTCTCACACGTGGCTGCA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	186.0 ± 79.92	224187.33 ± 20739.19
J3.DA10.1B	AGAAACAGTAAAAACTTTCA	CCGCGGGCGGCCGCTGCCGGGGCGCG	TTGACGCCTCCTTCTTTCTTAGGGATTGTGCTAGC	425.33 ± 29.86	349136.66 ± 15853.45
EP1	-	TCGGTAGCCTTTAGTCCATG	TTGACGCTGTATTCAGGCAAAGGGATTGTGCTAGC	1053.66 ± 82.92	10081.91 ± 940.10
EP2	-	TCGGTAGCCTTTAGTCCATG	TTGACAGTGCGTACGCAGGGAGGGATTGTGCTAGC	6170.33 ± 113.42	117299.0 ± 5279.40
EP3	-	TCGGTAGCCTTTAGTCCATG	TTGACGGTGAAGAGTATCAGAGGGATTGTGCTAGC	3516.66 ± 217.243	120075.33 ± 7740.25
EP4	-	TCGGTAGCCTTTAGTCCATG	TTGACAGCTCAGTGAGTAGTAGGGATTGTGCTAGC	5079.66 ± 417.91	66449.16 ± 6693.89
R2.E8.R206	(-81) TCGGCTCACTTATGCACGGC	CCGGCCCCCCCCGCTGCCGCGGGCCG	TTGACAAAGCTATGGCCGGCAGGGATTGTCACAGC	250.67 ± 21.2	6287.67 ± 670.47
R2.E8.R208	(-91) TCCCTGACCCTCGGCTCACT	CCGGCCCCCCCCGCTGCCGCGGGCCG	TTGACAAAGCTATGGCCGGCAGGGATTGTCACAGC	250.67 ± 21.2	3056.67 ± 414.77
R2.E8.R210	(-101) ACCCTTTCCTTCCCTGACCC	CCGGCCCCCCCCGCTGCCGCGGGCCG	TTGACAAAGCTATGGCCGGCAGGGATTGTCACAGC	250.67 ± 21.2	394.67 ± 56.89
R2.E8.R212	(-111) TTCCTCTCCTACCCTTTCCT	CCGGCCCCCCCCGCTGCCGCGGGCCG	TTGACAAAGCTATGGCCGGCAGGGATTGTCACAGC	250.67 ± 21.2	206.67 ± 4.51

Table S2: Dimerization domain affinity

Domain	K_d	Method of measurement	Ref
MS2:MCP	.0033uM	Filter binding assay	(Carey et al., 1983)
SYNZIP5: SYNZIP6	<.015uM	Fluorescence polarization	(Thompson et al., 2012)
PYL1:ABA	1 - 52uM	Isothermal titration calorimetry, Surface plasmon resonance	(Dupeux et al., 2011; Miyazono et al., 2009)
PYL1:ABA:ABI	.030uM	Isothermal titration calorimetry	(Dupeux et al., 2011)
GID1:GA	.2 - 4uM	Radioactivity assay with isotopically labeled GA, *In vitro* FRET binding assay, Surface plasmon resonance	(Miyamoto et al., 2012; Ueguchi-Tanaka et al., 2005; Yoshida et al., 2018)
GAI:GA:GID1	.180uM	Surface plasmon resonance	(Yoshida et al., 2018)

Table S3: Primers for promoter mutagenesis

Minimal Promoter	
oMP1	CGATTATAGATTGACRGCTAGCTCAGTCCTDGNNAYNGTGCTAGCGAATTCATTAAAGAG
oMP2	CGATTATAGATTGACTTGACANNNNNNNNNNNNNNAGGGATTGTNNNAGCGAATTCATTAAAGAG
oMP3_S2	CGCTAGCACAATCCCWNNNNNNSNNNNMRSYGTCAAGCCGGGAGAGCTGGTTCCATTGCGATTGCAGCCACGTGTGAGAGTT
oMP3_S2	CGCTAGCACAATCCCWNNNNNNSNNNNMRSYGTCAACGCCGCCCGGCAGCCGGCGCGGGCGCTGCAGCCACGTGTGAGAGTT
oMP3_S2	CGCTAGCACAATCCCWNNNNNNSNNNNMRSYGTCAACGCCGCGCGGCAGCCGCCGCCG

	CCGGTGCAGCCACGTGTGAGAGTT
UP-element	
oUP1	GCTAGCTGTCAAYNYTTTTTTAAAAAWWWWTNNNNNNTTGTGTCCAGAACGCTCCGTAG
oUP2	GCTAGCTGTCAAWWWWWWWWWWWWWWWWWWWWWWWTTGTGTCCAGAACGCTCCGTAG
oUP3	GCTAGCTGTCAAWWWHDWHDWHDWHDWWDWHWWHDWWWTTGTGTCCAGAACGCTCCGTAG
oUP4	GCTAGCTGTCAANNNNNNNNNNNNNNNNNNNNNNNNTTGTGTCCAGAACGCTCCGTAG
oUP5	GCTAGCTGTCAABBBBSSSSBBNNNNBSSSBSSSBSSSTTGTGTCCAGAACGCTCCGTAG
oUP6	GCTAGCTGTCAASSSSSSSSSSSSSSSSSSSSSSSSTTGTGTCCAGAACGCTCCGTAG
oUP6_S1	GCTAGCTGTCAASSSSSSSSSSSSSSSSSSSSSSSSTGCAGCCACGTGTGAGAGTTAG
scRNA target site	
oTS1	GCTCGTCTCCTCACTTCTCCTWWWWWWWWWWWWWWWWWWCCGGCCCCCCCCGCTGCCGCGGGCCGTTGACAGCTAGCTCAGTCCTAGG
oTS2	GCTCGTCTCCTCACTTCTCCTNNNNNNNNNNNNNNNNNNCCGGCCCCCCCCGCTGCCGCGGGCCGTTGACAGCTAGCTCAGTCCTAGG
oTS3	GCTCGTCTCCTCACTTCTCCTSSSSSSSSSSSSSSSSSSSCCGGCCCCCCCCGCTGCCGCGGGCCGTTGACAGCTAGCTCAGTCCTAGG
EL222 promoter	
oMPE	CCTCTTTAATGAATTCGCTNNNACAATCCCWNNNNNNSNNNNMRSYGTCAACATGGACTAAAGGCTACCTATAAA

Table S4: Plasmids used in this work

Plasmid	J-Promoter	scRNA target	UP Element	Minimal Promoter	RBS	CDS	gRNA	Terminator	res *	ori **
pJF143.J3	J3	J306	J3	J23117	Bujard	mRFP1	X	dbl term	A	S

pCK389.gRNA	X	X	X	Sp.Cas9, pTet, J23119	, Bujard, X	Sp.Cas9, MCP-SoxS (R93A,S101A)	J306, AAV, DA9	Dbl term, BBa_B1002, TrrnB	C	A
pJF182.gRNA	X	X	X	Sp.Cas9, J23107, J23119	, Bujard, X	Sp.Cas9, MCP-SoxS (R93A,S101A)	J306, AAV	dbl term, BBa_B1002, TrrnB	C	A
pDA010.188	X	X	X	J23107	Bujard	*Sp.* dCas9	X	ECK120033736	A	E
pRC029	X	X	X	J23107	Bujard	MCP-SoxS (R93A,S101A)	X	ECK120033736	A	E
pRC011	X	X	X	J23107	Bujard	MCP-SYNZIP6	X	ECK120033736	A	E
pRC012	X	X	X	J23107	Bujard	SoxS-SYNZIP5	X	ECK120033736	A	E
pRC025	X	X	X	J23107	Bujard	MCP-ABI	X	ECK120033736	A	E
pRC027	X	X	X	J23107	Bujard	SoxS-PYL1	X	ECK120033736	A	E
pRC042	X	X	X	J23107	Bujard	MCP-GAI	X	ECK120033736	A	E
pRC043	X	X	X	J23107	Bujard	SoxS-GID1	X	ECK120033736	A	E
pWS025.BLNNN-RFP	X	X	X, EL222_Binding_region	J23119, N	BBa_B0034, Bujard	EL222, mRFP1	X	BBa_B0015_dblT, dbl term	A	S

pDA010.EL222	X	X	X	J23106	BBa_B0034	EL222	X	BBa_B0015_dblT	A	E
pDA040.BLD7-mRFP	X	X	EL222_Binding_region	D7	X	mRFP1	X	ECK120033736	A	E
pDA040.BLD7-DA9	X	X	EL222_Binding_region	D7	X	X	DA9	ECK120033736	A	E
pDA040.BLD7-RR2	X	X	EL222_Binding_region	D7	X	X	RR2	ECK120033736	A	E
pDA303	J3	DA9	1	B	Bujard	mRFP1	X	ECK120033736	A	E
pDA304	J3	DA2	1	B	Bujard	mRFP1	X	ECK120033736	A	E
pDA305	J3	DA3	1	B	Bujard	mRFP1	X	ECK120033736	A	E
pDA306	J3	DA4	1	B	Bujard	mRFP1	X	ECK120033736	A	E
pDA307	J3	DA6	1	B	Bujard	mRFP1	X	ECK120033736	A	E
pDA309	J3	DA10	1	B	Bujard	mRFP1	X	ECK120033736	A	E
pDA310	J3	DA9	1	B	X	X	DA2	ECK120033736	A	E
pDA311	J3	DA9	1	B	X	X	DA3	ECK120033736	A	E
pDA312	J3	DA9	1	B	X	X	DA4	ECK120033736	A	E
pDA313	J3	DA9	1	B	X	X	DA6	ECK120033736	A	E

pDA314	J3	DA9	1	B	X	X	DA8	ECK120033736	A	E
pDA315	J3	DA2	1	B	X	X	DA10	ECK120033736	A	E
pDA316	J3	DA3	1	B	X	X	DA10	ECK120033736	A	E
pDA317	J3	DA4	1	B	X	X	DA10	ECK120033736	A	E
pDA318	J3	DA6	1	B	X	X	DA10	ECK120033736	A	E
pDA319	J3	DA8	1	B	X	X	DA10	ECK120033736	A	E
pDA320	J3	DA9	1	B	X	X	DA2	ECK120033736	A	E
pDA321	J3	DA2	1	B	X	X	DA3	ECK120033736	A	E
pDA322	J3	DA3	1	B	X	X	DA4	ECK120033736	A	E
pDA323	J3	DA4	1	B	X	X	DA6	ECK120033736	A	E
pDA324	J3	DA6	1	B	X	X	DA8	ECK120033736	A	E
pDA325	J3	DA8	1	B	X	X	DA10	ECK120033736	A	E
pRC014.(0-4)	R2	R206 (-81), R208 (-91), R210 (-101), R212 (-111)	E	8	Bujard	mRFP	X	ECK120033736	A	E

pCK956.gRNA	X	X	X	J23107, J23107, pTet	Bujard, Bujard, X	Sp.Cas9, MCP-SoxS (R93A,S101A)	J106, AAV	Dbl term, BBa_B1002, TrrnB	C	A
pDA506.gRNA	X	gRNA	1	B	Bujard	sfGFP	X	dbl term	K	S
pCK957	X	J106	1	B	X	X	DA4	Sht TermA	C, S	E, D
pCK958	X	J106, DA4	1	B	X	X	DA4, DA9	Sht TermA	C, S	E, D
pCK960	X	DA4	1	B	X	X	DA9	Sht TermA	C, S	E, D

*Resistance marker: C stands for chloramphenicol, A stands for ampicillin, S stands for spectinomycin, K stands for kanamycin

Origin of replication:. E stands for ColE1, A stands for p15A, and S stands for sc101, and D stand for CloDF13

Table S5: Deep cascade concentrations

Cascade/ Plasmid	**D1 -**	**D1 +**	**D2 -**	**D2 +**	**D3 -**	**D3 +**	**D4 -**	**D4 +**	**D5 -**	**D5 +**	**D6 -**	**D6 +**
pDA010.188	4	4	4	4	4	4	4	4	4	4	4	4
pRC029	4	4	4	4	4	4	4	4	4	4	4	4
pBT009.J1	0	.015	0	.015	0	.015	0	.015	0	.015	0	.015

.119.DA4												
pDA332			0.2	0.2	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
pDA320					0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
pDA315							0.34	0.34	0.34	0.34	0.34	0.34
pDA335									0.13	0.13	0.13	0.13
pDA336											.093	.093
pDA306	10	10										
pDA303			10	10								
pDA304					10	10						
pDA309							10	10				
pDA307									10	10		
pDA305											10	10

All plasmid concentrations are in nM.

Table S6: Component sequences

pRC011: J23107.Buj.MCP-SYNZIP6

tttacggctagctcagccctaggtattatgctagcGAATTCATTAAAGAGGAGAAAGGTACCatggggccc
gcttctaactttactcagttcgttctcgtcgacaatggcggaactggcgacgtgactgtcgccccaagcaacttcgctaacgg
gatcgctgaatggatcagctctaactcgcgttcacaggcttacaaagtaacctgtagcgttcgtcagagctctgcgcagaat
cgcaaatacaccatcaaagtcgaggtgcctaaaggcgcctggcgttcgtacttaaatatggaactaaccattccaattttcg
ccacgaattccgactgcgagcttattgttaaggcaatgcaaggtctcctaaaagatggaaacccgattccctcagcaatcg
cagcaaactccggcatctacGGTGGCGGAGGTAGCCAAAAAGTTGCGCAGCTGAAAAACCG
TGTTGCGTACAAACTGAAAGAAAACGCGAAGCTGGAGAACATCGTGGCGCGTCTG
GAAAACGACAATGCGAACCTGGAGAAAGACATTGCGAATCTCGAAAAGGACATCGC
AAATCTGGAACGTGACGTTGCGCGTTAAGCGGCCGCcacgcaaaaaaccccgcttcggcggg
gttttttcgc

pRC012: J23107.Buj.SoxS-SYNZIP5

tttacggctagctcagccctaggtattatgctagcGAATTCATTAAAGAGGAGAAAGGTACCATGTCC
CATCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGC
TTAACATTGATGTAGTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAAT
GTTCCGCACGGTGACGCATCAGACGCTTGGCGATTACATTCGCCAACGCCGCCTG
TTACTGGCCGCCGTTGAGTTGCGCACCACCGAGCGTCCGATTTTTGATATCGCAAT
GGACCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGCGTTTTCGCGCGGCAGTTT
GATCGCACTCCCGCGGATTATCGCCACCGCCTGGGTGGCGGAGGTAGCAACACCG
TTAAAGAACTGAAAAACTACATCCAGGAGCTGGAAGAGCGTAACGCTGAACTCAAA
AACCTGAAGGAACACCTGAAATTCGCAAAAGCGGAACTGGAATTCGAACTGGCGG
CTCACAAATTCGAGTAAGGCGCGCCcacgcaaaaaaccccgcttcggcggggttttttcgc

pRC025: J23107.Buj.MCP-ABI

tttacggctagctcagccctaggtattatgctagcGAATTCATTAAAGAGGAGAAAGGTACCatggggccc
gcttctaactttactcagttcgttctcgtcgacaatggcggaactggcgacgtgactgtcgccccaagcaacttcgctaacgg
gatcgctgaatggatcagctctaactcgcgttcacaggcttacaaagtaacctgtagcgttcgtcagagctctgcgcagaat
cgcaaatacaccatcaaagtcgaggtgcctaaaggcgcctggcgttcgtacttaaatatggaactaaccattccaattttcg
ccacgaattccgactgcgagcttattgttaaggcaatgcaaggtctcctaaaagatggaaacccgattccctcagcaatcg
cagcaaactccggcatctacGGTGGCGGAGGTAGCACGCGTGTGCCTTTGTATGGTTTTACT
TCGATTTGTGGAAGAAGACCTGAGATGGAAGcTGCTGTTTCGACTATACCAAGATTC
CTTCAATCTTCCTCTGGTTCGATGTTAGATGGTCGGTTTGATCCTCAATCCGCCGCT
CATTTCTTCGGTGTTTACGACGGCCATGGCGGTTCTCAGGTAGCGAACTATTGTAGA
GAGAGGATGCATTTGGCTTTGGCGGAGGAGATAGCTAAGGAGAAACCGATGCTCT
GCGATGGTGATACGTGGCTGGAGAAGTGGAAGAAAGCTCTTTTCAACTCGTTCCTG
AGAGTTGACTCGGAGATTGAGTCAGTTGCGCCGGAGACGGTTGGGTCAACGTCGG
TGGTTGCCGTTGTTTTCCCGTCTCACATCTTCGTCGCTAACTGCGGTGACTCTAGA
GCCGTTCTTTGCCGCGGCAAAACTGCACTTCCATTATCCGTTGACCATAAACCGGAT
AGAGAAGATGAAGCTGCGAGGATTGAAGCCGCAGGAGGGAAAGTGATTCAGTGGA
ATGGAGCTCGTGTTTTCGGTGTTCTCGCCATGTCGAGATCCATTGGCGATAGATACT
TGAAACCATCCATCATTCCTGATCCGGAAGTGACGGCTGTGAAGAGAGTAAAAGAA
GATGATTGTCTGATTTTGGCGAGTGACGGGGTTTGGGATGTAATGACGGATGAAGA
AGCGTGTGAGATGGCAAGGAAGCGGATTCTCTTGTGGCACAAGAAAAACGCGGTG
GCTGGGGATGCATCGTTGCTCGCGGATGAGCGGAGAAAGGAAGGGAAAGATCCTG
CGGCGATGTCCGCGGCTGAGTATTTGTCAAAGCTGGCGATACAGAGAGGAAGCAA
AGACAACATAAGTGTGGTGGTGGTTGATTTGAAGTAAGGCGCGCCcacgcaaaaaaccc
cgcttcggcggggttttttcgc

pRC027: J23107.Buj.SoxS-PYL1

tttacggctagctcagccctaggtattatgctagcGAATTCATTAAAGAGGAGAAAGGTACCATGTCC
CATCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGC
TTAACATTGATGTAGTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAAT
GTTCCGCACGGTGACGCATCAGACGCTTGGCGATTACATTCGCCAACGCCGCCTG
TTACTGGCCGCCGTTGAGTTGCGCACCACCGAGCGTCCGATTTTTGATATCGCAAT
GGACCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGCGTTTTCGCGCGGCAGTTT

GATCGCACTCCCGCGGATTATCGCCACCGCCTGGGTGGCGGAGGTAGCatgggtgggg
gcgcgccaactcaagacgaattcacccaactctcccaatcaatcgccgagttccacacgtaccaactcggtaacggccg
ttgctcatctctcctagctcagcgaatccacgcgccgccggaaacagtatggtccgtggtgagacgtttcgataggccaca
gatttacaaacacttcatcaaaagctgtaacgtgagtgaagatttcgagatgcgagtgggatgcacgcgcgacgtgaacg
tgataagtggattaccggcgaatacgtctcgagagagattagatctgttggacgatgatcggagagtgactgggtttagtat
aaccggtggtgaacataggctgaggaattataaatcggttacgacggttcatagatttgagaaagaagaagaagaaga
aaggatctggaccgttgttttggaatcttatgttgttgatgtaccggaaggtaattcggaggaagatacgagattgtttgctgat
acggttattagattgaatcttcagaaacttgcttcgatcactgaagctatgaacTAAGCGGCCGCCcgcaaaaaac
cccgcttcggcggggttttttcgc

pRC042: J23107.Buj.MCP-GAI

tttacggctagctcagccctaggtattatgctagcGAATTCATTAAAGAGGAGAAAGGTACCatggggccc
gcttctaactttactcagttcgttctcgtcgacaatggcggaactggcgacgtgactgtcgccccaagcaacttcgctaacgg
gatcgctgaatggatcagctctaactcgcgttcacaggcttacaaagtaacctgtagcgttcgtcagagctctgcgcagaat
cgcaaatacaccatcaaagtcgaggtgcctaaaggcgcctggcgttcgtacttaaatatggaactaaccattccaattttcg
ccacgaattccgactgcgagcttattgttaaggcaatgcaaggtctcctaaaagatggaaacccgattccctcagcaatcg
cagcaaactccggcatctacGGTGGCGGAGGTAGCATGAAGCGCGATCATCATCACCACCA
CCACCAGGATAAAAAGACGATGATGATGAATGAGGAAGATGATGGAAACGGGATGG
ACGAATTGCTGGCAGTGCTGGGATATAAGGTGCGTTCGTCCGAAATGGCAGATGTT
GCTCAGAAATTGGAGCAGTTAGAAGTAATGATGAGTAACGTTCAAGAAGATGATCTT
TCACAGTTAGCGACCGAAACTGTCCACTACAACCCTGCTGAGCTTTACACTTGGTT
GGACTCCATGCTTACCGATCTTAACtgacgcaaaaaaccccgcttcggcggggttttttcgc

pRC043: J23107.Buj.SoxS-GID1

tttacggctagctcagccctaggtattatgctagcGAATTCATTAAAGAGGAGAAAGGTACCATGTCC
CATCAGAAAATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGC
TTAACATTGATGTAGTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAAT
GTTCCGCACGGTGACGCATCAGACGCTTGGCGATTACATTCGCCAACGCCGCCTG
TTACTGGCCGCCGTTGAGTTGCGCACCACCGAGCGTCCGATTTTTGATATCGCAAT
GGACCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGCGTTTTCGCGCGGCAGTTT

GATCGCACTCCCGCGGATTATCGCCACCGCCTGGGTGGCGGAGGTAGCATGGCAG
CCTCCGACGAGGTAAATCTTATTGAGAGTCGTACCGTCGTTCCCTTGAATACTTGGG
TGTTGATCTCGAATTTCAAGGTCGCGTACAATATCTTACGCCGCCCGGATGGAACCT
TTAACCGTCACCTTGCAGAATATCTGGACCGCAAAGTTACAGCAAATGCTAATCCAG
TTGACGGTGTTTTCAGTTTTGACGTGCTGATTGATCGCCGTATCAACCTTCTGTCCC
GTGTCTATCGTCCTGCTTACGCCGATCAGGAGCAACCTCCATCCATTCTGGATCTG
GAAAAACCAGTGGATGGGGACATTGTCCCTGTCATCCTTTTTTTCCACGGGGGGTC
GTTCGCCCACTCGTCCGCCAACAGTGCGATCTACGACACTTTATGTCGTCGTCTTG
TCGGTCTTTGCAAATGCGTGGTCGTTTCCGTGAATTACCGTCGCGCTCCGGAGAAC
CCCTACCCATGTGCCTACGACGACGGATGGATTGCGTTAAATTGGGTTAATTCACGT
AGCTGGCTGAAAAGCAAGAAAGATTCGAAGGTTCACATTTTTTTAGCGGGCGATTCT
TCAGGAGGGAACATCGCTCATAATGTCGCATTGCGTGCAGGAGAGTCTGGCATCGA
TGTTCTGGGCAACATTTTACTGAACCCGATGTTTGGGGGGAACGAGCGCACAGAAT
CCGAGAAAAGCTTGGACGGGAAGTATTTCGTGACTGTTCGCGATCGTGACTGGTAT
TGGAAAGCGTTCTTGCCCGAGGGAGAGGACCGCGAGCACCCCGCATGCAACCCC
TTTTCACCTCGCGGAAAATCGCTGGAGGGGGTCAGTTTCCCAAAATCTTTAGTCGT
AGTAGCTGGCCTGGATCTGATCCGTGATTGGCAACTTGCGTATGCTGAAGGCCTTA
AGAAGGCTGGTCAAGAAGTAAAGCTGATGCACTTAGAGAAAGCTACGGTTGGCTTT
TATCTGTTACCAAATAACAATCACTTCCATAATGTGATGGATGAGATCTCCGCTTTCG
TTAATGCGGAATGCtgacgcaaaaaaccccgcttcggcggggttttttcgc

pDA010.EL222: J23106. BBaB0034.EL222
TTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCCTAGAGAAAGAGGAGAAATACT
AGATGTTGGATATGGGACAAGATCGGCCGATCGATGGAAGTGGGGCACCCGGGGC
AGACGACACACGCGTTGAGGTGCAACCGCCGGCGCAGTGGGTCCTCGACCTGAT
CGAGGCCAGCCCGATCGCATCGGTCGTGTCCGATCCGCGTCTCGCCGACAATCCG
CTGATCGCCATCAACCAGGCCTTCACCGACCTGACCGGCTATTCCGAAGAAGAATG
CGTCGGCCGCAATTGCCGATTCCTGGCAGGTTCCGGCACCGAGCCGTGGCTGAC
CGACAAGATCCGCCAAGGCGTGCGCGAGCACAAGCCGGTGCTGGTCGAGATCCT
GAACTACAAGAAGGACGGCACGCCGTTCCGCAATGCCGTGCTCGTTGCACCGATC
TACGATGACGACGACGAGCTTCTCTATTTCCTCGGCAGCCAGGTCGAAGTCGACGA

CGACCAGCCCAACATGGGCATGGCGCGCCGCGAACGCGCCGCGGAAATGCTCAA
GACGCTGTCGCCGCGCCAGCTCGAGGTTACGACGCTGGTGGCATCGGGCTTGCG
CAACAAGGAAGTGGCGGCCCGGCTCGGCCTGTCGGAGAAAACCGTCAAGATGCA
CCGCGGGCTGGTGATGGAAAAGCTCAACCTGAAGACCAGCGCCGATCTGGTGCG
CATTGCCGTCGAAGCCGGAATCTAAGGATCCAAACTCGAGTAAGGATCTCCAGGCA
TCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTT
GTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCT
GCGTTTATA

pDA040.BLD7-mRFP: EL222_Binding_region.D7.Bujard.mRFP
GGTAGCCTTTAGTCCATGTTGACGGTGAAGAGTATCAGAGGGATTGTGCTAGCGAA
TTCATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTT
CATgcgtttcaaagttcgtatggaaggttccgttaacggtcacgagttcgaaatcgaaggtgaaggtgaaggtcgtccgt
acgaaggtacccagaccgctaaactgaaagttaccaaaggtggtccgctgccgttcgcttgggacatcctgtccccgcag
ttccagtacggttccaaagcttacgttaaacacccggctgacatcccggactacctgaaactgtccttcccggaaggtttca
aatgggaacgtgttatgaacttcgaagacggtggtgttgttaccgttacccaggactcctccctgcaagacggtgagttcatc
tacaaagttaaactgcgtggtaccaacttcccgtccgacggtccggttatgcagaaaaaaaccatgggttgggaagcttcc
accgaacgtatgtacccggaagacggtgctctgaaaggtgaaatcaaaatgcgtctgaaactgaaagacggtggtcact
acgacgctgaagttaaaaccacctacatggctaaaaaaccggttcagctgccgggtgcttacaaaaccgacatcaaact
ggacatcacctcccacaacgaagactacaccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccggtgctt
aaggatccaaactcgagtaaggatctGTGCTTTTTTTaacgcatgagAAAGCCCCCGGAAGATCAC
CTTCCGGGGGCTTTtttattgcgc

Table S7: ANOVA analysis of combinatorial promoter screens

	Sum of Squares	**Degrees of Freedom**	**F Value**	***p* Value**
Minimal Promoter	23155.9	3	9.5	1.2e-4
UP-Element	315527.0	3	129.7	5.5e-18
Minimal Promoter & UP-Element	53678.2	9	7.4	1.0e-5
Residual	25954.2	32	-	-

Supplementary Methods

Methods S1: Plasmid Preparation for Cell-Free System

Plasmids intended for use in CFS were grown in culture volumes ~20 mL to ensure adequate yields for multiple cell-free reactions and 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).

Methods S2: CFS Blue-light CRISPRa/i modeling

The model was implemented using the text-based model definition language Antimony for Python 3.7. We introduced blue-light regulation as a piecewise function that modulates transcription of the sgRNA required for CRISPRi. We used linear functions of different slopes to capture the fast dimerization of the EL222 protein and binding to the DNA upon blight exposure as well as the slow unbinding in the absence of blue-light. Specifically, we the sgRNA basal transcription constant is modified with the following function:

$$0 \quad if \quad 0 < t < t_{delay}$$

$$\frac{k_{max}}{t_{ON}}\left(t - t_{delay}\right) \quad if \quad t_{delay} < t < t_{delay} + t_{ON}$$

$$k_{max} \; if \quad t_{delay} + t_{ON} < t < t_{delay} + t_{ON} + t_{expose}$$

$$\frac{-k_{max}}{t_{OFF}}\left(t - (t_{delay} + t_{ON} + t_{expose}\right) \quad if \quad t_{delay} + t_{ON} + t_{expose} < t < t_{delay} + t_{ON} + t_{expose} + t_{OFF}$$

Where k_{max} represents the transcription rate constant when EL222 is fully bound to the promoter, and t_{delay}, t_{ON}, t_{expose}, $and\ t_{OFF}$ represent the time delay for light exposure, the time for

EL222 dimerization and binding, the exposure time to light, and the time for EL222 unbinding upon removing the light source, respectively.

Methods S3: Quantification and Statistical Analysis

***E. coli* data analysis:**

Dynamic range:

Dynamic range was calculated as the ratio of measured RFP outputs without induction (0 nM aTc, or dark) and with induction (200nM aTc, or light):

$$DR = \frac{B^{\alpha_1} - B^0}{B^{\alpha_2} - B^0}$$

where:

B is RFP/OD_{600} measured at endpoint

α_1 is activated expression, with induction

α_2 is basal expression, without induction

0 is no RFP expression

Pareto optimality:

To identify the best-performing promoter variants belonging to the Pareto front, we compared the basal and activated RFP expression levels of each variant to all other variants. A variant belongs to the Pareto front if no other variant had both lower basal and higher activated expression levels:

$$v_o \in P(V) \;\; if\ there\ is\ no\ v\ such\ that\ (v^a > v_o^a \;\&\; v^b < v_o^b) \;\; for\ all\ v \in V$$

where:

V is the set of all promoter variants

v_o is a variant in V

v_o^a, v_o^b are the activated and basal expression levels of said variant

$P(V)$ is the set of promoter variants belonging to the Pareto front

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 RFP

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 and are calculated as:

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

where:

α is a specific CRISPRa/i circuit

Γ is constitutive expression

Fold change:

Fold change was calculated as the ratio of RFP values generated by CRISPRa in the presence of input scRNA compared to RFP values generated in the absence of input scRNA.

$$FC^{\alpha}(y) = \frac{B^{\alpha+}(t) - B^{\Gamma}(t)}{B^{\alpha-}(t) - B^{\Gamma}(t)}$$

where:

$\alpha +$ is CRISPRa with *y* nM input scRNA

$\alpha -$ is CRISPRa without input scRNA

Γ is constitutive expression

Time to maximum expression rate:

To calculate the time to maximum expression rate, 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 maximum expression is denoted as t_{max}.

$$t = t_{max} \; when \; \dot{B}^{\alpha}_{\Gamma}(t) = max(\dot{B}^{\alpha}_{\Gamma}(t))$$

where:

α is a specific CRISPRa/i circuit

Γ is constitutive expression

Change in time to maximum expression rate (Δt_{max}) is calculated by finding the difference in time to reach maximum production rate between the with and without input conditions.

$$\Delta t_{max} = t_{max}^{\alpha+} - t_{max}^{\alpha-}$$

Signal propagation efficiency:

Propagation efficiency of the CRISPRa cascade in CFS was calculated as the maximum fold change in cascade output ± input divided by the fold change provided by CRISPRa in the input layer at the same time point.

$$propagation = 100 \bullet \frac{max(FC^{\alpha}(y))}{FC^{\beta}(y)}$$

where:

α is CRISPRa cascade with y nM of scRNAs

β is CRISPRa with y nM of scRNAs

Signal delay:

Signal delay is calculated as the difference in time to reach the maximum fold change of the cascade between the cascade output and input layer.

$$t^{\alpha} = t_{max}^{\alpha} FC \ when \ FC^{\alpha}(y) = max(FC^{\alpha}(y))$$

$$t^{\beta} = t_{max}^{\beta} FC \ when \ FC^{\beta}(y) = max(FC^{\alpha}(y))$$

$$delay = t_{max}^{\beta} FC - t_{max}^{\alpha} FC$$

where:

α is CRISPRa cascade with y nM of scRNAs

β is CRISPRa with y nM of scRNAs

Methods S4: Relationship between signal delay and signal propagation

We define fraction of signal propagation at the *nth* node to be the product of the fraction of signal propagated at the previous nodes, namely:

$$fSP_n = \prod_i^n fSP_i$$

where the fraction of signal propagated by each node *i* is a function of the characteristic relative fold-change of node *i,* as well as the time of the reaction at which the signal propagates through node *i*:

$$fSP_i = rFC_i \cdot e^{\frac{-(\theta_i + t_o)}{\tau}}$$

where t_o is the reaction boot up time, and $\boldsymbol{\tau}$ is the characteristic time of the system. While seemingly simple, the exponential term accounts for the complex dynamics of cell-free expression and gRNA competition, and favors expression from earlier nodes.

The time of the reaction at which the signal propagates through the *nth* node can be estimated based on the fraction of the signal propagated through the *nth* node and the relative lifetime of the reaction:

$$\theta_n = rLT \cdot (1 - fSP_n)$$

where the relative lifetime of the reaction is the difference between the time to maximum fold activation of a one-layer cascade and the end of the reaction. With these equations and the characteristic relative fold-change of each node, both the signal delay and probation can be calculated iteratively.

Based on kinetic data, we set t_o and rLT to be ~2 hrs and ~5 hrs, respectively. In order to estimate $\boldsymbol{\tau}$, we fit the model to empirical signal propagation and delay data by minimizing the sum of residuals using the Nelder-Mead algorithm.

Methods S5: 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. For reactions containing ABA (Sigma, A4906) or GA, .1 µL of the small molecules were added alongside the plasmids. For reactions involving more than 5 plasmids, plasmids were mixed with an acoustic liquid handler robot (Echo Labcyte 525). 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. 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.

Methods S6: Plasmid and Library Construction

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 and libraries 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. Transformed cells were 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.

Methods S7: Optogenetic setup

The samples were placed at 37 °C or 29 °C in an incubator (Thermo Forma Orbital Shaker, Model #435) with the illumination source placed atop the incubator and irradiating inwards. The distance between the illumination source and the *E. coli* deepwell plates was 14 cm. CFS reactions were placed inside the incubator at 29 °C at a distance of 6 cm with the bottom of the wells facing the illumination source. In both cases, the dark conditions were kept inside a cardboard box inside the incubator. Endpoint plate reader measurements were conducted using a BioTek Synergy HTX.

Methods S8: *E. coli* experiments culturing and quantification conditions

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, covering with breathable membrane (Breathe Easier cat# Z763624) and shaking overnight at 37 °C at 1200 RPM on a Heidolph Titramax 1000. For inducible experiments, overnight cultures are subsequently diluted 1:40 into a fresh plate of EZ-RDM and supplemented with appropriate concentrations of aTc. Plate reader measurements were conducted using a BioTek Synergy HTX with a black flat bottom plate (Ref# 3631) using 100 µL of culture.

Methods S9: Plasmid and Library Preparation

Details regarding plasmid and library construction are presented in Methods S6. Plasmids were transformed into chemically competent NEB Turbo *E. coli.* 10 uL of the outgrowth with transformed libraries was diluted 1:20 with LB and plated onto LB-agar with carbenicillin to check library complexity. The remaining outgrowth was seeded into 5 mL of LB with carbenicillin or gentamicin. Cells were grown overnight ~16 hours at 37 °C. Single colonies were picked from plates and grown overnight in LB with carbenicillin. Single colonies and culture were sequence verified. Plasmids were isolated from subcultures using a DNA miniprep kit (QIAprep Spin Miniprep Kit) and Sanger sequenced (Genewiz inc.).

Methods S10: *E. coli* Experiments

dCas9, MCP-SoxS, and scRNA are on a p15A ori plasmid while reporter construct is located on a pSC101** ori plasmid. For experiments involving more than two plasmids, competent cells were first made from cells carrying the reporter plasmid and the CRISPRa plasmid (including either on- or off-target input scRNAs). The appropriate plasmids expressing internal scRNAs were transformed into the competent cells. Details regarding culturing conditions and quantification are provided in Methods S8.

Methods S11: Design of Promoter Region Libraries

3.1 Minimal Promoter Libraries: MP1 was designed by rationally mutagenizing specific bases that are known contacts of RNAP within the minimal promoter. MP2 was made by randomly mutagenizing within the intervening sequence. Since the libraries yielded

similar Pareto fronts, we combined these mutations into MP3, used in the sequential screening process (Table S3).

3.2 UP-Element Libraries: We designed five UP-element libraries mutagenizing the AT-rich *E. coli* consensus sequence with increasing GC-content. We generated 5 libraries from 0% to 100% GC-content, and a library representing the *E. coli* consensus sequence (Table S3).

3.3 scRNA Target Site Libraries: We generated three scRNA target site libraries with varying compositions of GC-content (0%, 50%, and 100%) (Table S3). These libraries were used in tandem with a GC-rich UP-element.

3.4 EL222 Minimal Promoter Libraries: Starting with the native *luxI* minimal promoter, we introduced rational mutations to make it resemble a synthetic activatable promoter (J23117). We then randomly mutagenized within the -10:-35 region (Table S3).

Methods S12: Cell-Free System Preparation

CFS was acquired from Arbor Biosciences (myTXTL). The CFS 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 sample homogeneity. Details about plasmid preparation are provided in Methods S1, and details about the CFS reaction are provided in Methods S5.

Methods S13: Optogenetic Experiments

E. coli cultures and CFS reactions were prepared as described above. The incubation conditions were modified to include a blue-light illumination source (UVP Visi-Blue UV

Transilluminator, 8 Watts, 460/470 nm). Details about optogenetic setup are provided in Methods S7.

Methods S14: CFS Blue-light CRISPRa/i modeling

The CFS blue-light CRISPRa/i model was expanded from the previously described CFS CRISPRa/i model [32]. The model constitutes a series of first order chemical reactions for protein and guide RNA production, CRISPR complex assembly, and DNA targeting. All model details are described in Methods S2.

Methods S15: Quantification and statistical analysis

7.1 Data analysis: Throughout this work all measured RFP levels in *E. coli* were normalized by measured OD600 with appropriate propagation of uncertainties. All metrics are described in Methods S3.

7.2 Statistics: Statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks in Figures indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.01, ∗∗∗: p-value < 0.001).

Chapter 3: Carbon-conserving Bioproduction of Malate in an *E. coli*-based Cell-Free System

Ryan A. L. Cardiff[1], Shaafique Chowdhury[2], Widianti Sugianto[3], Benjamin I. Tickman[3], Diego Alba Burbano[3], Pimphan A. Meyer[4], Margaret Cook[1], Brianne King[5], David Garenne[6], Alexander S. Beliaev[7,8,9], Vincent Noireaux[6], Pamela Peralta-Yahya*,[2,10], James M. Carothers*,[1,3]

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

2: School of Chemical and Biomolecular Engineering
Georgia Institute of Technology
Atlanta, GA 30332
United States

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

4: Oak Ridge National Laboratory
Oak Ridge, TN 37830
United States

5: Department of Chemistry
University of Washington
Seattle, WA 98195
United States

6: School of Physics and Astronomy
University of Minnesota
Minneapolis, MN 55455
United States

7: Pacific Northwest National Laboratory

Richland, WA 99354
United States

8: Centre for Agriculture and the Bioeconomy, School of Biological and Environmental Sciences
Queensland University of Technology, Gardens Point Campus
P.O. Box 2434, Brisbane 4001, Queensland, Australia

9: ARC Centre of Excellence in Synthetic Biology
Brisbane, Queensland, Australia.

10: School of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta, GA 30332
United States

*: Corresponding author(s)

jcaroth@uw.edu

pperalta-yahya@chemistry.gatech.edu

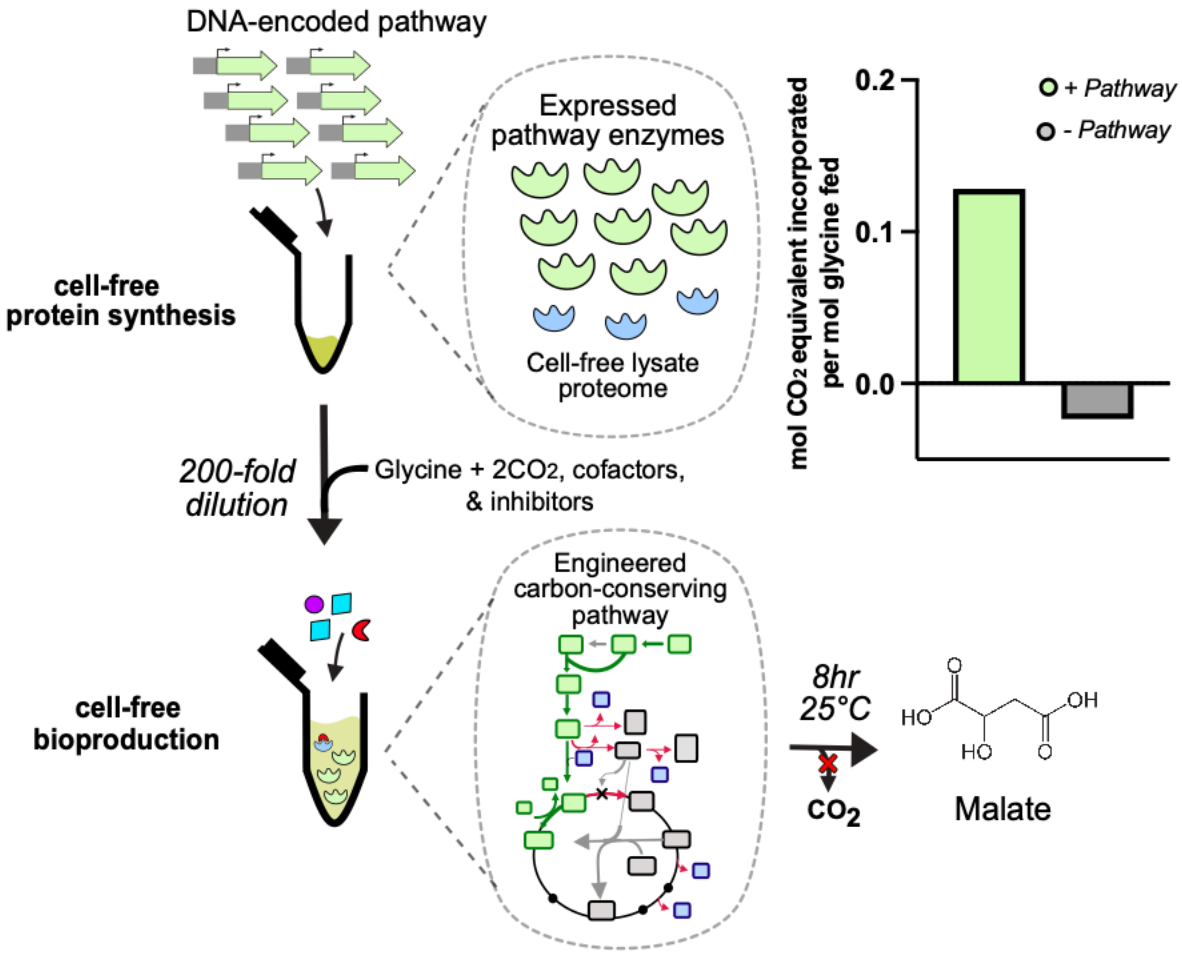
Published as a research article in Metabolic Engineering on April 8th, 2025. DOI: https://doi.org/10.1016/j.ymben.2025.03.020

Abstract

Formate, a biologically accessible form of CO_2, has attracted interest as a renewable feedstock for bioproduction. However, approaches are needed to investigate efficient routes for biological formate assimilation due to its toxicity and limited utilization by microorganisms. Cell-free systems hold promise due to their potential for efficient use of carbon and energy sources and compatibility with diverse feedstocks. However, bioproduction using purified cell-free systems is limited by costly enzyme purification, whereas lysate-based systems must overcome loss of flux to background reactions in the cell extract. Here, we engineer an *E. coli*-based system for an eight-enzyme pathway from DNA and incorporate strategies to regenerate cofactors and minimize loss of flux through background reactions. We produce the industrial di-acid malate from glycine, bicarbonate, and formate by engineering the carbon-conserving reductive TCA and formate assimilation pathways. We show that *in situ* regeneration of NADH drives metabolic flux towards malate, improving titer by 15-fold. Background reactions can also be reduced 6-fold by diluting the lysate following expression and introducing chemical inhibitors of competing reactions. Together, these results establish a carbon-conserving, lysate-based cell-free platform for malate production, producing 64 μM malate after 8 hours. This system conserves 43% of carbon otherwise lost as CO_2 through the TCA cycle and incorporates 0.13 mol CO_2 equivalents/mol glycine fed. Finally, techno-economic analysis of cell-free malate production from formate revealed that the high cost of lysate is a key challenge to the economic feasibility of the process, even

assuming efficient cofactor recycling. This work demonstrates the capabilities of cell-free expression systems for both the prototyping of carbon-conserving pathways and the sustainable bioproduction of platform chemicals.

Graphical Abstract



Introduction

Metabolic engineering has enabled the production of industrially relevant chemicals from renewable and low-cost one carbon (C1) feedstocks, including CO_2 and formate [1–3]. The use of renewable feedstocks offers a promising alternative to petroleum-based chemical synthesis for the sustainable production of platform chemicals [4,5]. However, the use of CO_2 and formate as feedstocks for common metabolic engineering chassis, such as *Escherichia coli,* has been limited by the low solubility of CO_2 and high formate toxicity [6,7]. Novel approaches are needed to efficiently incorporate C1 feedstocks into biologically-accessible chemicals.

The use of cell-free gene expression systems (CFE) as a platform for carbon-conserving metabolic engineering has the potential to address several of the limitations of microbial based bioproduction [8]. CFE are typically either prepared as a purified system, in which all the necessary components are individually purified and then reconstituted, or as a cell-based lysate [9,10]. Each CFE approach has its own advantages and weaknesses, but across systems the high cost of preparation limits their potential for metabolic engineering [11]. Lysate-based transcriptional and translational (TXTL) systems show promise for high-yield protein expression at lower costs than purified systems [12], while still enabling modification of reaction conditions, such as protein and chemical concentrations. CFE prepared from *E. coli* lysate contain high concentrations of cellular proteins and metabolites, including those required for transcriptional and translation, but lack the cell membrane and genomic DNA native to the cell [13]. This allows CFE to be leveraged as an open reaction vessel, where plasmid or linear DNA encoding pathways of interest can be added directly to the reaction mixture.

For lysate-based metabolic engineering, there are many outstanding questions with respect to the relevant expression levels of multi-enzyme pathways, compatibility of reaction conditions for gene expression and bioproduction, and the challenges associated with the endogenous proteome of the lysate. To date, the majority of *in vitro* carbon-fixing pathways have been assembled using purified enzyme systems to avoid interference from native enzymes present in a lysate-based system [14–16]. However, purified systems are limited by the high cost of biocatalyst preparation [14,17–19]. While less costly to prepare than purified enzyme systems, lysate-based systems face challenges from the presence of native pathways and an inability to easily regulate endogenous enzyme activity in CFE [20,21]. While the activity of the endogenous proteome in cell-free lysates is well-documented [18,22], effective interventions are still needed to minimize diversion of flux from competing reactions in the proteome. Developing systems to understand which enzymes and metabolic branches are important focal points for diversion is crucial for cell-free metabolic engineering and bioproduction more broadly.

Several carbon-fixing or carbon-conserving bioproduction pathways have been demonstrated using purified enzyme systems. Many systems take advantage of carboxylase-dependent cycles to iteratively fix carbon into a C3 or C4 product [14–16]. Carboxylase-dependent pathways are often easier to implement *in vitro* as they tend to avoid the need for unstable, oxygen-sensitive cofactors. However, these cyclical pathways tend to be energetically expensive in terms of NAD(P)H and ATP consumption compared to linear pathways, such as the reverse glycine cleavage pathway. However, linear C1-incorporating pathways may require the use of complex cofactors for key enzymes, such as vitamin B_{12}, ferredoxins, or rare metals [23–25]. Across

pathways, efficient regeneration of energy sources and cofactors is needed to improve reaction productivity and reduce costs for C1 assimilation pathways [4,15,21]. Novel approaches to reduce biocatalyst requirements and regulate metabolic activity will expand the potential of cell-free systems as a scalable platform for metabolic engineering.

In this work, we demonstrate a platform for the carbon-conserving bioproduction of malate in a lysate-based CFE. Microbial bioproduction of the industrial di-acid malate requires re-routing flux through either the oxidative or reductive TCA (rTCA) cycle. The rTCA cycle represents the most carbon-efficient route, with a theoretical yield of 1 mol malate/mol pyruvate with an extra carbon coming from bicarbonate; compared to a theoretical yield of 0.5 mol malate/mol pyruvate through the oxidative TCA cycle, which features two decarboxylation steps [26]. Microbial production through either pathway requires extensive genome engineering to introduce knockouts that will redirect metabolic flux towards malate accumulation. This often leads to cells that have significant growth defects, as the knocked out genes are typically involved in the TCA cycle or central carbon metabolism [26–28]. Cell-free bioproduction has the potential to overcome these limitations by decoupling growth and bioproduction. Therefore, we sought to develop a cell-free platform for malate production through the carbon-efficient rTCA pathway.

The presented pathway incorporates two carbon-fixation steps to generate the C4 di-acid malate from the C2 precursor glycine and C1 inputs formate and bicarbonate **(Figure 1)**. This system utilizes steps from both the formate assimilation pathway and reductive TCA cycle (rTCA). Recently, we have shown that the formate assimilation

pathway coupled with the reductive glycine pathway (rGCV) can be used to assimilate formate into glycine and serine [29]. This has shown to be the most energetically efficient of the natural C1 fixing pathways [29]. In the formate assimilation pathway, formate is first coupled with tetrahydrofolate (THF) to ultimately generate the one-carbon donor, 5,10-methylenetetrahydrofolate (5,10-CH_2-THF). Next, 5,10-CH_2-THF transfers the C1 to glycine to synthesize serine and recycle THF. Serine is then converted to pyruvate, where another CO_2 equivalent is incorporated through pyruvate carboxylase (*pyc*) to generate oxaloacetate. Oxaloacetate is then reduced to malate via malate dehydrogenase (*mdh*) using NADH. We find that the expression of formate dehydrogenase (*fdh*) to regenerate NADH helps drive metabolic flux towards malate by maintaining a high concentration of reducing equivalents.

Development of strategies to overcome limitations of lysate-based CFE metabolic engineering is crucial to the successful implementation of complex pathways. In TXTL systems, the optimal conditions for gene expression may not necessarily be optimal for biosynthesis. We separate the one-pot metabolic engineering platform into two parts by separating the cell-free protein synthesis reaction from the bioproduction reaction, allowing us to independently tune conditions for each process. By diluting protein synthesis reactions, we improve the volumetric efficiency of the bioproduction reaction and reduce the relative activity from endogenous enzymes in the lysate. We also show that metabolic flux from competing pathways can be blocked by the addition of small-molecule inhibitors directly to the bioproduction reaction. In addition to improving efficiency, this approach minimizes carbon loss through the TCA cycle without requiring engineering of the underlying strain. Together, these developments enable malate production from formate and glycine with a yield on glycine of 6.4% in an

eight-hour biosynthesis reaction. This work helps establish the capabilities of CFE as a platform for combined protein synthesis and metabolic engineering, enabling carbon-conserving bioproduction of platform chemicals.

Results

Metabolomics analysis of the endogenous metabolic activity of the CFE

While the lysis and filtration steps in the preparation of CFE lysates remove bulky membrane proteins and genomic DNA, many native protein complexes are still present at high concentrations [30]. We first sought to characterize the basal metabolic activity in the cell-free system by measuring the activity from endogenous enzymes in the system via targeted metabolomics.

To understand the relative carbon flux as a result of the endogenous enzymes present in the cell-free lysate, we spiked 1 mM of pathway intermediates into otherwise empty, plain cell-free reactions and measured the presence of relevant compounds, including malate, pyruvate, fumarate, glyoxylate, and succinate, over time **(Figure 2A)**. In these experiments, the “Plain CFE” reaction represents a cell-free reaction with no pathway intermediate spiked in. We found that, relative to a plain cell-free reaction with only water added, CFE with 1 mM malate spiked in accumulated high concentrations of fumarate, glyoxylate, succinate, and pyruvate **(Figure 2B, 2C)**, indicating that TCA enzymes are highly active in the cell-free lysate. We found that malate was rapidly converted to other TCA intermediates. In eight hours, a molar ratio of roughly 1:1:4:4 (fumarate:succinate:glyoxylate:pyruvate) was achieved. Additionally, there is a high interconversion between pyruvate and malate, as shown by the reactions spiked with

either compound having significant flux towards the other. With the addition of 1mM glycine, serine concentrations stay roughly constant at 25 μM **(Figure 2C)**. In contrast, spiking 1mM serine leads to a rapid accumulation of glycine, reaching 485 μM, or 48.5% conversion, after 8 hours. A thermodynamic analysis using Equilibrator indicated that glycine is more energetically favorable than serine (ΔG = +6.7 kJ/mol) [31], which may explain the larger flux towards glycine **(Figure S1)**. Qualitatively, the observation that spiking in serine leads to large amounts of glycine, while spiking in glycine leads to modest amounts of serine, **(Figure 2C)** agrees with the thermodynamic analysis. The decreasing concentrations seen for malate and glycine early in the reaction **(Figure 2C)** may be due to an initial surge in production that happens rapidly and is captured at the zero time point, which cannot be a true zero time point due to limitations in the time required for sample preparation (roughly 5-10 minutes). Therefore, the concentration of these metabolites likely increases rapidly before the first reading, then decreases slowly. It is also worth noting that the thermodynamic predictions given in Figure S1 assume equimolar concentrations of all chemical species in the reaction, which is not the case in a spiked-in reaction. Additionally, the activity of enzymes carrying out these conversions may be affected in an *in vitro* cell-free system [22], which could lead to behaviors that do not agree with thermodynamic predictions. These results highlight the large effects of the endogenous proteome on metabolite levels in lysate-based cell-free systems. For efficient bioproduction, strategies will be needed to counteract these effects.

Malate synthesis from the reductive TCA cycle

To engineer a carbon-efficient pathway for malate bioproduction, we began prototyping malate biosynthesis from the TCA cycle precursors oxaloacetate and pyruvate. We collected time-resolved measurements of malate production with cell-free expression of *E. coli* malate dehydrogenase (*mdh*) expressed from plasmid DNA with the p70 promoter. We found that in the case where *mdh* was overexpressed (+*mdh*), malate accumulated rapidly, achieving over 100% conversion within 10 minutes **(Figure S2A)**. Malate concentrations then rapidly decrease and reach baseline levels by 24 hours after the reaction start time in the +*mdh* condition. In the reaction without *mdh* overexpression (-*mdh*), we observed a gradual increase in malate levels over time **(Figure S2A)**. The difference between the +*mdh* and -*mdh* samples may be a result of the changes in reaction thermodynamics driven by differences in metabolite concentrations [32]. The +*mdh* reaction will have higher amounts of NAD^+ and malate early in the reaction due to the higher concentration of the *mdh* enzyme. This, in turn, could drive the *mdh* reaction in the oxidative direction and into the TCA cycle. It is worth noting that physiologically, the *mdh* reaction tends to run in the oxidative direction, despite the favorable thermodynamics for the reductive direction **(Figure S2B)** [33]. This pull towards oxaloacetate would result in a decrease in malate titers over time. High initial concentrations of malate could also have driven flux through the reversible fumarase enzyme, leading to a decrease in malate. Thermodynamic analysis predicts that the fumarase reaction prefers the formation of fumarate when malate concentrations are roughly 10-fold higher than fumarate (ΔG = -2.3 kJ/mol) **(Figure S2B)**.

Given that the CFE still maintains endogenous copies of TCA enzymes, the *-mdh* reaction may still be producing malate through the reductive *mdh* reaction, but at a slower rate than the *+mdh* reaction (as the reaction rate should be proportional to the enzyme concentration). Additionally, conversion of oxaloacetate to citrate is strongly favored thermodynamically (ΔG = -38.8 kJ/mol) **(Figure S2B)**, which could also lead to malate production via the oxidative TCA cycle in the case where *mdh* is not overexpressed.

We then questioned if coupling malate synthesis with *in situ* NADH regeneration would allow us to continuously drive flux towards malate in our CFE. To favor *mdh* towards malate, we incorporated a NADH regeneration system based on a previously reported formate dehydrogenase (*fdh*) mutant from *Starkeya novella* [34]. *fdh* can regenerate NADH and CO_2 from formate and NAD^+. Addition of NAD^+ and formate to CFE with *fdh* expressed resulted in near complete conversion of NAD^+ to NADH **(Figure 3A)**. We then found that upon addition of NADH and excess formate to reactions that had co-expressed *mdh* and *fdh*, malate concentrations were stable at 1000 μM, representing full conversion efficiency, up to 24 hours **(Figure 3A)**.

When *mdh* plasmid was omitted from the CFE, malate concentrations increased slowly over the reaction lifetime rather than accumulating quickly **(Figure S2)**. This is likely due to the presence of endogenous enzymes in the lysate that can natively produce malate. A proteomic analysis of *E. coli* cell-free lysate found that *mdh* and other TCA enzymes are among the most abundant enzymes in cell-free [30] **(Figure S3)**, supporting this hypothesis. Therefore, we questioned whether additional dilution of the

cell-free reaction following protein expression could improve the pathway-dependent production of malate by minimizing the effects of competing endogenous enzymes.

We tested three different dilutions of the CFE: 10 (our standard dilution factor), 50, and 200-fold. We found that the 200-fold dilution condition accumulated more malate than the other conditions when *mdh* was expressed **(Figure S4)**. The most likely explanation for the differences in malate production across the different dilution factors is that dilution of the cell-free lysate is more disruptive to the activity of competing TCA enzymes than it is to the overexpressed *mdh*. Dilution of the cell-free lysate results in the concentration of substrates and cofactors falling below relevant levels for competing enzymes with similar kinetic values. For example, citrate synthase (CS) has a similar reported Km for oxaloacetate as malate dehydrogenase (30 and 40 μM, respectively) [35,36]. However, CS requires acetyl-CoA as a substrate which would be present in concentrations far below the Km in the more dilute samples. NADH and ATP also serve as allosteric inhibitors of CS [37]. In the diluted reactions, the concentrations of NADH and ATP become much higher than the concentration of the CS enzyme, which would lead to a higher percentage of CS enzymes being bound and inhibited. Additionally, the dilution of the CFE may disrupt metabolon formation between TCA enzymes that accelerate malate loss [38]. We also found that the 200-fold dilution accumulated less malate than the other dilutions when *mdh* was not expressed **(Figure S4)**, suggesting the activity of native *mdh* has also fallen below relevant levels in those conditions.

Oxaloacetate, the direct C4 precursor of malate in our pathway, can be generated from C3 pyruvate and CO_2 using the enzyme pyruvate carboxylase (*pyc*) from *Rhizobium etli*. The *pyc* enzyme is a large enzyme complex made up of three distinct domains: the

ATP-consuming biotin carboxylase, carboxyltransferase, and biotin carboxyl carrier protein (BCCP) domain [39]. We coupled the pyruvate carboxylase reaction with *mdh* and *fdh* to produce malate from the C3 precursor, pyruvate **(Figure 3B)**. However, when we tested *pyc* activity from a 10-fold diluted CFE reaction, we observed no differential malate production in the conditions with and without *pyc*, with both conditions producing roughly 300 μM after four hours **(Figure 3C, S5)**. These results suggested that the engineered pathway may be outcompeted by the native carbon-catabolizing oxidative branch of the TCA cycle.

We hypothesized that oxaloacetate and pyruvate may be incorporated into the oxidative TCA cycle before being converted to malate by the heterologous pathway. Oxaloacetate and acetyl-CoA could be diverted into the endogenous oxidative TCA cycle by the enzyme citrate synthase, which is native to the lysate. The *pyc*-containing reactions are supplemented with acetyl-CoA, an allosteric activator shown to be essential for *pyc* activity [40]. We hypothesized that greater dilutions of the CFE may improve flux through our engineered pathway, as pyruvate carboxylase has a lower reported dissociation constant for acetyl-CoA than that of citrate synthase [36,40]. In the 10-fold diluted CFE condition, we did not see *pyc*-dependent malate production above the baseline levels generated by the endogenous lysate enzymes. In the 200-fold dilution condition, we observed increased *pyc*-dependent malate production compared to the baseline in the first hour of the reaction **(Figure S5)**, but after four hours the accumulated malate completely disappeared from both reactions **(Figure 3C)**. The large spike in malate may be due to an excess of NADH initially present in the system that drives flux in the reductive direction of *mdh*. From kinetic data, we found that titers of the

TCA cycle intermediates citrate and succinate increase over time in the reactions containing *pyc*, suggesting that the malate initially produced is converted to these compounds **(Figure 3B, S5B, S5C)**. Based on our previous thermodynamic analysis **(Figure S2)**, malate may be converted back to oxaloacetate, where oxaloacetate is then diverted into the oxidative TCA cycle by the citrate synthase enzyme.

Blocking TCA flux to improve pathway carbon-conservation

To minimize the loss of carbon through the oxidative TCA cycle, we investigated both targeted chemical inhibition of citrate synthase and re-routing of flux through overexpression of the glyoxylate shunt enzymes. Hydroxycitrate (HCT) is a potent competitive inhibitor of citrate synthase and ATP citrate lyase due to its structural similarity to citrate [41]. We tested the ability of HCT to improve flux through our engineered pathway by adding varying amounts of hydroxycitrate to cell-free reactions and measuring malate accumulation over time. In the highest HCT concentration tested (10 mM), we saw *pyc*-dependent accumulation of malate, reaching 212 ± 60 μM within four hours **(Figure S5A)**. Additionally, we measured the change in CO_2 concentrations in the headspace of our reactions, but found that this was not a reliable readout of reaction progress **(Figure S6)**.

In the reactions containing HCT, we also found that concentrations of citrate and succinate remain constant relative to reactions with no pathway expressed **(Figure S5B, S5C)**, suggesting that flux through the oxidative TCA cycle is being reduced effectively**.** Additionally, we observed accumulation of acetyl-CoA in the condition with the highest HCT concentration, implying that citrate synthase inhibition can prevent oxaloacetate consumption via the TCA cycle in cell-free reactions **(Figure S7A)**. When

combined, 200-fold dilution of the lysate and HCT inhibition lowered the background activity over 6-fold and improved pathway-dependent malate production 2.5-fold compared to 10-fold diluted reactions without TCA cycle inhibitor *($p < .05$)* **(Figure 3C) (Table 1).** In the conditions with 200-fold dilution and HCT inhibition, we see high concentrations of citrate that do not change depending on the presence of the pathway, indicating that the TCA cycle is blocked at this point **(Figure S7B)**. Additionally, we see no changes in succinate across any of the diluted conditions, suggesting that activity through the oxidative TCA cycle is low **(Figure S7B)**. Therefore, we believe that malate production via the oxidative TCA cycle, and therefore CO_2 loss through the TCA cycle, are low.

As a second strategy for minimizing carbon loss, we tested whether overexpression of the glyoxylate shunt enzymes could redirect oxaloacetate flux that enters the TCA cycle towards malate. The glyoxylate shunt consists of two enzymes, isocitrate lyase (*ICL*) and malate synthase (*MS*), to convert isocitrate and acetyl-CoA to succinate and malate without loss of any carbon [44]. We first confirmed that both enzymes in the glyoxylate shunt were active when fed with their respective substrates **(Figure S8A, S8B).** We then expressed the glyoxylate shunt genes along with the rTCA pathway. Here, we found that the malate produced when starting from pyruvate was lower with the glyoxylate shunt than without **(Figure S8C).** This may be due to the limited gene expression capacity of the cell-free system, or due to competition for acetyl-CoA between pyruvate carboxylase, which uses acetyl-CoA as an allosteric activator, and malate synthase. Based on these results**,** the glyoxylate shunt enzymes were not utilized in any of the studies described in the following sections.

Balancing cofactor and expression requirements to construct multi-enzyme pathways

In the previous section, we showed that inhibiting citrate synthase effectively blocked TCA flux, allowing malate to be produced from pyruvate through the carbon-conserving rTCA pathway. To explore approaches for producing malate from simpler substrates, we aimed to extend our pathway to start from serine, which can be synthesized from the C2 substrate glycine and the C1 substrate formate via the rGCV pathway. Yu and Liao previously showed that serine can be efficiently converted into pyruvate through the heterologous expression of serine dehydratase (*sda*) in *E. coli* [45]. We screened several isoforms of the *sda* enzyme from a variety of prokaryotic organisms in our standard CFE conditions and observed no catalytic activity **(Figure 4A)**. As most prokaryotic isoforms of *sda* contain an oxygen-sensitive iron-sulfur cluster [44], we attempted to obtain enzyme function by supplementing iron and reducing agents, including dithiothreitol (DTT) and ascorbic acid at the time of expression. However, we still observed no clear activity from any of the *sda* isoforms screened **(Figure 4A)**.

We next tested the promiscuous serine to pyruvate conversion activity of a pyridoxal 5'-phosphate (PLP) dependent threonine dehydratase (*tdcB*) from *E. coli* [45]. By adding exogenous serine into 200-fold diluted reactions, we were able to confirm serine-dependent *tdcB* activity, as *tdcB* expression produced 384 ± 101 μM pyruvate from 1 mM serine, compared to 65 ± 13 μM without *tdcB* expressed **(Figure 4B).** In the 10-fold diluted reaction condition, we did not observe differential production of pyruvate between the conditions with or without *tdcB* **(Figure S9)**. Based on these results and those in the previous section, we employed the 200-fold dilution condition for all subsequent reactions.

Having demonstrated *tdcB* activity in a single-step transformation, we then integrated *tdcB* with the rTCA pathway to produce malate from serine. We measured malate production in four reaction conditions with and without *tdcB* and/or *pyc*, with all conditions containing *mdh* and *fdh*. At four hours, we saw no malate production across all conditions. However, we observed pyruvate accumulation in the conditions without *pyc* (**Figure S10)**, suggesting *pyc* was effectively pulling pyruvate flux forward. By eight hours, malate titers reached 135 μM when both *pyc* and *tdcB* were expressed, compared to no malate accumulation in the conditions without both enzymes expressed **(Figure 4C).**

In multi-enzyme cell-free systems, the use of enzymes with numerous cofactor requirements can reduce pathway efficiency and increase costs [11,46]. Incorporation of ATP-regenerating systems in CFE can reduce cofactor requirements; however, we hypothesized that the addition of AMP as an allosteric activator for *tdcB* may interfere with ATP-dependent enzymes in the CFE [9,47]. Therefore, we aimed to identify an alternative to *tdcB* that was insensitive to AMP regulation. From the literature, we extracted sequence motifs from *tdcB* known to interact with AMP [48]. We then used multiple sequence alignment to choose other PLP-dependent enzymes that lacked these motifs and were predicted to be AMP-insensitive **(Figure 4D, Methods S1)** [49]. From this screen, we selected a serine dehydratase (*sds*) variant from the soil-dwelling amoeba *Dictyostelium discoideum* to characterize in CFE [50]. In the presence of AMP, the *sds* variant from *D. discoideum* produced similar levels of pyruvate as compared to *tdcB* from *E. coli* **(Figure 4E)**. In reactions without AMP, *sds* achieved 100% conversion of serine to pyruvate after only four hours, while production from *tdcB* was

indistinguishable from the background **(Figure 4E)**. We then tested the ability of each isoform to convert serine to malate via the rTCA cycle. Here, we found that *sds* resulted in 2-fold higher malate concentrations compared to *tdcB* in an eight hour reaction **(Figure 4E)**. Based on these results, subsequent pathway engineering was done with *sds* in place of *tdcB.*

Connecting the reductive TCA cycle to formate assimilation products

In a fully integrated system for carbon-conserving malate production, the combined formate assimilation, rGCV, and rTCA pathways would fix 2 CO_2 equivalents from formate and 2 from bicarbonate per malate produced. Glycine, the C2 product of the combined formate assimilation and rGCV pathways, can be converted to serine via the enzyme serine hydroxymethyltransferase (SHMT) encoded by the *glyA* gene from *E. coli*. *glyA* utilizes 5,10-CH_2-THF as a C1 donor to assimilate a second formate-derived carbon into glycine **(Figure 1)**. However, creating conditions for the rGCV pathway to run efficiently in the direction of glycine synthesis requires significant optimization of gene ratios and reaction tuning [51-55]. In this section, we prototype *glyA* activity and estimate the efficiency of carbon incorporation into malate from glycine and 2 CO_2 equivalents, 1 from formate and 1 from bicarbonate.

When fed with equimolar glycine and 5,10-CH_2-THF, *glyA* converted 27% of the glycine into serine in four hours (269 ± 41 μM) in a 200-fold diluted cell-free reaction **(Figure 5A)**. By incorporating *glyA* and *sds* with the rTCA genes, pathway-dependent malate production was observed directly from C2 glycine **(Figure 5B, top)**. Individual starting metabolites (pyruvate, serine, and glycine), were fed to batch reactions to analyze the efficiency of malate production from each step of the pathway. From pyruvate, we saw 1075 ± 36 μM malate in the presence of the pathway and 578 ± 113

μM with no pathway **(Figure 5B, left)**. When fed serine, the pathway produced 11-fold higher malate concentrations compared to the no-pathway control (273 ± 60 μM vs. 24 ± 8 μM) **(Figure 5B, left)**, suggesting the basal conversion of serine to pyruvate is low in CFE. From glycine, we saw 117 ± 6 μM malate produced with- and 62 ± 3 μM without- the pathway **(Figure 5B, left).**

We estimated the amount of CO_2 conserved in the engineered system using the pathway-dependent yield from the reactions fed with pyruvate. To calculate pathway-dependent yield, we took the difference in malate production with and without the engineered pathway. Then, to estimate conserved CO_2, we assumed a fixed reaction stoichiometry between malate produced and CO_2 equivalents conserved. Specifically, our calculations specify that for each pyruvate converted to malate, a maximum of 1 CO_2 equivalent can be fixed from bicarbonate via rTCA, and a maximum of 2 CO_2 equivalents can be released via the oxidative TCA cycle **(Table 1)**.

Given the pathway-dependent differences in malate titers measured above, we estimate that the engineered pathway conserved 497 μM of CO_2 compared to the oxidative TCA cycle when feeding pyruvate. This represents 50% of the potential conservation, assuming all of the malate accumulated above basal levels is produced through the engineered rTCA cycle **(Table 1).** The net result is that the introduction of the rTCA pathway increased malate yields by roughly two-fold (1075 vs. 578 μM malate) while reducing the estimated carbon loss as CO_2 from the oxidative TCA cycle by 43% (0.66 vs. 1.2 mM CO_2) **(Figure 5B, right) (Table 1).** Because the basal conversion of serine and glycine to malate from the endogenous lysate metabolism is low **(Figure 5B, left)**, we expect that the reduction in carbon loss remains similar when extending the engineered pathway to start from intermediates upstream from pyruvate.

Importantly, these calculations are meant to represent the CO_2 that is conserved relative to production through an un-engineered, plain cell-free system. Our system still contains many native CO_2-releasing enzymes involved in the production of alcohols, fatty acids, and other secondary metabolites [30]. Therefore, we focus our analysis specifically on the steps within the TCA cycle that are carbon-releasing or carbon-conserving. It is worth noting that previous work from other groups has shown that cell-free lysates can be further optimized to selectively downregulate or remove competing enzymes through heat inactivation, protein tagging, and CRISPR-based engineering [18, 19, 21], all of which would be viable strategies to avoid CO_2 loss in an engineered lysate.

Table 1: Carbon conservation for malate production through the rTCA pathway

	Through TCA	**Through rTCA**
CO_2/malate (mmol/mmol)	2	-1
Malate yield (mmol/mmol)	.58	.5
Carbon loss from pyruvate (mmol/mmol)	1.16	-.5
Total carbon loss (mmol/mmol)	1.16	1.16 - .5 = .66
Carbon loss (% of maximum)	$\frac{1.16}{2} \times 100 = 58\%$	$\frac{.66}{2} \times 100 = 33\%$
Reduction in carbon loss	$\frac{58 - 33}{58} \times 100 = 43\%$	

Table 1: Carbon conserved through the rTCA pathway is calculated for the production of malate from pyruvate. Data is shown in Figure 5B. The calculations assume a fixed stoichiometry for the moles of CO_2 released per mole of malate generated through either the TCA or rTCA cycles.

Prototyping cofactor regeneration to enable malate synthesis from formate

To construct the complete pathway for the conversion of formate and glycine to malate, 7 cofactors are required **(Figure 1)**, 4 of which are consumed during biosynthesis (ATP, NADPH, NADH, and THF). In the previous section, we showed that glycine and 5,10-CH_2-THF could be used to generate malate. In this section, we prototype the formate assimilation pathway for 5,10-CH_2-THF regeneration to produce malate from glycine and formate directly. 5,10-CH_2-THF can be generated from THF and formate in three enzymatic steps catalyzed by the ATP-dependent formate-THF ligase (*ftl*), methenyl-THF cyclohydrolase (*fch*), and the NADPH-dependent methylene-THF dehydrogenase (*mtdA*), all from *M. extorquens* [56] (**Figure 6A, left**). We previously saw that *fdh*-mediated regeneration of NADH was crucial for malate accumulation from the rTCA pathway. Accordingly, we prototyped a polyphosphate kinase (*ppk*) ATP regeneration system and a NADPH regenerating system employing *fdh* mutants with engineered specificity towards NADPH to regenerate cofactors consumed by the formate assimilation pathway.

To generate the biocompatible C1 donor 5,10-CH_2-THF *in situ*, we coupled the reductive TCA module with the three-enzyme formate assimilation pathway. In total, this represents an eight-enzyme pathway for malate production, all directly expressed from DNA in a one-pot reaction. Previously, we have assimilated formate into glycine and serine using the formate fixing module from *M. extorquens* in combination with the reverse glycine cleavage complex [29]. In this study, overexpression of the *M.extorquens* genes *ftl*, *fch* and *mtdA* for the formate assimilation pathway in the CFE improved 5,10-CH_2-THF accumulation 1.6-fold compared to lysate alone, reaching 178 ± 35 μM from 1 mM THF and 10 mM formate after four hours **(Figure 6A).**

To regenerate ATP, we first tested a polyphosphate kinase (*ppk*) from a previously published *Erysipelotrichaceae* bacterium [57] using inexpensive polyphosphate as a phosphate donor. The ppk used in this system is a Class III ppk2 enzyme, which has been shown to catalyze phosphorylation of both nucleoside mono- and di-phosphates [57]. We found that, when expressing *ppk* in CFE, we could generate 742 ± 3 μM ATP from 1000 μM AMP and 10 mM hexaphosphate in a four hour reaction **(Figure S11)**. However, when we incorporated *ppk* along with the biosynthetic pathway for serine to malate conversion, we found that ATP regeneration did not improve titers over an eight hour reaction. Therefore, we concluded that ATP availability is not limiting flux through the rTCA pathway in these conditions **(Figure S11).** We tested ATP regeneration in the ATP-dependent formate assimilation pathway under standard conditions (1 mM ATP, 1 mM NADPH, 1 mM THF) and found that 5,10-CH_2-THF accumulation was abolished, despite the ATP regeneration system itself remaining functional **(Figure S11).** This result indicates that the ATP regeneration system introduces pathway incompatibilities in our system, and so it was excluded in future experiments.

To regenerate NADPH consumed by *mtdA* during formate assimilation, we tested two NADPH-selective *fdh* variants for orthogonal NADH and NADPH regeneration *in situ*. Disappointingly, we were not able to see NADPH accumulation in reactions fed with 1 mM $NADP^+$ after screening variants from *A. thaliana* and *Candida methylica* **(Figure S12)** [58,59]. This may be due to poor expression or low activity of these enzymes, resulting in them not being able to compensate for native NADPH consumption in the lysate. Thus, our final reaction set-up for cell-free malate production consisted of the

eight-enzyme system integrating rTCA, formate assimilation, and NADH regeneration (**Figures 1A and 6B**).

We applied the formate assimilation pathway for 5,10-CH_2-THF regeneration to produce malate from C2 glycine and the C1 compounds formate and bicarbonate directly. We co-expressed the formate assimilation module with the rTCA cycle and *fdh* for NADH regeneration. Reactions fed with 10 mM formate, 10 mM bicarbonate, 1 mM THF, and 1 mM glycine produced 64 ± 37 μM of malate in an eight-hour reaction **(Figure 6B),** which is about half of the malate produced from 1 mM 5,10-CH_2-THF fed in directly with 1 mM glycine (117 ± 6 μM) **(Figure 5B)**. This decrease in malate production is likely due to low efficiency through the formate assimilation pathway. Conversion of THF to 5,10-CH_2-THF through the formate assimilation pathway proceeded with only 18% conversion efficiency after four hours **(Figure 6A).** Therefore, less 5,10-CH_2-THF was available as a substrate for *glyA*, which, in turn, led to decreased malate production. The reactions coupling formate assimilation, rTCA cycle and *fdh* also produced 67 ± 13 μM succinate and .47 ± .14 mM citrate as side products **(Figure S13)**. With 2 CO_2 equivalents fixed per malate produced, integrating the formate assimilation pathway with the rTCA cycle process enabled incorporation of 0.13 moles of CO_2 equivalents per mole of glycine fed.

Techno-economic analysis of cell-free bioproduction of malate

Despite a number of recent successes of large-scale cell-free biomanufacturing [60-62], there remains a perception that the commercial realization of CFE as a metabolic engineering platform is necessarily limited by the costs of catalyst preparation and chemical requirements. To evaluate the economic feasibility of industrializing cell-free

bioproduction, we used our platform for malate bioproduction from formate, bicarbonate, and glycine as a basis for performing techno-economic analysis (TEA). TEA is a valuable tool to provide insights into the challenges and opportunities to reduce costs associated with bioproduction.

The use of low-cost feedstocks and cofactor regeneration systems will be crucial to achieving economically competitive cell-free metabolic engineering [11]. In our previous work, we have shown that glycine and serine can be produced using cell-free based biocatalysts solely from formate through the rGCV pathway with a yield on formate of up to 30% [29]. In this work, we showed malate production from serine with a yield of 27% **(Figure 5B)**. With further optimization, it should be possible to integrate the rGCV and rTCA pathways in CFE and enable malate production from formate directly. Therefore, our TEA assumes that malate can be produced using formate and bicarbonate as the primary feedstocks, with 20% conversion efficiency of formate as the base assumption. Additionally, we demonstrate that enzymatic recycling of NADH and THF can be used to reduce the need for exogenous cofactors; and thus our TEA assumes a continuous process with the ability to regenerate crucial cofactors *in situ*.

The chemical process block diagram is given in **Figure 7A** and base case assumptions are defined in **Table S2**. Our TEA uses the "n^{th}" plant assumption, which assumes that similar plants have already been established and therefore does not account for excess costs associated with a first-time process (**Methods S2)**. The TEA revealed that malate produced from formate via a cell-free process could reach a minimum selling price (MSP) of $9.6/kg with the given assumptions **(Figure 7B, Table S3, Methods S2)**. This price is only 5.4-fold higher than the current market price for

fossil-fuel derived malate ($1.8/kg) [63], suggesting that further technological improvements could result in cost-competitive bioproduction. Additionally, we considered the case where formate and glycine are co-utilized as inputs for the cell-free production of malate. In this case, our analysis showed that the cost of malate production only increased by 8%, up to $10.4/kg **(Figure 7B)**.

We found that operating costs, specifically the cell-free reagent price, are the most significant cost drivers **(Figure 7B)**. Process costs of bioproduction can help be offset by co-purifying and selling valuable co-products. In our system, we saw consistent accumulation of succinate and citrate **(Figure S13)**, organic acids with large defined markets. Therefore, our analysis calculates the MSP of malate by summing the input costs and subtracting the price of the sellable co-products **(Figure 7B)**. The cost of the cell-free lysate at $120/L[11] accounts for nearly 100% of the process costs, with feedstock accounting for roughly 10% of the total cost **(Figure 7B)**. This is in stark contrast to traditional cell-based bioproduction platforms, in which feedstock costs account for up to 60% of total costs [65].

Cell-free systems have the potential for near-stoichiometric conversions due to the high control over reaction conditions. Therefore, improving conversion efficiency is a promising strategy to help CFE reach economic viability in coming years. Our experimental results indicated that there is substantial diversion of pyruvate to other metabolites (**Figure 2)**, suggesting that conversion efficiencies and the MSP could be improved through lysate engineering and reaction optimization to reduce loss of carbon through competing pathways. Our analysis shows that improving conversion efficiency

of this proof-of-concept system to 95% would be needed to enable a competitive MSP **(Figure 7C).**

Beyond optimizing cell-free reaction conditions, improvements in other technical areas will be crucial to lower costs associated with cell-free metabolic engineering and bioproduction from renewable formate. Cell-free lysate preparation is a major driver of the cost for cell-free bioproduction. We found that reducing the cost of lysate preparation from $120/L to $25/L alone could reduce the MSP of malate to $2.0/kg **(Figure 7C)**, making it competitive with petroleum-derived production. As the price of labor is the primary cost-driver of cell-free preparation, scaling its production could help reduce associated costs [12]. The efficiency of electrochemical CO_2 conversion to formate is expected to increase by up to 50% in coming years [66,67], which could further reduce the production cost of malate to $2.1/kg **(Figure 7C)**. Metabolic engineering efforts to produce higher value compounds, such as fatty acids derived from acetyl-CoA or other organic acid derivatives, would also permit a higher MSP. While not economically competitive in its current state, this TEA shows promise for the sustainable production of platform chemicals using CFE.

Discussion

The use of formate and other renewable C1 feedstocks has great potential for sustainable and low-cost bioproduction of commodity chemicals [1,7,66,68]. Here, we demonstrate production of the central TCA metabolite, malate, using an “off-the-shelf” CFE without additional genetic engineering. These are widely accessible through commercial vendors and have been broadly adopted for genetic circuit engineering and enzyme overexpression and prototyping [69-72]. While there are many examples of malate

production in microbial systems via the rTCA cycle, these typically still utilize refined glucose as a feedstock and introduce significant growth defects to the engineered strain [26,73,74]. Additionally, naturally C1-assimilating organisms tend to be difficult to engineer and are limited in their uptake and incorporation of C1 feedstocks [7]. In this work, we use unique aspects of CFE to enable carbon-conserving production of the industrial di-acid malate that would be very difficult to achieve in microbial cell-based fermentation. Specifically, we were able to 1) minimize background reactions by diluting the CFE, 2) minimize flux through carbon-inefficient oxidative TCA with no strain engineering by introducing chemical inhibitors, and 3) overcome cellular transport limitations to utilize formate as a substrate by using CFE. Collectively, these aspects of CFE allowed us to redirect flux, minimize background reactions, and maintain stable titers of malate while conserving 43% of the carbon that would have been otherwise lost via the TCA cycle. The approaches developed in this work should be widely applicable to metabolic engineering efforts for other small molecules in CFE. This work establishes a benchmark for the potential of CFE for carbon-conserving prototyping and metabolic engineering.

Decoupled enzyme expression and bioconversion is not a new idea to the field of bioproduction [75]. For cell-based fermentation, this often involves growing the cells to a sufficient biomass before inducing heterologous protein expression and bioconversion. Traditionally, purified enzyme systems have decoupled these stages by separate expression and enrichment of individual enzymes before combining biocatalysts and chemicals for bioconversion [14,15,52]. By contrast, we achieve decoupling in a one-pot reaction mixture, in which all enzymes are expressed from DNA in a single reaction

before being combined with substrates and cofactors. Additionally, our lysate-based cell-free reaction operates efficiently with roughly 100-fold less enzyme than comparable *in vitro* systems [14–16] (**Table 1**).

While one-pot enzyme expression and bioconversion reduces process complexities, it faces the challenge of competition from endogenous enzymes left in the CFE. Enzymes from the TCA cycle have been shown to be highly active in lysates prepared from *E. coli* [76], creating a challenge for efficiently routing carbon flux into desired products. This represents a challenge for efficient cell-free bioproduction, as the TCA cycle is the major source of carbon loss in *E. coli,* typically accounting for over 50% of all CO_2 loss in the cell [77]. By utilizing the reductive TCA cycle, one CO_2 equivalent is incorporated per malate produced from pyruvate. By avoiding decarboxylations through the oxidative TCA cycle, an additional 2 mol CO_2/mol malate is conserved. Starting from pyruvate, our system reduced carbon loss through the TCA cycle by an estimated 43% compared to malate that was produced by the endogenous metabolism of the lysate **(Table 1, Figure 5B).** We also show that an additional CO_2 equivalent can be incorporated into malate via formate and glycine by integrating the formate assimilation pathway with the reductive TCA cycle.

Crucial to these accomplishments were the findings that CFE can be diluted following protein expression and that inexpensive inhibitors added directly to reactions to reduce competition from residual enzymes present in the lysate. Specifically, we found that the inhibition of citrate synthase following CFE dilution improved our malate titers 14-fold in a four hour reaction **(Figure S5)**. Dilution of the CFE presents an efficient method to reduce the effect of endogenous enzymes and study engineered

pathways that would not be possible in a microbial system. Direct addition of inhibitors into a bioproduction reaction would also likely not be feasible in a microbial host, as the inhibitor may introduce toxicity or be limited in its uptake through the cell membrane. Together, these interventions improved pathway-dependent bioproduction and reduced loss of carbon through CO_2.

The utility of CFE for rapid prototyping and optimization of *in vivo* metabolic engineering programs has been demonstrated widely [69,70,78]. The development of CFE as a metabolic engineering platform for carbon-conserving biosynthesis, however, faces many systems-level challenges. For one, the utility of cell-free reactions has been limited by the inability to regenerate necessary cofactors and the loss of activity from pathway enzymes. In our system, we found that malate accumulation could be extended up to 24 hours, compared to 30 minutes, by diluting the CFE following gene expression and regenerating NADH to drive flux towards malate **(Figure 2A)**. However, we found that malate titers peaked at 10 hours when starting from serine, and then decreased until reaching baseline levels **(Figure S10),** suggesting that the engineered pathway had lost function, or that the produced malate is siphoned off to other reactions in the cell lysate. We also found that *in situ* ATP regeneration did not improve malate yields in our system **(Figure S11)**, although efficient regeneration of ATP may be necessary to drive flux in other engineered pathways. Broadly, ATP regeneration systems are well studied for cell-free gene expression [79-81], while the development of generalizable, cost-effective regeneration systems applicable to biosynthetic pathways are still needed [11,82]. Traditionally, *in vitro* reactions have relied on supplementing purified enzymes and cofactors throughout the reaction to extend lifetimes [14,15]. Notably,

our system reaches comparable reaction lifetimes to previous *in vitro* systems without the supplementation of additional enzymes and cofactors. Generalizable strategies to improve enzyme stability and regenerate cofactors *in situ* will greatly reduce the barriers to achieving cell-free bioproduction at scale.

Other considerations to optimize engineered pathways in CFE are the balancing of cofactor compatibilities, reaction conditions, and enzyme activities, among other parameters [8,18,83]. As the number of steps in an engineered pathway increases, incompatibilities with respect to cofactor and reaction condition requirements can often arise [15,17,45]. In this work, we found that rational screening of enzyme isoforms could help alleviate pathway incompatibilities to construct larger pathways **(Figure 4D, 4E).** Additionally, some degree of reaction screening in different buffers and DNA template ratios was necessary to construct our eight-enzyme pathway **(Figures S14-S17)**. The performance of this system with minimal optimization demonstrates the high potential for carbon-conserving pathway engineering with high efficiency in CFE. A much more exhaustive and systematic survey of reaction conditions would be necessary to truly optimize the pathway performance. While the high dimensionality of the design space makes predicting optimal conditions *a priori* difficult, model-driven engineering may help bridge experimentation and machine learning to accelerate systems-level optimization of bioproduction platforms [84–86]. Furthermore, the incorporation of isotope labeling experiments along with mechanistic and machine learning models would greatly advance our understanding of carbon utilization and the behavior of complex enzyme networks in cell-free systems.

For our base case TEA analysis, we assumed 100% cofactor recycling, which otherwise accounts for 97% of the cost for a bench-top scale CFE conversion of formate to serine [29]. Significant improvements in cofactor regeneration will be needed to be able to negate such large costs. The drop in malate concentrations after 10 hours **(Figure S10)** signifies the need for additional product removal unit operations for a successful transition from batch production to a continuous production modelled in our TEA analysis. Taking these caveats into account, our work demonstrates an innovative carbon-conserving alternative for large scale bioproduction of malic acid.

The global malic acid market is estimated to be around 100,000 metric tons annually, produced almost entirely from petrochemicals [87]. Cell-free production of malate using electrochemically generated formate has the potential to reach a higher carbon efficiency than microbial production, which has failed to achieve commercial scale to date. Less than 10% of all CO_2 waste streams generated by U.S. bioethanol plants would be needed to satisfy the global malic acid market, assuming just a 5% yield on formate [90]. In turn, this would avoid roughly 400,000 tons of CO_2 emissions annually associated with traditional synthesis [91]. Our work demonstrates construction of carbon negative pathways by assimilation of formate and bicarbonate, as well as carbon conservation steps in the rTCA cycle to produce central metabolites such as pyruvate and malate, which themselves can be a starting point for other metabolic pathways. Cell free biocatalysis allows metabolic pathways to be stitched together in a modular nature by addition of appropriate enzymes. Pyruvate, an intermediate in our engineered pathway, can be routed through acetyl-coA to produce a variety of commodity chemicals such as fatty acids, terpenes, alcohols or polymers [19, 82, 86]. Given the rapid growth of the

field in recent years, cell-free bioproduction from renewable formate may soon be able to address key limitations of microbial fermentation and replace petroleum-based chemical production.

Methods

1. Plasmid Preparation

Plasmids expressing pathway genes were cloned in *E. coli* NEB Turbo cells. All PCR amplification of genomic DNA used Phusion DNA polymerase. Primers were synthesized by IDT and gBlocks were synthesized by Twist Biosciences. Both primers and gBlocks were resuspended with nuclease-free water. Plasmid assembly was achieved using 5X In-Fusion HD mastermix (Takara). Assembled plasmids were plated onto LB-agar plates with 100 µg/mL carbenicillin. Transformed cells were grown overnight at 30 °C. Single colonies were picked from plates and grown overnight in LB shaking at 30 °C with 100 µg/mL carbenicillin.

Plasmids were isolated from subcultures using a DNA miniprep kit (QIAprep Spin Miniprep Kit) and sequenced with Sanger (Genewiz inc.) or full-plasmid sequencing (Primordium) to verify correctly assembled plasmids. Plasmids were grown in culture volumes of ~100 mL to ensure adequate yields for multiple cell-free reactions. Plasmids were further purified using a PCR purification kit (Invitrogen PureLink, Cat. K310001) and eluted with nuclease-free water. Plasmid concentrations were quantified via spectrophotometry (Nanodrop 2000c, Cat. ND-2000C).

2. Cell-Free Protein Synthesis Reactions

The cell-free system was acquired from Arbor Biosciences (myTXTL). 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. For reactions containing three or fewer genes, reactions were assembled on ice from the CFE, purified DNA, and necessary cofactors. The CFE was pipette mixed and added to each PCR tube in 7.5 µL for a final volume of 10 µL. These PCR tubes were incubated overnight at 30C. For reactions involving more than three genes, plasmids and cofactors were mixed with an acoustic liquid handler robot (Echo Labcyte 525) into Labcyte 384-well destination plates (001-14555). The 384-well plates were then incubated at 30°C overnight.

3. Cell-Free Bioproduction Reactions

Cell-free bioproduction reactions were mixed in 25 µL containing 2.5 µL of CFE-expressed enzymes. CFE-expressed enzymes were diluted in 10 mM Tris pH 8 prior to adding if they were diluted beyond 10-fold in the final reaction. Bioproduction reactions were done in 50 mM HEPES pH 8. For reactions containing three or fewer enzymes, reactions were assembled by hand from the CFE-expressed enzymes and necessary substrates and cofactors. Detailed information about sample preparation can be found in Methods S3. For reactions involving more than three enzymes, enzymes and chemicals were mixed with an acoustic liquid handler robot (Echo Labcyte 525) into 96-well V-bottom plates (Costar, Cat. 3363). The plates were sealed with a foil adhesive (Thermo, AB0626) and the reactions were run for 4 - 8 hours at room temperature. In

general, reactions containing three or fewer enzymes were run for four hours, while larger reactions were run for eight hours. This is due largely to the slow reaction progress of the *pyc* enzyme, which led to us allowing longer reactions when using *pyc* as an intermediate step. At the end of the reaction, samples were quenched with 2 volumes of acetonitrile to denature proteins. The entire sample volume was then filtered using 96-well .2 µm Supor membrane plates (Cytiva: 8019) and spun for 15 min at max speed. Samples were then either analyzed via LC/MS or stored at -20C until ready for analysis.

4. Metabolite quantification with LC/MS

Samples were analyzed via Agilent 6530 LC/Q-TOF in negative mode using a BEH Amide 50 mm column (Waters, 186004800). Standard curves were prepared by spiking known amounts of metabolites into diluted CFE and HEPES. For LC/MS, the aqueous phase was LC/MS grade water and the organic phase was 95/5 acetonitrile/water with 10mM ammonium acetate and .04% v/v ammonium hydroxide. The % aqueous/organic gradient was run as follows: hold at 5/95 for 2.5 minutes, move to 33.5/66.5 over 5 minutes, 40/60 over 1 minute, hold 40/60 for 1 minute, then return to 5/95 over 1 minute. The flow rate was held at 0.5 mL/min.

5. Statistics

Statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks in Figures indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.005).

Figures

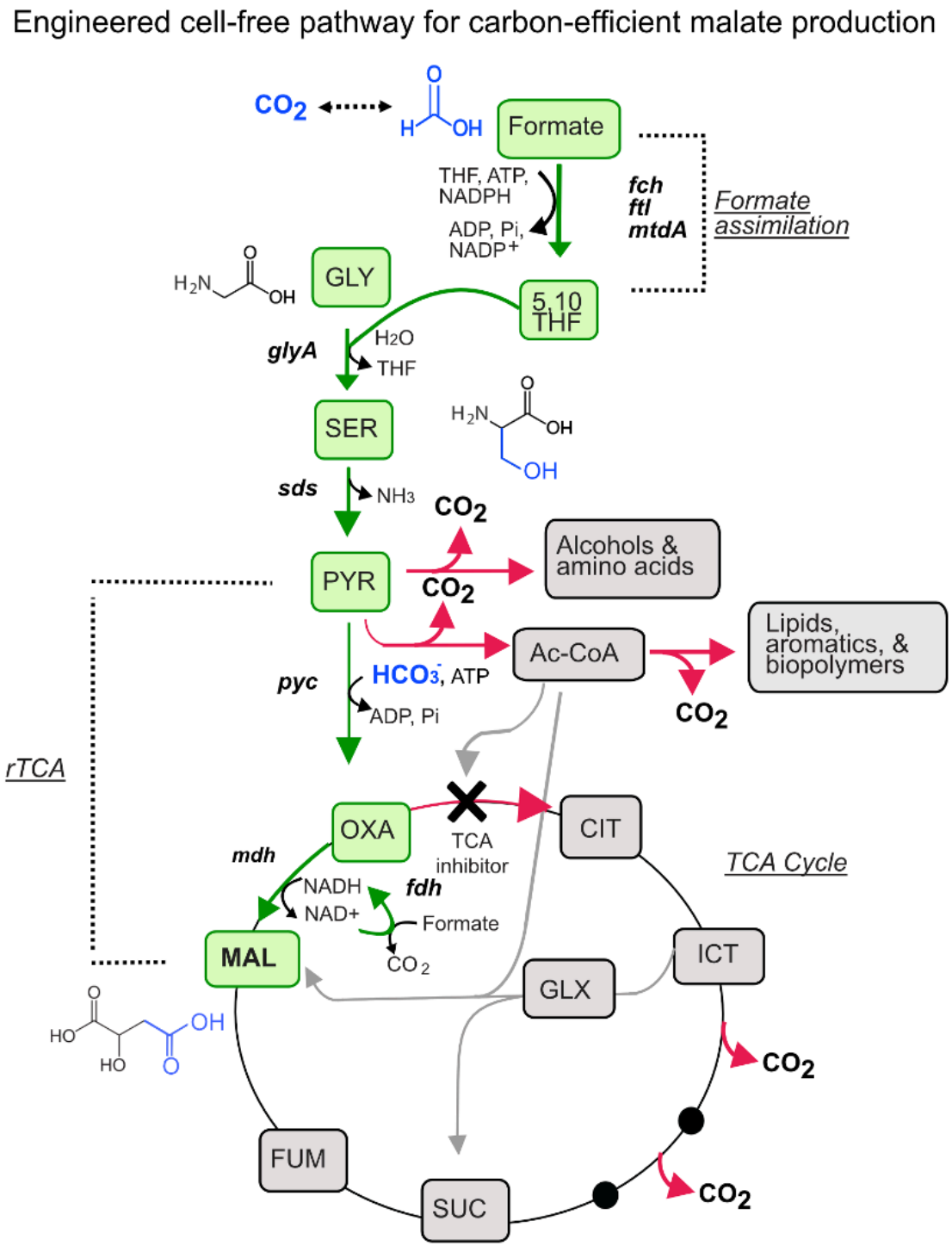

Engineered cell-free pathway for carbon-efficient malate production
CO2
Formate
THF, ATP, NADPH
ADP, Pi, NADP+
fch
ftl
mtdA
Formate assimilation
GLY
5,10 THF
glyA
H2O
THF
SER
sds
NH3
CO2
Alcohols & amino acids
PYR
CO2
Ac-CoA
Lipids, aromatics, & biopolymers
CO2
pyc
HCO3-, ATP
ADP, Pi
rTCA
OXA
TCA inhibitor
CIT
TCA Cycle
mdh
NADH
NAD+
fdh
Formate
CO2
MAL
GLX
ICT
CO2
FUM
SUC
CO2

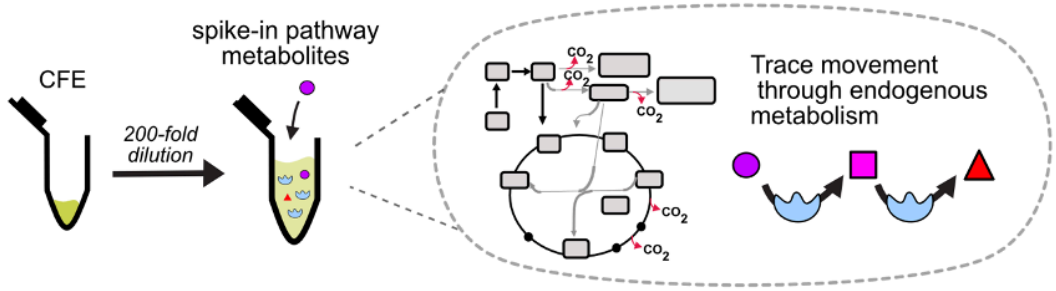
Figure 1. Overview of the *E. coli*-based cell-free system for malate production

A. Pathway diagram for conversion of formate and glycine into malate. Formate is incorporated into glycine to form serine via the four-step formate assimilation pathway. Serine is then converted to malate via the three-step reductive TCA cycle. High NADH concentrations are maintained by *in situ* regeneration from *fdh* to bias *mdh* flux towards malate.

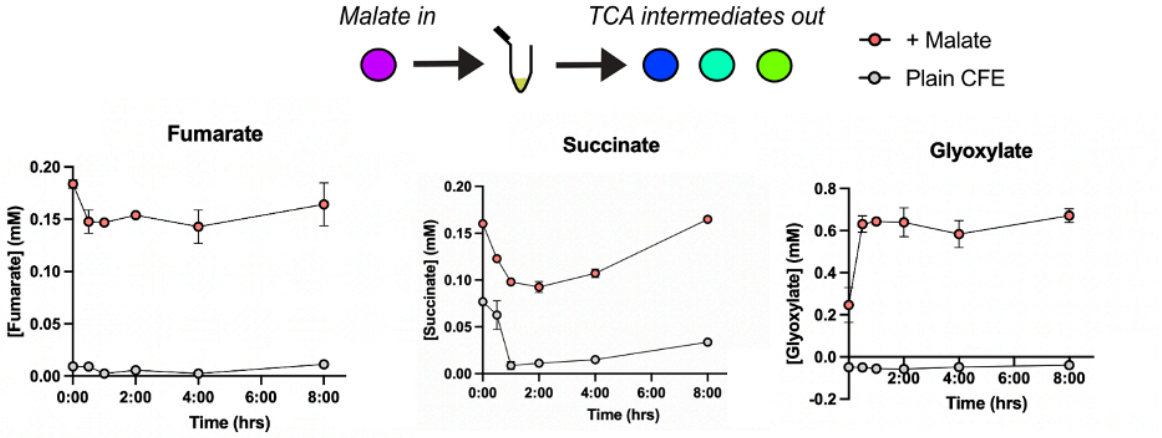
Gene abbreviations: *ftl*, formate-tetrahydrofolate ligase; *fch*, methenyltetrahydrofolate cyclohydrolase; *mtdA*, methylenetetrahydrofolate dehydrogenase ($NADP^+$); *glyA*, glycine hydroxymethyltransferase; *sds,* serine dehydratase; *pyc*, pyruvate carboxylase; *mdh,* malate dehydrogenase; *fdh*, formate dehydrogenase.

Metabolite abbreviations: THF, tetrahydrofolate; 5,10-THF, 5,10-methylene tetrahydrofolate; GLY, glycine; SER, serine; PYR, pyruvate; OXA, oxaloacetate; MAL, malate; SUC, succinate; ICT, isocitrate; CIT, citrate; GLX, glyoxylate; FUM, fumarate; Ac-CoA, acetyl-CoA.

A. Analyzing the endogenous proteome of *E. coli* lysate-based CFE



B. Tracing the movement of malate through the TCA cycle



C. Endogenous interconversion of pathway metabolites

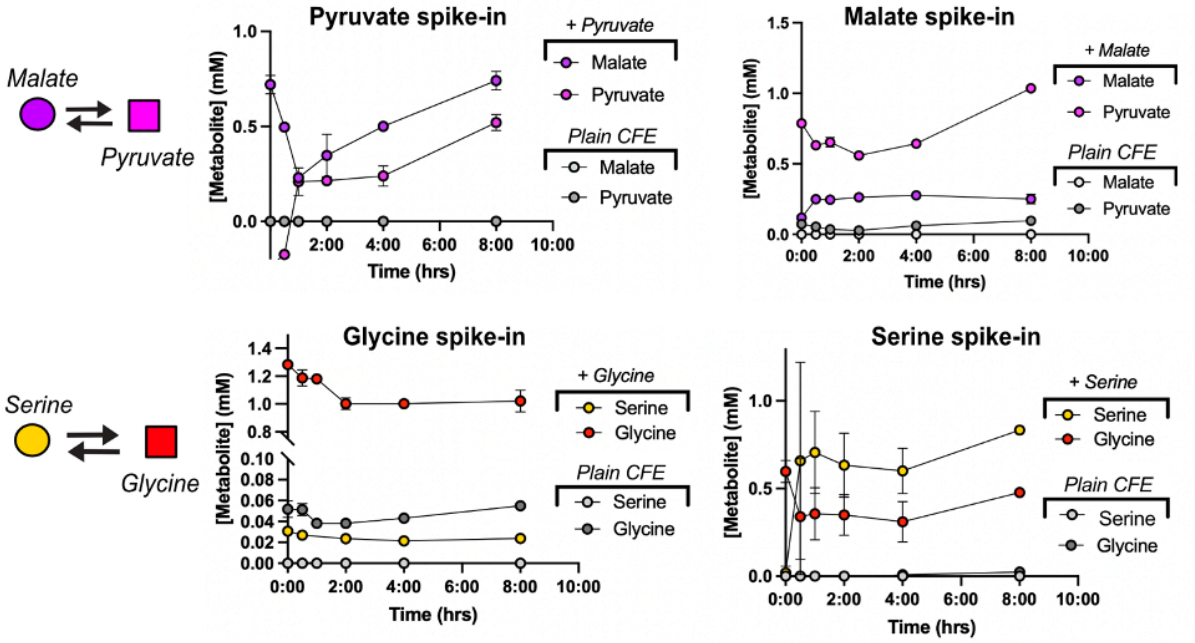


Figure 2. Analyzing the endogenous proteome of E. coli lysate-based CFE

A. CFE reactions are set up with no genes expressed. Following overnight incubation, reactions are diluted 200-fold before pathway metabolites are added or excluded ("Plain CFE"). Metabolite concentrations are analyzed over eight hours.
B. Flux through different TCA metabolites are traced over time for CFE reactions with or without malate added. Fumarate, succinate, and glyoxylate are shown as malate-derived metabolites with high flux.
C. Interconversion of pathway metabolites is measured to understand the activity of endogenous enzymes in the system. For reactions with malate or pyruvate added, both malate and pyruvate are measured. For reactions with serine or glycine added, both serine and glycine are measured.

   For all panels, values represent the mean ± standard deviation of two technical replicates.

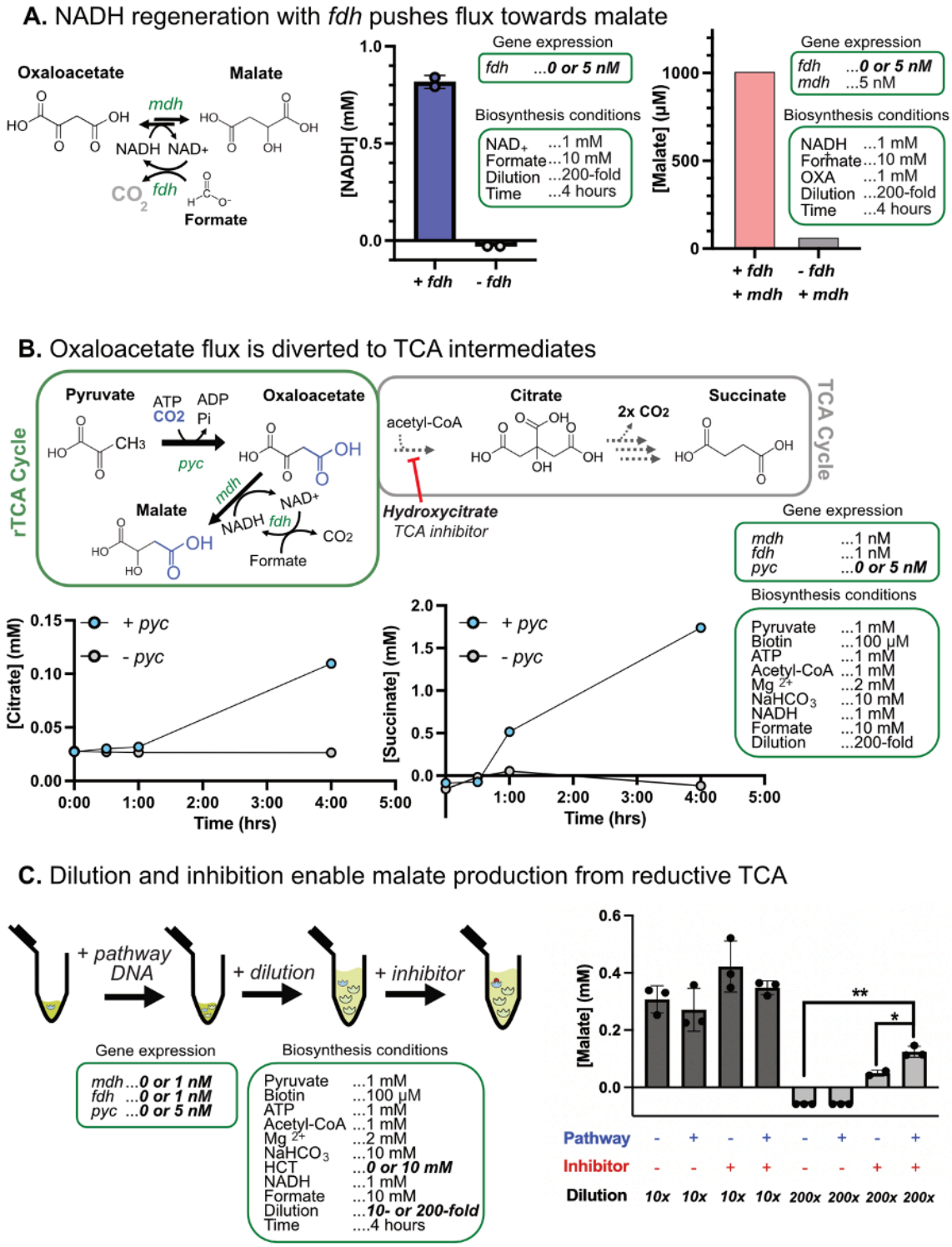

A. NADH regeneration with fdh pushes flux towards malate
Oxaloacetate
Malate
mdh
NADH NAD+
fdh
CO2
Formate
[NADH] (mM)
Gene expression
fdh ...0 or 5 nM
Biosynthesis conditions
NAD+ ...1 mM
Formate ...10 mM
Dilution ...200-fold
Time ...4 hours
+ fdh
- fdh
[Malate] (µM)
Gene expression
fdh ...0 or 5 nM
mdh ...5 nM
Biosynthesis conditions
NADH ...1 mM
Formate ...10 mM
OXA ...1 mM
Dilution ...200-fold
Time ...4 hours
+ fdh + mdh
- fdh + mdh
B. Oxaloacetate flux is diverted to TCA intermediates
rTCA Cycle
Pyruvate
ATP ADP
CO2 Pi
Oxaloacetate
pyc
mdh
Malate
NADH fdh NAD+
CO2
Formate
TCA Cycle
Citrate
2x CO2
Succinate
acetyl-CoA
Hydroxycitrate
TCA inhibitor
Gene expression
mdh ...1 nM
fdh ...1 nM
pyc ...0 or 5 nM
Biosynthesis conditions
Pyruvate ...1 mM
Biotin ...100 µM
ATP ...1 mM
Acetyl-CoA ...1 mM
Mg 2+ ...2 mM
NaHCO3 ...10 mM
NADH ...1 mM
Formate ...10 mM
Dilution ...200-fold
[Citrate] (mM)
+ pyc
- pyc
Time (hrs)
[Succinate] (mM)
+ pyc
- pyc
Time (hrs)
C. Dilution and inhibition enable malate production from reductive TCA
+ pathway DNA
+ dilution
+ inhibitor
Gene expression
mdh ...0 or 1 nM
fdh ...0 or 1 nM
pyc ...0 or 5 nM
Biosynthesis conditions
Pyruvate ...1 mM
Biotin ...100 µM
ATP ...1 mM
Acetyl-CoA ...1 mM
Mg 2+ ...2 mM
NaHCO3 ...10 mM
HCT ...0 or 10 mM
NADH ...1 mM
Formate ...10 mM
Dilution ...10- or 200-fold
Time ....4 hours
[Malate] (mM)
**
*
Pathway - + - + - + - +
Inhibitor - - + + - - + +
Dilution 10x 10x 10x 10x 200x 200x 200x 200x

Figure 3. Optimization of the reductive TCA cycle

A. High malate levels are maintained by co-expression of *fdh* and *mdh*. Oxaloacetate is reduced to malate via *mdh* using NADH as the reducing equivalent. NADH can be regenerated from NAD^+ and formate by *fdh* to maintain flux towards malate and prevent the reverse reaction of *mdh.* **Left:** NADH concentrations in reactions containing 1 mM NAD^+, 10 mM formate, and CFE either with or without 5 nM *fdh* expression. Values represent the mean ± standard deviation of two technical replicates. **Right:** Malate concentrations in reactions containing 1 mM NADH, 1 mM oxaloacetate, and CFE with either 5 nM each of *mdh* and *fdh* or just *mdh*. Values shown represent one technical replicate.

B. *pyc* expression results in diversion of oxaloacetate flux to TCA intermediates. **Top:** Oxaloacetate and acetyl-CoA, an allosteric activator for *pyc*, can be fed into the TCA cycle via the irreversible citrate synthase (CS) enzyme. Entry into TCA can be inhibited by addition of the small molecule CS inhibitor, hydroxycitrate (HCT). Citrate **(bottom, left)** and succinate **(bottom, right)** concentrations are shown over time for reactions with or without pyruvate carboxylase (*pyc*) expressed. All reactions are diluted 200-fold for biosynthesis. Reactions contain 1 mM pyruvate, 10 mM HCO_3^-, 1 mM ATP, 2 mM $Mg(CH_3COO)_2$, 1 mM acetyl-CoA, 100 µM biotin, 1 mM NADH, and 10 mM formate. - *pyc* reactions express 1 nM of *mdh* and *fdh*, + *pyc* reactions also express 5 nM of *pyc*. Values shown represent one technical replicate.

C. CFE dilution and TCA inhibition enables carbon-efficient malate production from pyruvate. *pyc* incorporates one CO_2 equivalent into pyruvate to form oxaloacetate and requires ATP, Mg^{2+}, biotin, and acetyl-CoA for activation. Malate concentrations are measured after four hours for reactions with different conditions. DNA for pathway enzymes, additional CFE dilution, and the inhibitor hydroxycitrate are changed between conditions. All reactions contain 1 mM pyruvate, 10 mM HCO_3^-, 10 mM HCT, 1 mM acetyl-CoA, 100 μM biotin, 1 mM ATP, 2 mM $Mg(CH_3COO)_2$, 1 mM NADH, and 10 mM formate. Reactions contain CFE with or without 1 nM *mdh*, 1 nM *fdh*, and 5 nM *pyc*. Values represent the mean ± standard deviation of three technical replicates.

For all panels, statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.005).

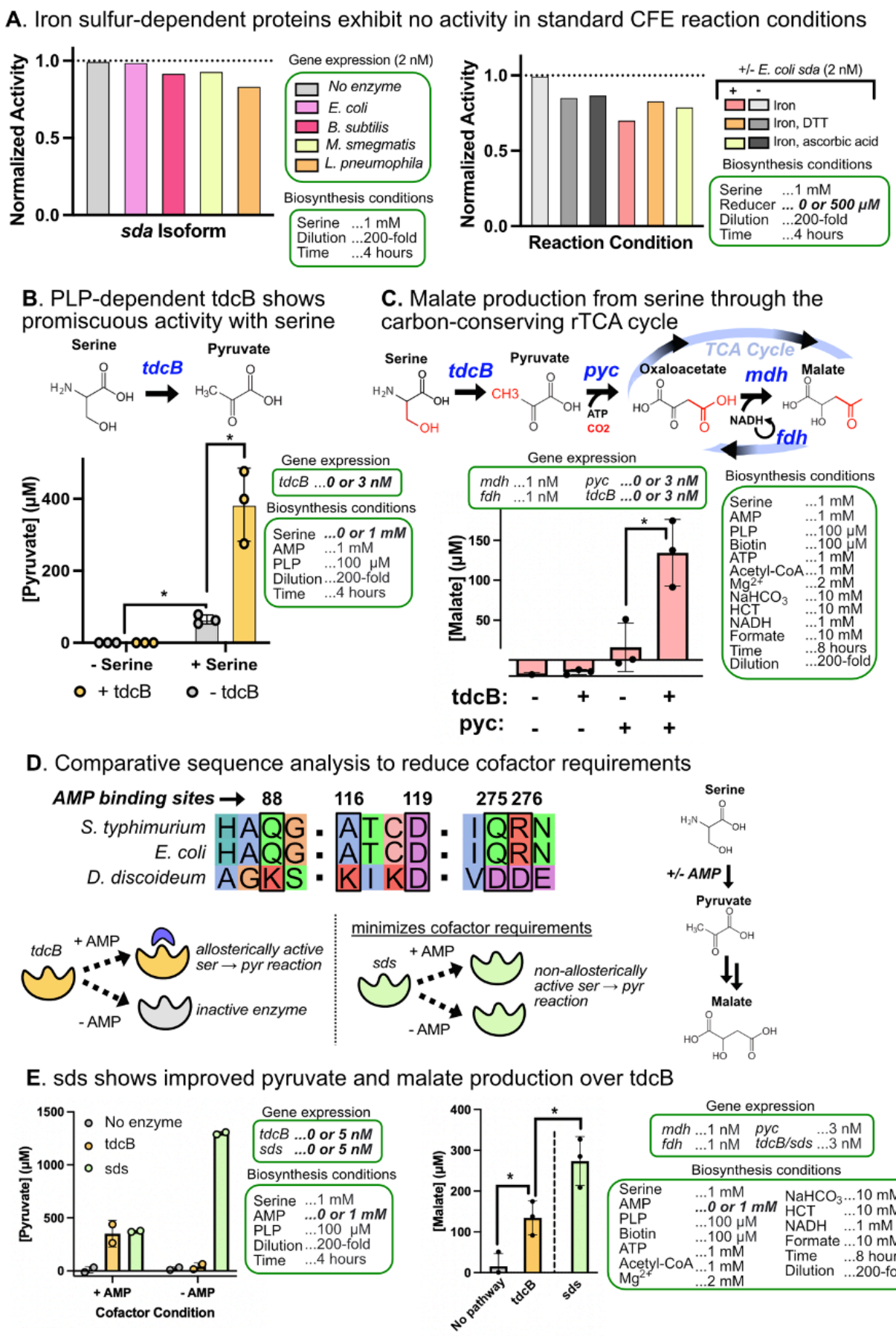

A. Iron sulfur-dependent proteins exhibit no activity in standard CFE reaction conditions
Normalized Activity
1.0
0.5
0.0
sda Isoform
Gene expression (2 nM)
No enzyme
E. coli
B. subtilis
M. smegmatis
L. pneumophila
Biosynthesis conditions
Serine ...1 mM
Dilution ...200-fold
Time ...4 hours
Reaction Condition
+/- E. coli sda (2 nM)
Iron
Iron, DTT
Iron, ascorbic acid
Serine ...1 mM
Reducer ... 0 or 500 µM
Dilution ...200-fold
Time ...4 hours
B. PLP-dependent tdcB shows promiscuous activity with serine
Serine
Pyruvate
tdcB
[Pyruvate] (µM)
400
200
0
- Serine
+ Serine
+ tdcB
- tdcB
Gene expression
tdcB ...0 or 3 nM
Biosynthesis conditions
Serine ...0 or 1 mM
AMP ...1 mM
PLP ...100 µM
Dilution ...200-fold
Time ...4 hours
C. Malate production from serine through the carbon-conserving rTCA cycle
Serine
tdcB
Pyruvate
pyc
ATP
CO2
Oxaloacetate
mdh
Malate
NADH
fdh
TCA Cycle
Gene expression
mdh ...1 nM
fdh ...1 nM
pyc ...0 or 3 nM
tdcB ...0 or 3 nM
[Malate] (µM)
150
100
50
tdcB: - + - +
pyc: - - + +
Biosynthesis conditions
Serine ...1 mM
AMP ...1 mM
PLP ...100 µM
Biotin ...100 µM
ATP ...1 mM
Acetyl-CoA ...1 mM
Mg2+ ...2 mM
NaHCO3 ...10 mM
HCT ...10 mM
NADH ...1 mM
Formate ...10 mM
Time ...8 hours
Dilution ...200-fold
D. Comparative sequence analysis to reduce cofactor requirements
AMP binding sites → 88 116 119 275 276
S. typhimurium HAQG ATCD IQRN
E. coli HAQG ATCD IQRN
D. discoideum AGKS KIKD VDDE
tdcB
+ AMP
- AMP
allosterically active ser → pyr reaction
inactive enzyme
minimizes cofactor requirements
sds
+ AMP
- AMP
non-allosterically active ser → pyr reaction
Serine
+/- AMP
Pyruvate
Malate
E. sds shows improved pyruvate and malate production over tdcB
[Pyruvate] (µM)
1500
1000
500
0
No enzyme
tdcB
sds
+ AMP
- AMP
Cofactor Condition
Gene expression
tdcB ...0 or 5 nM
sds ...0 or 5 nM
Biosynthesis conditions
Serine ...1 mM
AMP ...0 or 1 mM
PLP ...100 µM
Dilution ...200-fold
Time ...4 hours
[Malate] (µM)
400
300
200
100
0
No pathway
tdcB
sds
Gene expression
mdh ...1 nM
fdh ...1 nM
pyc ...3 nM
tdcB/sds ...3 nM
Biosynthesis conditions
Serine ...1 mM
AMP ...0 or 1 mM
PLP ...100 µM
Biotin ...100 µM
ATP ...1 mM
Acetyl-CoA ...1 mM
Mg2+ ...2 mM
NaHCO3 ...10 mM
HCT ...10 mM
NADH ...1 mM
Formate ...10 mM
Time ...8 hours
Dilution ...200-fold

Figure 4: Characterization of serine deaminase function in CFE

A. Screening of iron-sulfur dependent serine deaminase (*sda*) variants in lysate-based CFE. **Left:** Isoforms from *E. coli, B. subtilis, M. smegmatis,* and *L. pneuomophila* were chosen from the literature based on favorable kinetic values and previous expression in *E. coli*. **Right:** Reaction conditions were screened to try to rescue activity of iron-sulfur enzymes in CFE. Iron was added to assist formation of the iron-sulfur cluster, and dithiothreitol or ascorbic acid were added to create a reducing environment. All genes were expressed at 5 nM and chemicals added at 500 μM. Values shown represent one technical replicate.
B. Pyruvate production is enabled using a promiscuous threonine deaminase (*tdcB)* from *E. coli. tdcB* belongs to the class of PLP-utilizing deaminases, avoiding the need for iron-sulfur proteins. Pyruvate concentrations are shown for reactions either with or without serine and *tdcB.* All reactions contain 100 μM PLP and 1 mM AMP. Values represent the mean ± standard deviation of three technical replicates.
C. Serine is converted to malate using a four enzyme pathway in which *tdcB* converts serine to pyruvate, and pyruvate is carboxylated into oxaloacetate via *pyc*. Oxaloacetate is then reduced to malate by *mdh*, with cofactor regeneration from *fdh* to drive flux towards malate. Malate concentrations are shown for eight hour reactions containing combinations of *tdcB* and/or *pyc*, with all reactions containing *fdh* and *mdh*. Values represent the mean ± standard deviation of three technical replicates.

D. Identification of an isoform via comparative sequence analysis removes the need for additional cofactors and reduces pathway incompatibilities. AMP binding sites found for the *tdcB* enzyme from *S. typhimurium* are compared against the *E. coli* isoform sequence and the *sds* sequence.

E. Comparison of *sds* and *tdcB* for production of pyruvate (**left)** and malate (**right**). **Left:** Use of the *sds* isoform shows improved pyruvate production relative to *tdcB* in the absence of AMP. Pyruvate concentrations are shown for four hour reactions containing 1 mM serine, 100 μM PLP, and either with or without 1 mM AMP. Values represent the mean ± standard deviation of two technical replicates. **Right:** Comparison of *tdcB* and *sds* for serine to malate conversion via the four-enzyme pathway. Data shown for *tdcB* and *sds* are reproduced from Figures 4C and 5B, respectively. Values represent the mean ± standard deviation of three technical replicates.

   For all panels, statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.005).

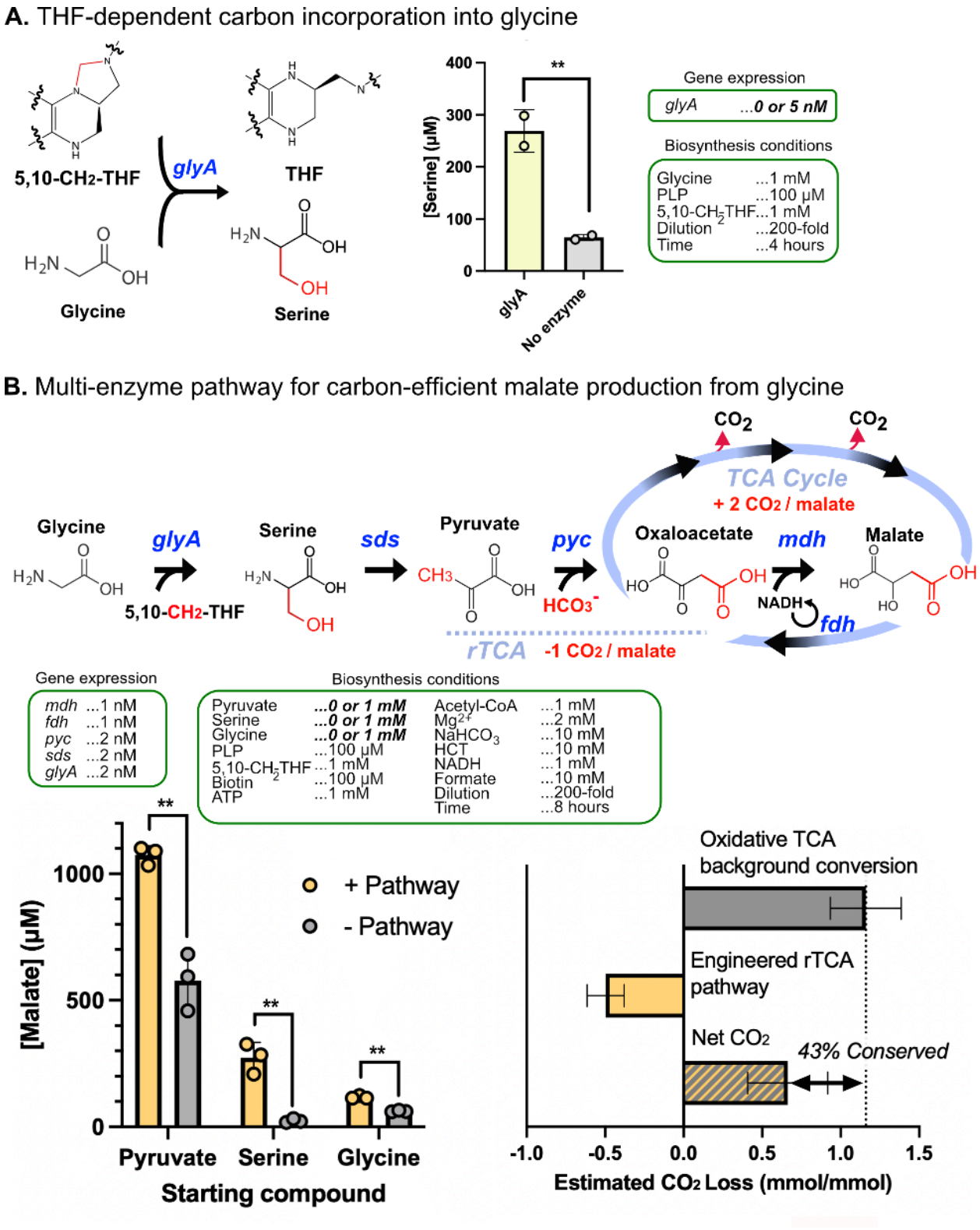


Figure 5. Conversion of glycine to malate.

A. C1-incorporation into glycine, mediated by a THF intermediate. 5,10-methylenetetrahydrofolate (5,10-CH_2-THF) acts as a C1 donor to convert glycine and water into serine, transforming 5,10-CH_2-THF back to THF in the process. This process is catalyzed by the PLP-dependent *glyA* enzyme. Serine concentrations are shown for reactions containing 1 mM glycine, 1 mM 5,10-CH_2-THF, 100 µM PLP in conditions with and without 5 nM *glyA* expression. Values represent the mean ± standard deviation of two technical replicates.

B. Multi-enzyme pathway enables carbon-efficient malate production from glycine in CFE. The pathway consists of five enzymatic steps, including *in situ* NADH regeneration. Incorporated carbons are shown in the schematic in red. Left: Malate concentrations are shown for multi-step reactions with different starting compounds either with or without the pathway expressed. Right: CO_2 conserved through the introduction of the rTCA pathway is shown for pyruvate to malate conversion. CO_2 conserved is estimated using the pathway-dependent conversion and assuming a fixed reaction stoichiometry between malate produced and CO_2 equivalents conserved. Pathway-dependent conversion represents the added conversion efficiency above the background level from the pathway enzymes. Our calculations assume that for each pyruvate converted to malate, a maximum of 1 CO_2 equivalent can be fixed from bicarbonate via rTCA, and a maximum of 2 CO_2 equivalents can be released via the oxidative TCA cycle. For both panels, reactions contain all necessary chemical cofactors for the

five-step pathway. Values represent the mean ± standard deviation of three technical replicates.

For all panels, statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.005).

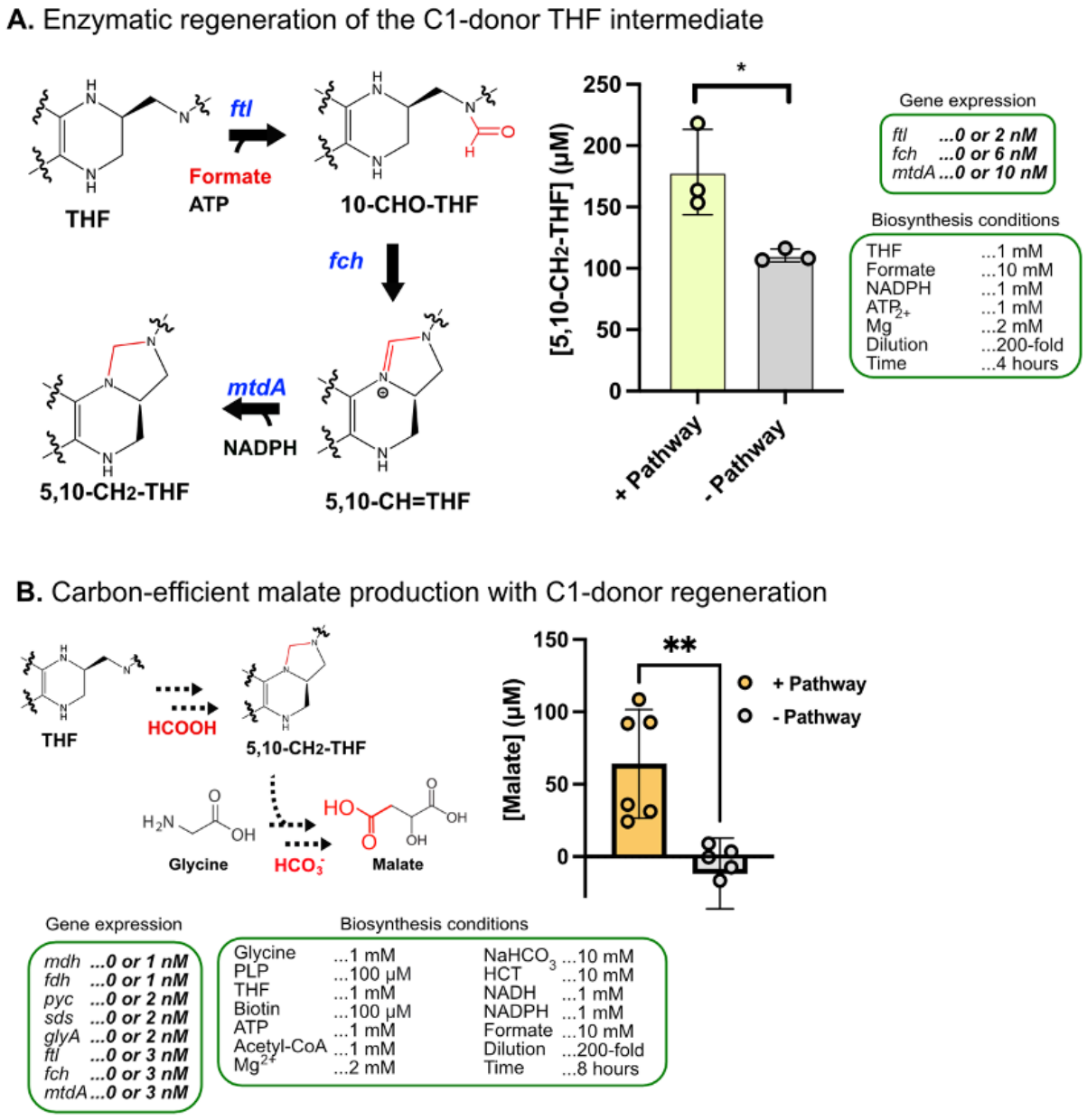


Figure 6. Regeneration of THF cofactor drives carbon-efficient malate production.

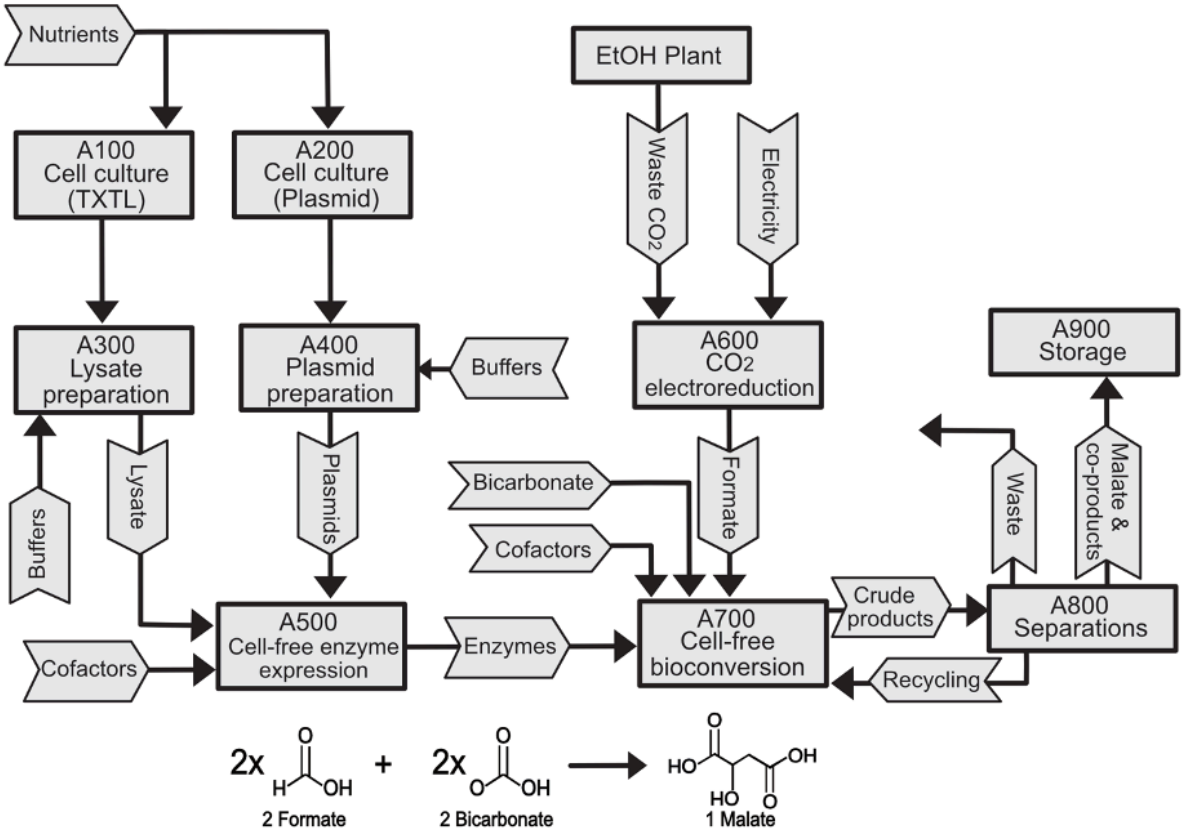
A. Enzymatic pathway for formate assimilation into THF to form 5,10-CH_2-THF. The pathway consists of three enzymes and utilizes one ATP and one NADPH equivalent. THF can be recycled back into 5,10-CH_2-THF after C1 transfer to glycine. 5,10-CH_2-THF concentration is shown for reactions fed with THF,

formate, NADPH, and ATP. Reactions contain CFE either with or without the formate assimilation pathway. Values represent the mean ± standard deviation of three technical replicates.

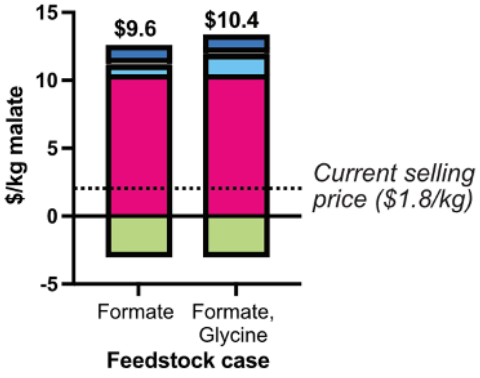
B. Carbon-efficient cell-free conversion of glycine and formate to malate. The pathway consists of eight enzymes for THF recycling, NADH regeneration, and malate synthesis. Two carbon equivalents are incorporated into malate. Malate concentration is shown for reactions fed with all necessary chemical cofactors and starting from formate and glycine. Reactions contain CFE either with or without the full pathway. In this experiment, cell-free bioproduction reactions were run at 250 µl volume. Values represent the mean ± standard deviation of six technical replicates.

For all panels, statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.005).

A. Process block diagram for malate bioproduction in a cell-free expression system



B. Cost contributions to cell-free bioproduction



Cost factor	% of total cost: Formate	Formate, Glycine
Nutrients for enzyme prod.	109%	101%
Nutrients for cell-free conv.	7%	14%
Other chemicals	1%	1%
Utility and labor	4%	4%
Capital related	10%	9%
Co-products	-31%	-29%

C. TEA of price sensitivity to process efficiency

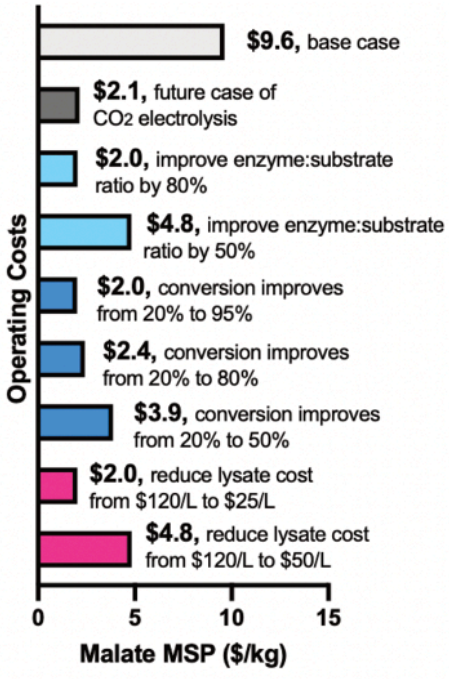


Figure 7: Techno-economic analysis of cell-free bioproduction of malate.

A. Process block flow diagram to produce malate by cell free bioconversion. The cell-free bioproduction plant is assumed to be located next to a bio-ethanol plant. Carbon dioxide effluent from the ethanol plant is fed to a low temperature electrolysis process converting the CO_2 stream to formic acid. An average CO_2 flow rate of 227 thousand metric ton per year is used to estimate the process scale [88]. Economic analysis and technical assumptions of the CO_2 conversion step have been established by a multi-national lab team [64]. Capital and operating costs of the cell free process are assumed to be comparable to the enzyme production process published by NREL [89]. Cost assumptions of the biological conversion of formic acid to malic acid is taken from previous work and outlined in Methods S2 [64].
B. Contributions of different inputs to the cost of cell-free malate bioproduction are shown. Two cases are considered, one where formate is used as the sole carbon source, and one where formate and glycine are co-fed into the bioproduction platform. The total cost is broken down by area. Nutrients for enzyme production includes the costs of the cell-free lysate, DNA preparation, cofactors required for protein expression. Nutrients for cell-free conversion includes the cost of feedstocks and cofactors required for bioconversion. The other chemicals category includes chemicals used for pH adjustments, separations, and purification. The co-purification of co-products succinate and citrate are included

to offset costs associated with the process. The current selling price of malate produced from fossil fuels ($1.8/kg) is shown with a dashed line.

C. Techno-economic analysis of price sensitivity to operating costs and process efficiency. Operating costs, including cost of formate from CO_2 electrolysis, enzyme turnover, conversion efficiency, and cell-free reagent cost are varied to understand their impacts on the final product cost. TEA suggests that operating costs were the primary drivers affecting malate MSP. The effect of improvements in various operating costs on the final MSP of malate are shown. Assumptions for TEA are given in Table S2 and Methods S2.

References

(1) Collas, F.; Dronsella, B. B.; Kubis, A.; Schann, K.; Binder, S.; Arto, N.; Claassens, N. J.; Kensy, F.; Orsi, E. Engineering the Biological Conversion of Formate into Crotonate in *Cupriavidus Necator*. *Metab. Eng.* **2023**, *79*, 49–65. https://doi.org/10.1016/j.ymben.2023.06.015.

(2) Köpke, M.; Mihalcea, C.; Bromley, J. C.; Simpson, S. D. Fermentative Production of Ethanol from Carbon Monoxide. *Curr. Opin. Biotechnol.* **2011**, *22* (3), 320–325. https://doi.org/10.1016/j.copbio.2011.01.005.

(3) Humphreys, C. M.; Minton, N. P. Advances in Metabolic Engineering in the Microbial Production of Fuels and Chemicals from C1 Gas. *Curr. Opin. Biotechnol.* **2018**, *50*, 174–181. https://doi.org/10.1016/j.copbio.2017.12.023.

(4) Bergquist, P. L.; Siddiqui, S.; Sunna, A. Cell-Free Biocatalysis for the Production of Platform Chemicals. *Front. Energy Res.* **2020**, *8*.

(5) Takkellapati, S.; Li, T.; Gonzalez, M. A. An Overview of Biorefinery Derived Platform Chemicals from a Cellulose and Hemicellulose Biorefinery. *Clean Technol. Environ. Policy* **2018**, *20* (7), 1615–1630. https://doi.org/10.1007/s10098-018-1568-5.

(6) Wang, J.; Anderson, K.; Yang, E.; He, L.; Lidstrom, M. E. Enzyme Engineering and in Vivo Testing of a Formate Reduction Pathway. *Synth. Biol.* **2021**, *6* (1), ysab020. https://doi.org/10.1093/synbio/ysab020.

(7) Cotton, C. A.; Claassens, N. J.; Benito-Vaquerizo, S.; Bar-Even, A. Renewable Methanol and Formate as Microbial Feedstocks. *Curr. Opin. Biotechnol.* **2020**, *62*, 168–180. https://doi.org/10.1016/j.copbio.2019.10.002.

(8) Lim, H. J.; Kim, D.-M. Cell-Free Metabolic Engineering: Recent Developments and Future Prospects. *Methods Protoc.* **2019**, *2* (2), 33. https://doi.org/10.3390/mps2020033.

(9) Garenne, D.; Thompson, S.; Brisson, A.; Khakimzhan, A.; Noireaux, V. The All- *E. Coli* TXTL Toolbox 3.0: New Capabilities of a Cell-Free Synthetic Biology Platform. *Synth. Biol.* **2021**, ysab017. https://doi.org/10.1093/synbio/ysab017.

(10) Lavickova, B.; Maerkl, S. J. A Simple, Robust, and Low-Cost Method To Produce the PURE Cell-Free System. *ACS Synth. Biol.* **2019**, *8* (2), 455–462. https://doi.org/10.1021/acssynbio.8b00427.

(11) Rasor, B. J.; Vögeli, B.; Landwehr, G. M.; Bogart, J. W.; Karim, A. S.; Jewett, M. C. Toward Sustainable, Cell-Free Biomanufacturing. *Curr. Opin. Biotechnol.* **2021**, *69*, 136–144. https://doi.org/10.1016/j.copbio.2020.12.012.

(12) Sun, Z. Z.; Hayes, C. A.; Shin, J.; Caschera, F.; Murray, R. M.; Noireaux, V. Protocols for Implementing an Escherichia Coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *JoVE J. Vis. Exp.* **2013**, No. 79, e50762. https://doi.org/10.3791/50762.

(13) Garamella, J.; Marshall, R.; Rustad, M.; Noireaux, V. The All E. Coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synth. Biol.* **2016**, *5* (4), 344–355. https://doi.org/10.1021/acssynbio.5b00296.

(14) Schwander, T.; Schada von Borzyskowski, L.; Burgener, S.; Cortina, N. S.; Erb, T. J. A Synthetic Pathway for the Fixation of Carbon Dioxide in Vitro. *Science* **2016**, *354* (6314), 900–904. https://doi.org/10.1126/science.aah5237.

(15) Luo, S.; Lin, P. P.; Nieh, L.-Y.; Liao, G.-B.; Tang, P.-W.; Chen, C.; Liao, J. C. A Cell-Free Self-Replenishing CO2-Fixing System. *Nat. Catal.* **2022**, *5* (2), 154–162. https://doi.org/10.1038/s41929-022-00746-x.

(16) Luo, S.; Diehl, C.; He, H.; Bae, Y.; Klose, M.; Claus, P.; Cortina, N. S.; Fernandez, C. A.; Schulz-Mirbach, H.; McLean, R.; Ramírez Rojas, A. A.; Schindler, D.; Paczia, N.; Erb, T. J. Construction and Modular Implementation of the THETA Cycle for Synthetic CO2 Fixation. *Nat. Catal.* **2023**, *6* (12), 1228–1240. https://doi.org/10.1038/s41929-023-01079-z.

(17) Bat-Erdene, U.; Billingsley, J. M.; Turner, W. C.; Lichman, B. R.; Ippoliti, F. M.; Garg, N. K.; O'Connor, S. E.; Tang, Y. Cell-Free Total Biosynthesis of Plant Terpene Natural Products Using an Orthogonal Cofactor Regeneration System. *ACS Catal.* **2021**, *11* (15), 9898–9903. https://doi.org/10.1021/acscatal.1c02267.

(18) Dinglasan, J. L. N.; Doktycz, M. J. Rewiring Cell-Free Metabolic Flux in *E. Coli* Lysates Using a Block-Push-Pull Approach. *Synth. Biol.* **2023**, ysad007. https://doi.org/10.1093/synbio/ysad007.

(19) Garcia, D. C.; Dinglasan, J. L. N.; Shrestha, H.; Abraham, P. E.; Hettich, R. L.; Doktycz, M. J. A Lysate Proteome Engineering Strategy for Enhancing Cell-Free Metabolite Production. *Metab. Eng. Commun.* **2021**, *12*, e00162. https://doi.org/10.1016/j.mec.2021.e00162.

(20) Richardson, K. N.; Black, W. B.; Li, H. Aldehyde Production in Crude Lysate- and Whole Cell-Based Biotransformation Using a Noncanonical Redox Cofactor System. *ACS Catal.* **2020**, *10* (15), 8898–8903. https://doi.org/10.1021/acscatal.0c03070.

(21) Rasor, B. J.; Yi, X.; Brown, H.; Alper, H. S.; Jewett, M. C. An Integrated in Vivo/in Vitro Framework to Enhance Cell-Free Biosynthesis with Metabolically Rewired Yeast Extracts. *Nat. Commun.* **2021**, *12* (1), 5139. https://doi.org/10.1038/s41467-021-25233-y.

(22) Wagner, L.; Jules, M.; Borkowski, O. What Remains from Living Cells in Bacterial Lysate-Based Cell-Free Systems. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 3173–3182. https://doi.org/10.1016/j.csbj.2023.05.025.

(23) Santos Correa, S.; Schultz, J.; Lauersen, K. J.; Soares Rosado, A. Natural Carbon Fixation and Advances in Synthetic Engineering for Redesigning and Creating New Fixation Pathways. *J. Adv. Res.* **2023**, *47*, 75–92. https://doi.org/10.1016/j.jare.2022.07.011.

(24) Zelcbuch, L.; Lindner, S. N.; Zegman, Y.; Vainberg Slutskin, I.; Antonovsky, N.; Gleizer, S.; Milo, R.; Bar-Even, A. Pyruvate Formate-Lyase Enables Efficient Growth of Escherichia Coli on Acetate and Formate. *Biochemistry* **2016**, *55* (17), 2423–2426. https://doi.org/10.1021/acs.biochem.6b00184.

(25) Song, Y.; Lee, J. S.; Shin, J.; Lee, G. M.; Jin, S.; Kang, S.; Lee, J.-K.; Kim, D. R.; Lee, E. Y.; Kim, S. C.; Cho, S.; Kim, D.; Cho, B.-K. Functional Cooperation of the Glycine Synthase-Reductase and Wood–Ljungdahl Pathways for Autotrophic Growth of Clostridium Drakei. *Proc. Natl. Acad. Sci.* **2020**, *117* (13), 7516–7523. https://doi.org/10.1073/pnas.1912289117.

(26) Jiang, Y.; Zheng, T.; Ye, X.; Xin, F.; Zhang, W.; Dong, W.; Ma, J.; Jiang, M. Metabolic Engineering of Escherichia Coli for L-Malate Production Anaerobically. *Microb. Cell Factories* **2020**, *19* (1), 165. https://doi.org/10.1186/s12934-020-01422-0.

(27) Moon, S. Y.; Hong, S. H.; Kim, T. Y.; Lee, S. Y. Metabolic Engineering of Escherichia Coli for the Production of Malic Acid. *Biochem. Eng. J.* **2008**, *40* (2), 312–320. https://doi.org/10.1016/j.bej.2008.01.001.
(28) Zhang, X.; Wang, X.; Shanmugam, K. T.; Ingram, L. O. L-Malate Production by Metabolically Engineered Escherichia Coli. *Appl. Environ. Microbiol.* **2011**, *77* (2), 427–434. https://doi.org/10.1128/AEM.01971-10.
(29) Chowdhury, S.; Westenberg, R.; Wennerholm, K.; Cardiff, R.A.L.; Beliaev, A.S.; Noireaux, V.; Carothers, J.M.; Peralta-Yahya, P. Carbon Negative Synthesis of Amino Acids Using a Cell-free-Based-Biocatalyst. *ACS Synth. Biol.* **2024.** https://doi.org/10.1021/acssynbio.4c00359.
(30) Garenne, D.; Beisel, C. L.; Noireaux, V. Characterization of the All- *E. coli* Transcription-translation System myTXTL by Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2019**, *33* (11), 1036–1048. https://doi.org/10.1002/rcm.8438.
(31) Noor, E.; Haraldsdóttir, H. S.; Milo, R.; Fleming, R. M. T. Consistent Estimation of Gibbs Energy Using Component Contributions. *PLOS Comput. Biol.* **2013**, *9* (7), e1003098. https://doi.org/10.1371/journal.pcbi.1003098.
(32) Holm, A. K.; Blank, L. M.; Oldiges, M.; Schmid, A.; Solem, C.; Jensen, P. R.; Vemuri, G. N. Metabolic and Transcriptional Response to Cofactor Perturbations in *Escherichia Coli*. *J. Biol. Chem.* **2010**, *285* (23), 17498–17506. https://doi.org/10.1074/jbc.M109.095570.
(33) de Lorenzo, L.; Stack T. M. M.; Fox, K. M.; Walstrom, K. M. Catalytic mechanism and kinetics of malate dehydrogenase. *Essays Biochem.* **2024,** 68(2), 73—82. doi: 10.1042/EBC20230086
(34) Partipilo, M.; Whittaker, J. J.; Pontillo, N.; Coenradij, J.; Herrmann, A.; Guskov, A.; Slotboom, D. J. Biochemical and Structural Insight into the Chemical Resistance and Cofactor Specificity of the Formate Dehydrogenase from Starkeya Novella. *FEBS J.* **2023**, *290* (17), 4238–4255. https://doi.org/10.1111/febs.16871.
(35) Anderson, D.H.; Duckworth, H.W.; In vitro mutagenesis of Escherichia coli citrate synthase to clarify the locations of ligand binding sites. *J Biol Chem.* **1988** 15;263(5):2163-9. PMID: 3276685.
(36) Yin, Y.; Kirsch, J.F.; Identification of functional paralog shift mutations: conversion of Escherichia coli malate dehydrogenase to a lactate dehydrogenase. *Proc Natl Acad Sci* **2007** Oct 30;104(44):17353-7. doi: 10.1073/pnas.0708265104
(37) Venkat, S.; Chen, H.; McGuire, P.; Stahman, A.; Gan, Q.; Fan, C. Characterizing Lysine Acetylation of Escherichia Coli Type II Citrate Synthase. *FEBS J.* **2019**, *286* (14), 2799–2808. https://doi.org/10.1111/febs.14845.
(38) Bulutoglu, B.; Garcia, K. E.; Wu, F.; Minteer, S. D.; Banta, S. Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. *ACS Chem. Biol.* **2016**, *11* (10), 2847–2853. https://doi.org/10.1021/acschembio.6b00523.
(39) St. Maurice, M.; Reinhardt, L.; Surinya, K. H.; Attwood, P. V.; Wallace, J. C.; Cleland, W. W.; Rayment, I. Domain Architecture of Pyruvate Carboxylase, a Biotin-Dependent Multifunctional Enzyme. *Science* **2007**, *317* (5841), 1076–1079.
(40) Adina-Zada, A.; Zeczycki, T. N.; Attwood, P. V. Regulation of the Structure and Activity of Pyruvate Carboxylase by Acetyl CoA. *Arch. Biochem. Biophys.* **2012**, *519* (2), 118–130. https://doi.org/10.1016/j.abb.2011.11.015.

(41) Schlachter, C. R.; Klapper, V.; Radford, T.; Chruszcz, M. Comparative Studies of Aspergillus Fumigatus 2-Methylcitrate Synthase and Human Citrate Synthase. *Biol. Chem.* **2019**, *400* (12), 1567–1581. https://doi.org/10.1515/hsz-2019-0106.
(42) Westenberg, R.; Peralta-Yahya, P. Toward Implementation of Carbon-Conservation Networks in Nonmodel Organisms. *Curr. Opin. Biotechnol.* **2023**, *81*, 102949. https://doi.org/10.1016/j.copbio.2023.102949.
(43) Yu, H.; Liao, J. C. A Modified Serine Cycle in Escherichia Coli Coverts Methanol and CO2 to Two-Carbon Compounds. *Nat. Commun.* **2018**, *9* (1), 3992. https://doi.org/10.1038/s41467-018-06496-4.
(44) Grabowski, R.; Hofmeister, A. E.; Buckel, W. Bacterial L-Serine Dehydratases: A New Family of Enzymes Containing Iron-Sulfur Clusters. *Trends Biochem. Sci.* **1993**, *18* (8), 297–300. https://doi.org/10.1016/0968-0004(93)90040-t.
(45) Shizuta, Y.; Nakazawa, A.; Tokushige, M.; Hayaishi, O. Studies on the Interaction between Regulatory Enzymes and Effectors. *The Journal of Biological Chemistry.* **1968,** *244,* 1883-1889.
(46) Claassens, N. J.; Burgener, S.; Vögeli, B.; Erb, T. J.; Bar-Even, A. A Critical Comparison of Cellular and Cell-Free Bioproduction Systems. *Curr. Opin. Biotechnol.* **2019**, *60*, 221–229. https://doi.org/10.1016/j.copbio.2019.05.003.
(47) Diez, M. D. A.; Aleanzi, M. C.; Iglesias, A. A.; Ballicora, M. A. A Novel Dual Allosteric Activation Mechanism of Escherichia Coli ADP-Glucose Pyrophosphorylase: The Role of Pyruvate. *PLOS ONE* **2014**, *9* (8), e103888. https://doi.org/10.1371/journal.pone.0103888.
(48) Simanshu, D. K.; Savithri, H. S.; Murthy, M. R. N. Crystal Structures of Salmonella Typhimurium Biodegradative Threonine Deaminase and Its Complex with CMP Provide Structural Insights into Ligand-Induced Oligomerization and Enzyme Activation *. *J. Biol. Chem.* **2006**, *281* (51), 39630–39641. https://doi.org/10.1074/jbc.M605721200.
(49) Madeira, F.; Madhusoodanan, N.; Lee, J.; Eusebi, A.; Niewielska, A.; Tivey, A. R. N.; Lopez, R.; Butcher, S. The EMBL-EBI Job Dispatcher Sequence Analysis Tools Framework in 2024. *Nucleic Acids Res.* **2024**, *52* (W1), W521–W525. https://doi.org/10.1093/nar/gkae241.
(50) Glöckner, G.; Eichinger, L.; Szafranski, K.; Pachebat, J. A.; Bankier, A. T.; Dear, P. H.; Lehmann, R.; Baumgart, C.; Parra, G.; Abril, J. F.; Guigó, R.; Kumpf, K.; Tunggal, B.; Cox, E.; Quail, M. A.; Platzer, M.; Rosenthal, A.; Noegel, A. A.; Barrell, B. G.; Rajandream, M.-A.; Williams, J. G.; Kay, R. R.; Kuspa, A.; Gibbs, R.; Sucgang, R.; Muzny, D.; Desany, B.; Zeng, K.; Zhu, B.; de Jong, P.; Dingermann, T.; Gerisch, G.; Philippsen, P.; Schleicher, M.; Schuster, S. C.; Winckler, T. Seq uence and Analysis of Chromosome 2 of Dictyostelium Discoideum. *Nature* **2002**, *418* (6893), 79–85. https://doi.org/10.1038/nature00847.
(51) Hong, Y.; Ren, J.; Zhang, X.; Wang, W.; Zeng, A.-P. Quantitative Analysis of Glycine Related Metabolic Pathways for One-Carbon Synthetic Biology. *Curr. Opin. Biotechnol.* **2020**, *64*, 70–78. https://doi.org/10.1016/j.copbio.2019.10.001.
(52) Ren, J.; Wang, W.; Nie, J.; Yuan, W.; Zeng, A.-P. Understanding and Engineering Glycine Cleavage System and Related Metabolic Pathways for C1-Based Biosynthesis. In *One-Carbon Feedstocks for Sustainable Bioproduction*; Zeng, A.-P., Claassens, N. J.,

Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, 2022; pp 273–298. https://doi.org/10.1007/10_2021_186.
(53) Xu, Y.; Ren, J.; Wang, W.; Zeng, A. Improvement of Glycine Biosynthesis from One-carbon Compounds and Ammonia Catalyzed by the Glycine Cleavage System in Vitro. *Eng. Life Sci.* **2021**, *22* (1), 40–53. https://doi.org/10.1002/elsc.202100047.
(54) Wu, R.; Li, F.; Cui, X.; Li, Z.; Ma, C.; Jiang, H.; Zhang, L.; Zhang, Y.-H. P. J.; Zhao, T.; Zhang, Y.; Li, Y.; Chen, H.; Zhu, Z. Enzymatic Electrosynthesis of Glycine from CO2 and NH3. *Angew. Chem. Int. Ed.* **2023**, *62* (14), e202218387. https://doi.org/10.1002/anie.202218387.
(55) Zhang, X.; Li, M.; Xu, Y.; Ren, J.; Zeng, A.-P. Quantitative Study of H Protein Lipoylation of the Glycine Cleavage System and a Strategy to Increase Its Activity by Co-Expression of LplA. *J. Biol. Eng.* **2019**, *13* (1), 32. https://doi.org/10.1186/s13036-019-0164-5.
(56) Crowther, G. J.; Kosály, G.; Lidstrom, M. E. Formate as the Main Branch Point for Methylotrophic Metabolism in Methylobacterium Extorquens AM1. *J. Bacteriol.* **2008**, *190* (14), 5057–5062. https://doi.org/10.1128/jb.00228-08.
(57) Tavanti, M.; Hosford, J.; Lloyd, R. C.; Brown, M. J. B. ATP Regeneration by a Single Polyphosphate Kinase Powers Multigram-Scale Aldehyde Synthesis in Vitro. *Green Chem.* **2021**, *23* (2), 828–837. https://doi.org/10.1039/D0GC03830J.
(58) Ihara, M.; Kawano, Y.; Urano, M.; Okabe, A. Light Driven CO2 Fixation by Using Cyanobacterial Photosystem I and NADPH-Dependent Formate Dehydrogenase. *PLOS ONE* **2013**, *8* (8), e71581. https://doi.org/10.1371/journal.pone.0071581.
(59) Özgün, G. P.; Ordu, E. B.; Tütüncü, H. E.; Yelboğa, E.; Sessions, R. B.; Gül Karagüler, N. Site Saturation Mutagenesis Applications on Candida Methylica Formate Dehydrogenase. *Scientifica* **2016**, *2016*, 4902450. https://doi.org/10.1155/2016/4902450.
(60) Sherkhanov, S.; Korman, T. P.; Chan, S.; Faham, S.; Liu, H.; Sawaya, M. R.; Hsu, W.-T.; Vikram, E.; Cheng, T.; Bowie, J. U. Isobutanol Production Freed from Biological Limits Using Synthetic Biochemistry. *Nat. Commun.* **2020**, *11* (1), 4292. https://doi.org/10.1038/s41467-020-18124-1.
(61) Zawada, J. F.; Burgenson, D.; Yin, G.; Hallam, T. J.; Swartz, J. R.; Kiss, R. D. Cell-Free Technologies for Biopharmaceutical Research and Production. *Curr. Opin. Biotechnol.* **2022**, *76*, 102719. https://doi.org/10.1016/j.copbio.2022.102719.
(62) You, C.; Shi, T.; Li, Y.; Han, P.; Zhou, X.; Zhang, Y.-H. P. An in Vitro Synthetic Biology Platform for the Industrial Biomanufacturing of Myo-Inositol from Starch. *Biotechnol. Bioeng.* **2017**, *114* (8), 1855–1864. https://doi.org/10.1002/bit.26314.
(63) Claassens, N. J.; Satanowski, A.; Bysani, V. R.; Dronsella, B.; Orsi, E.; Rainaldi, V.; Yilmaz, S.; Wenk, S.; Lindner, S. N. Engineering the Reductive Glycine Pathway: A Promising Synthetic Metabolism Approach for C1-Assimilation. In *One-Carbon Feedstocks for Sustainable Bioproduction*; Zeng, A.-P., Claassens, N. J., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, 2022; pp 299–350. https://doi.org/10.1007/10_2021_181.
(64) *Malic Acid, DL - Chemical Economics Handbook (CEH) | S&P Global*. https://www.spglobal.com/commodityinsights/en/ci/products/malic-acid-chemical-economics-handbook.html (accessed 2024-04-05).

(65) Jiang, Y.; Ye, X.; Zheng, T.; Dong, W.; Xin, F.; Ma, J.; Jiang, M. Microbial Production of L-Malate from Renewable Non-Food Feedstocks. *Chin. J. Chem. Eng.* **2021**, *30*, 105–111. https://doi.org/10.1016/j.cjche.2020.10.017.

(66) Huang, Z.; Grim, R. G.; Schaidle, J. A.; Tao, L. The Economic Outlook for Converting CO2 and Electrons to Molecules. *Energy Environ. Sci.* **2021**, *14* (7), 3664–3678. https://doi.org/10.1039/D0EE03525D.

(67) Jouny, M.; Luc, W.; Jiao, F. General Techno-Economic Analysis of CO2 Electrolysis Systems. *Ind. Eng. Chem. Res.* **2018**, *57* (6), 2165–2177. https://doi.org/10.1021/acs.iecr.7b03514.

(68) Bar-Even, A. Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry* **2016**, *55* (28), 3851–3863. https://doi.org/10.1021/acs.biochem.6b00495.

(69) Vögeli, B.; Schulz, L.; Garg, S.; Tarasava, K.; Clomburg, J. M.; Lee, S. H.; Gonnot, A.; Moully, E. H.; Kimmel, B. R.; Tran, L.; Zeleznik, H.; Brown, S. D.; Simpson, S. D.; Mrksich, M.; Karim, A. S.; Gonzalez, R.; Köpke, M.; Jewett, M. C. Cell-Free Prototyping Enables Implementation of Optimized Reverse β-Oxidation Pathways in Heterotrophic and Autotrophic Bacteria. *Nat. Commun.* **2022**, *13* (1), 3058. https://doi.org/10.1038/s41467-022-30571-6.

(70) Karim, A. S.; Dudley, Q. M.; Juminaga, A.; Yuan, Y.; Crowe, S. A.; Heggestad, J. T.; Garg, S.; Abdalla, T.; Grubbe, W. S.; Rasor, B. J.; Coar, D. N.; Torculas, M.; Krein, M.; Liew, F. (Eric); Quattlebaum, A.; Jensen, R. O.; Stuart, J. A.; Simpson, S. D.; Köpke, M.; Jewett, M. C. In Vitro Prototyping and Rapid Optimization of Biosynthetic Enzymes for Cell Design. *Nat. Chem. Biol.* **2020**, *16* (8), 912–919. https://doi.org/10.1038/s41589-020-0559-0.

(71) Tickman, B. I.; Burbano, D. A.; Chavali, V. P.; Kiattisewee, C.; Fontana, J.; Khakimzhan, A.; Noireaux, V.; Zalatan, J. G.; Carothers, J. M. Multi-Layer CRISPRa/i Circuits for Dynamic Genetic Programs in Cell-Free and Bacterial Systems. *Cell Syst.* **2021**, *0* (0). https://doi.org/10.1016/j.cels.2021.10.008.

(72) Alba Burbano, D.; Cardiff, R. A. L.; Tickman, B. I.; Kiattisewee, C.; Maranas, C. J.; Zalatan, J. G.; Carothers, J. M. Engineering Activatable Promoters for Scalable and Multi-Input CRISPRa/i Circuits. *Proc. Natl. Acad. Sci.* **2023**, *120* (30), e2220358120. https://doi.org/10.1073/pnas.2220358120.

(73) Dong, X.; Chen, X.; Qian, Y.; Wang, Y.; Wang, L.; Qiao, W.; Liu, L. Metabolic Engineering of Escherichia Coli W3110 to Produce L-Malate. *Biotechnol. Bioeng.* **2017**, *114* (3), 656–664. https://doi.org/10.1002/bit.26190.

(74) Sun, W.; Jiang, B.; Zhao, D.; Pu, Z.; Bao, Y. Integration of Metabolic Pathway Manipulation and Promoter Engineering for the Fine-Tuned Biosynthesis of Malic Acid in Bacillus Coagulans. *Biotechnol. Bioeng.* **2021**, *118* (7), 2597–2608. https://doi.org/10.1002/bit.27780.

(75) Xie, D. Continuous Biomanufacturing with Microbes — Upstream Progresses and Challenges. *Curr. Opin. Biotechnol.* **2022**, *78*, 102793. https://doi.org/10.1016/j.copbio.2022.102793.

(76) Jewett, M. C.; Calhoun, K. A.; Voloshin, A.; Wuu, J. J.; Swartz, J. R. An Integrated Cell-free Metabolic Platform for Protein Production and Synthetic Biology. *Mol. Syst. Biol.* **2008**, *4* (1), 220. https://doi.org/10.1038/msb.2008.57.
(77) Zhou, H.; Zhang, Y.; Long, C. P.; Xia, X.; Xue, Y.; Ma, Y.; Antoniewicz, M. R.; Tao, Y.; Lin, B. A Citric Acid Cycle-Deficient Escherichia Coli as an Efficient Chassis for Aerobic Fermentations. *Nat Commun* **2024**, *15* (1), 2372. https://doi.org/10.1038/s41467-024-46655-4.
(78) Krüger, A.; Mueller, A. P.; Rybnicky, G. A.; Engle, N. L.; Yang, Z. K.; Tschaplinski, T. J.; Simpson, S. D.; Köpke, M.; Jewett, M. C. Development of a Clostridia-Based Cell-Free System for Prototyping Genetic Parts and Metabolic Pathways. *Metab. Eng.* **2020**, *62*, 95–105. https://doi.org/10.1016/j.ymben.2020.06.004.
(79) Calhoun, K. A.; Swartz, J. R. Energy Systems for ATP Regeneration in Cell-Free Protein Synthesis Reactions. In *In Vitro Transcription and Translation Protocols*; Grandi, G., Ed.; Humana Press: Totowa, NJ, 2007; pp 3–17. https://doi.org/10.1007/978-1-59745-388-2_1.
(80) Kim, H.-C.; Kim, D.-M. Methods for Energizing Cell-Free Protein Synthesis. *J. Biosci. Bioeng.* **2009**, *108* (1), 1–4. https://doi.org/10.1016/j.jbiosc.2009.02.007.
(81) Caschera, F.; Noireaux, V. A Cost-Effective Polyphosphate-Based Metabolism Fuels an All *E. Coli* Cell-Free Expression System. *Metab. Eng.* **2015**, *27*, 29–37. https://doi.org/10.1016/j.ymben.2014.10.007.
(82) Bowie, J. U.; Sherkhanov, S.; Korman, T. P.; Valliere, M. A.; Opgenorth, P. H.; Liu, H. Synthetic Biochemistry: The Bio-Inspired Cell-Free Approach to Commodity Chemical Production. *Trends Biotechnol.* **2020**, *38* (7), 766–778. https://doi.org/10.1016/j.tibtech.2019.12.024.
(83) Dudley, Q. M.; Karim, A. S.; Jewett, M. C. Cell-Free Metabolic Engineering: Biomanufacturing beyond the Cell. *Biotechnol. J.* **2015**, *10* (1), 69–82. https://doi.org/10.1002/biot.201400330.
(84) HamediRad, M.; Chao, R.; Weisberg, S.; Lian, J.; Sinha, S.; Zhao, H. Towards a Fully Automated Algorithm Driven Platform for Biosystems Design. *Nat. Commun.* **2019**, *10* (1), 5150. https://doi.org/10.1038/s41467-019-13189-z.
(85) Radivojević, T.; Costello, Z.; Workman, K.; Garcia Martin, H. A Machine Learning Automated Recommendation Tool for Synthetic Biology. *Nat. Commun.* **2020**, *11* (1), 4879. https://doi.org/10.1038/s41467-020-18008-4.
(86) Martin, J. P.; Rasor, B. J.; DeBonis, J.; Karim, A. S.; Jewett, M. C.; Tyo, K. E. J.; Broadbelt, L. J. A Dynamic Kinetic Model Captures Cell-Free Metabolism for Improved Butanol Production. *Metab. Eng.* **2023**, *76*, 133–145. https://doi.org/10.1016/j.ymben.2023.01.009.
(87) Wei, Z.; Xu, Y.; Xu, Q.; Cao, W.; Huang, H.; Liu, H. Microbial Biosynthesis of L-Malic Acid and Related Metabolic Engineering Strategies: Advances and Prospects. *Front. Bioeng. Biotechnol.* **2021**, *9*. https://doi.org/10.3389/fbioe.2021.765685.
(88) Irwin, S. CO2 Production by the U.S. Ethanol Industry and the Potential Value of Sequestration. *Farmdoc Dly.* **2024**, *14* (34).
(89) Mangili, P. V.; Dias, R. F.; Santos, L. S.; Prata, D. M. COMPARISON OF THE CO2 EMISSIONS FROM TWO MALEIC ANHYDRIDE PRODUCTION PROCESSES

THROUGH COMPUTATIONAL SIMULATION. *Lat. Am. Appl. Res. - Int. J.* **2019**, *49* (1), 1–6. https://doi.org/10.52292/j.laar.2019.289.

(90) Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, P.; Lukas, J.; Olthof, B.; Worley, M.; Sexton, D.; Dudgeon, D. *Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover*; NREL/TP-5100-47764, 1013269; 2011; p NREL/TP-5100-47764, 1013269. https://doi.org/10.2172/1013269.

(91) Dutta, A.; Mukarakate, C.; Iisa, K.; Wang, H.; Talmadge, M.; Santosa, D.; Harris, K.; Baddour, F.; Hartley, D.; Cai, H.; Ou, L.; Schaidle, J.; Griffin, M. *Ex Situ Catalytic Fast Pyrolysis of Lignocellulosic Biomass to Hydrocarbon Fuels: 2020 State of Technology*; NREL/TP-5100-80291, 1805204, MainId:42494; 2021; p NREL/TP-5100-80291, 1805204, MainId:42494. https://doi.org/10.2172/1805204.

Supplementary Figures

Supplementary Tables

Table S1: Bioproduction & carbon efficiency calculations

Table S2: Base case TEA assumptions

Table S3: Base case TEA results

Table S4: DNA sequences used in this study

Supplementary Methods

Methods S1: Comparative sequence analysis of enzyme isoforms

Methods S2: Techno-economic analysis

Methods S3: Preparation of cell-free bioproduction reagents and reactions

Methods S4: CO_2 quantification with GC/MS

Supplementary References

Supplementary Figures

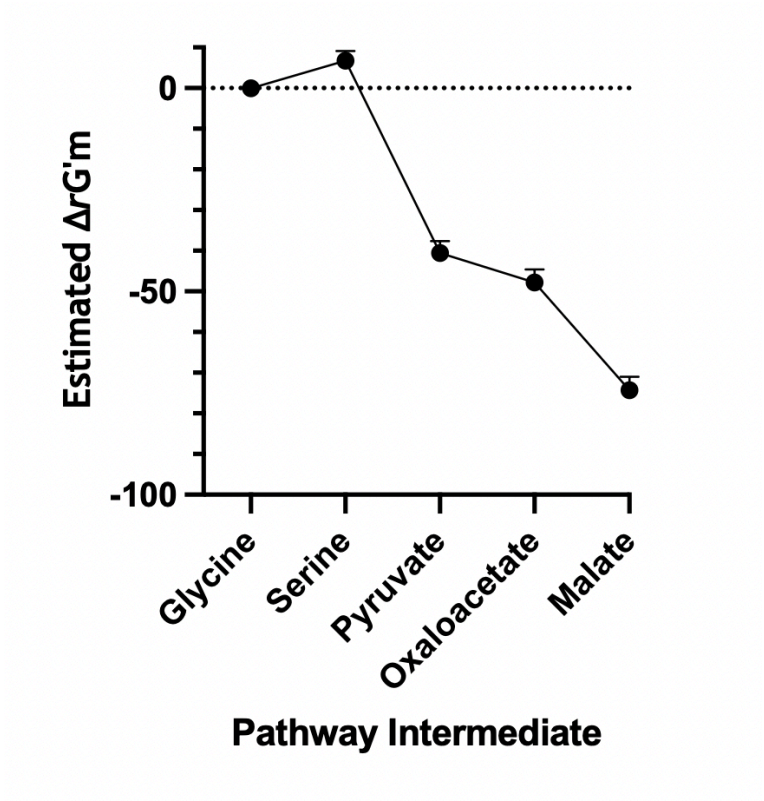


Figure S1. Thermodynamics for the synthesis of malate from glycine. The ΔG'ᵐ of each step was calculated using eQuilibriator assuming a standard concentration of 1mM for all reactants.

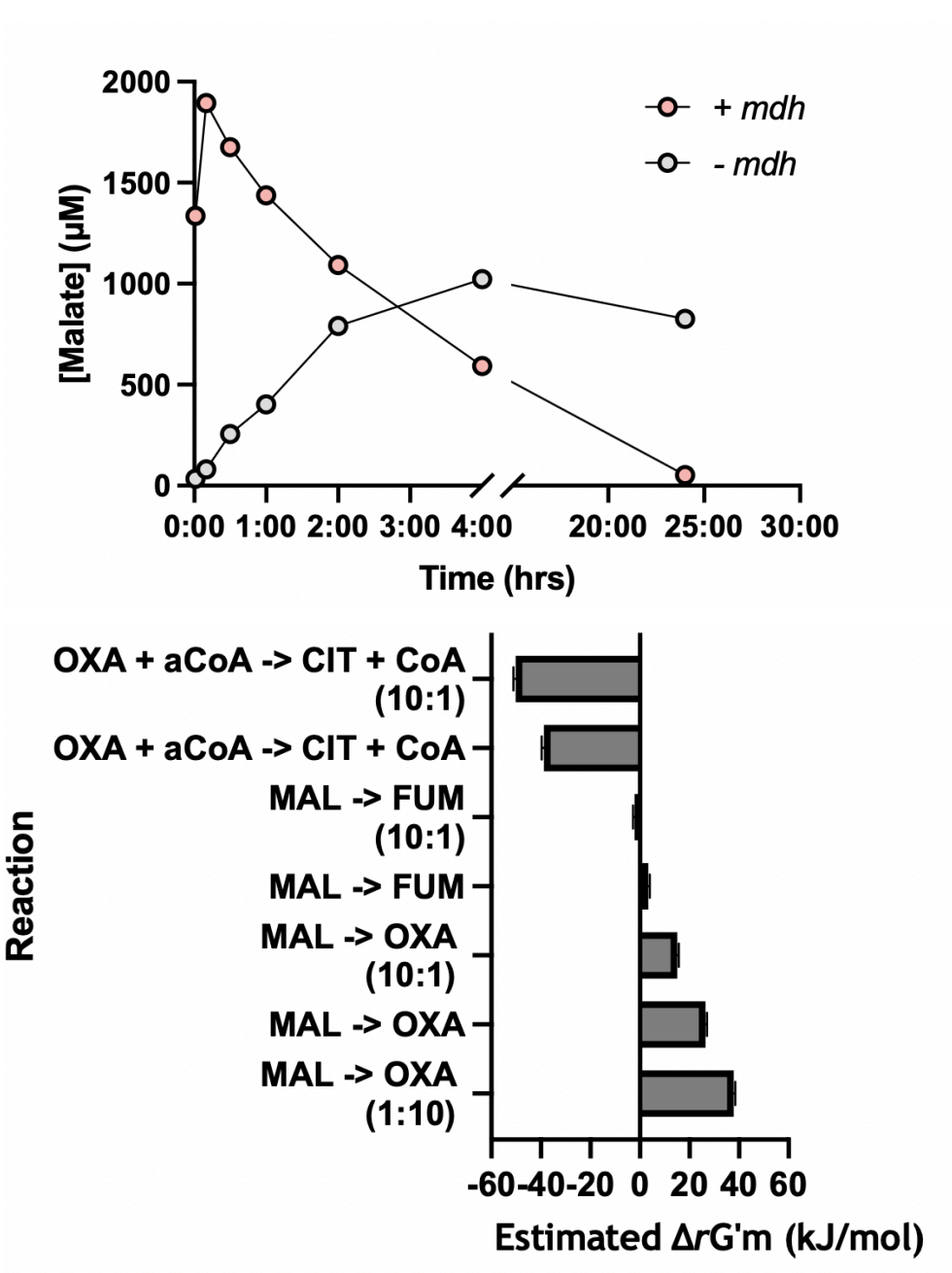


Figure S2. *mdh*-dependent malate concentrations over time. **A.** Malate concentrations are shown over 24 hours for CFE with or without *mdh* overexpression. Reactions contain oxaloacetate and NADH at 1mM each. *mdh* is expressed at 5 nM for the '+ *mdh*' condition. CFE reactions are diluted 10-fold in the 50 mM HEPES pH 8. Time points

were collected at 0, 10, 30, 60, 120, 240, and 1440 minutes. Values shown represent one technical replicate. **B.** Thermodynamic predictions are given for relevant malate- and oxaloacetate-consuming reactions to identify sources for non-specific malate production and loss. The $\Delta G'^{m}$ for each reaction was calculated using eQuilibriator. For reactions with the labels (10:1) or (1:10), predictions were adjusted to assume either a 10:1 or 1:10 molar ratio of substrates to products. All other reactions assume a standard concentration of 1mM for all reactants. Metabolite abbreviations: OXA, oxaloacetate; MAL, malate; CIT, citrate; FUM, fumarate; aCoA, acetyl-CoA.

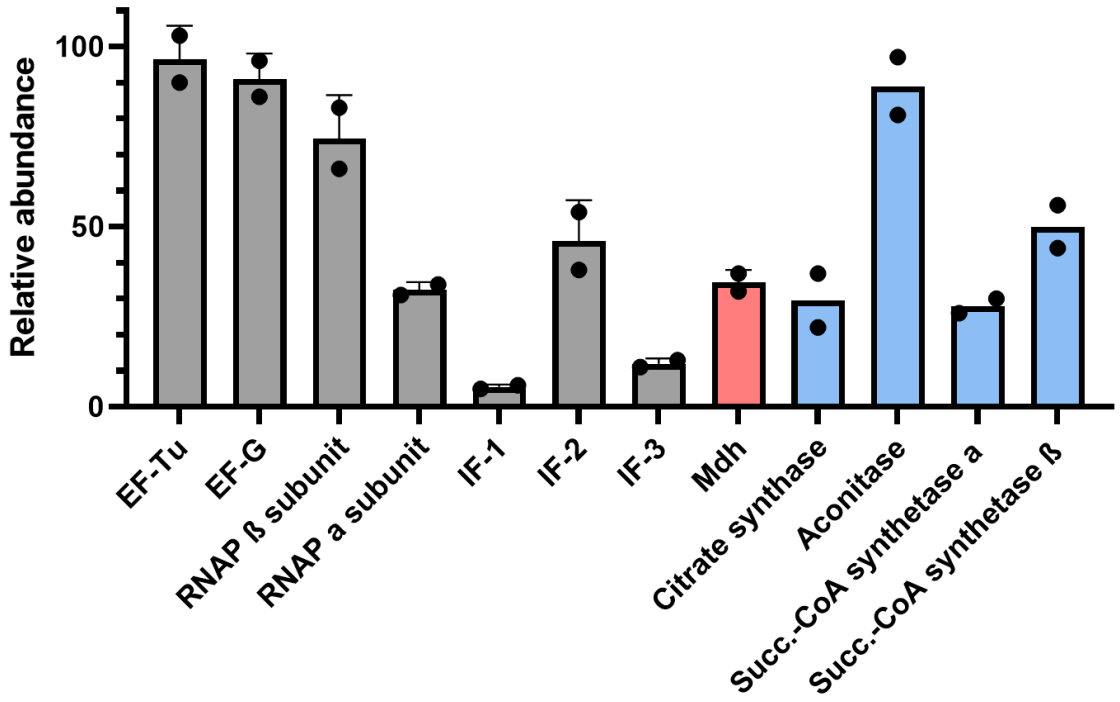


Figure S3. Proteome analysis of TCA enzymes in an *E. coli*-based CFE. Published proteomic data from CFE was analyzed to compare relative amounts of TCA cycle enzymes to highly expressed enzymes required for key transcription and translation processes.

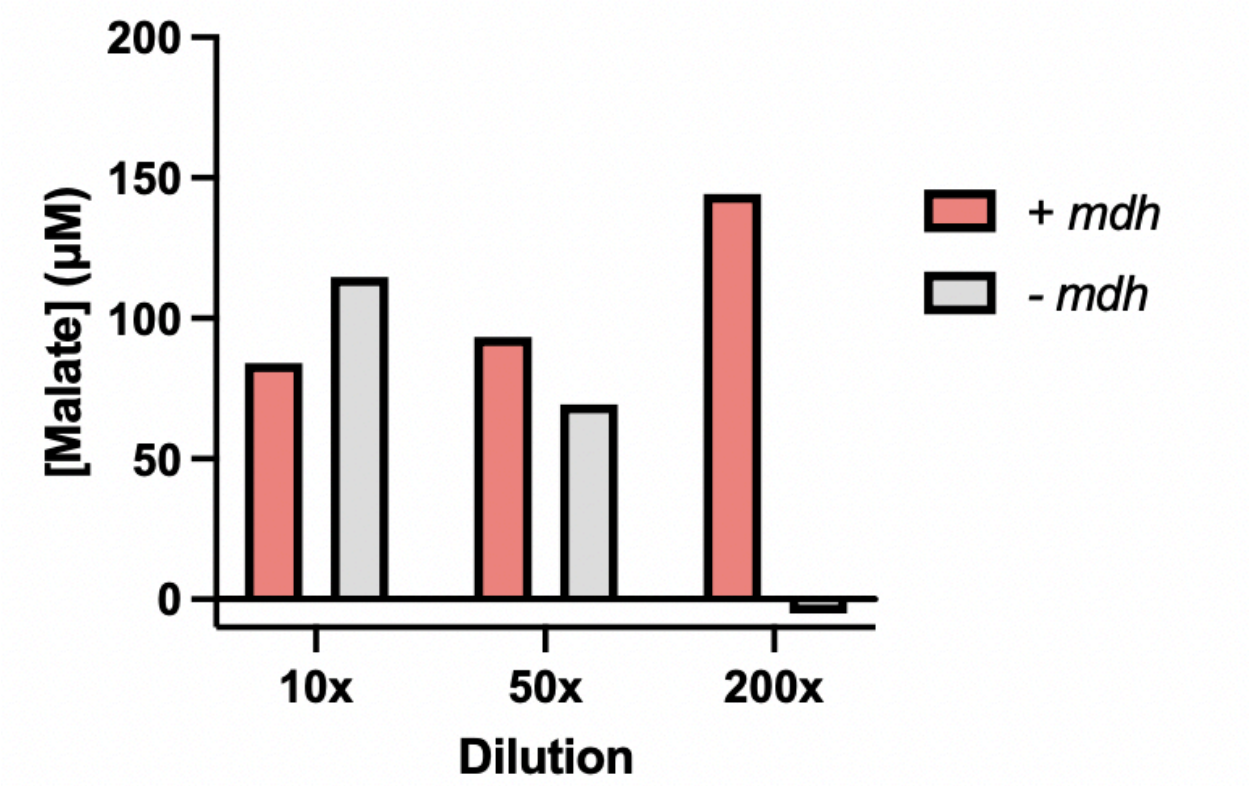


Figure S4. Effect of dilution on *mdh*-dependent malate production. Malate concentrations are shown for reactions expressing *mdh* and *fdh* or *fdh* alone. Reactions were diluted 10, 50, or 200-fold into a buffer containing 1mM NADH, 10mM formate, and 1mM oxaloacetate. Malate concentrations were measured after four hours. Values shown represent one technical replicate.

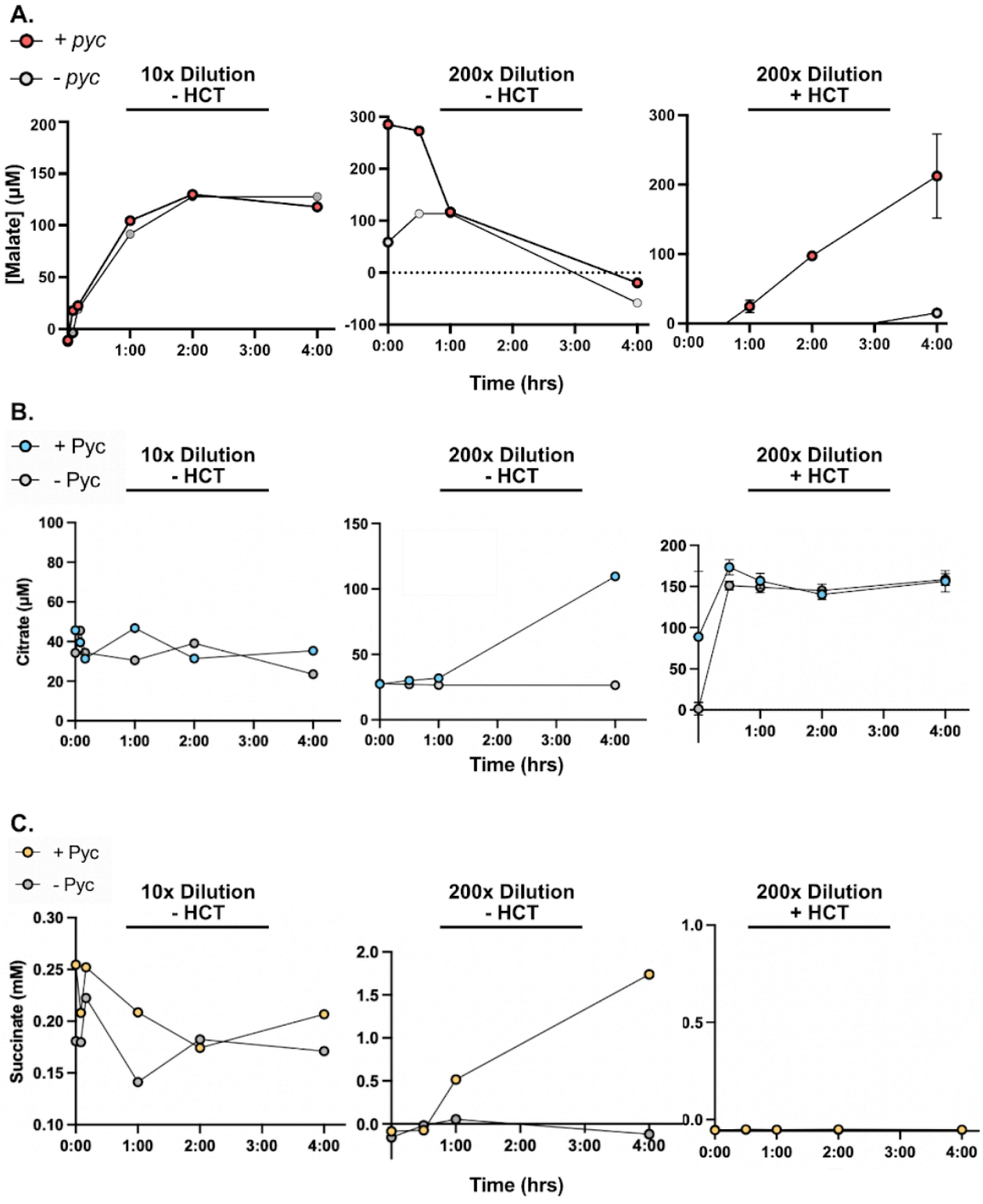

A.
+ pyc
- pyc
10x Dilution
- HCT
200x Dilution
- HCT
200x Dilution
+ HCT
[Malate] (µM)
Time (hrs)
B.
+ Pyc
- Pyc
10x Dilution
- HCT
200x Dilution
- HCT
200x Dilution
+ HCT
Citrate (µM)
Time (hrs)
C.
+ Pyc
- Pyc
10x Dilution
- HCT
200x Dilution
- HCT
200x Dilution
+ HCT
Succinate (mM)
Time (hrs)

Figure S5. Time-course measurements of interventions to improve *pyc*-dependent malate production. Concentrations are shown for **(A)** malate, **(B)** citrate, and **(C)** succinate over four hours for CFE reactions diluted either 10- or 200-fold and with or without hydroxycitrate (HCT) as an inhibitor of the TCA cycle. Reactions in the left panels contain 0.5mM pyruvate, 1mM HCO_3^-, 1mM ATP, 1mM $Mg(CH_3COO)_2$, 1mM acetyl-CoA, 1mM NADH, 10mM formate. Reactions in middle and right panels contain 1mM pyruvate, 5mM HCO_3^-, 1mM ATP, 2mM $Mg(CH_3COO)_2$, 1mM acetyl-CoA, 1mM NADH, 15mM formate. Reactions containing HCT contain 2.5mM. In the left and middle panels, values represent one technical replicate. In the right panels, values represent the mean ± standard deviation of three technical replicates.

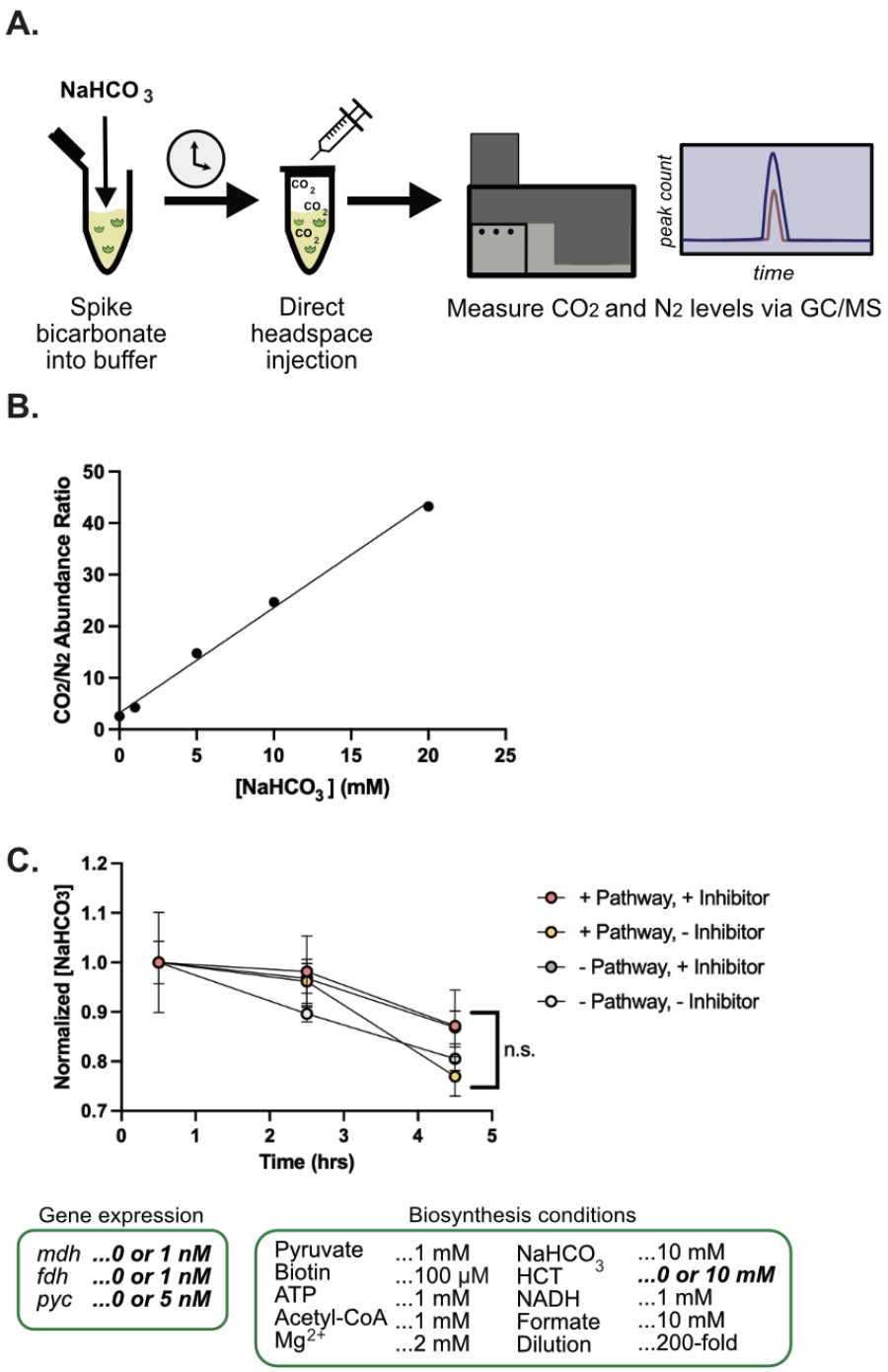


Figure S6. Measurement of CO_2 in reaction headspace. **A.** Bicarbonate is spiked into cell-free reactions during the biosynthesis step. We then measured CO_2 in the headspace of our reactions via GC/MS at multiple time points after mixing reactions.

Reaction headspace was sampled via direct headspace injection into the GC/MS **(Methods S4). B.** We first showed there is a linear relationship between the amount of bicarbonate spiked into the reaction buffer and the CO_2 measured in the headspace based on the quickly established equilibrium between HCO_3^- and CO_2 in the liquid and gas phases. We normalized the CO_2 peak abundance by the N_2 signal to account for variations in manual injections. **C.** We prepared reactions for pyruvate to malate conversion using our standard conditions and measured the reaction after 0, 2, and 4 hours to investigate whether we could measure differential depletion of CO_2 as a readout for the reaction progress. However, we found no significantly different changes in CO_2 levels relative to the starting time point for any of the conditions tested. This may be due to the low amount of bicarbonate that is converted in the reaction relative to the total amount in the system. A 20% conversion of pyruvate to malate would only correspond to a 2% change in the concentration of the bicarbonate in the reaction, as bicarbonate is added at a 10-fold higher concentration than pyruvate. Bicarbonate levels are kept in excess due to the high Km of the enzyme for bicarbonate (~10 mM). These small differences, in combination with fluctuations in CO_2 levels resulting from other pathways in the lysate, may be difficult to detect via GC/MS. Therefore, moving forward, we relied on product accumulation, measured via LC/MS, to assess the carbon efficiency of the system. Values represent the mean ± standard deviation of two technical replicates.

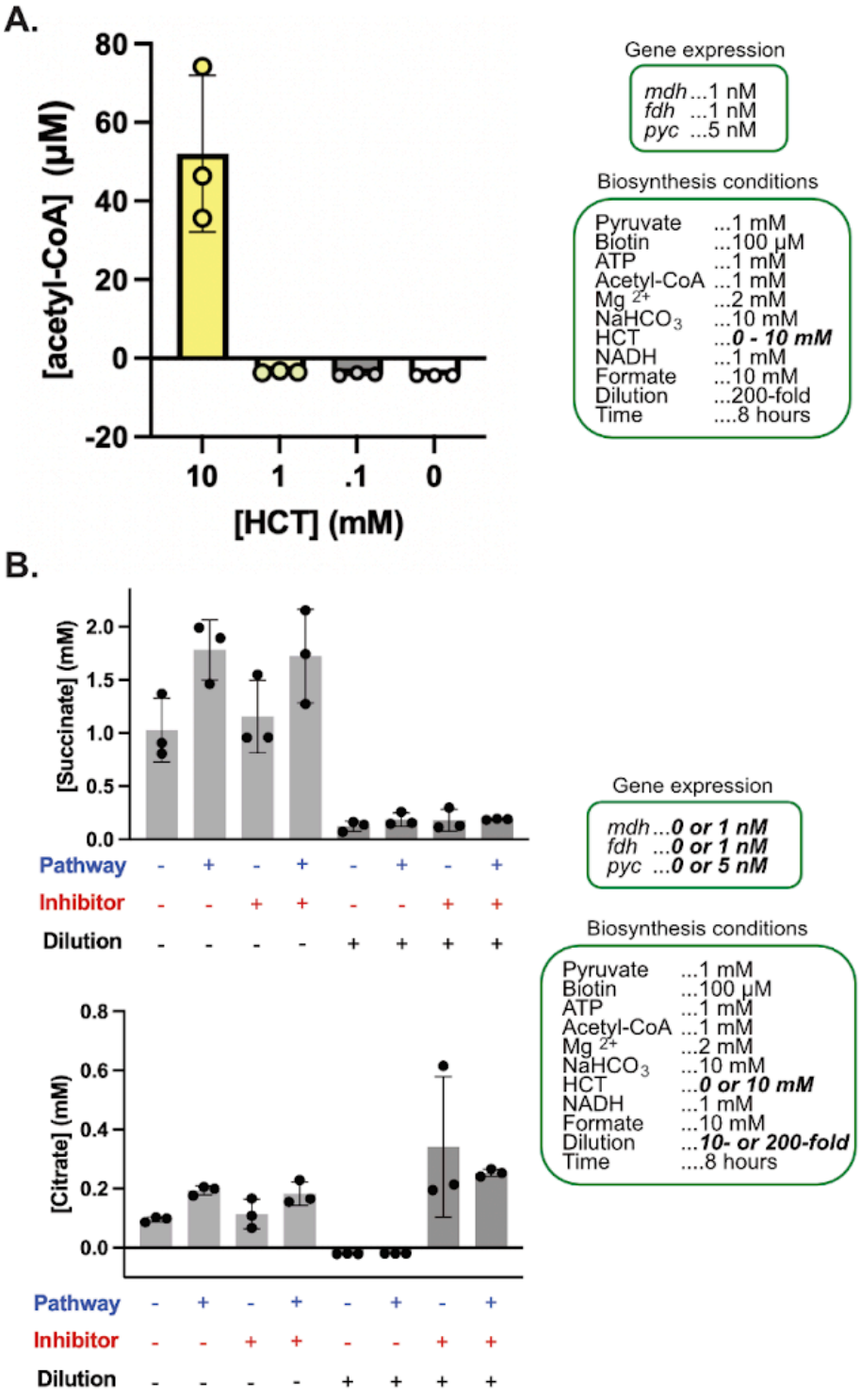

A.
80
60
40
20
0
-20
[acetyl-CoA] (μM)
10
1
.1
0
[HCT] (mM)
Gene expression
mdh ...1 nM
fdh ...1 nM
pyc ...5 nM
Biosynthesis conditions
Pyruvate ...1 mM
Biotin ...100 μM
ATP ...1 mM
Acetyl-CoA ...1 mM
Mg 2+ ...2 mM
NaHCO3 ...10 mM
HCT ...0 - 10 mM
NADH ...1 mM
Formate ...10 mM
Dilution ...200-fold
Time ....8 hours
B.
[Succinate] (mM)
2.0
1.5
1.0
0.5
0.0
Pathway - + - + - + - +
Inhibitor - - + + - - + +
Dilution - - - - + + + +
[Citrate] (mM)
0.8
0.6
0.4
0.2
0.0
Pathway - + - + - + - +
Inhibitor - - + + - - + +
Dilution - - - - + + + +
Gene expression
mdh ...0 or 1 nM
fdh ...0 or 1 nM
pyc ...0 or 5 nM
Biosynthesis conditions
Pyruvate ...1 mM
Biotin ...100 μM
ATP ...1 mM
Acetyl-CoA ...1 mM
Mg 2+ ...2 mM
NaHCO3 ...10 mM
HCT ...0 or 10 mM
NADH ...1 mM
Formate ...10 mM
Dilution ...10- or 200-fold
Time ....8 hours

Figure S7. Effect of hydroxycitrate and dilution on accumulation of TCA intermediates.

A. Acetyl-CoA accumulation is hydroxycitrate dependent. Acetyl-CoA concentrations are shown after four hours for reactions with different concentrations of hydroxycitrate (HCT) for reactions diluted 200-fold. Reactions express *pyc*, *mdh*, and *fdh*, and contain all of the needed substrates and cofactors for pyruvate to malate conversion. Acetyl-CoA and oxaloacetate, the product of pyc, are substrates for citrate synthase. The accumulation of acetyl-CoA at 10mM of HCT suggests that citrate synthase is partially inhibited at this concentration. **B.** Concentrations are shown for TCA intermediates succinate (top) and citrate (bottom) after four hours for reactions with different conditions. DNA for pathway enzymes, additional CFE dilution, and the inhibitor hydroxycitrate are changed between conditions. Data shown in these panels are from the same experiments as in Figure 3C. Dilution of the CFE reduces accumulation of TCA intermediates, suggesting that activity of endogenous TCA enzymes is lessened in these conditions. In diluted reactions containing HCT, we see accumulation of high concentrations of citrate. The high levels of citrate may be because hydroxycitrate is a stronger inhibitor for the downstream TCA enzymes aconitase and isocitrate dehydrogenase than citrate synthase. This results in build up of citrate, preventing loss of CO_2 through the TCA cycle. For both intermediates, the concentration does not change depending on the presence of the pathway, suggesting that malate production via the oxidative TCA cycle is low. For all panels, values represent the mean ± standard deviation of three technical replicates.

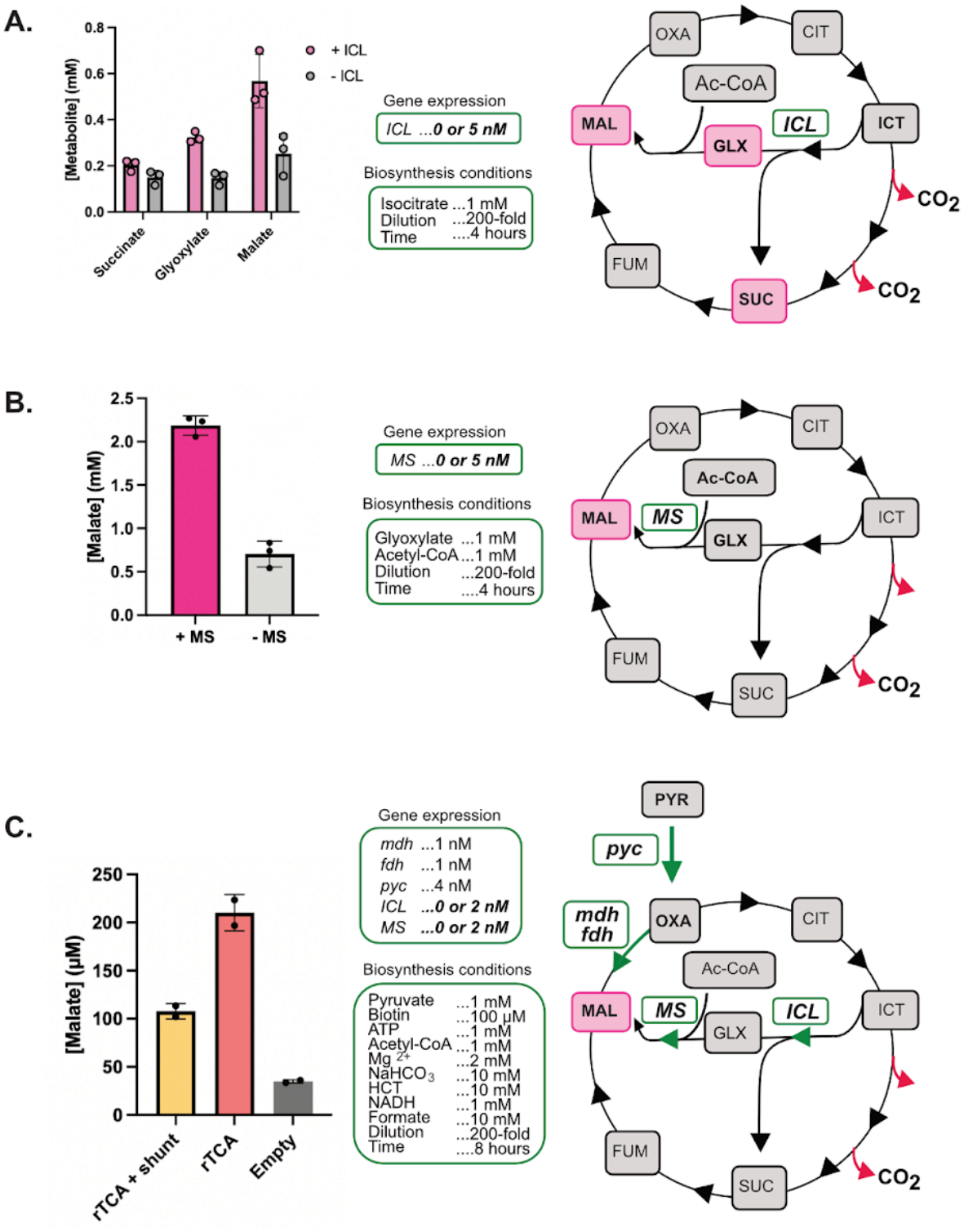


Figure S8. Validation of glyoxylate shunt as a complementary carbon-conserving pathway in CFE. **A.** Isocitrate lyase (ICL) was individually expressed in CFE and

reacted with its substrate, isocitrate. Concentrations are given for the direct products, succinate and glyoxylate, as well as malate. **B.** Malate synthase (MS) was individually expressed in CFE and reacted with its substrates, glyoxylate and acetyl-CoA. Malate concentrations were measured after four hours. **C.** Reactions containing either the rTCA module (*pyc*, *mdh*, and *fdh*) or rTCA + shunt (*pyc*, *mdh*, *fdh*, *ICL, MS*) were measured after eight hours. Malate concentrations may be lower in the rTCA + shunt due to the limited gene expression capacity of the cell-free system or due to consumption of acetyl-CoA, an allosteric activator for *pyc*, from the shunt pathway. For all panels, values represent the mean ± standard deviation of at least two technical replicates.

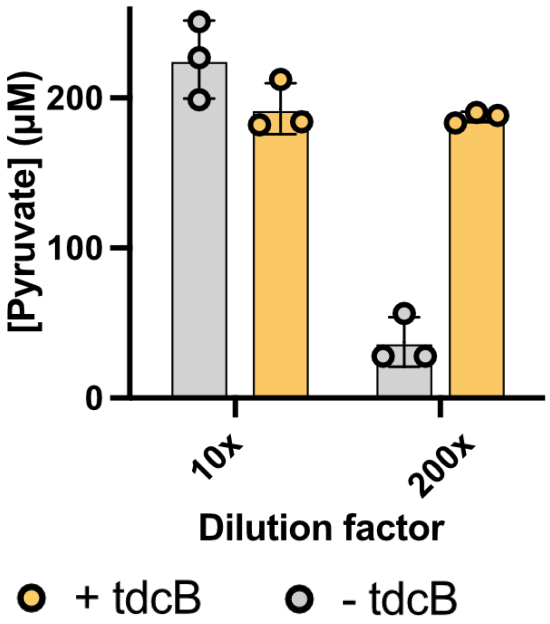


Figure S9. Dilution of CFE improves pathway-specific pyruvate production. Pyruvate concentrations are shown for reactions containing serine, PLP, and AMP. CFE reactions containing *tdcB* or a no enzyme control are diluted 10- or 200-fold in the final reaction. Pyruvate is measured after four hours. Values represent the mean ± standard deviation of three technical replicates.

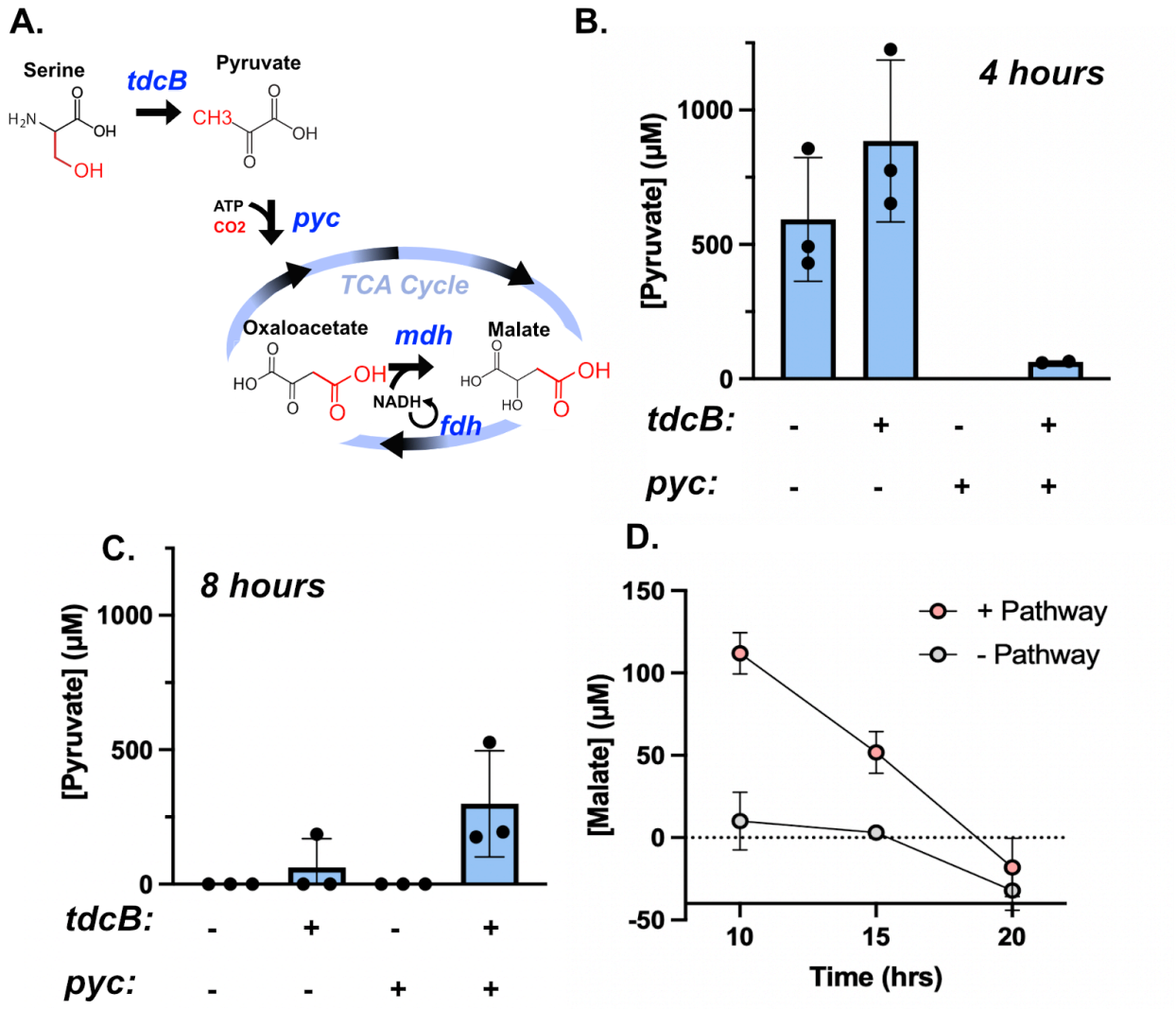


Figure S10. Time-course serine to malate bioconversion via a four-enzyme pathway. **A.** Serine is converted to malate using a four enzyme pathway in which *tdcB* converts serine to pyruvate, and pyruvate is carboxylated into oxaloacetate via *pyc*. Oxaloacetate is then reduced to malate by *mdh*, with cofactor regeneration from *fdh* to drive flux towards malate. For all panels, reactions contain *mdh* and *fdh* expressed at .1 and 1.5 nM, respectively. *pyc* and *tdcB* are either included at 2.5 nM or excluded in each condition. **B.** Pyruvate concentrations are shown at four hours. In reactions with *pyc*, we see a large decrease in pyruvate titers, suggesting it is being converted to oxaloacetate.

C. Pyruvate concentrations are shown at eight hours. The buildup of pyruvate in the conditions with *tdcB* expressed suggests that conversion from pyc may not be able to match the rate of production from *tdcB*. **D.** Malate concentrations are shown at 10, 15, and 20 hours when the pathway is expressed. This data is from a different experiment to that shown in panels B and C. The decrease in malate over time may be due to instability of cofactors or enzymes in our engineered pathway. For all panels, values represent the mean ± standard deviation of at least two technical replicates.

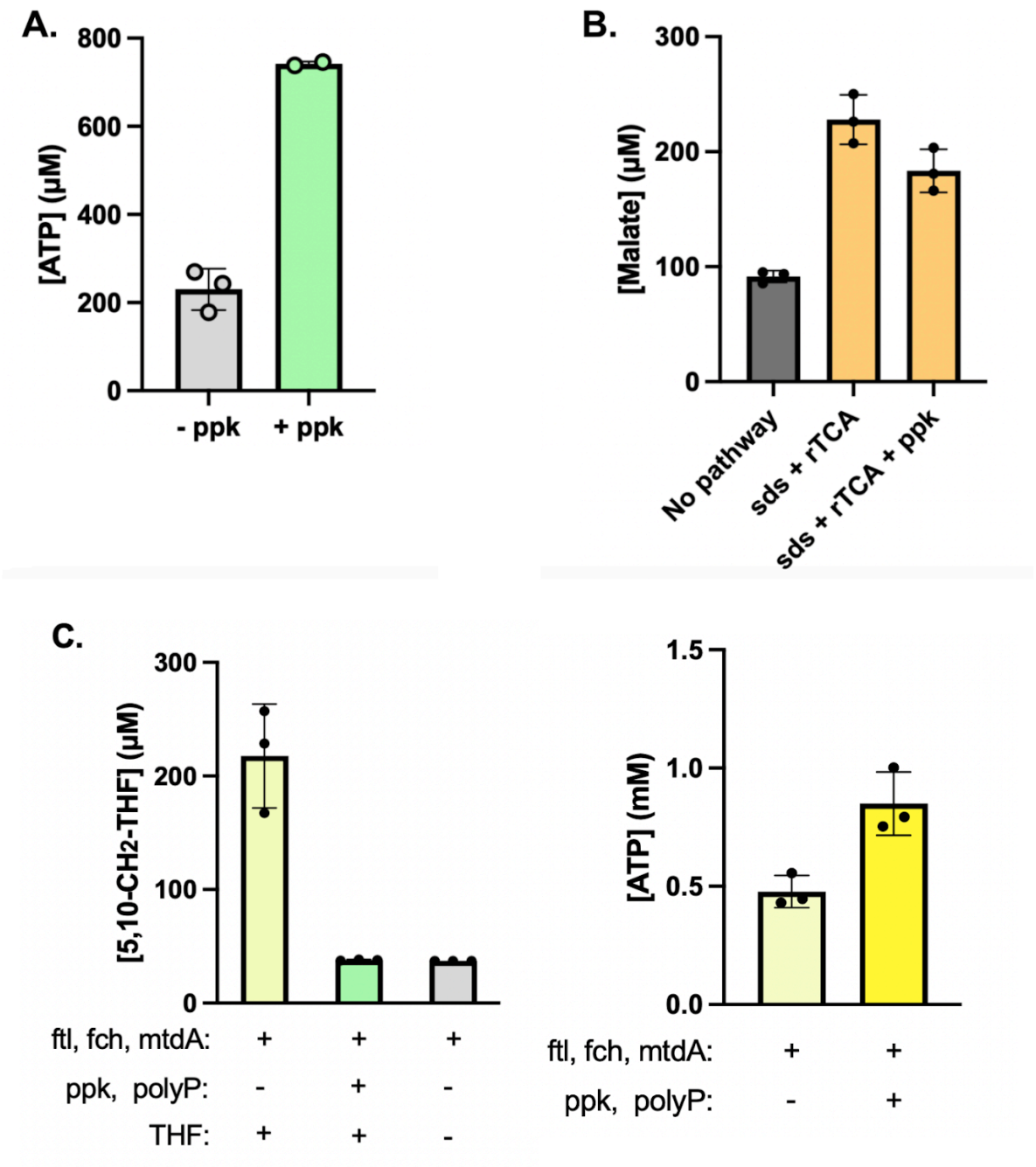


Figure S11. Effect of ATP-regenerating ppk on cell-free biosynthesis. A. ATP concentrations are shown for reactions containing polyphosphate kinase from a *Erysipelotrichaceae* bacterium (*ppk12*). Reactions contain 1mM AMP and 10mM polyphosphate. Reactions are collected after four hours. B. Malate concentrations are

shown for reactions containing sds + rTCA (*pyc*, *mdh*, and *fdh*) with or without ppk. The similar malate concentrations between the conditions with or without ppk suggest that ATP availability may not be limiting in our system for malate production. C. 5,10-CH_2-THF and ATP concentrations are shown for reactions containing the formate assimilation pathway, consisting of the ftl, fch, and mtdA enzymes, and either including or excluding the ATP regeneration pathway. On the left, reactions either include or exclude the ATP regeneration system shown in Panel A, or THF as a negative control. The loss of 5,10-CH_2-THF accumulation in the presence of the ATP regeneration system suggests that it may interfere with function of the formate assimilation pathway. Panels A, B, and C represent separate experiments. For all panels, values represent the mean ± standard deviation of at least two technical replicates.

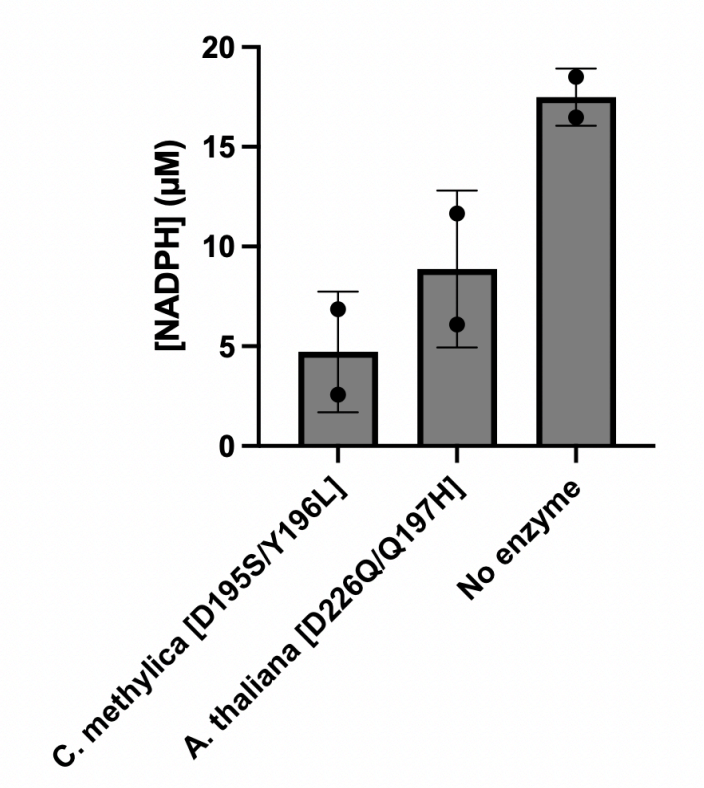


Figure S12. Lack of activity from NADPH-specific *fdh* variants in CFE. NADPH concentrations are shown for three CFE reactions containing either a NADPH-specific formate dehydrogenase (fdh) variant or no enzyme. We did not observe differential NADPH levels for either enzyme tested. This may suggest that these enzymes are not well-expressed in our system, or that NADPH is being consumed rapidly by endogenous enzymes in the CFE. Reactions are fed with 1mM $NADP^+$ and 10mM formate. Reactions are measured after four hours. Values represent the mean ± standard deviation of two technical replicates.

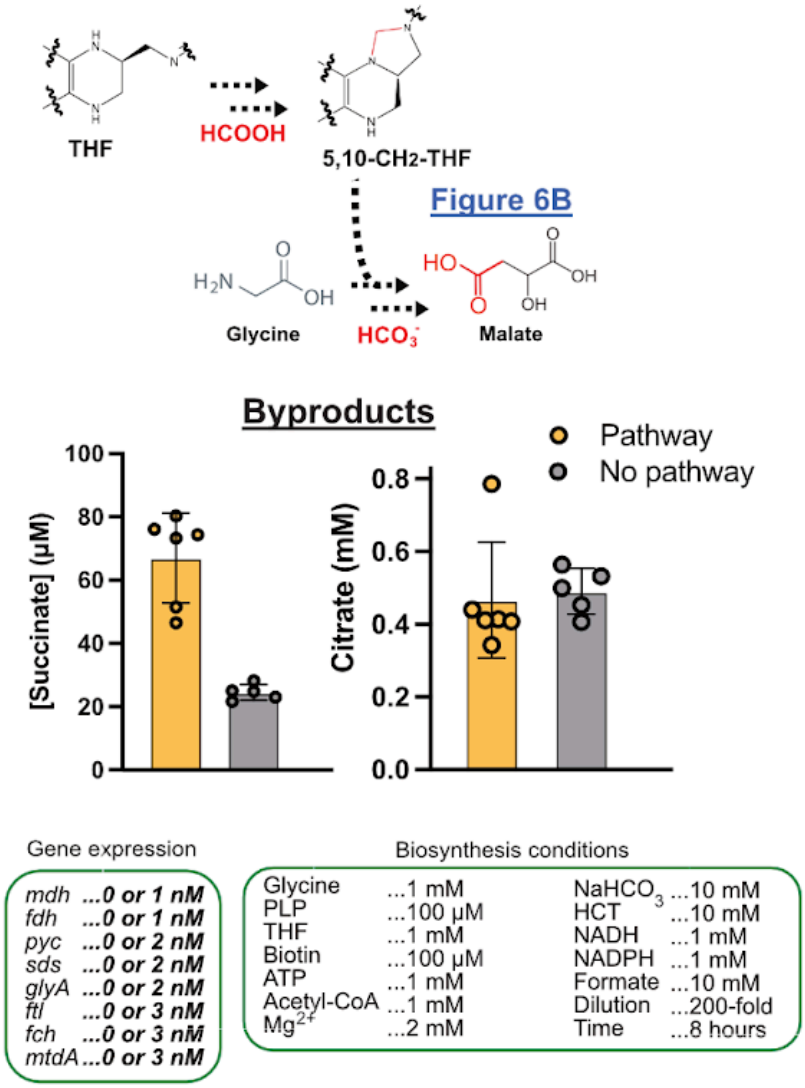


Figure S13. Side product formation through the assembled formate assimilation and rTCA pathways. Succinate (left) and citrate (right) concentrations are shown for reactions designed for glycine, formate, and bicarbonate conversion to malate. Malate concentrations for the same experiment are shown in Figure 6B. This experiment shows that TCA intermediates accumulate despite the addition of hydroxycitrate. There are roughly equimolar amounts of malate and succinate produced in the sample with the pathway expressed. We see accumulation of citrate regardless of the presence of our

pathway. The high levels of citrate may be because hydroxycitrate is a stronger inhibitor for the downstream TCA enzymes aconitase and isocitrate dehydrogenase than citrate synthase. This results in citrate or isocitrate becoming the stopping point in the TCA cycle. The literature reports that hydroxycitrate inhibits aconitase and isocitrate dehydrogenase with K_i values of roughly .3 mM [1,2], while hydroxycitrate inhibits citrate synthase with a K_i of roughly 10 mM [3]. Therefore, at 10 mM, hydroxycitrate is likely providing much stronger inhibition of downstream TCA enzymes than CS, leading to a buildup of citrate. This, in turn, would lead to product inhibition of CS [4], further decreasing its activity.

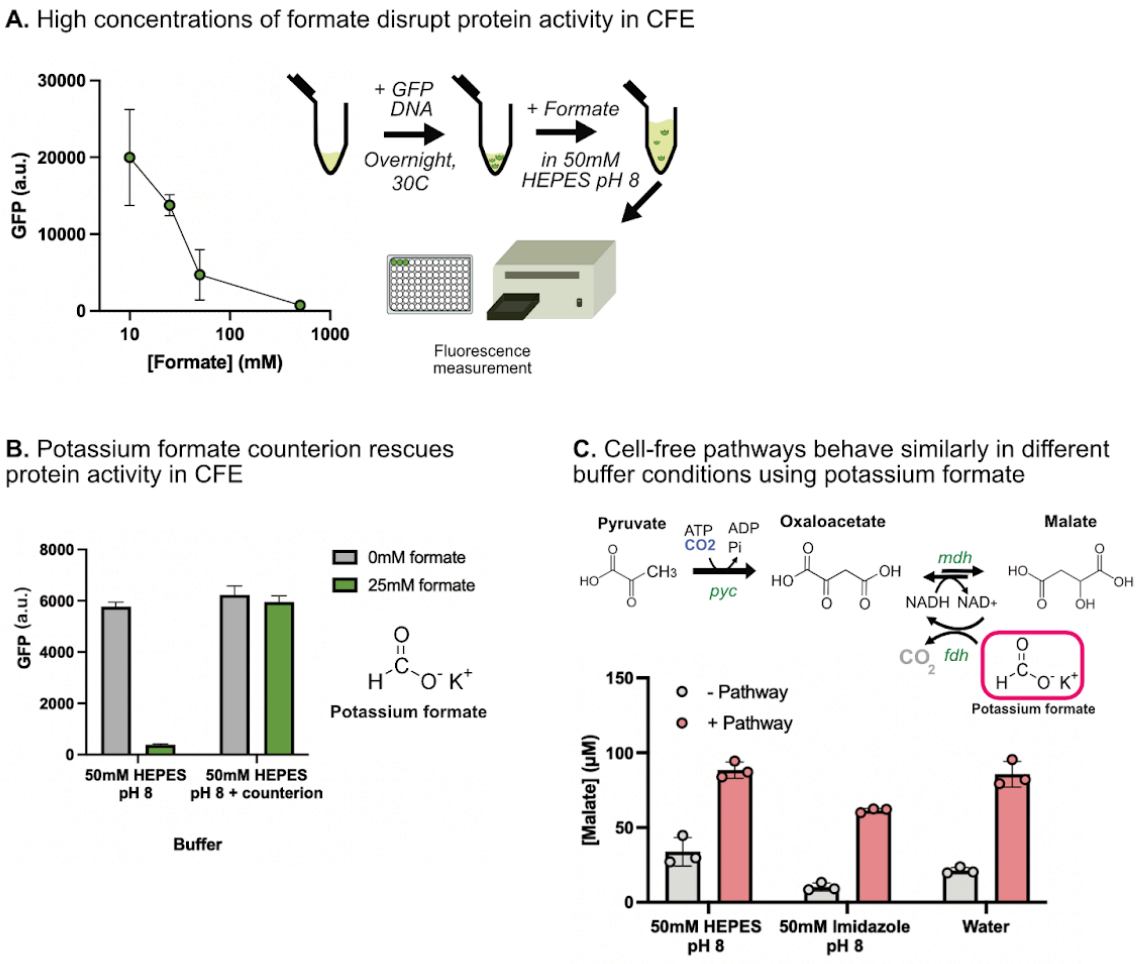


Figure S14. Buffered formic acid maintains CFE activity. **A.** GFP expression is poisoned by high concentrations of formic acid. GFP levels are shown for cell-free reactions spiked with increasing concentrations of formate. Cell-free reactions expressing GFP were incubated overnight at 30C. After 16-24 hours, cell-free reactions were diluted in 50 mM HEPES pH 8 and split into different replicates. Concentrated formic acid was then added and the reactions were incubated for 10 minutes at room temperature. The fluorescence from the reactions was then measured via plate reader using the excitation wavelength at 485 nm and emission wavelength at 528 nm. **B.** Formic acid is buffered with a potassium counterion to maintain CFE activity by minimizing pH change in the

reaction. Cell-free reactions were prepared as in Panel A. Formate was added at 25 mM either as formic acid or as potassium formate. **C.** Cell-free pathways behave similarly in different buffer conditions using potassium formate. Malate titers are shown for the three-enzyme system for pyruvate to malate transformation in three different buffer conditions: 50 mM HEPES pH 8, 50 mM imidazole pH 8, and water. We found that the buffer used for dilution of the cell-free reaction did not have a significant effect on malate titers when formate was buffered. This result suggests that providing a counterion for formate effectively protects from large changes in pH upon its addition to the reaction.

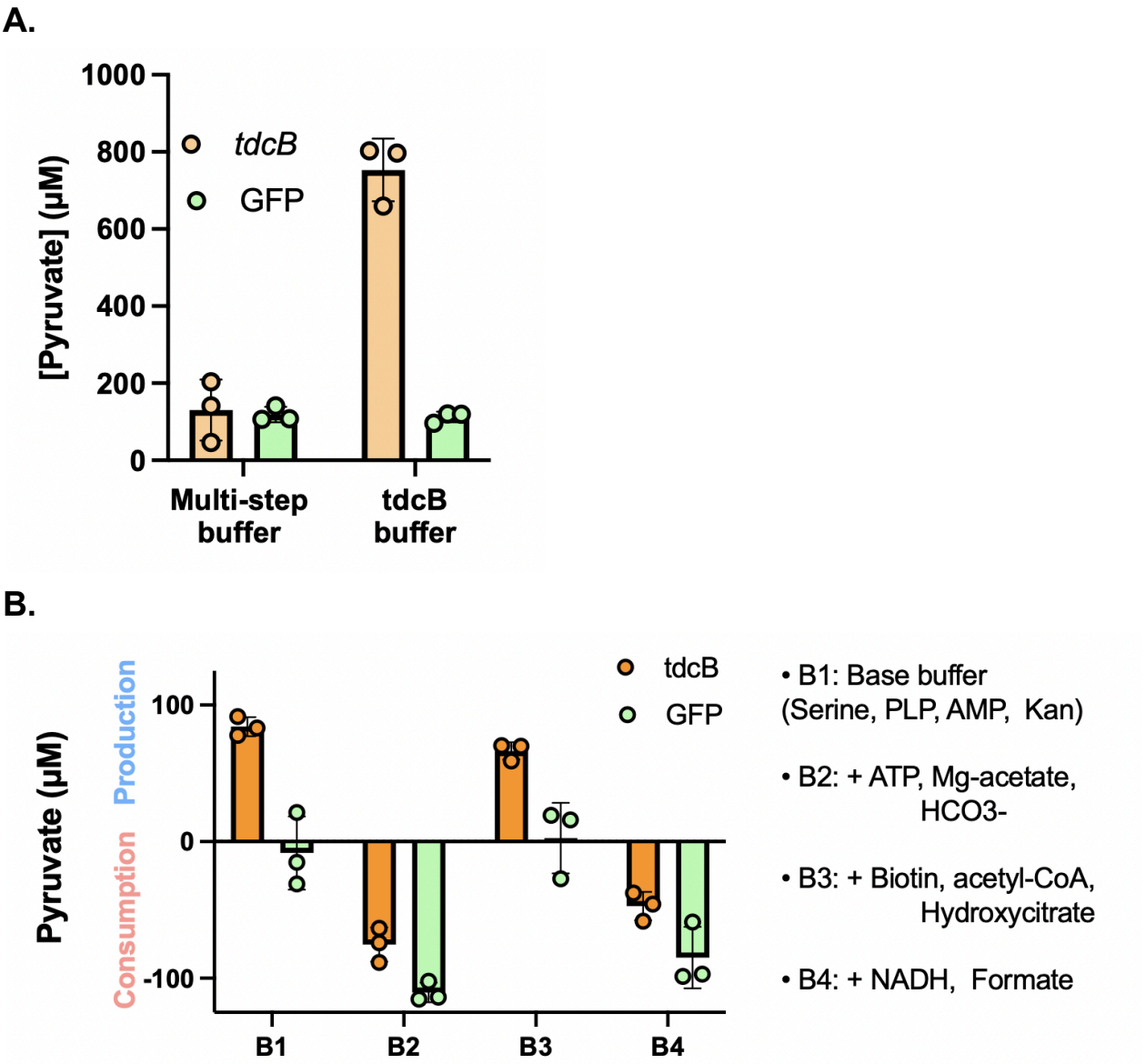


Figure S15. Pyruvate accumulation in the presence of downstream cofactors. To investigate potential pathway incompatibilities, we measured pyruvate accumulation from serine using *tdcB* in two buffer conditions. Pyruvate concentrations are shown in the presence of the cofactors required for downstream steps of the pathway.

A. CFE expressing *tdcB* is compared to a reaction expressing GFP in conditions containing either *tdcB* buffer (TB) (1mM AMP, 1mM PLP) or with a multi-step buffer (MSB) with the cofactors required for *pyc* and *mdh* as well (1mM AMP, 1mM PLP, 1mM

ATP, 2mM $Mg(CH_3COO)_2$, 1mM biotin, 1mM acetyl-CoA, 2.5mM hydroxycitrate, 1mM NADH, 10mM formate). Under MSB conditions, we see no significant pyruvate accumulation after four hours.

B. Additional buffer conditions are tested to understand which cofactors lead to higher consumption of pyruvate in our system, with each buffer containing different combinations of cofactors required for either *mdh* or *pyc* function. B1 contains the necessary cofactors for *tdcB* function (1mM AMP, 1mM PLP). B2 contains B1 with 1mM ATP, 2mM $Mg(CH_3COO)_2$, and 10mM HCO_3^-. B3 contains B1 with 1mM biotin, 1mM acetyl-CoA, and 2.5mM hydroxycitrate. B4 contains B1 with 1mM NADH and 10mM formate. As our CFE contains a significant amount of pyruvate initially, production and consumption are measured relative to the amount of pyruvate present at 0 hours. We found that pyruvate was being consumed faster than it was produced in the buffers containing ATP (B2) and NADH (B4), but not in B1 and B3. These results suggest that the addition of certain compounds that are broadly used throughout metabolism may accelerate the siphoning of pyruvate into side products.

For both panels, reactions are measured after four hours. Values represent the mean ± standard deviation of three technical replicates.

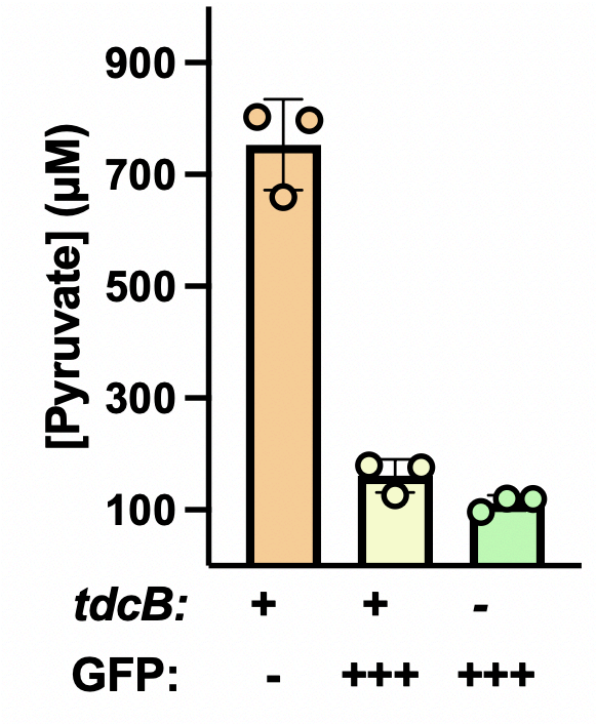


Figure S16. Effect of competing DNA expression on *tdcB* activity. In one-pot cell-free gene expression reactions, there is a tradeoff between the relative amounts of protein that can be expressed from different DNA templates. We demonstrated this tradeoff by measuring pyruvate accumulation via *tdcB* when expressed with or without deGFP as a competing plasmid. Pyruvate concentrations are shown for reactions containing plasmid DNA for *tdcB*, GFP, or both. For reactions with *tdcB*, plasmid DNA is added at 3 nM. GFP plasmid is added at 9 nM. We found that pyruvate production levels from *tdcB* were highly sensitive to the amount of deGFP plasmid added in the system. These results highlight that high expression requirements must be balanced with the total amount of DNA added to the system so that each enzyme can be expressed at relevant levels.

Reactions are measured after four hours. Values represent the mean ± standard deviation of three technical replicates.

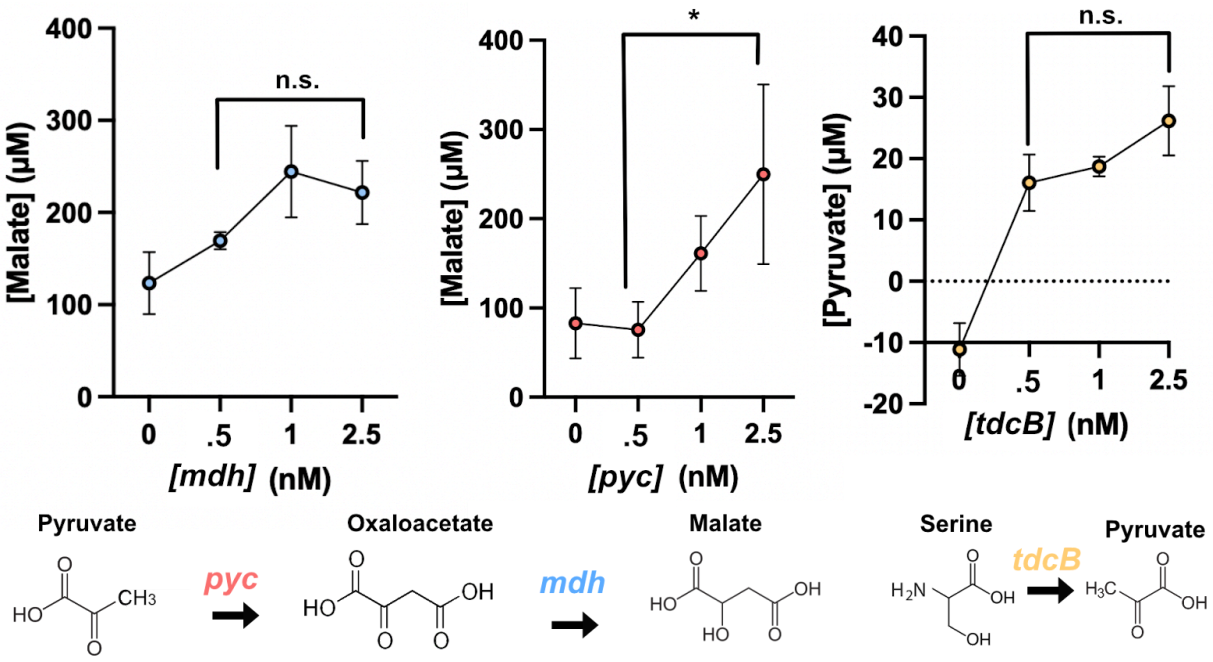


Figure S17. Effect of titrating plasmid DNA on bioconversion efficiency. Product concentrations are shown for reactions titrating either *mdh*, *pyc*, or *tdcB* plasmid DNA to understand which reaction steps were most sensitive to enzyme concentrations. For the *mdh* and *pyc* titrations, *fdh* is held constant at 1 nM. *mdh* was held constant at 1 nM for the *pyc* titration and *pyc* was held constant at 2.5 nM for the *mdh* titration. For both, pyruvate was used as the feedstock at 1 mM with all of the relevant cofactors provided. For the *tdcB* titration, the single step transformation of serine to pyruvate was measured and no other enzymes were expressed. For the single-step *mdh* and *tdcB* reactions, we found that product titers were not significantly different over a five-fold increase in plasmid concentration. These results indicate that low concentrations of enzyme-encoding DNA are sufficient for these reactions, and expression resources can be allocated towards other enzymes. In the *pyc* titration, malate continued to increase

up to 2.5 mM of plasmid DNA, suggesting that *pyc* concentration may be limiting in our system and should be highly expressed.

Reactions are measured after eight hours. Values represent the mean ± standard deviation of at least two technical replicates.

Supplementary Tables

Table S1. Bioproduction & carbon efficiency calculations

A. Effect of dilution & inhibition on malate production (Figure 3B)

	- pathway - dilution - inhibitor	**+ pathway - dilution - inhibitor**	**- pathway + dilution + inhibitor**	**+ pathway + dilution + inhibitor**
Malate (μM)	308	348	50	125
Production from pathway	$\frac{348}{308} = 1.1$		$\frac{125}{50} = 2.5$	
Reduction in background reactions	$\frac{308}{50} = 6.2$			

B. Malate yield on formate (Figure 6B)

Malate concentration (mM)	**Malate concentration (g/L)**	**Formate concentration (mM)**	**Formate concentration (g/L)**	**Yield (g/g)**
.064	.0085	10	.46	.018

C. Malate yield on glycine (Figure 6B)

Malate concentration (mM)	**Malate concentration (g/L)**	**Glycine concentration (mM)**	**Glycine concentration (g/L)**	**Yield (g/g)**
.064	.0085	1	.087	.098

D. Carbon fixation efficiency (Figure 6B)

Malate through pathway (mM)	Malate without pathway (mM)	Difference in malate production (mM)	CO2/malate (mol/mol)	CO2 fixation (mM)	Maximum CO2 fixation (mM)	CO_2 fixation efficiency (%)
.064	-.011	.075	2	.150	2	7.5

E. Enzyme usage

TXTL expression capacity (mg/ml)	TXTL dilution in final reaction vol.	Total expressed enzyme concentration (mg/ml)	Conc. from Luo 2023 (mg/ml)	Conc. from Luo 2022 (mg/ml)	Conc. from Schwander 2016 (mg/ml)
4	200	.02	10.5	10.4	3.0

Table S2. Base case TEA assumptions [5–7]

Formic acid production via CO2 electrolysis	CO_2 Electrolysis Current Density, 140 mA/cm Onstream Factor, 40% Electricity cost, $0.068/kWh CO_2 Single-Pass Conversion, 20% CO_2 Electrolysis Cell Voltage, 3.5 V CO_2 Price, $40/metric ton CO_2 Formic Acid Faradaic Efficiency, 94%
Enzyme to formic acid ratio	1.031 (by mass)
Formic acid conversion	20%
Cell free lysate cost	$120/L
Enzyme yield	44.75 g/L

Table S3. Base case TEA results

Total project investment, TPI (Excluding the electrolysis process)	MM$ 724
Uninstalled equipment (Excluding the electrolysis process)	MM$ 192
Malic acid production	36,510 metric tons/year
Calculated malic acid price using 2020 US$	$9.63/kg

Table S4. DNA sequences used in this study

Gene	Sequence
mdh_Ec	ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGGTATTGGCCAGGCGCTT GCACTACTGTTAAAAACCCAACTGCCTTCAGGTTCAGAACTCTCTCTG TATGATATCGCTCCTGTGACTCCCGGTGTGGCTGTCGATCTGAGCCAT ATCCCTACTGCTGTGAAAATCAAAGGTTTTTCTGGTGAAGATGCGACT CCGGCGCTGGAAGGCGCAGATGTCGTTCTTATCTCTGCAGGCGTACGG CGTAAACCGGGTATGGATCGTTCCGACCTGTTTAACGTTAACGCCGGC ATCGTGAAAAACCTGGTACAGCAAGTTGCGAAAACCTGCCCGAAAGCG TGCATTGGTATTATCACTAACCCGGTTAACACCACAGTTGCAATTGCT GCTGAAGTGCTGAAAAAAGCCGGTGTTTATGACAAAAACAAACTGTTG GGCGTTACCACGCTGGATATCATTCGTTCCAACACCTTTGTTGCGGAA CTGAAAGGCAAACAGCCTGGCGAAGTTGAAGTGCCGGTTATTGGCGGT CACTCTGGTGTTACCATTCTGCCGCTGCTGTCACAGGTTCCTGGCGTT AGTTTTACCGAGCAGGAAGTCGCTGATCTGACCAAACGCATCCAGAAC GCGGGTACTGAAGTGGTTGAAGCGAAGGCCGGTGGCGGGTCTGCAACC CTGTCTATGGGCCAGGCAGCTGCACGTTTTGGTCTGTCTCTGGTTCGT GCACTGCAGGGCGAACAAGGCGTTGTCGAATGTGCCTACGTTGAAGGC GACGGTCAGTACGCCCGTTTCTTCTCTCAACCGCTGCTGCTGGGTAAA AACGGCGTGGAAGAGCGTAAATCTATCGGTACCCTGAGCGCATTTGAA CAGAACGCGCTGGAAGGTATGCTGGATACGCTGAAGAAAGATATCGCC CTGGGGCAAGAGTTCGTTAATAAGTAA
fdh_Sn	ATGGCCAAAATTTTATGCGTATTATACGATGATCCGGTCGATGGCTAC CCAAAAACCTATGCACGCGATGACTTGCCGAAGATCGACCACTATCCG GGGGGTCAGACACTGCCGACGCCAAAGGCGATTGACTTCACTCCTGGT GCATTGTTGGGGTCCGTAAGCGGTGAGTTGGGATTACGCAAATACCTT GAGGCTAACGGTCATACGTTTGTCGTGACGAGTGACAAAGACGGGCCG GACTCAGTGTTTGAACGCGAACTGGTGGACGCGGACGTTGTAATTTCA CAGCCTTTCTGGCCGGCTTATTTAACCCCGGAGCGCATTGCAAAAGCC AAGAATCTTAAGCTGGCTTTGACTGCTGGTATTGGTTCGGACCATGTG GATCTGCAGTCTGCAATTGACCGTGGCATCACAGTGGCAGAGGTCACT TATTGTAATTCCATTAGTGTTGCGGAGCATGTAGTAATGATGATTCTG GGGTTAGTTCGCAACTACATCCCTTCCCATGACTGGGCTCGTAAGGGC GGATGGAACATCGCGGATTGCGTCGAGCATTCGTATGACCTTGAGGGT ATGACAGTGGGCTCAGTCGCCGCTGGCCGTATTGGCCTTGCTGTGCTG CGCCGTTTAGCGCCCTTCGATGTCAAATTGCATTATACGGATCGTCAT CGCTTACCAGAAGCCGTCGAGAAGGAACTGGGTTTAGTTTGGCACGAC ACACGCGAGGACATGTACCCACATTGTGATGTCGTCACGTTAAACGTC CCCCTTCATCCTGAAACCGAGCACATGATTAACGACGAAACTTTAAAG CTTTTTAAACGCGGTGCGTACATTGTAAATACGGCTCGCGGGAAGTTG GCTGACCGTGATGCTATTGTACGTGCCATTGAATCCGGACAACTGGCT GGGTATGCTGGGGATGTGTGGTTCCCTCAACCAGCACCAAAGGACCAT CCATGGCGTACCATGAAATGGGAGGGAATGACACCTCACATCAGTGGG

	ACGTCACTTAGCGCGCAAGCCCGTTACGCTGCTGGGACACGCGAAATT TTAGAATGCTTTTTTGAGGGGCGTCCGATTCGCGATGAATACCTGATC GTGCAAGGCGGCGCGTTGGCCGGGACCGGCGCACATTCGTACTCTAAA GGGAACGCTACGGGTGGAAGTGAGGAGGCGGCTAAGTTTAAAAAAGCT GGATAA
pyc_Re	ATGCCGATTAGTAAAATTTTGGTTGCAAATCGTAGTGAGATTGCTATC CGTGTCTTTCGCGCGGCTAATGAACTGGGGATTAAGACCGTTGCGATT TGGGCCGAAGAGGACAAGCTGGCGTTGCACCGCTTCAAAGCTGACGAA TCGTACCAGGTTGGTCGCGGTCCTCATCTGGCCCGCGATCTGGGCCCG ATCGAGAGTTATCTGAGCATTGACGAAGTTATCCGCGTAGCTAAGCTG TCCGGTGCTGATGCTATCCACCCCGGCTATGGTCTTTTATCTGAGTCT CCAGAATTTGTGGATGCGTGCAACAAAGCCGGGATCATTTTCATTGGT CCGAAAGCCGACACGATGCGCCAGTTGGGAAACAAGGTCGCGGCCCGC AATCTGGCTATTTCCGTAGGCGTCCCGGTAGTGCCTGCTACAGAACCG CTGCCCGACGATATGGCAGAAGTAGCTAAGATGGCCGCGGCAATTGGG TATCCAGTGATGCTGAAGGCATCATGGGGAGGAGGAGGACGCGGTATG CGCGTAATTCGTTCAGAAGCGGATCTTGCGAAGGAAGTTACAGAGGCC AAACGCGAAGCCATGGCGGCGTTTGGTAAAGACGAGGTGTACCTGGAG AAATTGGTAGAACGCGCACGCCACGTCGAGAGTCAAATTTTAGGCGAT ACCCATGGCAACGTCGTGCACTTATTCGAACGCGATTGCTCAGTGCAA CGTCGTAATCAAAAGGTTGTGGAGCGTGCACCCGCTCCATATCTGTCC GAAGCACAACGTCAAGAACTGGCCGCATACAGTCTGAAAATTGCAGGC GCTACTAACTATATCGGCGCCGGAACCGTGGAATACTTAATGGACGCC GACACGGGGAAGTTCTATTTTATCGAGGTTAACCCTCGTATTCAGGTA GAACACACAGTGACCGAGGTGGTGACGGGGATCGACATCGTTAAAGCC CAGATTCATATTTTAGACGGCGCCGCCATTGGGACTCCCCAGAGTGGG GTCCCTAATCAAGAAGACATTCGTCTGAATGGTCACGCTTTGCAATGC CGCGTAACGACAGAGGATCCAGAACACAACTTCATCCCGGACTATGGC CGCATCACAGCCTATCGCTCGGCCAGTGGTTTCGGTATCCGCTTGGAC GGTGGAACGTCCTACAGTGGCGCCATCATCACCCGTTATTACGACCCA TTGCTGGTCAAAGTCACTGCTTGGGCACCCAACCCATTAGAAGCAATT TCTCGTATGGACCGCGCCTTACGTGAGTTCCGTATCCGCGGTGTTGCG ACGAACCTTACGTTCTTGGAAGCTATCATCGGCCACCCTAAGTTTCGT GACAATAGCTATACCACGCGCTTTATCGACACTACCCCCGAATTATTT CAACAGGTGAAGCGCCAGGACCGCGCCACTAAGTTGCTGACTTATCTT GCAGACGTAACAGTTAATGGGCATCCGGAAGCTAAGGACCGCCCCAAG CCTCTTGAAAACGCCGCGCGCCCCGTGGTACCGTATGCCAATGGCAAC GGTGTCAAGGACGGCACTAAGCAATTACTTGACACATTAGGGCCGAAA AAGTTTGGCGAATGGATGCGTAACGAGAAACGTGTCCTTTTGACAGAT ACCACAATGCGCGACGGCCATCAATCCTTGTTGGCCACCCGCATGCGC ACTTATGATATTGCACGTATCGCTGGGACCTATTCGCACGCCCTTCCC AACTTGTTGAGTCTGGAGTGCTGGGGGGGTGCCACTTTCGATGTGTCG ATGCGCTTCCTTACCGAGGACCCTTGGGAGCGCTTGGCTCTGATTCGT GAGGGGGCGCCAAATCTTCTGCTTCAAATGTTACTGCGTGGCGCCAAT GGGGTAGGATATACGAACTACCCTGACAACGTCGTCAAATACTTCGTC

	CGCCAAGCTGCTAAAGGAGGAATCGATCTGTTTCGCGTTTTCGACTGT TTAAATTGGGTGGAGAATATGCGTGTGTCTATGGATGCGATTGCCGAG GAGAATAAGCTTTGCGAAGCTGCCATTTGTTATACGGGGGATATTCTG AATTCCGCCCGTCCTAAGTATGACTTGAAATACTATACTAACCTTGCG GTCGAATTGGAGAAGGCTGGAGCTCACATTATCGCTGTCAAGGACATG GCAGGATTGCTGAAGCCTGCCGCAGCAAAAGTGCTTTTCAAGGCTCTT CGTGAAGCAACGGGTCTGCCAATTCATTTCCATACCCATGATACCAGC GGAATCGCAGCTGCTACTGTGTTAGCGGCGGTGGAAGCAGGAGTAGAT GCCGTAGACGCAGCTATGGACGCCCTGAGTGGGAACACGTCGCAACCA TGCCTGGGCTCAATCGTAGAAGCCCTTTCAGGTTCAGAACGCGATCCA GGATTGGACCCTGCGTGGATTCGTCGCATTTCTTTTTACTGGGAGGCT GTTCGTAACCAATACGCAGCCTTTGAGAGTGACTTGAAGGGCCCTGCT AGCGAAGTCTACTTGCATGAAATGCCCGGAGGTCAATTTACAAATCTG AAGGAGCAAGCCCGCTCTCTGGGCTTAGAAACTCGCTGGCACCAAGTA GCACAGGCGTACGCGGATGCGAACCAGATGTTCGGCGATATCGTTAAG GTAACGCCGTCGAGCAAAGTCGTTGGCGACATGGCGTTGATGATGGTT TCTCAGGATCTGACTGTGGCGGATGTTGTCAGCCCAGATCGCGAGGTA AGCTTCCCTGAGTCAGTCGTAAGTATGCTTAAAGGAGACCTGGGACAG CCGCCTTCGGGGTGGCCGGAAGCTTTACAAAAGAAAGCTCTGAAAGGC GAGAAACCGTACACTGTACGTCCTGGTAGTTTGCTTAAGGAAGCCGAT CTGGACGCCGAGCGTAAGGTTATCGAGAAGAAATTGGAGCGCGAGGTG AGTGATTTTGAATTCGCAAGTTATTTAATGTATCCAAAGGTGTTTACA GATTTTGCCCTTGCATCAGATACTTATGGTCCTGTATCAGTTTTGCCC ACCCCTGCCTATTTTTATGGTTTAGCCGATGGAGAGGAGCTTTTTGCC GATATCGAAAAAGGCAAGACTCTGGTGATTGTTAATCAGGCAGTAAGC GCCACGGACAGCCAAGGCATGGTTACAGTATTCTTTGAGTTAAACGGC CAGCCCCGTCGTATCAAAGTACCTGACCGCGCTCATGGAGCTACGGGA GCCGCCGTTCGCCGCAAAGCAGAACCGGGGAACGCCGCTCATGTTGGG GCCCCGATGCCGGGGGTAATTAGCCGTGTTTTTGTTTCGTCAGGACAG GCAGTAAATGCCGGGGACGTGCTTGTTTCTATCGAGGCGATGAAAATG GAAACCGCAATCCATGCGGAGAAGGATGGTACGATCGCTGAAGTATTG GTTAAAGCTGGCGATCAAATCGATGCTAAGGACCTGCTGGCGGTGTAT GGGGGTTAA
sda_Ec	ATGATCTCGCTGTTCGATATGTTCAAAGTAGGTATCGGACCTAGCAGC AGCCATACCGTCGGGCCGATGAAAGCCGGGAAACAATTTGTCGATGAC TTGGTAGAGAAAGGTTTATTAGACTCCGTAACACGCGTGGCCGTGGAC GTATATGGGTCCCTTTCATTGACAGGAAAGGGGCACCACACCGATATC GCCATCATTATGGGGCTTGCAGGAAACGAACCAGCGACCGTGGATATC GACTCTATTCCGGGGTTTATCCGCGATGTCGAGGAGCGTGAGCGCCTT TTGCTGGCTCAGGGTCGCCACGAGGTTGATTTCCCCCGTGACAATGGT ATGCGTTTCCACAATGGCAACCTTCCGCTTCACGAAAACGGCATGCAG ATTCACGCATACAACGGGGACGAAGTAGTGTACTCTAAGACATATTAC TCCATTGGCGGGGGATTTATTGTTGATGAGGAACATTTCGGACAAGAT GCGGCAAATGAAGTGTCCGTACCTTATCCATTTAAGTCAGCCACGGAA TTGCTTGCCTACTGCAACGAGACTGGATATTCCTTGTCTGGCTTAGCG

	ATGCAGAATGAGCTGGCTTTGCACTCAAAGAAGGAGATTGATGAGTAC TTTGCGCACGTGTGGCAAACCATGCAAGCGTGTATCGATCGCGGAATG AACACAGAAGGGGTCTTACCGGGACCGTTACGTGTCCCACGTCGTGCC TCGGCTCTGCGTCGTATGCTTGTATCATCTGATAAATTATCTAATGAC CCCATGAATGTTATTGACTGGGTGAATATGTTCGCTCTGGCAGTGAAT GAAGAAAATGCGGCCGGGGGACGTGTAGTTACTGCTCCGACAAACGGT GCCTGTGGAATTGTGCCTGCGGTTCTTGCTTACTACGACCATTTCATC GAGTCGGTCTCCCCGGACATCTACACACGTTATTTCATGGCCGCTGGC GCAATTGGTGCCCTTTACAAGATGAACGCGTCGATTTCGGGGGCCGAG GTAGGGTGCCAAGGCGAGGTGGGAGTGGCTTGCTCAATGGCAGCGGCC GGCTTGGCGGAATTGCTGGGAGGAAGTCCAGAGCAGGTCTGCGTCGCG GCCGAAATCGGTATGGAGCACAATCTGGGGTTAACCTGTGATCCGGTC GCGGGGCAGGTCCAAGTGCCTTGTATTGAGCGCAATGCTATTGCGTCT GTGAAAGCTATCAATGCTGCCCGCATGGCGCTGCGTCGTACTTCCGCG CCACGTGTATCTCTGGATAAAGTTATTGAAACCATGTATGAAACCGGG AAGGACATGAACGCCAAGTATCGCGAAACCTCTCGTGGGGGACTTGCA ATCAAAGTGCAGTGCGATTGA
sda_Bsub	ATGAAATACCGCAGTGTATTCGATATTATTGGCCCAGTCATGATTGGT CCTTCAAGTTCTCACACTGCTGGTGCAGCCCGTATTGGGCGTGTGGCC CGCTCGTTGTTCGGTCGCGAGCCTGAGCGCATTATTGTTTCGTTTTAT GGATCATTCGCAGAGACGTATAAAGGCCATGGAACCGATGTCGCCATT ATTGGAGGCTTACTTGATTTTGACACGTTTGATGAGCGCATCAAGACG GCGATCCAGATCGCGGAAGCCAAAGGCATCGACATCGAGTTCCGCGTG GAGGATGCGGTCCCGGTACATCCTAACACAGCCAAGATTACGATTTCA GATGAAAAAGGTGAACTTGAACTTACGGGGATTTCGATCGGCGGGGGA AAAATTGAAATCACTGAATTAAACGGATTTGAATTACGCCTTAGCGGG AATCATCCAGCTATCCTTGTAGTGCACAACGATAAATTCGGGACAATT GCTGGAGTCGCTAACGTGCTTGCGAAGTTTAGTATTAATGTAGGACAC ATGGAAGTCGCCCGTAAAGACATTGGTCAACTGGCCCTTATGACGATC GAAGTAGACCAGAACATTGACGACCATATTCTGGACGAACTTTCAAAA TTGCCCAACATTATTCAGGTTACTAAGATTGCAGATTGA
sda_Msmeg	ATGGATCTGGTTACCTTAGATGACATTTCGGGTGCAGCTGCCCGTATT GCGGCCGATATCGTCCGTACACCTTTATTGGCGGCCGATTGGGGGGAT CCTCGCTGTCCATTATGGCTTAAAGCTGAAACACTGCAGCCCATCGGC GCATTCAAAATTCGCGGTGCTTTTAACGCTCTGGGTCGCCTTGACACG CACACGCGCGCACGTGGCGTAGTGGCATATTCTTCGGGGAATCATGCC CAGGCGGTAGCATATGCAGCGGCTGCCTATGGTGTTCCCGCCCACATT GTCATGCCAGAGGAGACACCAGCCGTCAAAGTCGAGGCAACTCGTCGT CGTGGAGCCCATGTGGTGCTGTGTGGTGCCGGTGAACGTGAACGCACC GCCGCAGAGTTAGTGGAGAAGACGGGAGCCGTACTGATCCCGCCTTTC GACCACCCCGATATTATCGCTGGTCAGGGGACGATTGGAATTGAGATT GCGGAGGATCTTCCCGAATTGGCTACAGTGTTGATCCCGGTAAGCGGC GGGGGTTTGGCGAGTGGGATTGGAACTGCCATCCGCGCGTTGCGTCCT AAGGCAAAAATTTTCGCAGTAGAGCCGGAGTTAGCGGCGGATACCGCA GAGAGCTTAGCGTTGGGAAGTATTGTCGAGTGGCCAGTTGCCAAGCGC

	AACCGCACAATTGCAGACGGTTTACGTTCCACACCATCAGAGTTAACG TTTGCCCACCTGCGCCAAGTCATTGATGACGTCATTACGGTTTCGGAG GATGAAATCCGTTCAGCAGTACGTGAGCTGGCATTGCGTGCGCGTTTG GTTGCAGAACCGTCCGGAGCGGTATCTCTTGCCGGGTATCGCAAGGCG GCTTTACCTGATGGCAGTGCGGTTGCCATCGTGAGCGGGGGGAACATT GAACCGGCACAACTTGCGGCTATCCTGGCTGGCGGATAA
sda_Lpneu	ATGAATATTTCTGTCTTCGATTTGTTCAGTATCGGGATTGGCCCTTCT TCATCCCACACTGTCGGCCCAATGTTGGCAGCAAATGCCTTCCTTCAA CTTCTTGAGCAGAAGAACCTTTTCGATAAAACGCAACGCGTAAAAGTG GAGCTGTATGGCAGTTTGGCGCTGACCGGCAAAGGCCATGGAACCGAC AAGGCGATCTTGAACGGTCTGGAAAACAAGGCCCCTGAGACAGTGGAC CCGGCCTCCATGATCCCACGCATGCACGAGATCTTAGATTCCAATTTG CTGAACCTGGCTGGGAAGAAGGAGATTCCCTTTCACGAAGCAACCGAT TTCCTGTTCTTGCAAAAAGAACTGCTTCCAAAGCACTCAAATGGCATG CGCTTCTCGGCCTTCGATGGGAATGCCAACCTGCTTATTGAACAGGTG TATTATAGTATCGGCGGCGGGTTCATTACAACAGAAGAGGATTTTGAT AAGTCCACTACTGATACTAATCCACCTCCTTATCCTTTTGCGACAGCA ACTGAGTTACTTAAATTATGCAAAAAGCATCATCTTACGATCGCTGAA CTGATGCTTGTTAATGAGAAGACTTGGCGTAGTTCTGCTGAAATCCAT AAAGGTATCCTTGACATTGCTAAGGTAATGGACGATTGTATCAACAAT GGTTGCAAACATGATGGTGTCCTGCCGGGAGGCTTAAATTTAAAACGC CGCGCTCCCGACTTGTACCGCAAACTTATTGAGCAGAAGGGAGTGAAA TCTGTATTTGAGCAGTCTGACATCATGAATCACTTAAACTTGTACGCC ATGGCGGTCAATGAGGAAAATGCAGCTGGCGGCCGTATCGTGACCGCA CCTACAAATGGGGCGGCAGGCATCATCCCGGCGGTACTGAAGTATTGC CAACAGGCCCACGACCGCATGTCGAATGAGGACATTTACACCTATTTC TTGACCGCTGCGGCTATTGGCATTCTGTATAAGAAGGGCGCGTCAATT TCCGGAGCGGAAGTAGGTTGTCAGGGTGAAGTAGGAGTAGCATCCTCT ATGGCAGCCGCGGGACTTACTGCAGTTCTGGGAGGGACCATCGAACAA GTGGAGAACGCGGCTGAGATCGCCATGGAGCACCATCTGGGAATGACG TGCGATCCTGTCTTGGGCTTAGTACAGATCCCGTGTATCGAGCGCAAC GCCATGGGTGCCGTTAAGGCTGTTAATGCGACACGCATGGCGTTAATC GGAGATGGGCAGCACCAAATTTCCCTGGATAAGGTGATCAAGACGATG AAGCAGACCGGCATGGACATGCAGTCAATTTACAAGGAAACGTCTATG GGTGGTCTGGCAGTGAACCTTCCAGAGTGCTAA
TdcB_Ec	atgcatattacatacgatctgccggttgctattgatgacattattgaa gcgaaacaacgactggctgggcgaatttataaaacaggcatgcctcgc tccaactattttagtgaacgttgcaaaggtgaaatattcctgaagttt gaaaatatgcagcgtacgggttcatttaaaattcgtggcgcatttaat aaattaagttcactgaccgatgcggaaaaacgcaaaggcgtggtggcc tgttctgcgggcaaccatgcgcaaggggtttccctctcctgcgcgatg ctgggtatcgacggtaaagtggtgatgccaaaaggtgcgccaaaatcc aaagtagcggcaacgtgcgactactccgcagaagtcgttctgcatggt gataacttcaacgacactatcgctaaagtgagcgaaattgtcgaaatg gaaggccgtatttttatcccaccttacgatgatccgaaagtgattgct

	ggccagggaacgattggtctggaaattatggaagatctctatgatgtc gataacgtgattgtgccaattggtggtggcggtttaattgctggtatt gcggtggcaattaaatctattaacccgaccattcgtgttattggcgta cagtctgaaaacgttcacggcatggcggcttctttccactccggagaa ataaccacgcaccgaactaccggcaccctggcggatggttgtgatgtc tcccgcccgggtaatttaacttacgaaatcgttcgtgaattagtcgat gacatcgtgctggtcagcgaagacgaaatcagaaacagtatgattgcc ttaattcagcgcaataaagtcgtcaccgaaggcgcaggcgctctggca tgtgctgcattattaagcggtaaattagaccaatatattcaaaacaga aaaaccgtcagtattatttccggcggcaatatcgatctttctcgcgtc tctcaaatcaccggtttcgttgacgcttaa
sds_Dd	ATGGGACTGAAGACTACAATCAGCCCTGACTCATCGTTTGATAAAATC ATCAATACTAAGACGAGTCCTCCTCTGCATATCAATAGTCCCATGTTG GAGTCTTTGGCCTTAAGCAAATTATTCAAGGAGGAAAACGCGAAAGTC TGGATGAAGGTAGACGCTCTTCAACCTTCTGGATCGTTTAAGATCCGT GGCGTAGGATTGCTTTGCAACCAGTTGCTGAAGGAAAAGAAATCTAAA AACGAAGAGGCCCACTTCATCTGTTCTTCGGGCGGAAACGCGGGCAAG TCAGTGGCGTACGCAGGACGTAAACTTAACGTTAAGACCACTATCGTG CTTCCAAACACCATTCCGGAGGCGACCATCGAGAAGATTAAAGATGAG GGCGCTAATGTTATTGTGCACGGAACCATCTGGGATGAGGCGAACACC TTTGCGCTTGAGCTTGCTGAGAAAGAGGGCTGCACTGACTGCTACATT CATCCCTTCGATCACCCTCTTCTTTGGGAGGGCCACAGTACCATGATT GATGAGATTTATCAAGATGTCCAGAACGGAGTATGCGAGAAGCCAGAC GTAATTCTGTTCTCAGTGGGTGGCGGCGGCATGATGATCGGCATCTTA CAAGGTTTAGACCGCTACGGATGGAACGATATTCCGATTGTAACCGTG GAGACGGTCGGTAGCCACTCTTTTTGGAAATCCTTCCAAGAAAAACAA CTTACTAAATTAGACGTAAGTGAGGTCACCAGTGTGATCAAAACATTA TCCACGCGTTCAGTTTGCTCAGAAGCATGGGAGATTTCTAAACGCTTT AATATTAAACCCATTCTTGTCACCGACCGTGACGCTGTAGACGCATGC TTAAAGTTCGTTGACGACGAACGCATTCTGGTCGAGCCTAGTTGCGGT GCTACTCTGTCGGTGCTTTACTCGAAGAAATTGTCTACTTTGTTAGAT ATTAACTCTAAAAACATCTTAACCATTGTATGTGGTGGCAACGGGACC AGTATTCTTCAGTTGAACGACTTATTGCAGACGCTGCCGAAGtaa
glyA_Ec	atgttaaagcgtgaaatgaacattgccgattatgatgccgaactgtgg caggctatggagcaggaaaaagtacgtcaggaagagcacatcgaactg atcgcctccgaaaactacaccagcccgcgcgtaatgcaggcgcagggt tctcagctgaccaacaaatatgctgaaggttatccgggcaaacgctac tacggcggttgcgagtatgttgatatcgttgaacaactggcgatcgat cgtgcgaaagaactgttcggcgctgactacgctaacgtccagccgcac tccggctcccaggctaactttgcggtctacaccgcgctgctggaacca ggtgataccgttctgggtatgaacctggcgcatggcggtcacctgact cacggttctccggttaacttctccggtaaactgtacaacatcgttcct tacggtatcgatgctaccggtcatatcgactacgccgatctggaaaaa caagccaaagaacacaagccgaaaatgattatcggtggtttctctgca tattccggcgtggtggactgggcgaaaatgcgtgaaatcgctgacagc

	atcggtgcttacctgttcgttgatatggcgcacgttgcgggcctggtt gctgctggcgtctacccgaacccggttcctcatgctcacgttgttact accaccactcacaaaaccctggcgggtccgcgcggcggcctgatcctg gcgaaaggtggtagcgaagagctgtacaaaaaactgaactctgccgtt ttccctggtggtcagggcggtccgttgatgcacgtaatcgccggtaaa gcggttgctctgaaagaagcgatggagcctgagttcaaaacttaccag cagcaggtcgctaaaaacgctaaagcgatggtagaagtgttcctcgag cgcggctacaaagtggtttccggcggcactgataaccacctgttcctg gttgatctggttgataaaaacctgaccggtaaagaagcagacgccgct ctgggccgtgctaacatcaccgtcaacaaaaacagcgtaccgaacgat ccgaagagcccgtttgtgacctccggtattcgtgtaggtactccggcg attacccgtcgcggctttaaagaagccgaagcgaaagaactggctggc tggatgtgtgacgtgctggacagcatcaatgatgaagccgttatcgag cgcatcaaaggtaaagttctcgacatctgcgcacgttacccggtttac gca
MtdA_Mex	atgtctaagaaactgctctttcagtttgacactgatgcaactccgtct gtatttgacgttgttgttggctatgacggcggtgcagaccatattact ggctatggcaatgttactcccgacaatgttggcgcatatgttgacggc actatttatactcgtggaggcaaagagaaacagtctacagcaatcttt gttggcggcggcgacatggcagcaggcgagcgggtatttgaggcagta aagaagcgtttctttggcccgtttcgcgtttcttgtatgctggattct aatggctctaatactactgcagcagcaggcgttgcactcgttgttaaa gcagcaggcggctctgttaaaggcaagaaagcagttgttctcgcaggt actggtccggttggtatgcgctctgcagctctgttagccggcgagggc gcagaggttgttctgtgtgggcgcaaactcgacaaagcacaggcagca gcagattctgttaataaacgcttcaaagttaatgttactgcagcagag actgcagacgacgcatctcgcgcagaggccgtgaaaggcgcacatttt gtctttactgcaggtgcaattggccttgaactgctgccgcaggcagca tggcagaatgagtcttctattgaaattgtggccgattataatgcacag ccgccgctcggcattggcgggattgatgcaactgacaaaggcaaagaa tatggcggaaaacgcgcatttggtgcgctcggcattggcggcttgaaa ctcaaactgcatcgcgcatgtattgcaaaactgtttgagtcttctgaa ggtgtatttgatgcagaggagatttataaactggcaaaagaaatggca tga
Fch_Mex	atggctggcaatgagactattgaaacattcttggacggcctggcatca tctgctccgactcccggcggcggcggtgcagcagcaatttctggcgca atgggcgcagcacttgtttctatggtttgcaatcttactattggcaag aagaaatatgttgaggttgaggcagacttaaaacaggttctggagaaa tctgaaggcctgcgccgcactctcactggcatgattgcagacgacgtt gaagcctttgacgcagttatgggcgcttatgggctgccgaagaatact gacgaagagaaagcagcacgcgcagcaaagattcaagaggcactcaaa actgcaactgacgttccgctcgcatgttgtcgcgtttgtcgcgaggtt attgatctggcagagattgttgcagagaaaggcaatctcaatgttatt tctgatgcaggcgttgcagtgctctctgcttatgcaggtctgcgctct gctgcacttaatgtctatgtaaatgcaaaaggcctcgacgaccgcgca

	tttgcagaggagcggcttaaagagctggagggcctactggctgaggca ggtgcactcaatgagcgaatttatgagactgttaaatctaaagtgaat tga
Ftl_Mex	atgccgagcgatattgaaattgcacgcgctgctactctgaaaccgatt gcgcaagttgcggagaaactgggtattccggacgaggctcttcataat tatggcaaacatatcgctaaaatcgaccatgactttattgcttctctt gagggtaaaccagagggcaaacttgttctggttactgctatttcgccg actccagctggcgagggcaaaactactactgttggtctgggcgat gctctcaaccgcattggcaaacgtgctgttatgtgtctgcgcgagccc tctctcggcccctgttttggcatgaaaggcggcgctgctggtggcggc aaagctcaggttgttccgatggagcagattaatctgcacttcaccggc gattttcacgctattacttctgctcactctctcgctgctgctctgatt gataaccatatttattgggctaacgaactgaatattgacgttcgccgc attcattggcgccgcgttgttgatatgaacgatcgggctctgcgcgct attaatcagtctctcggcggcgttgctaatggctttccgcgcgaggat gggtttgacattactgttgcttctgaggttatggctgtgttttgcctc gccaagaatctggctgatcttgaggagcggctcggccgcattgttatt gcagaaactcgcgatcgcaaaccggttactctggctgatgttaaagct actggcgctatgactgttctgctcaaggatgctcttcagccgaatctc gtgcagactctggagggcaacccggctctgattcacggcggcccgttt gctaacattgctcatggctgtaactcggttattgctactcgcactggc ctgcggctcgctgactatactgttactgaggctggctttggcgctgat ctcggcgctgagaaattcattgatattaaatgtcgccagactggcctc aagccctctgctgttgttattgttgctacgattcgcgctctcaaaatg catggcggcgttaacaagaaagatctccaggctgagaatctggatgcg ctggagaaaggttttgcaaatcttgagcgccatgttcacaatgttcgc tcttttggcctgccggttgttgttggtgttaaccacttctttcaggat actgatgctgagcatgttcggttgaaagaactgtgccgcgatcggctt caggttgaggctattacttgtaagcattgggctgagggcggcgcaggc gcagaagcactggcacaggcagttgttaaactggctgaaggcgagcag aaaccgctgacttttgcatatgagaccgaaactaagattactgacaag attaaggcaattgctactaaactgtatggtgctgctgatattcagatt gagtctaaagccgccactaagctcgctggcttcgagaaagatggctat ggtaagctgccggtctgtatggccaagactcaatattcattttctact gatccgactcttatgggcgctccctctggtcatctggtttctgtgcgc gatgttcgcctctctgctggcgctggcttcgttgttgttatttgtggt gagattatgaccatgccgggtctgccgaaggttccagcagcagatact attcgcctcgatgctaacggtcagattgatgggctgttctag
ppk12	ATGATTAACATCTACAAGATCGACAAACTTAACAACTTTAACCTGAAT AACCACAAGACCGACGACTACTCTCTTTGTAAGGATAAGGATACTGCT CTGGAGCTTACTCAGAAGAACATCCAGAAAATTTATGATTATCAGCAA AAATTGTACGCCGAAAAAAAAGAAGGATTGATCATTGCATTCCAAGCA ATGGATGCCGCCGGTAAGGATGGAACTATCCGTGAGGTGTTAAAAGCC CTGGCTCCCCAAGGAGTGCACGAGAAGCCCTTCAAGTCCCCATCGAGT ACCGAGCTTGCCCACGATTACTTGTGGCGCGTACACAATGCCGTACCT

	GAAAAAGGTGAGATCACAATTTTTAACCGTAGTCATTACGAAGACGTA TTGATTGGAAAGGTTAAGGAGCTTTATAAATTCCAAAACAAAGCGGAT CGCATTGACGAGAACACGGTAGTTGACAATCGTTATGAGGATATTCGT AACTTTGAGAAATACTTATATAATAATTCAGTCCGCATTATCAAAATT TTTTTGAACGTGTCTAAAAAAGAACAGGCCGAACGTTTCTTATCGCGT ATTGAGGAACCGGAAAAGAACTGGAAATTTTCCGACAGTGATTTCGAG GAGCGTGTATATTGGGACAAATATCAACAGGCTTTCGAAGATGCAATT AACGCAACGTCTACGAAGGATTGTCCTTGGTACGTCGTGCCTGCTGAT CGCAAATGGTATATGCGCTATGTTGTTTCCGAAATCGTGGTCAAGACC TTAGAAGAGATGAACCCTAAGTACCCTACCGTCACCAAAGAGACCCTG GAGCGTTTTGAAGGTTATCGTACAAAGCTTCTGGAAGAGTATAACTAC GACTTGGATACTATCCGTCCCATCGAGAAA
fdh_At	atgaaacaggcgagcagtggagacagtaaaaagatcgtgggagtgttc tataaggccaacgagtacgccaccaagaacccgaacttcctgggctgt gtagagaatgcgctgggcattcgggactggctggaatctcaaggtcac cagtacatcgtgacggacgataaggagggaccagattgcgagctggag aagcacattccagatttacacgtgttaatttcaacgccgttccatcct gcgtatgtcacggccgagcgcattaagaaggcaaaaaacttaaagctg ctgctgacggcgggcatcggaagtgaccacattgacctgcaggcggcg gcagcggcaggtttaacggtggccgaagtgactggaagcaatgtggta agtgtcgcagaggatgaactgatgcggatactgatcttaatgcgcaat ttcgtccctggttataaccaagttgtcaaaggcgagtggaatgtggcc ggcatagcataccgcgcgtatgacctggagggtaaaacgatcggcact gtcggagctggcagaattggcaagttgttattacaaagactgaagccc tttggatgtaacctgctgtaccacgatcgccttcagatggcaccagaa ctggaaaaggagactggagccaagtttgtggaagacttgaatgaaatg cttccgaagtgtgacgttatcgtgatcaatatgccattaacggaaaag acgcgtggaatgttcaataaggagttaattggaaaacttaagaagggt gtgttaattgtgaacaatgcccgtggtgcgattatggagcgccaagcc gttgtggacgctgtcgagtcaggtcatatcggaggttactcaggcgat gtctgggacccgcaacctgctccgaaggatcacccgtggcgatatatg ccaaatcaagcgatgactccgcacacgagtggaacgacaattgacgcg caattaagatatgctgctgggacaaaggatatgctagaacgctatttt aagggcgaggactttcctactgaaaactacatcgttaaggatggcgag ttagcgccacaataccgatga
fdh_Cm	ATGAAGATTGTACTTGTCTTGTATGACGCAGGTAAGCACGCAGCGGAC GAGGAGAAGTTGTATGGTTGCACCGAAAACAAGTTAGGGATTGCAAAC TGGCTGAAGGATCAGGGACATGAACTGATTACGACTAGCGACAAAGAA GGTGAAACCAGCGAGCTTGATAAACATATCCCCGACGCGGATATTATC ATTACGACGCCATTCCATCCTGCCTATATCACGAAAGAGCGTCTTGAT AAGGCGAAGAATCTTAAGAGTGTCGTAGTCGCAGGCGTCGGTTCGGAC CACATCGACCTTGATTACATCAACCAAACAGGGAAAAAGATTTCGGTA CTTGAGGTGACAGGTTCAAATGTTGTCTCTGTAGCTGAGCATGTCGTA ATGACGATGTTAGTGCTGGTGCGCAACTTCGTGCCCGCTCATGAACAA ATTATCAACCACGATTGGGAGGTTGCAGCGATTGCCAAGGACGCCTAT

	GATATTGAGGGTAAGACCATCGCTACGATCGGTGCCGGTCGTATCGGC TACCGCGTTCTTGAGCGTTTACTGCCATTCAATCCCAAGGAGTTATTA TATTACTCTCTTCAAGCTCTTCCTAAAGAAGCAGAGGAGAAAGTGGGG GCACGTCGTGTCGAAAATATTGAAGAGTTGGTCGCTCAGGCTGACATC GTCACCGTGAACGCGCCATTACACGCCGGCACAAAGGGTCTTATTAAC AAGGAGCTGCTGAGCAAGTTTAAGAAAGGTGCGTGGCTGGTAAATACG GCACGTGGGGCTATTTGTGTCGCGGAAGACGTTGCCGCAGCTCTGGAG TCAGGTCAGTTGCGTGGATATGGCGGGGATGTATGGTTCCCTCAACCA GCGCCTAAGGATCACCCTTGGCGTGATATGCGTAACAAATACGGCGCA GGTAACGCAATGACTCCGCACTACTCAGGGACAACACTTGACGCACAG ACACGTTATGCGGAGGGAACCAAAAACATTTTGGAAAGCTTTTTCACG GGTAAATTTGATTACCGCCCCCAAGACATCATCCTTTTGAATGGCGAA TACGTTACTAAGGCATACGGAAAGCATGACAAGAAATGA
deGFP	AGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACG TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCA CCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGT CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG GCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACG TCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCA AGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC

Supplementary Methods

Methods S1. Comparative sequence analysis of enzyme isoforms

BRENDA Enzyme Database and UniProt were searched for PLP-dependent L-serine ammonia-lyase (EC 4.3.1.17) and threonine ammonia-lyase (EC 4.3.1.19) isoforms from single-celled organisms. Isoforms were down-selected based on previous examples of functional enzyme expression in *E. coli*. Multiple sequence analysis (MSA) to determine the presence of AMP binding sites was conducted using Clustal Omega. Protein sequences were gathered from UniProt. Analysis of homology across organisms was performed manually.

Methods S2. Technoeconomic analysis

The analysis was carried out assuming electrocatalytically generated formate from CO_2 as the sole feedstock and with *in situ* regeneration of cofactors. In our TEA model, we also use the “nth” plant assumption, in which the cost assumptions reflect a future time when the technology is well established, and several plants have already been built and are operating. This simplification allows us to avoid accounting for start-up costs and focus on the cost of operation. The specific “nth” cost and financial assumptions are listed elsewhere [8]. Additional technical assumptions to perform the TEA are listed in **Table S2**. A process block flow diagram of the CFE to produce malate is shown in **Figure 7A**. The process economic results are presented in terms of minimum selling price (MSP) on a dollar per kg basis, which is determined by using a discounted cash flow analysis method. All costs in this article are presented on a 2020 constant US dollar basis.

Malate has a market price of $1.8/kg in 2016 [9]. Its U.S. consumption is 15,100 metric tons/year and 16,900 metric tons/year in 2016 and 2021, respectively, with an annual growth rate of 2-3% [9]. Based on the average ethanol plant scale in the U.S., electrochemical conversion efficiency, and product titers, malate production from the cell free process is estimated at 36,510 metric tons/year. Therefore, the effluent CO_2 stream from a single ethanol plant could already fulfill the U.S. malate market. TEA suggests that the MSP of malic acid from this technology is about $9.63/kg **(Table S3)**.

Methods S3. Preparation of cell-free bioproduction reagents and reactions

For reactions including *fdh*, formate was prepared as a potassium formate solution to buffer the reaction. For reactions including *sda*, iron was prepared as ferrous (Fe^{2+}) ammonium sulfate and ferric (Fe^{3+}) ammonium sulfate and mixed at equimolar concentrations to account for both possible oxidation states of iron-sulfur cluster proteins. Dithiothreitol (DTT) and ascorbic acid were prepared fresh as 25 mM aqueous stocks. For reactions including *pyc,* hydroxycitrate was sourced from MedChemExpress (HY-*W009156*) as hydroxycitric acid tripotassium salt. Magnesium was prepared as magnesium acetate.

For biosynthesis reactions, master mixes were prepared by adding concentrated stocks of all substrates and cofactors required for each enzyme either by hand or using an acoustic liquid handler robot (Echo Labcyte 525). Reaction conditions for each experiment are illustrated in the figures. Master mixes were split into 22.5 µL aliquots and 2.5 µL of CFE-expressed enzymes were added to each reaction. For negative control or “no pathway” reactions, 5 nM of *deGFP* was expressed in CFE and added to

reactions to maintain similar levels of endogenous proteins and metabolites as the reactions containing pathway enzymes. For time course reactions, larger cell-free biosynthesis reactions were split into individual 25 uL technical replicates. Technical replicates were sacrificed for each time point.

Methods S4. CO_2 quantification with GC/MS

Samples were analyzed via Agilent 5973 GC/MS using a HP-5MS 5% Phenyl Methyl Siloxane column. An isocratic method was used, maintaining the oven temperature at 30°C over 4 minutes. The front inlet temperature was set to 250°C at a pressure of 9.5 psi. The column flow was maintained at .8 mL/min using acetone as the solvent. The split-splitless inlet flow rate (H_2) was maintained at 11.2mL/min.

For reactions measured via GC/MS, reactions were carried out in 100 µL volume in GC/MS vials. 0.5 mL of the reaction headspace was injected into the GC/MS using a GC syringe needle.

Supplemental References

(1) Cheema-Dhadli, S.; Halperin, M. L.; Leznoff, C.C. Inhibition of Enzymes Which Interact with Citrate by (-)Hydroxycitrate and 1,2,3,-Tricarboxybenzene. *Eur J. Biochem.* **1973,** 38(1), 98-102. DOI: 10.1111/j.1432-1033.1973.tb03038.x

(2) Jena, B.S.; Jayaprakasha, G.K.; Singh, R.P.; Sakariah, K.K. Chemistry and Biochemistry of (−)-Hydroxycitric Acid from Garcinia*. Journal of Agricultural and Food Chemistry*. **2002,** 50(1), 10-22. https://doi.org/10.1021/jf010753k

(3) Schlachter, C. R.; Klapper, V.; Radford, T.; Chruszcz, M. Comparative Studies of Aspergillus Fumigatus 2-Methylcitrate Synthase and Human Citrate Synthase. *Biol. Chem.* **2019**, *400* (12), 1567–1581. https://doi.org/10.1515/hsz-2019-0106.

(4) Smith, C. M.; Williamson, J. R. Inhibition of Citrate Synthase by Succinyl-CoA and Other Metabolites. *FEBS Lett.* **1971**, *18* (1), 35–38. https://doi.org/10.1016/0014-5793(71)80400-3.

(5) Huang, Z.; Grim, R. G.; Schaidle, J. A.; Tao, L. The Economic Outlook for Converting CO2 and Electrons to Molecules. *Energy Environ. Sci.* **2021**, *14* (7), 3664–3678. https://doi.org/10.1039/D0EE03525D.

(6) Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, P.; Lukas, J.; Olthof, B.; Worley, M.; Sexton, D.; Dudgeon, D. *Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover*; NREL/TP-5100-47764, 1013269; 2011; p NREL/TP-5100-47764, 1013269. https://doi.org/10.2172/1013269.

(7) Rasor, B. J.; Vögeli, B.; Landwehr, G. M.; Bogart, J. W.; Karim, A. S.; Jewett, M. C. Toward Sustainable, Cell-Free Biomanufacturing. *Curr. Opin. Biotechnol.* **2021**, *69*, 136–144. https://doi.org/10.1016/j.copbio.2020.12.012.

(8) Dutta, A.; Mukarakate, C.; Iisa, K.; Wang, H.; Talmadge, M.; Santosa, D.; Harris, K.; Baddour, F.; Hartley, D.; Cai, H.; Ou, L.; Schaidle, J.; Griffin, M. *Ex Situ Catalytic Fast Pyrolysis of Lignocellulosic Biomass to Hydrocarbon Fuels: 2020 State of Technology*; NREL/TP-5100-80291, 1805204, MainId:42494; 2021; p NREL/TP-5100-80291, 1805204, MainId:42494. https://doi.org/10.2172/1805204.

(9) *Malic Acid, DL - Chemical Economics Handbook (CEH) | S&P Global*. https://www.spglobal.com/commodityinsights/en/ci/products/malic-acid-chemical-economics-handbook.html (accessed 2024-04-05).

Chapter 4: Systems-Level Modeling for CRISPR-Based Metabolic Engineering

Ryan A. L. Cardiff[1,2,3] *, James M. Carothers[1,3], Jesse G. Zalatan[1,2], Herbert M. Sauro[1, 4] *

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

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

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

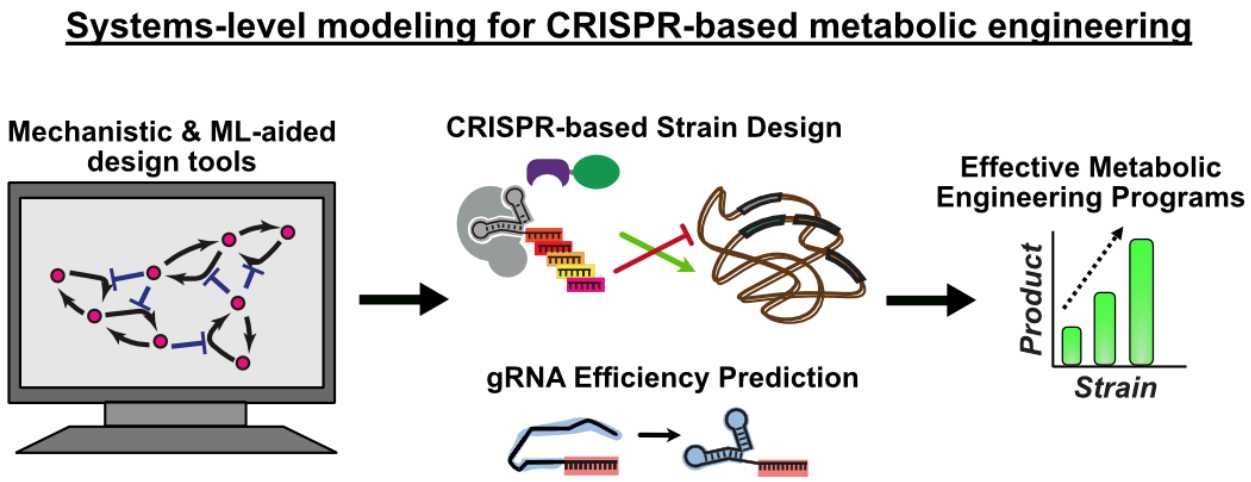
4: Department of Bioengineering
University of Washington
Seattle, WA 98195
United States

*: Corresponding authors
cardiffr@uw.edu
707-338-3971

hsauro@uw.edu
206-880-8093

Published as a review article in American Chemical Society (ACS) Synthetic Biology on August 9th, 2024. DOI: https://doi.org/10.1021/acssynbio.4c00053

Graphical Abstract



Abstract

The CRISPR-Cas system has enabled the development of sophisticated, multi-gene metabolic engineering programs through the use of guide RNA-directed activation or repression of target genes. To optimize biosynthetic pathways in microbial systems, we need improved models to inform design and implementation of transcriptional programs. Recent progress has resulted in new modeling approaches for identifying gene targets and predicting the efficacy of guide RNA targeting. Genome-scale and flux balance models have successfully been applied to identify targets for improving biosynthetic production yields using combinatorial CRISPR-interference (CRISPRi) programs. The advent of new approaches for tunable and dynamic CRISPR activation (CRISPRa) promises to further advance these engineering capabilities. Once appropriate targets are identified, guide RNA prediction models can lead to increased efficacy in gene targeting. Developing improved models and incorporating approaches from machine learning may be able to overcome current limitations and greatly expand the capabilities of CRISPR-Cas9 tools for metabolic engineering.

Introduction

The CRISPR-Cas9 system has enabled sophisticated, multi-gene metabolic engineering programs in a variety of organisms [1,2,3]. Catalytically inactive Cas9 (dCas9) has proven to be a powerful tool for gene regulation due to its ability for programmable RNA-guided DNA binding, allowing it to be used for CRISPR-interference (CRISPRi) or CRISPR-activation (CRISPRa) [4,5]. CRISPRi blocks transcription by recruiting dCas9 or a transcriptional repressor to the promoter or open reading frame (ORF) of a gene of interest, while CRISPRa increases transcription through recruitment of a transcriptional activator to the promoter region. dCas9 can be recruited to a specific DNA sequence through the use of modified guide RNAs (gRNAs) known as single guide RNAs (sgRNAs) or scaffold RNAs (scRNAs) that recognize targets based on Watson-Crick base pairing [6]. scRNAs mediate CRISPRa by acting as sgRNAs with RNA hairpins appended at the 3' end to recruit activator domains to the CRISPR complex [6,7]. Orthogonal gRNAs can be used for multi-gene metabolic engineering platforms capable of targeting an arbitrary set of endogenous and heterologous targets [8].

Metabolic engineering is currently limited by our incomplete understanding of the native gene regulatory networks in the cell and our inability to predictably regulate target gene expression [9]. For effective metabolic engineering programs, models are needed to both identify gene targets and implement effective transcriptional programs. Recently, systems-level modeling has sought to overcome these limitations. The applications of biological modeling for CRISPR-based metabolic engineering typically fall into one of two main categories: (1) using modeling tools to predict favorable gene targets [10–12], and (2) gRNA design to improve on-target efficiency and reduce off-target effects [13–15]. Constraint-based genome-scale modeling is a commonly used approach for suggesting CRISPRi gene knockdown targets predicted to yield improved titers [12,16]. Several recent theoretical and experimental works have shown CRISPRi targeting to be effective for redirecting metabolic flux [1,12,16,17]. Additionally, the development of microbial CRISPRa has enabled activation of targets in the genome that can direct flux towards a pathway of interest [5,18]. Algorithms that can simultaneously recommend targets for up- and down-

regulation will greatly advance our control of metabolic flux. Once targets have been identified, effective gRNAs are needed to implement the desired perturbation. gRNA design tools typically make predictions from biophysical models that determine the free energies of RNA folding and RNA:DNA hybridization, genome-wide screens that rely on empirical data-driven approaches, or some combination of the two methods. In both cases, models have been used to learn rules for sgRNA sequence design to enable highly efficient genome targeting for both metabolic engineering and gene editing applications. In this review, we will discuss the current state and outstanding limitations of CRISPR-based metabolic engineering, as well as how systems level modeling tools will help advance the field.

Genome-scale modeling for CRISPR-based metabolic engineering

There is a prevailing question in metabolic engineering as to which genes to perturb and how much to up- or down-regulate their expression to achieve a desired phenotype. Engineers aim to divert flux away from competing pathways and maximize flux through their desired pathway. However, strongly repressing genes involved in essential pathways, such as central carbon metabolism or amino acid biosynthesis, can have deleterious effects on the organism including growth defects, undesired mutations [19,20]. Similarly, overexpression of some pathways can lead to metabolic burden or toxicity [21,22]. Therefore, engineers have turned to model-guided metabolic engineering platforms to provide insight into which genes to target and how strongly to activate or repress them to maximize product yield and accelerate the design-build-test-learn (DBTL) cycle.

Genome-scale metabolic models (GEMs) are mathematical representations of metabolic networks based on experimental data and biochemical assumptions such as steady-state. Flux balance analysis (FBA) is one method to predict the movement of metabolites through a GEM. Algorithms such as FBA make use of a user-defined objective function and other constraints defined by the GEM to provide insight into phenotypic effects of genetic perturbations [23]. FBA is based on constructing the objective function, assuming steady-state, and the use of linear programming to solve

for a flux distribution. The flux distribution output fulfills the constraints derived by the GEM and optimizes the user-defined objective function [24,25]. The solver will maximize or minimize the objective function depending on the user definition. The choice of the objective function has a large effect on the output flux distribution and therefore the closeness of the distribution to measured data [26,27]. In industrial biotechnology, algorithms derived from FBA have long been used to identify genetic interventions that maximize product yield [28–30].

Traditionally, genetic interventions are implemented through strain engineering to create knockouts or heterologously overexpress pathway genes. However, there are recent examples of more complex tools that can recommend optimal enzyme levels [31,32] or targets for dynamic regulation [33,34]. CRISPR systems may be particularly well-suited for implementing recommendations from these tools due to their tunable and dynamic capabilities. Additionally, there is growing attention in the use of the CRISPRa/i system to regulate gene targets identified through flux balance analysis [17,35–37]. Still, there is a gap in our ability to reliably predict the degree to which a CRISPRa/i perturbation will affect gene expression, and the degree to which changes in expression will affect enzyme activity for a given gene [38,39].

Although there are many methods able to recommend multiplexed interventions from GEMs, slow strain engineering cycles often limits the ability to broadly explore these design spaces [28,40]. The use of CRISPR tools to implement perturbations may address this limitation. The demonstrated ability to simultaneously activate and repress multiple genes using CRISPRa/i in microorganisms is a fairly recent development. Lian and colleagues were the first to demonstrate a tri-functional CRISPR system; taking advantage of activation, interference, and deletion in yeast [2]. Their system achieved significantly improved titers of two metabolic engineering targets through combinatorial perturbations of up to 36 genes. Similar tri-functional systems have yet to be demonstrated in other model organisms for biosynthesis. However, other experimental work has shown that combining CRISPRa/i control can help improve titers over a single method of regulation [2,8,41]. There are several notable examples of strain design algorithms that use genome-scale models to recommend perturbations to multiple genes through multiplexed activation, repression, and knockouts [28–30,40,42–44]. These tools

are particularly well suited for use with multi-functional CRISPR-based metabolic engineering programs as perturbations can be quickly implemented in combination. However, none of these strain design tools have been validated with CRISPR-based perturbations. Instead, most strain engineering has been achieved through gene knockouts and heterologous overexpression. The use of multi-functional CRISPR tools could greatly accelerate the prototyping of strains recommended by these algorithms, providing insight into the effects of finely-tuned and combined perturbations.

To take greater advantage of the tunability accessible with CRISPR tools, current strain design tools may need to be adapted. Updating constraint-based genome-scale models that focus solely on deletion or knockdowns to allow activation through a given pathway should be straightforward [28,42]. CRISPR-Cas tools have made it faster and easier to implement multi-functional metabolic engineering programs. Therefore, engineers should be encouraged to use computational tools that match their experimental capabilities.

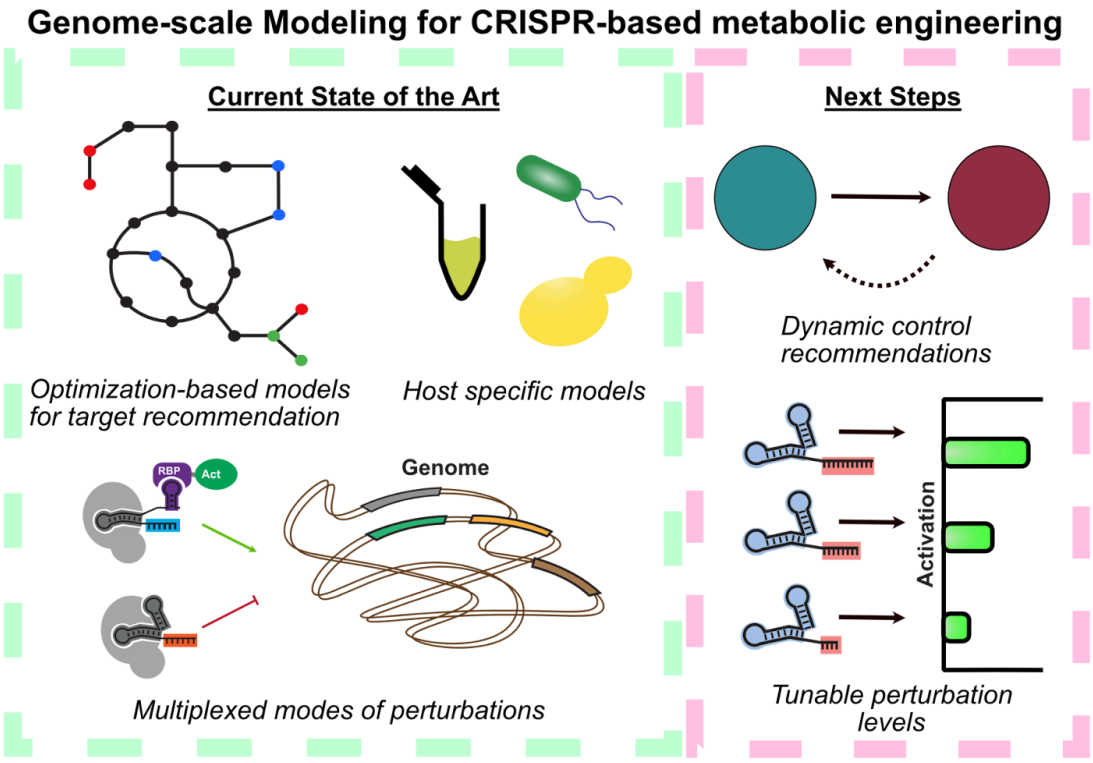


Figure 1. Current tools and opportunities for improvement of genome-scale models for metabolic engineering.

In recent years, the development of CRISPR tools has enabled programmable regulation at genomic targets in a wide array of novel microbes [45]. However, identifying relevant targets for metabolic engineering can be challenging for microbes with industrially useful but understudied metabolisms. To address this issue, computational tools can be used to generate novel metabolic models. Model reconstruction can be achieved using species-specific genomic data and "top-down" or "bottom-up" reconstruction methods, or through a combination of these approaches [46–49]. Typically, "top-down" approaches use a universal template model to assist with gap-filling, while a "bottom-up" approach may rely more on sequence homology and experimental data to derive metabolic network topologies [46–49]. Crucially, these tools minimize the need for manual construction and are broadly applicable to diverse bacterial species. Recent work has demonstrated that the development of genome-scale models can be transformative for realizing the potential for metabolic engineering for non-model organisms with unique, desirable metabolic properties [50]. For example, a previously published GEM and thermodynamic model of the Wood-Ljungdahl pathway in *Clostridium ljungdahlii* provided novel insights on autotrophic metabolism and potential pathways to high-value products generated from syngas. This model then informed the first example of CRISPRi-based metabolic engineering in *Clostridium ljungdahlii* several years later [51–53].

To circumvent the need for a new model of each unique organism's metabolism, several groups have developed machine learning-based (ML) strain design tools. These tools are able to recommend targets to improve bioproduction and accelerate the DBTL cycle without a full mechanistic understanding of the biological system [31,54,55]. Both the Automated Recommendation Tool (ART) and EVOLVE use Bayesian ensemble approaches to make strain recommendations with minimal training instances [31,54]. Bayesian models often outperform other ML algorithms on datasets that are small or noisy, making them a common choice for biological datasets [55]. From initial experimentation, the user provides input parameters such as transcriptome data or

promoter strengths, as well as the observed output, e.g. product yield. The model then returns predicted output levels for different input profiles. In a direct comparison of ART and EVOLVE, the tools showed similar overall performances for recommendations of promoter strengths for a five-gene tryptophan overproduction strain [54].

While a meaningful step towards realizing an organism-agnostic model for metabolic engineering, machine learning approaches typically require a great deal of input data to generate a predictive model. Moreover, current ML based methods are generally unable to provide a rationale for their predictions. The difficulty in automated and high-throughput data collection and analysis is a major challenge to the practical use of machine learning methods for synthetic biology. An integrative tool that combines machine learning techniques and genome-scale modeling may offer the predictive power of machine learning algorithms with minimal training data [54]. This could theoretically be achieved by feeding fluxes determined from the literature and GEM simulations as inputs to existing machine learning tools. In addition, by combining ML with mechanistic modeling, better insights can be achieved to understand why a given set of perturbations fails to achieve a desired phenotype.

Cell-free systems (CFS) have recently gained popularity as a means for rapidly prototyping metabolic engineering programs and improving experimental throughput [56,57]. The use of CFS as a metabolic engineering platform offers many advantages over traditional *in vivo* systems, such as reduced carbon diversion towards endogenous metabolism and potential for exotic and toxic chemistries [58–60]. However, models of CFS tend to focus on protein expression and resource competition and lack consideration for the presence of central metabolism and endogenous enzymes in the CFS [61–66]. Recently, Martin et al. constructed a kinetic model of butanol production in a non-steady state CFS [67]. Their model incorporated over 200 reactions, metabolites, and inhibitor-enzyme pairs to accurately capture interactions between heterologous and endogenous metabolic networks and offer unintuitive predictions to improve butanol production. Their model was effective in identifying aldehyde dehydrogenase as the step with the most flux control using minimal training data, and lending insights into the effect of pairwise perturbations on butanol production. While these insights may have been possible without a large ensemble modeling approach, a significant advantage of

the ensemble is that the scope of the model may lend itself well for being easily applied to other cell-free metabolic engineering efforts with minimal training data. Still, the process of generating individual knockout strains and cell-free lysates can be time-intensive. To address this limitation, other work from this group has explored the use of multiplexed CRISPR tools in a *S. cerevisiae* CFS to improve production of 2,3-butanediol [68]. The use of CRISPR tools allowed the authors to rapidly assess combinatorial perturbations in the CFS without the need for extensive engineering of the underlying lysate strain. Further work to integrate these two efforts by screening model-informed combinatorial CRISPR perturbations may greatly accelerate DBTL cycles for cell-free metabolic engineering.

Metabolic engineers have long been interested in leveraging dynamic control to separate growth and production phases, maintain healthy cell growth, and ultimately improve product yields [69,70]. Dynamic regulation can describe simple systems that rely on addition of inducers to initiate gene expression, and autonomous systems that regulate flux based on sensing the internal state of the cell [69,71]. As chemical inducers are often too expensive for large-scale fermentation, there is growing interest in the rational design of self-regulating gene expression programs [19]. For effective dynamic regulation, metabolic engineers may require simultaneous overexpression and repression at several gene targets [70,72]. Engineers must also avoid redundant genetic sequences, such as the repeated use of inducible promoters, as they may be prone to recombination [73]. The programmability and orthogonality of the CRISPR system addresses these challenges, and has enabled significant improvements in the ability of synthetic biologists to dynamically control genes in response to environmental stimuli [74–76].

Despite the advantages of CRISPR tools, it remains difficult to implement regulation at the correct time and intensity during fermentation. For example, dynamic control in response to the presence of a toxic intermediate must be rapid and sensitive to prevent accumulation while maintaining flux through the desired pathway [69,71,77]. To date, most algorithms recommending perturbations have been limited to static control, leaving them unable to provide valuable insight as to when and how to implement dynamic control [78]. A model for dynamic control would be difficult to construct in practice

as it may require further parameterization to understand how implementing perturbations at different stages of growth affects metabolism [79]. The integration of mathematical optimization with kinetic models of gene expression will improve our ability to predictably implement dynamic control for metabolic engineering programs using CRISPR-Cas systems [36,80].

Model-informed guide RNA design

gRNA design tools typically aim to maximize either gRNA specificity or efficiency. gRNA specificity is important to ensure that a given gRNA does not bind off-target in the genome, causing undesired growth defects or changes to metabolism. Tools for gRNA efficiency also ensure that the designed gRNA will fold properly, based on predicted RNA structure and sequence context. There are several recent reviews discussing gRNA design in great detail with focuses on machine learning and web-based tools [81,82]. In this review, we focus on models with the greatest impact for constructing multi-gene biosynthetic programs. Therefore, special attention is paid to models that give consideration to systems-level design, CRISPRa/i efficiency, gRNA competition, and portability to microorganisms.

There have been many models developed to predict gRNA performance, quantified by off- and on-target effects, using combinations of machine learning [38,82–84], kinetic [85,86], and thermodynamic models [14,85,87]. While early gRNA prediction models have been around for nearly a decade, it was only recently that large-scale models were updated to include functionality for CRISPRa/i targeting in addition to genome editing [88]. For CRISPRa, predicting high-performing guides may be challenging as targetable positions are limited to PAMs in the promoter region. CRISPRi target positions are more flexible, but knockdowns closer to the TSS have been shown to be more effective [89]. Therefore, there is significant interest in computational models that can predict ideal target sites and guide RNA sequences for a gene of interest.

The target sequence has been shown to have large effects on the degree of repression or activation achievable with CRISPRa/i control, in part due to both primary sequence composition and secondary structure of the gRNA [9,90,91]. CRISPRa/i control is

highly sensitive to target sequence composition in both prokaryotic and eukaryotic systems. However, models that rely solely on gRNA sequence information are vastly outperformed by those that incorporate sequence context and structural information as well, highlighting the complexity of effective gRNA design [83,92]. gRNA efficiency is highly influenced by sequence context, both with respect to the sequence surrounding the PAM site [87] and the local structure at the target site [93,94]. Farasat & Salis (2016) demonstrated the importance of incorporating sequence context and structural information into gRNA design by developing an integrative biophysical model of CRISPR-Cas9 activity. Their model incorporates statistical thermodynamics and kinetics with next-gen sequencing data on on- and off-target specificity to build a system-wide understanding of Cas9 on- and off-target effects in the genome [86]. Their mechanistic model includes kinetic representations of gRNA expression and folding, as well as thermodynamic representations of gRNA-Cas9 binding and Cas9-DNA binding. With this model, they unveil novel insights about differences in Cas9 off-target activity across organisms that previous studies focusing on gRNA sequence failed to identify. According to their model, the PAM-proximal seed region is the most important determinant for dCas9 binding affinity. Other kinetic models have found the free-energy change from the formation of the R-loop needed for DNA displacement is crucial for dictating Cas9 efficiency [95]. These results may be important considerations for the design of effective gRNAs, especially for targets with few available PAM sites.

A great deal about the tunability of CRISPR systems has been learned through the development and implementation of gRNA models. The use of truncated and mismatched gRNAs to tune the level of response in CRISPRa/i systems [96], a rule now leveraged often in metabolic engineering applications [97], was described in the CRISPRscan model [98,99]. The use of mismatched sgRNAs for systematic titration of gene expression has further been modeled by Jost et al. [100]. When using truncations or mismatches to tune CRISPR activity, the potential for off-target effects increases as gRNA specificity decreases. These models provide important rules for the design of effective guides.

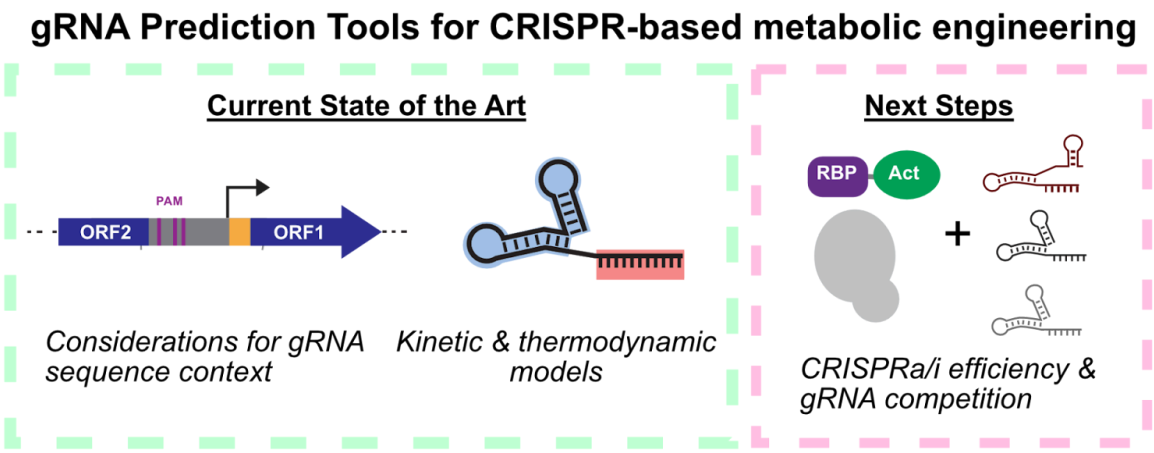


Figure 2. Current considerations and opportunities for improvement of gRNA design tools for metabolic engineering.

Many gRNA models developed for eukaryotic systems are not portable to prokaryotic systems, and vice-versa. Differences in genome organization and rules for effective interference and activation limit the generalizability of these models across branches of life. Wang & Zhang sought to overcome this limitation by developing a deep learning model based on RNA folding parameters, melting temperatures, and experimental data from bacterial cells to accurately predict on- and off-target activities for genome targeting sgRNAs and then use transfer learning to apply their model to eukaryotic systems as well [84]. Transfer learning is a form of machine learning useful to transfer parameters optimized from one dataset to another model with limited data. However, their experimental results confirmed previous findings [101,102] that gRNA activity prediction models developed for eukarya or prokarya are not appropriate for the other, likely due to differences in chromatin structure and genome organization. For microbial metabolic engineering, this results in a limited number of reliable gRNA design tools. Human cell lines are the most popular chassis for the creation of guide prediction models and software. However, these models are generally not transferable to metabolic engineering contexts, which typically require an understanding of CRISPR performance in bacteria or yeast. There is a clear need in the field for a gRNA prediction software suited towards metabolic engineers; that can accurately predict on- and off-target gRNA performance in relevant model organisms.

Beyond predicting gRNA efficiency for a single gene target, several models have investigated the impact of resource competition on the efficacy of CRISPR perturbations. Specifically, the decrease in CRISPR efficiency that occurs as a result of dCas9 resource competition from expressing multiple orthogonal gRNAs is well-characterized [95,103–105]. Clamons & Murray proposed a straightforward model for resource competition in which the expression of competing gRNAs reduces the available pool of CRISPR complexes able to perform CRISPRi at each target site by roughly 1/N, where N is the number of gRNAs. The Clamons & Murray model suggests that CRISPRa is more tolerant to resource competition than CRISPRi, however subsequent work has challenged this claim [106]. In practice, CRISPRa is likely similarly affected by resource competition as CRISPRi because the number of activator complexes in the cell may be limiting. Recently, Barbier *et al.* showed that differences in gRNA binding affinities between scRNAs and sgRNAs leads to difficulty in predicting the behavior of combined CRISPRa/i programs in bacteria [107]. Through the use of a mathematical model for dCas9 resource competition and experimental validation, they demonstrated that using gRNAs with similar structures reduces differences in gRNA binding affinity and improves the function of multi-gRNA programs. These results suggest that the use of computational tools for predicting gRNA folding and dCas9 binding affinity will be crucial for constructing effective multi-gene CRISPRa/i metabolic engineering programs.

While resource competition in CRISPRa and CRISPRi has been modeled for basic regulatory networks, the effect of dCas9 and activator resource competition on metabolic engineering productivity remains understudied. This is due in part to a lack of integrative models that consider the capabilities of CRISPR systems while making strain recommendations. For example, a model that can account for gRNA competition as the number of genomic targets increases may help engineers prioritize which sets of genes to target. Incorporating a simple gRNA competition model into strain recommendation tools would be straightforward; however, more complex gRNA considerations such as target sequence context may be more difficult to integrate. Moving forward, these integrative models will be an important consideration when building large combinatorial metabolic engineering platforms that aim to simultaneously up- and down- regulate

multiple genes, as there is an inherent tradeoff between the number of targets and the degree of perturbation.

Outlook and Future Directions

Significant advances in metabolic engineering have been achieved through the use of model-informed CRISPR transcriptional control programs. Models have enabled the prediction of effective gene targets for perturbation and the rational design of high performing gRNA sequences. Still, there are many limitations to existing tools that, if overcome, would yield substantial benefits to the field of metabolic engineering. Current models are helpful for identifying targets for up- or down-regulation [10–12,16,17,35,36,80,108], but cannot effectively predict the optimal amount of intervention, as total activation or repression of a gene is rarely the optimal perturbation [19,37,55]. Furthermore, CRISPRa/i tools will enable improved dynamic control over metabolic engineering programs [22]. As dynamic control becomes more prevalent, models able to integrate kinetic predictions with strain optimization will prove increasingly valuable.

CRISPR-Cas systems offer several advantages over other gene regulation tools that make them particularly well-suited for implementing model-driven recommendations. Compared to other tools such as ZF or TALE proteins, CRISPR tools have more ease of programmability, higher modularity and multiplexing capabilities, and lower off-target effects [109,110]. Additionally, over the last several years, CRISPR-Cas tools have been developed in a diverse range of microorganisms, expanding the scope of available chassis for metabolic engineering beyond traditional hosts. While numerous systems are available for implementing genetic interventions, CRISPR-based tools are uniquely able to accelerate engineering and DBTL cycles for strain design informed by genome-scale modeling [109].

Synthetic biologists have recently become interested in how machine learning can advance metabolic engineering goals [111]. The design of effective gRNAs is a well-suited problem for machine learning. The value of machine learning for gRNA design has been demonstrated by the use of a deep learning algorithm to predict gRNA on- and off-target effects for dCas13, a relatively understudied Cas variant [112]. Yu et al.

recently developed a machine learning approach using publicly available depletion screens for prediction of CRISPRi guide efficiency that outperformed previous models. In addition, their model highlights the importance of gene expression levels and gene-specific features in making accurate predictions, providing novel insights into the understanding of CRISPRi screens for future work [38]. There are already many existing datasets and large libraries that can easily be screened in a high-throughput fashion if linked to a fitness score [113,114] or biosensor output [115,116].

For strain design, the use of machine learning is less straightforward. Ideally, machine learning-based strain design algorithms would be fit on multi-omics datasets, including transcriptome, proteome, and metabolome data [31,117]. Minimally, metabolomics data would be needed to relate strain genotypes to production phenotypes [97,117]. Alternatively, metabolite-responsive biosensors can be used to generate input data, but this approach is restricted by the space of available biosensors and their respective dynamic ranges [54,118]. Additionally, biosensor outputs do not necessarily correlate strongly with measured metabolite production [54]. In any case, the time, labor, and costs required to collect sufficient perturbation data to create an effective model is limiting [111]. Porting a model to another organism, or the same organism with a novel heterologous pathway, may lead to a decrease in predictive power without the collection of new training data. Alternatively, the use of mechanistic models leads to hypothesis-driven experimentation, which is more cost and time efficient in the near- and long-term. Additionally, there is still interest in understanding the underlying mechanisms behind genotype/phenotype relationships. The use of machine learning hides these relationships, which can introduce challenges when trying to understand how or why machine-learning aided design fails.

The use of constraint-based genome-scale models has been shown to be useful for predicting the phenotypic effect of knockouts. However, such models are less informative when making predictions based on more subtle changes to protein expression, which may be achieved through CRISPR control. Metabolic pathways include gene regulatory networks, and allosteric and feedback control, both of which play critical roles in shaping an organism’s phenotype. For example, changing the expression level of a single gene will change metabolite levels which, through feedback

mechanisms, will cause changes in protein expression at other sites, resulting in further downstream effects. Such effects cannot be easily captured by current constraint-based models.

Ideally, more detailed kinetic models could be developed to address challenges like feedback effects and regulatory networks, allowing metabolic engineers to make more precise predictions. While kinetic models are difficult to build, there is a growing list of success stories. For example, Martin *et al.* [67] built an in vitro pathway model of butanol production consisting of over 200 reactions, van Niekerk *et al.* [119] developed a detailed and validated model of energy metabolism in *P. falciparum*, and Millard *et al.* [120] built a validated kinetic model of core metabolism in *E. coli*. As kinetic models become larger, there is also growing importance in the reproducibility of these models [121], as a considerable amount of intellectual effort and expense is used to develop them.

Other modeling approaches include building approximate kinetic models based on lin-log kinetics [122,123]. Approximate kinetic models are simpler than full mechanistic models, while also giving engineers direct access to the pathway sensitivities through metabolic control analysis [67,122]. However, a disadvantage of lin-log models is that their predictive power drops if changes in gene expression are too large. Nevertheless, these models may still be able to predict the most promising sites for CRISPR perturbations.

As the available toolbox for metabolic engineering using CRISPR/Cas tools grows, the need for quantitative and predictive models to inform experiments will become increasingly valuable. Integrative models capable of suggesting genetic perturbations, selecting effective gRNAs, and composing selections into large multi-gene CRISPRa/i programs will greatly improve our ability to rapidly survey combinatorial perturbations and accelerate the DBTL cycle.

Acknowledgements

We thank Diego Alba Burbano, Benjamin I. Tickman, Cholpisit (Ice) Kiattisewee, and all members of the Carothers and Zalatan groups for advice, materials, and comments on the manuscript.

Competing interests

J.G.Z. and J.M.C are members of the Wayfinder Biosciences scientific advisory board. R.A.L.C., J.G.Z, and J.M.C. are inventors on patents and/or patent applications filed by the University of Washington that describe CRISPRa/i tools in prokaryotic systems.

Funding information

This work was supported by US National Science Foundation (NSF) Award MCB 2225632 MCB (J.M.C. and J.G.Z.) and US Department of Energy (DOE) Award DE-SC0023091 (J.M.C., J.G.Z., and H.M.S.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Author Information

Corresponding Authors

Ryan A. L. Cardiff - *Molecular Engineering & Sciences Institute, Center for Synthetic Biology, Department of Chemistry, Department of Chemical Engineering, University of Washington, Seattle, WA 98195, United States,* https://orcid.org/0009-0001-0530-9010, Email: cardiffr@uw.edu

Herbert M. Sauro - *Molecular Engineering & Sciences Institute, Center for Synthetic Biology, Department of Bioengineering, University of Washington, Seattle, WA 98195, United States,* https://orcid.org/0000-0002-3659-6817, Email: hsauro@uw.edu

Authors

Jesse G. Zalatan - *Molecular Engineering & Sciences Institute, Center for Synthetic Biology, and Department of Chemistry, University of Washington, Seattle, WA 98195, United States,* https://orcid.org/0000-0002-1458-0654, Email: zalatan@uw.edu

James M. Carothers - *Molecular Engineering & Sciences Institute, Center for Synthetic Biology, Department of Chemical Engineering, University of Washington, Seattle, WA 98195, United States,* https://orcid.org/0000-0001-6728-7833, Email: jcaroth@uw.edu

Author Contributions

R.A.L.C. and H.M.S. wrote the manuscript with input and supervision by J.M.C. and J.G.Z.

References

(1) Li, S.; Jendresen, C. B.; Grünberger, A.; Ronda, C.; Jensen, S. I.; Noack, S.; Nielsen, A. T. Enhanced Protein and Biochemical Production Using CRISPRi-Based Growth Switches. *Metab. Eng.* **2016**, *38*, 274–284. https://doi.org/10.1016/j.ymben.2016.09.003.

(2) Lian, J.; HamediRad, M.; Hu, S.; Zhao, H. Combinatorial Metabolic Engineering Using an Orthogonal Tri-Functional CRISPR System. *Nat. Commun.* **2017**, *8* (1), 1688. https://doi.org/10.1038/s41467-017-01695-x.

(3) Kiattisewee, C.; Dong, C.; Fontana, J.; Sugianto, W.; Peralta-Yahya, P.; Carothers, J. M.; Zalatan, J. G. Portable Bacterial CRISPR Transcriptional Activation Enables Metabolic Engineering in Pseudomonas Putida. *Metab. Eng.* **2021**, *66*, 283–295. https://doi.org/10.1016/j.ymben.2021.04.002.

(4) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.

(5) Bikard, D.; Jiang, W.; Samai, P.; Hochschild, A.; Zhang, F.; Marraffini, L. A. Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System. *Nucleic Acids Res.* **2013**, *41* (15), 7429–7437. https://doi.org/10.1093/nar/gkt520.

(6) 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.; Lim, W. A. Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **2015**, *160* (1), 339–350. https://doi.org/10.1016/j.cell.2014.11.052.

(7) Dong, C.; Fontana, J.; Patel, A.; Carothers, J. M.; Zalatan, J. G. Synthetic CRISPR-Cas Gene Activators for Transcriptional Reprogramming in Bacteria. *Nat. Commun.* **2018**, *9* (1), 2489. https://doi.org/10.1038/s41467-018-04901-6.

(8) McCarty, N. S.; Graham, A. E.; Studená, L.; Ledesma-Amaro, R. Multiplexed CRISPR Technologies for Gene Editing and Transcriptional Regulation. *Nat. Commun.* **2020**, *11* (1), 1281. https://doi.org/10.1038/s41467-020-15053-x.

(9) Fontana, J.; Sparkman-Yager, D.; Zalatan, J. G.; Carothers, J. M. Challenges and Opportunities with CRISPR Activation in Bacteria for Data-Driven Metabolic Engineering. *Curr. Opin. Biotechnol.* **2020**, *64*, 190–198. https://doi.org/10.1016/j.copbio.2020.04.005.

(10) McAnulty, M. J.; Yen, J. Y.; Freedman, B. G.; Senger, R. S. Genome-Scale Modeling Using Flux Ratio Constraints to Enable Metabolic Engineering of Clostridial Metabolism in Silico. *BMC Syst. Biol.* **2012**, *6* (1), 42. https://doi.org/10.1186/1752-0509-6-42.

(11) Cautha, S. C.; Gowen, C. M.; Lussier, F.-X.; Gold, N. D.; Martin, V. J. J.; Mahadevan, R. Model-Driven Design of a Saccharomyces Cerevisiae Platform Strain with Improved Tyrosine Production Capabilities. *IFAC Proc. Vol.* **2013**, *46* (31), 221–226. https://doi.org/10.3182/20131216-3-IN-2044.00066.

(12) Banerjee, D.; Eng, T.; Lau, A. K.; Sasaki, Y.; Wang, B.; Chen, Y.; Prahl, J.-P.; Singan, V. R.; Herbert, R. A.; Liu, Y.; Tanjore, D.; Petzold, C. J.; Keasling, J. D.; Mukhopadhyay, A. Genome-Scale Metabolic Rewiring Improves Titers Rates and

Yields of the Non-Native Product Indigoidine at Scale. *Nat. Commun.* **2020**, *11* (1), 5385. https://doi.org/10.1038/s41467-020-19171-4.
(13) Li, W.; Xu, H.; Xiao, T.; Cong, L.; Love, M. I.; Zhang, F.; Irizarry, R. A.; Liu, J. S.; Brown, M.; Liu, X. S. MAGeCK Enables Robust Identification of Essential Genes from Genome-Scale CRISPR/Cas9 Knockout Screens. *Genome Biol.* **2014**, *15* (12), 554. https://doi.org/10.1186/s13059-014-0554-4.
(14) 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.; Root, D. E. Optimized sgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9. *Nat. Biotechnol.* **2016**, *34* (2), 184–191. https://doi.org/10.1038/nbt.3437.
(15) Labun, K.; Montague, T. G.; Gagnon, J. A.; Thyme, S. B.; Valen, E. CHOPCHOP v2: A Web Tool for the next Generation of CRISPR Genome Engineering. *Nucleic Acids Res.* **2016**, *44* (W1), W272–W276. https://doi.org/10.1093/nar/gkw398.
(16) Kim, S. K.; Seong, W.; Han, G. H.; Lee, D.-H.; Lee, S.-G. CRISPR Interference-Guided Multiplex Repression of Endogenous Competing Pathway Genes for Redirecting Metabolic Flux in Escherichia Coli. *Microb. Cell Factories* **2017**, *16* (1), 188. https://doi.org/10.1186/s12934-017-0802-x.
(17) Kozaeva, E.; Volkova, S.; Matos, M. R. A.; Mezzina, M. P.; Wulff, T.; Volke, D. C.; Nielsen, L. K.; Nikel, P. I. Model-Guided Dynamic Control of Essential Metabolic Nodes Boosts Acetyl-Coenzyme A–Dependent Bioproduction in Rewired Pseudomonas Putida. *Metab. Eng.* **2021**, *67*, 373–386. https://doi.org/10.1016/j.ymben.2021.07.014.
(18) Chavez, A.; Scheiman, J.; Vora, S.; Pruitt, B. W.; Tuttle, M.; P R Iyer, E.; Lin, S.; Kiani, S.; Guzman, C. D.; Wiegand, D. J.; Ter-Ovanesyan, D.; Braff, J. L.; Davidsohn, N.; Housden, B. E.; Perrimon, N.; Weiss, R.; Aach, J.; Collins, J. J.; Church, G. M. Highly Efficient Cas9-Mediated Transcriptional Programming. *Nat. Methods* **2015**, *12* (4), 326–328. https://doi.org/10.1038/nmeth.3312.
(19) Holtz, W. J.; Keasling, J. D. Engineering Static and Dynamic Control of Synthetic Pathways. *Cell* **2010**, *140* (1), 19–23. https://doi.org/10.1016/j.cell.2009.12.029.
(20) Schober, A. F.; Mathis, A. D.; Ingle, C.; Park, J. O.; Chen, L.; Rabinowitz, J. D.; Junier, I.; Rivoire, O.; Reynolds, K. A. A Two-Enzyme Adaptive Unit within Bacterial Folate Metabolism. *Cell Rep.* **2019**, *27* (11), 3359-3370.e7. https://doi.org/10.1016/j.celrep.2019.05.030.
(21) Zhang, F.; Carothers, J. M.; Keasling, J. D. Design of a Dynamic Sensor-Regulator System for Production of Chemicals and Fuels Derived from Fatty Acids. *Nat. Biotechnol.* **2012**, *30* (4), 354–359. https://doi.org/10.1038/nbt.2149.
(22) Gupta, A.; Reizman, I. M. B.; Reisch, C. R.; Prather, K. L. J. Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit. *Nat. Biotechnol.* **2017**, *35* (3), 273–279. https://doi.org/10.1038/nbt.3796.
(23) Fang, X.; Lloyd, C. J.; Palsson, B. O. Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications. *Nat. Rev. Microbiol.* **2020**, *18* (12), 731–743. https://doi.org/10.1038/s41579-020-00440-4.
(24) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat. Biotechnol.* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.

(25) Lewis, N. E.; Nagarajan, H.; Palsson, B. O. Constraining the Metabolic Genotype-Phenotype Relationship Using a Phylogeny of in Silico Methods. *Nat. Rev. Microbiol.* **2012**, *10* (4), 291–305. https://doi.org/10.1038/nrmicro2737.
(26) Schuetz, R.; Kuepfer, L.; Sauer, U. Systematic Evaluation of Objective Functions for Predicting Intracellular Fluxes in Escherichia Coli. *Mol. Syst. Biol.* **2007**, *3* (1), 119. https://doi.org/10.1038/msb4100162.
(27) García Sánchez, C. E.; Torres Sáez, R. G. Comparison and Analysis of Objective Functions in Flux Balance Analysis. *Biotechnol. Prog.* **2014**, *30* (5), 985–991. https://doi.org/10.1002/btpr.1949.
(28) Pharkya, P.; Maranas, C. D. An Optimization Framework for Identifying Reaction Activation/Inhibition or Elimination Candidates for Overproduction in Microbial Systems. *Metab. Eng.* **2006**, *8* (1), 1–13. https://doi.org/10.1016/j.ymben.2005.08.003.
(29) *OptDesign: Identifying Optimum Design Strategies in Strain Engineering for Biochemical Production | ACS Synthetic Biology*. https://pubs.acs.org/doi/10.1021/acssynbio.1c00610 (accessed 2024-03-25).
(30) *Redirector: Designing Cell Factories by Reconstructing the Metabolic Objective | PLOS Computational Biology*. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002882 (accessed 2024-03-25).
(31) Radivojević, T.; Costello, Z.; Workman, K.; Garcia Martin, H. A Machine Learning Automated Recommendation Tool for Synthetic Biology. *Nat. Commun.* **2020**, *11* (1), 4879. https://doi.org/10.1038/s41467-020-18008-4.
(32) Sabzevari, M.; Szedmak, S.; Penttilä, M.; Jouhten, P.; Rousu, J. Strain Design Optimization Using Reinforcement Learning. *PLOS Comput. Biol.* **2022**, *18* (6), e1010177. https://doi.org/10.1371/journal.pcbi.1010177.
(33) van Rosmalen, R. P.; Smith, R. W.; Martins dos Santos, V. A. P.; Fleck, C.; Suarez-Diez, M. Model Reduction of Genome-Scale Metabolic Models as a Basis for Targeted Kinetic Models. *Metab. Eng.* **2021**, *64*, 74–84. https://doi.org/10.1016/j.ymben.2021.01.008.
(34) Venayak, N.; von Kamp, A.; Klamt, S.; Mahadevan, R. MoVE Identifies Metabolic Valves to Switch between Phenotypic States. *Nat. Commun.* **2018**, *9* (1), 5332. https://doi.org/10.1038/s41467-018-07719-4.
(35) *Bidirectional titration of yeast gene expression using a pooled CRISPR guide RNA approach*. PNAS. https://www.pnas.org/doi/abs/10.1073/pnas.2007413117 (accessed 2022-02-28).
(36) Dusad, V.; Thiel, D.; Barahona, M.; Keun, H. C.; Oyarzún, D. A. Opportunities at the Interface of Network Science and Metabolic Modeling. *Front. Bioeng. Biotechnol.* **2021**, *8*, 1570. https://doi.org/10.3389/fbioe.2020.591049.
(37) Liu, Z.; Robson, P.; Cheng, A. Simultaneous Multifunctional Transcriptome Engineering by CRISPR RNA Scaffold. bioRxiv June 22, 2022, p 2022.06.21.497089. https://doi.org/10.1101/2022.06.21.497089.
(38) Yu, Y.; Gawlitt, S.; de Andrade e Sousa, L. B.; Merdivan, E.; Piraud, M.; Beisel, C. L.; Barquist, L. Improved Prediction of Bacterial CRISPRi Guide Efficiency from Depletion Screens through Mixed-Effect Machine Learning and Data Integration. *Genome Biol.* **2024**, *25* (1), 13. https://doi.org/10.1186/s13059-023-03153-y.

(39) Wang, T.; Guan, C.; Guo, J.; Liu, B.; Wu, Y.; Xie, Z.; Zhang, C.; Xing, X.-H. Pooled CRISPR Interference Screening Enables Genome-Scale Functional Genomics Study in Bacteria with Superior Performance. *Nat. Commun.* **2018**, *9* (1), 2475. https://doi.org/10.1038/s41467-018-04899-x.
(40) Mahadevan, R.; von Kamp, A.; Klamt, S. Genome-Scale Strain Designs Based on Regulatory Minimal Cut Sets. *Bioinformatics* **2015**, *31* (17), 2844–2851. https://doi.org/10.1093/bioinformatics/btv217.
(41) Shaw, W. M.; Studená, L.; Roy, K.; Hapeta, P.; McCarty, N. S.; Graham, A. E.; Ellis, T.; Ledesma-Amaro, R. Inducible Expression of Large gRNA Arrays for Multiplexed CRISPRai Applications. *Nat. Commun.* **2022**, *13* (1), 4984. https://doi.org/10.1038/s41467-022-32603-7.
(42) Shen, F.; Sun, R.; Yao, J.; Li, J.; Liu, Q.; Price, N. D.; Liu, C.; Wang, Z. OptRAM: In-Silico Strain Design via Integrative Regulatory-Metabolic Network Modeling. *PLOS Comput. Biol.* **2019**, *15* (3), e1006835. https://doi.org/10.1371/journal.pcbi.1006835.
(43) Kim, J.; Reed, J. L. OptORF: Optimal Metabolic and Regulatory Perturbations for Metabolic Engineering of Microbial Strains. *BMC Syst. Biol.* **2010**, *4* (1), 53. https://doi.org/10.1186/1752-0509-4-53.
(44) Razaghi-Moghadam, Z.; Nikoloski, Z. GeneReg: A Constraint-Based Approach for Design of Feasible Metabolic Engineering Strategies at the Gene Level. *Bioinformatics* **2021**, *37* (12), 1717–1723. https://doi.org/10.1093/bioinformatics/btaa996.
(45) Call, S. N.; Andrews, L. B. CRISPR-Based Approaches for Gene Regulation in Non-Model Bacteria. *Front. Genome Ed.* **2022**, *4*, 892304. https://doi.org/10.3389/fgeed.2022.892304.
(46) Machado, D.; Andrejev, S.; Tramontano, M.; Patil, K. R. Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities. *Nucleic Acids Res.* **2018**, *46* (15), 7542–7553. https://doi.org/10.1093/nar/gky537.
(47) Zimmermann, J.; Kaleta, C.; Waschina, S. Gapseq: Informed Prediction of Bacterial Metabolic Pathways and Reconstruction of Accurate Metabolic Models. *Genome Biol.* **2021**, *22* (1), 81. https://doi.org/10.1186/s13059-021-02295-1.
(48) Mendoza, S. N.; Olivier, B. G.; Molenaar, D.; Teusink, B. A Systematic Assessment of Current Genome-Scale Metabolic Reconstruction Tools. *Genome Biol.* **2019**, *20* (1), 158. https://doi.org/10.1186/s13059-019-1769-1.
(49) Henry, C. S.; DeJongh, M.; Best, A. A.; Frybarger, P. M.; Linsay, B.; Stevens, R. L. High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models. *Nat. Biotechnol.* **2010**, *28* (9), 977–982. https://doi.org/10.1038/nbt.1672.
(50) Greene, J.; Daniell, J.; Köpke, M.; Broadbelt, L.; Tyo, K. E. J. Kinetic Ensemble Model of Gas Fermenting Clostridium Autoethanogenum for Improved Ethanol Production. *Biochem. Eng. J.* **2019**, *148*, 46–56. https://doi.org/10.1016/j.bej.2019.04.021.
(51) Woolston, B. M.; Emerson, D. F.; Currie, D. H.; Stephanopoulos, G. Rediverting Carbon Flux in Clostridium Ljungdahlii Using CRISPR Interference (CRISPRi). *Metab. Eng.* **2018**, *48*, 243–253. https://doi.org/10.1016/j.ymben.2018.06.006.

(52) Schuchmann, K.; Müller, V. Autotrophy at the Thermodynamic Limit of Life: A Model for Energy Conservation in Acetogenic Bacteria. *Nat. Rev. Microbiol.* **2014**, *12* (12), 809–821. https://doi.org/10.1038/nrmicro3365.
(53) Nagarajan, H.; Sahin, M.; Nogales, J.; Latif, H.; Lovley, D. R.; Ebrahim, A.; Zengler, K. Characterizing Acetogenic Metabolism Using a Genome-Scale Metabolic Reconstruction of Clostridium Ljungdahlii. *Microb. Cell Factories* **2013**, *12* (1), 118. https://doi.org/10.1186/1475-2859-12-118.
(54) Zhang, J.; Petersen, S. D.; Radivojevic, T.; Ramirez, A.; Pérez-Manríquez, A.; Abeliuk, E.; Sánchez, B. J.; Costello, Z.; Chen, Y.; Fero, M. J.; Martin, H. G.; Nielsen, J.; Keasling, J. D.; Jensen, M. K. Combining Mechanistic and Machine Learning Models for Predictive Engineering and Optimization of Tryptophan Metabolism. *Nat. Commun.* **2020**, *11* (1), 4880. https://doi.org/10.1038/s41467-020-17910-1.
(55) HamediRad, M.; Chao, R.; Weisberg, S.; Lian, J.; Sinha, S.; Zhao, H. Towards a Fully Automated Algorithm Driven Platform for Biosystems Design. *Nat. Commun.* **2019**, *10* (1), 5150. https://doi.org/10.1038/s41467-019-13189-z.
(56) Karim, A. S.; Dudley, Q. M.; Juminaga, A.; Yuan, Y.; Crowe, S. A.; Heggestad, J. T.; Garg, S.; Abdalla, T.; Grubbe, W. S.; Rasor, B. J.; Coar, D. N.; Torculas, M.; Krein, M.; Liew, F. (Eric); Quattlebaum, A.; Jensen, R. O.; Stuart, J. A.; Simpson, S. D.; Köpke, M.; Jewett, M. C. In Vitro Prototyping and Rapid Optimization of Biosynthetic Enzymes for Cell Design. *Nat. Chem. Biol.* **2020**, *16* (8), 912–919. https://doi.org/10.1038/s41589-020-0559-0.
(57) Krüger, A.; Mueller, A. P.; Rybnicky, G. A.; Engle, N. L.; Yang, Z. K.; Tschaplinski, T. J.; Simpson, S. D.; Köpke, M.; Jewett, M. C. Development of a Clostridia-Based Cell-Free System for Prototyping Genetic Parts and Metabolic Pathways. *Metab. Eng.* **2020**, *62*, 95–105. https://doi.org/10.1016/j.ymben.2020.06.004.
(58) Grubbe, W. S.; Rasor, B. J.; Krüger, A.; Jewett, M. C.; Karim, A. S. Cell-Free Styrene Biosynthesis at High Titers. *Metab. Eng.* **2020**, *61*, 89–95. https://doi.org/10.1016/j.ymben.2020.05.009.
(59) Kay, J. E.; Jewett, M. C. A Cell-Free System for Production of 2,3-Butanediol Is Robust to Growth-Toxic Compounds. *Metab. Eng. Commun.* **2020**, *10*, e00114. https://doi.org/10.1016/j.mec.2019.e00114.
(60) Dudley, Q. M.; Karim, A. S.; Jewett, M. C. Cell-Free Metabolic Engineering: Biomanufacturing beyond the Cell. *Biotechnol. J.* **2015**, *10* (1), 69–82. https://doi.org/10.1002/biot.201400330.
(61) Marshall, R.; Noireaux, V. Quantitative Modeling of Transcription and Translation of an All-E. Coli Cell-Free System. *Sci. Rep.* **2019**, *9* (1), 11980. https://doi.org/10.1038/s41598-019-48468-8.
(62) Tickman, B. I.; Burbano, D. A.; Chavali, V. P.; Kiattisewee, C.; Fontana, J.; Khakimzhan, A.; Noireaux, V.; Zalatan, J. G.; Carothers, J. M. Multi-Layer CRISPRa/i Circuits for Dynamic Genetic Programs in Cell-Free and Bacterial Systems. *Cell Syst.* **2021**, *0* (0). https://doi.org/10.1016/j.cels.2021.10.008.
(63) Moore, S. J.; MacDonald, J. T.; Wienecke, S.; Ishwarbhai, A.; Tsipa, A.; Aw, R.; Kylilis, N.; Bell, D. J.; McClymont, D. W.; Jensen, K.; Polizzi, K. M.; Biedendieck, R.; Freemont, P. S. Rapid Acquisition and Model-Based Analysis of Cell-Free Transcription–Translation Reactions from Nonmodel Bacteria. *Proc. Natl. Acad.*

Sci. **2018**, *115* (19), E4340–E4349. https://doi.org/10.1073/pnas.1715806115.
(64) Lehr, F.-X.; Hanst, M.; Vogel, M.; Kremer, J.; Göringer, H. U.; Suess, B.; Koeppl, H. Cell-Free Prototyping of AND-Logic Gates Based on Heterogeneous RNA Activators. *ACS Synth. Biol.* **2019**, *8* (9), 2163–2173. https://doi.org/10.1021/acssynbio.9b00238.
(65) Koch, M.; Faulon, J.-L.; Borkowski, O. Models for Cell-Free Synthetic Biology: Make Prototyping Easier, Better, and Faster. *Front. Bioeng. Biotechnol.* **2018**, *6*, 182. https://doi.org/10.3389/fbioe.2018.00182.
(66) Vilkhovoy, M.; Horvath, N.; Shih, C.-H.; Wayman, J. A.; Calhoun, K.; Swartz, J.; Varner, J. D. Sequence Specific Modeling of E. Coli Cell-Free Protein Synthesis. *ACS Synth. Biol.* **2018**, *7* (8), 1844–1857. https://doi.org/10.1021/acssynbio.7b00465.
(67) Martin, J. P.; Rasor, B. J.; DeBonis, J.; Karim, A. S.; Jewett, M. C.; Tyo, K. E. J.; Broadbelt, L. J. A Dynamic Kinetic Model Captures Cell-Free Metabolism for Improved Butanol Production. *Metab. Eng.* **2023**, *76*, 133–145. https://doi.org/10.1016/j.ymben.2023.01.009.
(68) Rasor, B. J.; Yi, X.; Brown, H.; Alper, H. S.; Jewett, M. C. An Integrated in Vivo/in Vitro Framework to Enhance Cell-Free Biosynthesis with Metabolically Rewired Yeast Extracts. *Nat. Commun.* **2021**, *12* (1), 5139. https://doi.org/10.1038/s41467-021-25233-y.
(69) Brockman, I. M.; Prather, K. L. J. Dynamic Metabolic Engineering: New Strategies for Developing Responsive Cell Factories. *Biotechnol. J.* **2015**, *10* (9), 1360–1369. https://doi.org/10.1002/biot.201400422.
(70) Wu, Y.; Chen, T.; Liu, Y.; Lv, X.; Li, J.; Du, G.; Ledesma-Amaro, R.; Liu, L. CRISPRi Allows Optimal Temporal Control of N-Acetylglucosamine Bioproduction by a Dynamic Coordination of Glucose and Xylose Metabolism in Bacillus Subtilis. *Metab. Eng.* **2018**, *49*, 232–241. https://doi.org/10.1016/j.ymben.2018.08.012.
(71) Hartline, C. J.; Schmitz, A. C.; Han, Y.; Zhang, F. Dynamic Control in Metabolic Engineering: Theories, Tools, and Applications. *Metab. Eng.* **2021**, *63*, 126–140. https://doi.org/10.1016/j.ymben.2020.08.015.
(72) Wu, Y.; Chen, T.; Liu, Y.; Tian, R.; Lv, X.; Li, J.; Du, G.; Chen, J.; Ledesma-Amaro, R.; Liu, L. Design of a Programmable Biosensor-CRISPRi Genetic Circuits for Dynamic and Autonomous Dual-Control of Metabolic Flux in Bacillus Subtilis. *Nucleic Acids Res.* **2020**, *48* (2), 996–1009. https://doi.org/10.1093/nar/gkz1123.
(73) Watt, V. M.; Ingles, C. J.; Urdea, M. S.; Rutter, W. J. Homology Requirements for Recombination in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (14), 4768–4772.
(74) Nihongaki, Y.; Kawano, F.; Nakajima, T.; Sato, M. Photoactivatable CRISPR-Cas9 for Optogenetic Genome Editing. *Nat. Biotechnol.* **2015**, *33* (7), 755–760. https://doi.org/10.1038/nbt.3245.
(75) Cunningham-Bryant, D.; Sun, J.; Fernandez, B.; Zalatan, J. G. CRISPR–Cas-Mediated Chemical Control of Transcriptional Dynamics in Yeast. *ChemBioChem* **2019**, *20* (12), 1519–1523. https://doi.org/10.1002/cbic.201800823.
(76) Kundert, K.; Lucas, J. E.; Watters, K. E.; Fellmann, C.; Ng, A. H.; Heineike, B. M.; Fitzsimmons, C. M.; Oakes, B. L.; Qu, J.; Prasad, N.; Rosenberg, O. S.; Savage, D. F.; El-Samad, H.; Doudna, J. A.; Kortemme, T. Controlling CRISPR-Cas9 with

Ligand-Activated and Ligand-Deactivated sgRNAs. *Nat. Commun.* **2019**, *10* (1), 2127. https://doi.org/10.1038/s41467-019-09985-2.
(77) Fontana, J.; Voje, W. E.; Zalatan, J. G.; Carothers, J. M. Prospects for Engineering Dynamic CRISPR–Cas Transcriptional Circuits to Improve Bioproduction. *J. Ind. Microbiol. Biotechnol.* **2018**, *45* (7), 481–490. https://doi.org/10.1007/s10295-018-2039-z.
(78) Lee, J. W.; Na, D.; Park, J. M.; Lee, J.; Choi, S.; Lee, S. Y. Systems Metabolic Engineering of Microorganisms for Natural and Non-Natural Chemicals. *Nat. Chem. Biol.* **2012**, *8* (6), 536–546. https://doi.org/10.1038/nchembio.970.
(79) Maia, P.; Rocha, M.; Rocha, I. In Silico Constraint-Based Strain Optimization Methods: The Quest for Optimal Cell Factories. *Microbiol. Mol. Biol. Rev.* **2015**, *80* (1), 45–67. https://doi.org/10.1128/mmbr.00014-15.
(80) Liu, D.; Mannan, A. A.; Han, Y.; Oyarzún, D. A.; Zhang, F. Dynamic Metabolic Control: Towards Precision Engineering of Metabolism. *J. Ind. Microbiol. Biotechnol.* **2018**, *45* (7), 535–543. https://doi.org/10.1007/s10295-018-2013-9.
(81) Motoche-Monar, C.; Ordoñez, J. E.; Chang, O.; Gonzales-Zubiate, F. A. gRNA Design: How Its Evolution Impacted on CRISPR/Cas9 Systems Refinement. *Biomolecules* **2023**, *13* (12), 1698. https://doi.org/10.3390/biom13121698.
(82) Konstantakos, V.; Nentidis, A.; Krithara, A.; Paliouras, G. CRISPR–Cas9 gRNA Efficiency Prediction: An Overview of Predictive Tools and the Role of Deep Learning. *Nucleic Acids Res.* **2022**, *50* (7), 3616–3637. https://doi.org/10.1093/nar/gkac192.
(83) Yu, Y.; Gawlitt, S.; de Andrade e Sousa, L. B.; Merdivan, E.; Piraud, M.; Beisel, C.; Barquist, L. *Improved Prediction of Bacterial CRISPRi Guide Efficiency through Data Integration and Automated Machine Learning*; preprint; Bioinformatics, 2022. https://doi.org/10.1101/2022.05.27.493707.
(84) Wang, L.; Zhang, J. Prediction of sgRNA On-Target Activity in Bacteria by Deep Learning. *BMC Bioinformatics* **2019**, *20* (1), 517. https://doi.org/10.1186/s12859-019-3151-4.
(85) Eslami-Mossallam, B.; Klein, M.; Smagt, C. V. D.; Sanden, K. V. D.; Jones, S. K.; Hawkins, J. A.; Finkelstein, I. J.; Depken, M. A Kinetic Model Predicts SpCas9 Activity, Improves off-Target Classification, and Reveals the Physical Basis of Targeting Fidelity. *Nat. Commun.* **2022**, *13* (1), 1367. https://doi.org/10.1038/s41467-022-28994-2.
(86) Farasat, I.; Salis, H. M. A Biophysical Model of CRISPR/Cas9 Activity for Rational Design of Genome Editing and Gene Regulation. *PLoS Comput. Biol.* **2016**, *12* (1), e1004724. https://doi.org/10.1371/journal.pcbi.1004724.
(87) Corsi, G. I.; Qu, K.; Alkan, F.; Pan, X.; Luo, Y.; Gorodkin, J. CRISPR/Cas9 gRNA Activity Depends on Free Energy Changes and on the Target PAM Context. *Nat. Commun.* **2022**, *13* (1), 3006. https://doi.org/10.1038/s41467-022-30515-0.
(88) Labun, K.; Montague, T. G.; Krause, M.; Torres Cleuren, Y. N.; Tjeldnes, H.; Valen, E. CHOPCHOP v3: Expanding the CRISPR Web Toolbox beyond Genome Editing. *Nucleic Acids Res.* **2019**, *47* (W1), W171–W174. https://doi.org/10.1093/nar/gkz365.
(89) Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; Qi, L. S.; Kampmann,

M.; Weissman, J. S. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **2014**, *159* (3), 647–661. https://doi.org/10.1016/j.cell.2014.09.029.
(90) Smith, J. D.; Suresh, S.; Schlecht, U.; Wu, M.; Wagih, O.; Peltz, G.; Davis, R. W.; Steinmetz, L. M.; Parts, L.; St.Onge, R. P. Quantitative CRISPR Interference Screens in Yeast Identify Chemical-Genetic Interactions and New Rules for Guide RNA Design. *Genome Biol.* **2016**, *17* (1), 45. https://doi.org/10.1186/s13059-016-0900-9.
(91) Kocak, D. D.; Josephs, E. A.; Bhandarkar, V.; Adkar, S. S.; Kwon, J. B.; Gersbach, C. A. Increasing the Specificity of CRISPR Systems with Engineered RNA Secondary Structures. *Nat. Biotechnol.* **2019**, *37* (6), 657–666. https://doi.org/10.1038/s41587-019-0095-1.
(92) Haeussler, M.; Schönig, K.; Eckert, H.; Eschstruth, A.; Mianné, J.; Renaud, J.-B.; Schneider-Maunoury, S.; Shkumatava, A.; Teboul, L.; Kent, J.; Joly, J.-S.; Concordet, J.-P. Evaluation of Off-Target and on-Target Scoring Algorithms and Integration into the Guide RNA Selection Tool CRISPOR. *Genome Biol.* **2016**, *17* (1), 148. https://doi.org/10.1186/s13059-016-1012-2.
(93) Chen, Y.; Zeng, S.; Hu, R.; Wang, X.; Huang, W.; Liu, J.; Wang, L.; Liu, G.; Cao, Y.; Zhang, Y. Using Local Chromatin Structure to Improve CRISPR/Cas9 Efficiency in Zebrafish. *PLoS ONE* **2017**, *12* (8), e0182528. https://doi.org/10.1371/journal.pone.0182528.
(94) Jensen, K. T.; Fløe, L.; Petersen, T. S.; Huang, J.; Xu, F.; Bolund, L.; Luo, Y.; Lin, L. Chromatin Accessibility and Guide Sequence Secondary Structure Affect CRISPR-Cas9 Gene Editing Efficiency. *FEBS Lett.* **2017**, *591* (13), 1892–1901. https://doi.org/10.1002/1873-3468.12707.
(95) Zhang, S.; Voigt, C. A. Engineered dCas9 with Reduced Toxicity in Bacteria: Implications for Genetic Circuit Design. *Nucleic Acids Res.* **2018**, *46* (20), 11115–11125. https://doi.org/10.1093/nar/gky884.
(96) Van Hove, B.; De Wannemaeker, L.; Missiaen, I.; Maertens, J.; De Mey, M. Taming CRISPRi: Dynamic Range Tuning through Guide RNA Diversion. *New Biotechnol.* **2023**, *77*, 50–57. https://doi.org/10.1016/j.nbt.2023.07.001.
(97) Fontana, J.; Sparkman-Yager, D.; Faulkner, I.; Cardiff, R.; Kiattisewee, C.; Walls, A.; Primo, T. G.; Kinnunen, P. C.; Martin, H. G.; Zalatan, J. G.; Carothers, J. M. Guide RNA Structure Design Enables Combinatorial CRISPRa Programs for Biosynthetic Profiling. bioRxiv November 17, 2023, p 2023.11.17.567465. https://doi.org/10.1101/2023.11.17.567465.
(98) Moreno-Mateos, M. A.; Vejnar, C. E.; Beaudoin, J.-D.; Fernandez, J. P.; Mis, E. K.; Khokha, M. K.; Giraldez, A. J. CRISPRscan: Designing Highly Efficient sgRNAs for CRISPR-Cas9 Targeting in Vivo. *Nat. Methods* **2015**, *12* (10), 982–988. https://doi.org/10.1038/nmeth.3543.
(99) Ding, W.; Zhang, Y.; Shi, S. Development and Application of CRISPR/Cas in Microbial Biotechnology. *Front. Bioeng. Biotechnol.* **2020**, *8*, 711. https://doi.org/10.3389/fbioe.2020.00711.
(100) Jost, M.; Santos, D. A.; Saunders, R. A.; Horlbeck, M. A.; Hawkins, J. S.; Scaria, S. M.; Norman, T. M.; Hussmann, J. A.; Liem, C. R.; Gross, C. A.; Weissman, J. S. Titrating Gene Expression Using Libraries of Systematically Attenuated CRISPR

Guide RNAs. *Nat. Biotechnol.* **2020**, *38* (3), 355–364. https://doi.org/10.1038/s41587-019-0387-5.
(101) Guo, J.; Wang, T.; Guan, C.; Liu, B.; Luo, C.; Xie, Z.; Zhang, C.; Xing, X.-H. Improved sgRNA Design in Bacteria via Genome-Wide Activity Profiling. *Nucleic Acids Res.* **2018**, *46* (14), 7052–7069. https://doi.org/10.1093/nar/gky572.
(102) Cui, L.; Bikard, D. Consequences of Cas9 Cleavage in the Chromosome of Escherichia Coli. *Nucleic Acids Res.* **2016**, *44* (9), 4243–4251. https://doi.org/10.1093/nar/gkw223.
(103) Huang, H.-H.; Bellato, M.; Qian, Y.; Cárdenas, P.; Pasotti, L.; Magni, P.; Del Vecchio, D. dCas9 Regulator to Neutralize Competition in CRISPRi Circuits. *Nat. Commun.* **2021**, *12* (1), 1692. https://doi.org/10.1038/s41467-021-21772-6.
(104) Chen, P.-Y.; Qian, Y.; Del Vecchio, D. A Model for Resource Competition in CRISPR-Mediated Gene Repression. In *2018 IEEE Conference on Decision and Control (CDC)*; 2018; pp 4333–4338. https://doi.org/10.1109/CDC.2018.8619016.
(105) Clamons, S.; Murray, R. *Modeling Predicts That CRISPR-Based Activators, Unlike CRISPR-Based Repressors, Scale Well with Increasing gRNA Competition and dCas9 Bottlenecking*; preprint; Synthetic Biology, 2019. https://doi.org/10.1101/719278.
(106) Manoj, K.; Del Vecchio, D. Emergent Interactions Due to Resource Competition in CRISPR-Mediated Genetic Activation Circuits. In *2022 IEEE 61st Conference on Decision and Control (CDC)*; IEEE: Cancun, Mexico, 2022; pp 1300–1305. https://doi.org/10.1109/CDC51059.2022.9993376.
(107) Barbier, I.; Kusumawardhani, H.; Chauhan, L.; Harlapur, P. V.; Jolly, M. K.; Schaerli, Y. Synthetic Gene Circuits Combining CRISPR Interference and CRISPR Activation in E. Coli: Importance of Equal Guide RNA Binding Affinities to Avoid Context-Dependent Effects. *ACS Synth. Biol.* **2023**, *12* (10), 3064–3071. https://doi.org/10.1021/acssynbio.3c00375.
(108) Fang, L.; Fan, J.; Luo, S.; Chen, Y.; Wang, C.; Cao, Y.; Song, H. Genome-Scale Target Identification in Escherichia Coli for High-Titer Production of Free Fatty Acids. *Nat. Commun.* **2021**, *12* (1), 4976. https://doi.org/10.1038/s41467-021-25243-w.
(109) Liu, R.; Bassalo, M. C.; Zeitoun, R. I.; Gill, R. T. Genome Scale Engineering Techniques for Metabolic Engineering. *Metab. Eng.* **2015**, *32*, 143–154. https://doi.org/10.1016/j.ymben.2015.09.013.
(110) Gaj, T.; Gersbach, C. A.; Barbas, C. F., III. ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering. *Trends Biotechnol.* **2013**, *31* (7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004.
(111) Antonakoudis, A.; Barbosa, R.; Kotidis, P.; Kontoravdi, C. The Era of Big Data: Genome-Scale Modelling Meets Machine Learning. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3287–3300. https://doi.org/10.1016/j.csbj.2020.10.011.
(112) Wessels, H.-H.; Stirn, A.; Méndez-Mancilla, A.; Kim, E. J.; Hart, S. K.; Knowles, D. A.; Sanjana, N. E. Prediction of On-Target and off-Target Activity of CRISPR–Cas13d Guide RNAs Using Deep Learning. *Nat. Biotechnol.* **2023**, 1–10. https://doi.org/10.1038/s41587-023-01830-8.
(113) de Bakker, V.; Liu, X.; Bravo, A. M.; Veening, J.-W. CRISPRi-Seq for Genome-Wide Fitness Quantification in Bacteria. *Nat. Protoc.* **2022**, *17* (2),

252–281. https://doi.org/10.1038/s41596-021-00639-6.
(114) McGlincy, N. J.; Meacham, Z. A.; Reynaud, K. K.; Muller, R.; Baum, R.; Ingolia, N. T. A Genome-Scale CRISPR Interference Guide Library Enables Comprehensive Phenotypic Profiling in Yeast. *BMC Genomics* **2021**, *22* (1), 205. https://doi.org/10.1186/s12864-021-07518-0.
(115) Mormino, M.; Lenitz, I.; Siewers, V.; Nygård, Y. Identification of Acetic Acid Sensitive Strains through Biosensor-Based Screening of a Saccharomyces Cerevisiae CRISPRi Library. *Microb. Cell Factories* **2022**, *21* (1), 214. https://doi.org/10.1186/s12934-022-01938-7.
(116) Wang, J.; Li, C.; Jiang, T.; Yan, Y. Biosensor-Assisted Titratable CRISPRi High-Throughput (BATCH) Screening for over-Production Phenotypes. *Metab. Eng.* **2023**, *75*, 58–67. https://doi.org/10.1016/j.ymben.2022.11.004.
(117) St. John, P. C.; Bomble, Y. J. Approaches to Computational Strain Design in the Multiomics Era. *Front. Microbiol.* **2019**, *10*. https://doi.org/10.3389/fmicb.2019.00597.
(118) Kaczmarek, J. A.; Prather, K. L. J. Effective Use of Biosensors for High-Throughput Library Screening for Metabolite Production. *J. Ind. Microbiol. Biotechnol.* **2021**, *48* (9–10), kuab049. https://doi.org/10.1093/jimb/kuab049.
(119) van Niekerk, D. D.; du Toit, F.; Green, K.; Palm, D.; Snoep, J. L. A Detailed Kinetic Model of Glycolysis in Plasmodium Falciparum-Infected Red Blood Cells for Antimalarial Drug Target Identification. *J. Biol. Chem.* **2023**, *299* (9), 105111. https://doi.org/10.1016/j.jbc.2023.105111.
(120) Millard, P.; Smallbone, K.; Mendes, P. Metabolic Regulation Is Sufficient for Global and Robust Coordination of Glucose Uptake, Catabolism, Energy Production and Growth in Escherichia Coli. *PLOS Comput. Biol.* **2017**, *13* (2), e1005396. https://doi.org/10.1371/journal.pcbi.1005396.
(121) Shin, J.; Porubsky, V.; Carothers, J.; Sauro, H. M. Standards, Dissemination, and Best Practices in Systems Biology. *Curr. Opin. Biotechnol.* **2023**, *81*, 102922. https://doi.org/10.1016/j.copbio.2023.102922.
(122) John, P. C. S.; Strutz, J.; Broadbelt, L. J.; Tyo, K. E. J.; Bomble, Y. J. Bayesian Inference of Metabolic Kinetics from Genome-Scale Multiomics Data. *PLOS Comput. Biol.* **2019**, *15* (11), e1007424. https://doi.org/10.1371/journal.pcbi.1007424.
(123) McNaughton, A. D.; Bredeweg, E. L.; Manzer, J.; Zucker, J.; Munoz Munoz, N.; Burnet, M. C.; Nakayasu, E. S.; Pomraning, K. R.; Merkley, E. D.; Dai, Z.; Chrisler, W. B.; Baker, S. E.; St. John, P. C.; Kumar, N. Bayesian Inference for Integrating Yarrowia Lipolytica Multiomics Datasets with Metabolic Modeling. *ACS Synth. Biol.* **2021**, *10* (11), 2968–2981. https://doi.org/10.1021/acssynbio.1c00267.

Chapter 5: CRISPR-Cas tools for simultaneous transcription & translation control in bacteria

Ryan A. L. Cardiff[1], Ian D. Faulkner[2], Juliana G. Beall[3], James M. Carothers[1,2]*, Jesse G. Zalatan[1,2,3]*

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: Department of Chemistry
University of Washington
Seattle, WA 98195
United States

*Corresponding authors

jcaroth@uw.edu
206-221-4902

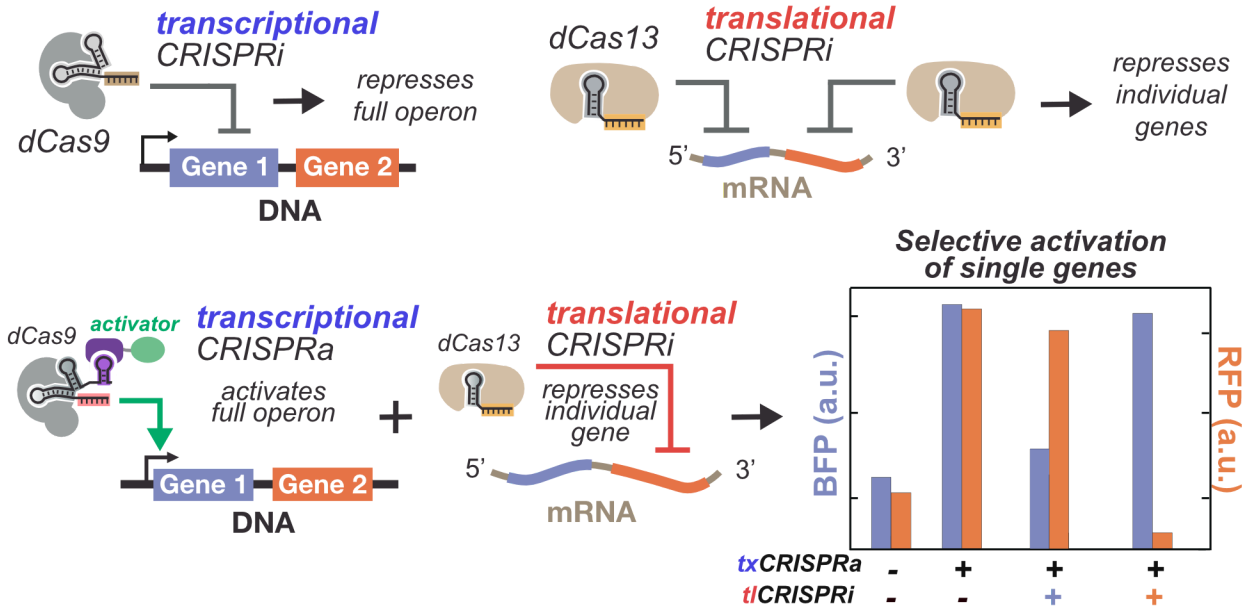
zalatan@uw.edu
206-543-1670

Published as a research article in Nucleic Acids Research on April 13th, 2024. DOI: https://doi.org/10.1093/nar/gkae275

Abstract

Robust control over gene translation at arbitrary mRNA targets is an outstanding challenge in microbial synthetic biology. The development of tools that can regulate translation will greatly expand our ability to precisely control genes across the genome. In *E. coli*, most genes are contained in multi-gene operons, which are subject to polar effects where targeting one gene for repression leads to silencing of other genes in the same operon. These effects pose a challenge for independently regulating individual genes in multi-gene operons. Here, we use CRISPR-dCas13 to address this challenge. We find dCas13-mediated repression exhibits up to 6-fold lower polar effects compared to dCas9. We then show that we can selectively activate single genes in a synthetic multi-gene operon by coupling dCas9 transcriptional activation of an operon with dCas13 translational repression of individual genes within the operon. We also show that dCas13 and dCas9 can be multiplexed for improved biosynthesis of a medically-relevant human milk oligosaccharide. Taken together, our findings suggest that combining transcriptional and translational control can access effects that are difficult to achieve with either mode independently. These combined tools for gene regulation will expand our abilities to precisely engineer bacteria for biotechnology and perform systematic genetic screens.

Programmable transcriptional and translational regulation



Graphical Abstract

Introduction

CRISPR-Cas tools have enabled programmable regulation of gene expression in bacteria, allowing control of genomic functions and optimization of biosynthetic production pathways (1–3). To date, most CRISPR-Cas tools that activate, repress, or knockout gene targets in bacteria have been implemented at the DNA level. Despite their enormous potential, these DNA targeting systems have important limitations in their ability to precisely up- or down-regulate individual bacterial gene targets. One major challenge is that targeting one gene in an operon also affects the genes upstream or downstream in the same transcriptional unit (4). In *E. coli*, 68% of all genes are contained in operons containing two or more genes, with some operons containing as many as 14 genes (5, 6). The prevalence of these multi-gene operons poses a challenge for the regulation of individual genes across the genome. Programmable CRISPR-Cas tools that target RNA could potentially overcome this challenge by regulating translation of individual genes.

Programmable RNA targeting with Cas13 has been applied to a variety of applications in eukaryotic and bacterial systems (7–9). When assembled with a CRISPR RNA (crRNA), Cas13 acts as a crRNA-guided RNA-targeting nuclease. The crRNA contains a direct repeat (DR) sequence with a hairpin structure that is bound by Cas13. The crRNA also contains a variable length spacer sequence that specifies the target RNA (10). In mammalian systems, Cas13 has been used for genome-wide knockdown screens, targeted RNA silencing, and nucleic acid detection (8, 11, 12). The use of catalytically inactive dCas13 fused to an effector domain has also enabled RNA editing and splicing applications (9, 13–15). In *E. coli*, the use of dCas13 for translational

repression (CRISPRi) and activation (CRISPRa) of endogenous transcripts has been demonstrated recently (16–19). Despite this progress, research and applications of Cas13 as a tool for regulating gene expression in bacteria are still limited compared to eukaryotic systems.

Here, we investigate dCas13 as a tool for programmable gene regulation in bacteria. Using catalytically inactive dCas13, we first systematically characterize translational silencing (tlCRISPRi) in *E. coli* with a reporter gene assay. We demonstrate tunable repression by varying dCas13 expression levels and guide RNA (gRNA) length, and we characterize sensitivity to target site position and transcript abundance. We then demonstrate that translational-CRISPRi (tlCRISPRi) with dCas13 can preferentially silence specific genes in a multi-gene operon. Although there are still detectable effects on other genes in the operon, these effects are substantially smaller than those seen from transcriptional CRISPRi (txCRISPRi) with dCas9. We then show that we can preferentially activate individual genes in a multi-gene operon by activating the entire operon with transcriptional CRISPRa (txCRISPRa), then downregulating individual genes using tlCRISPRi. Finally, we demonstrate that multiplexed CRISPRa/i tools operating at both the transcriptional and translational levels can be combined to improve bioproduction of an important human milk oligosaccharide (HMO) relative to transcriptional control alone.

Materials & Methods

Bacterial Strains and Plasmid Constructs

E. coli K-12 MG1655 was the parent strain for all fluorescent reporter experiments. JM109, containing a *ΔlacZ* knockout, was used for LNT biosynthesis experiments. JM109 was a gift from Joachim Messing (Addgene plasmid #49761) (20). CK024 was constructed by integration of Sp.pCas9-dCas9 into MG1655 genome at *arsB* site using the previously described method (21). Strain details/genotypes are described in Table S1. All PCR fragments were amplified with Phusion DNA Polymerase (Thermo-Fisher Scientific) for Infusion Cloning (Takara Bio). Plasmids were transformed into chemically competent NEB Turbo *E. coli* (New England Biolabs) cells, plated on LB-agar, and cultured in LB media supplied with the appropriate antibiotics used in the following concentrations: 100 μg/mL Carbenicillin, 25 μg/mL Chloramphenicol, 30 μg/mL Kanamycin, 100 μg/mL Spectinomycin. All plasmid constructs were confirmed by Sanger sequencing (GENEWIZ). Selected plasmids and complete plasmid maps will be deposited on Addgene (https://www.addgene.org/Jesse_Zalatan/).

All plasmids used in this study are described in Table S2. Relevant DNA sequences are given in the Supplementary Materials. (d)Cas13d and crRNA sequences were cloned from existing plasmids (pXR001 and pXR002) (8). dCas9 plasmids were cloned as described previously (3, 22). Each CRISPR system was expressed in a p15A vector when used alone (Table S2). dCas9 and dCas13 were expressed from the J23107 promoter, except for where otherwise noted. Attempts to clone the Cas proteins under the stronger J23119 promoter failed, likely due to high cellular burden from the expression of the enzymes. In Figures 3 and 4, dCas9 and MCP-SoxS (R93A, S101A) (abbreviated MCP-SoxS) were expressed in a pRSF1010 vector to avoid compatibility issues (Table S2) (23). For the multiplexed dCas9/dCas13 experiments in Figure 5,

dCas9 was moved to a genome integrated copy (CK024 in Table S1). For the CRISPRa experiment in Figure 5, MCP-SoxS was expressed from the BBa_J23107 promoter on the p15a vector (http://parts.igem.org). For CRISPRa reporter experiments, the guide RNAs for dCas9 and dCas13 were expressed from the strong BBa_J23119 promoter, either in the same plasmid with the Cas-carrying plasmid or in a separate ColE1 plasmid (Table S3). All fluorescent reporters contain the J3 array of CRISPR sites upstream from the minimal promoter. All CRISPRi reporters were expressed from the BBa_J23110 minimal promoter on a pSC101** plasmid, unless specified. All CRISPRa reporters and biosynthetic enzymes were expressed from the BBa_J23117 minimal promoter on a pSC101** to provide a greater dynamic range for gene activation.

Plate Reader Experiments

Single colonies from LB plates were inoculated in 400 µL of EZ-RDM (Teknova) supplemented with the appropriate antibiotics and grown in 96-deep-well plates at 37 °C with shaking overnight 900 rpm on a Heidolph titramax 1000. 100 µL of the overnight culture were transferred into flat, clear-bottomed black 96-well plates (Corning) and the OD_{600} and fluorescence were measured in a Biotek Synergy HTX plate reader. For mRFP1 (referred to as mRFP) 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. For mTagBFP2 (referred to as mBFP) detection, the excitation wavelength was 400 nm and emission wavelength was 450 nm. At least three biological replicates were used for all experiments unless stated otherwise.

For kinetic experiments, overnight cultures in stationary phase were subcultured to OD_{600} 0.1 in 200 µL of EZ-RDM (Teknova) supplemented with the appropriate antibiotics. Cultures were grown in flat, clear-bottomed black 96-well plates (Corning) at 37 °C in a Biotek Synergy HTX plate reader set to shake at 1200 RPM. Surrounding wells were filled with 200 µL of water to maintain humidity. The OD_{600} and fluorescence were measured every 30 minutes for 16 hours.

LNT Production

LNT producing constructs were made using the medium-lacY, medium-lgtA, high-CvGalT strain (JM109 cells containing the pJF234.21 and pIDF96C plasmids) as a starting point (24). Single colonies from LB-agar plates were inoculated in 2 mL EZ-RDM (Teknova) with 10 g/L glucose, 2 g/L lactose and supplemented with appropriate antibiotics. Cultures were grown in 14 mL glass culture tubes because we found that previous protocols using polypropylene tubes (24) can have detrimental effects on production. Cultures were incubated at 37 °C with shaking for 48 h. A culture time of 48 h was used to allow adequate time for chemical production after the cells reach stationary phase, as has been done in previous HMO metabolic engineering efforts (25, 26). 500 µL of supernatant from each culture were loaded onto 10 kDa microcentrifuge filters (Millipore) and spun for 20 min at 14000 rcf.

Samples were analyzed via Agilent 6530 LC/Q-TOF in negative mode using a Poroshell 120 C18 150 mm column. We previously used HPLC analysis for detection of LNT (24). In this work, we used LCMS for improved signal resolution. A standard curve was prepared by spiking known amounts of LNT (Elicityl) into supernatants derived from cultures of MG1655 or JM109 E. coli transformed with empty vectors. For LC/MS, the

aqueous phase was water with 0.1% formic acid and the organic phase was acetonitrile with 0.1% formic acid. The % aqueous/organic gradient was run as follows: hold at 3/97 for 1 minute, 10/90 over 1 minute, 40/60 over 3 minutes, hold 40/60 for 2 minutes, then return to 3/97 over 2.5 minutes. The flow rate was held at 0.3 mL/min. The changes in growth conditions and analysis method that we used here resulted in higher LNT titers for the parent strain (medium-lacY, medium-lgtA, high-CvGalT) than previously reported (24).

Statistics

Statistical significance was calculated using two-tailed unpaired Welch's *t*-tests. Asterisks in Figures indicate a statistically significant difference (∗: p-value < 0.05, ∗∗: p-value < 0.005).

Results

Optimizing tlCRISPRi repression in E. coli

To repress translation in *E. coli*, we need to deliver dCas13 and a CRISPR RNA (crRNA) that targets the mRNA of a gene of interest. We used the catalytically inactive version of RfxCas13d, derived from *Ruminococcus flavefaciens* XPD3002 (8). RfxCas13d (referred to as dCas13 from here) was used as it was shown to have greater knockdown efficiency than Cas13a and Cas13b in mammalian cells (8). We provided crRNAs that either targeted an mRFP fluorescent reporter or were non-targeting in the *E. coli* genome as a control. Although crRNAs targeting the RBS have been shown to be more effective for dCas13-mediated knockdown, many RBS sequences often share

similar sequences (16, 27). Therefore, we designed a crRNA that targets downstream of the Shine-Dalgarno sequence, overlapping with the 3' end of the RBS and the 5' end of the open reading frame (ORF). When expressed without dCas13, the crRNA targeting mRFP gave 1.4-fold repression compared to the parental strain (Figure 1A), an effect which has previously been reported for Cas13 systems (16). This crRNA-dependent effect may be due to direct binding of the mRNA target, forming a dsRNA structure and interfering with mRNA function (28, 29). When dCas13 was expressed with the crRNA, the overall tlCRISPRi effect improved to 2.9-fold (Figure 1A). The crRNAs tested used a DR from a previously described Cas13d isoform from an uncultured *Ruminococcus* strain (*Ur*Cas13) with broad specificity (8). We also tested the cognate DR for RfxCas13d and found improved tlCRISPRi when targeting mRFP for repression (Figure S1). We then tested another dCas13 isoform with its cognate crRNA, dCas13a from *Leptotrichia buccalis* (30). We found that dCas13d was more active, consistent with previous results in mammalian cells (8) (Figure S2). We focused on these Cas13d and Cas13a variants as they are the most well characterized of the available Cas13 variants (31, 32). Together, these results show that dCas13 can be used with a crRNA for programmable translational repression.

To further investigate the effects of the crRNA alone on translation, we measured tlCRISPRi for a fluorescent reporter as a function of crRNA spacer sequence length both in the presence and absence of dCas13. crRNAs were truncated 2 bp at a time from an initial spacer length of 28 nt. Spacers were truncated from the 5' end, although previous work has shown that truncations from either the 5' or 3' end follow similar qualitative trends (12). We found that the crRNA-alone effect results in roughly a 2-fold

repression when crRNA length varied between 16 - 28 nt (Figure 1B), similar to the effects observed in the previous experiment (Figure 1A). With dCas13 present, we observed repression effects ranging from 1.5-fold up to 18-fold, with the strongest repression from longer crRNAs (≥20 nt) (Figure 1B). Together, these results demonstrate that the both dCas13 and the crRNA are required for maximal repression, the natural crRNA length of 26-28 nt is optimal for dCas13-mediated repression, and varying the spacer length from 16-28 nt does not change the crRNA-alone effects.

Having demonstrated simple tlCRISPRi, we sought to further optimize the dCas13 expression levels. Because dCas9 has been reported to be toxic to *E. coli* cells at high expression levels (33, 34), we cloned dCas13 under the anhydrotetracycline (aTc) inducible Tet promoter. aTc titrations showed that *E. coli* is tolerant to relatively high dCas13 concentrations (Figure S3). We found that tlCRISPRi repression improves up to 200 nM aTc induction (Figure 1C). We then cloned dCas13 under weak and medium strength constitutive promoters and observed a similar trend where stronger expression gave greater repression (Figure 1D).

We then compared tlCRISPRi from catalytically-inactive dCas13d to the wild-type, nuclease-active Cas13d. We found that dCas13d exhibits stronger mRFP repression than Cas13d (Figures S3 & S4). Additionally, both Cas13d and dCas13d maintain their tlCRISPRi activity after two rounds of subculturing and induction (Figure S3). To test whether there were any growth burden effects from Cas13 expression, we measured growth kinetics for strains with dCas13d or Cas13d in three conditions: no crRNA, non-targeting crRNA, and on-target crRNA. We found that in the absence of cRNA, both catalytically-active Cas13d and inactive dCas13d exhibit growth defects

relative to the empty vector control. In the presence of a non-targeting crRNA, the strain with Cas13d exhibits a larger growth defect while strains with dCas13d return to normal growth. This behavior may be due to off-target effects that activate collateral RNA cleavage from Cas13d, resulting in a larger growth defect (35, 36). In the presence of an on-target crRNA, both Cas13d and dCas13d show no growth defect. This behavior may be due to the reduced cellular burden when the highly-expressed target transcript is repressed. Overall, dCas13 and Cas13 exhibit similar tlCRISPRi activity. However, dCas13 may still be preferable for targeted translational knockdowns due to its reduced growth burden and lack of collateral activity.

To better understand crRNA targeting rules for tlCRISPRi, we designed four crRNAs targeting different positions of the mRFP transcript. We found that as the distance from the 5' end of the mRNA transcript increased, the tlCRISPRi efficiency decreased (Figure 2A). This observation is in agreement with previous findings that have found targeting the unstructured ribosome binding site (RBS) to be most effective for dCas13-mediated knockdowns (16). This result is also consistent with similar distance-dependent effects in other CRISPR systems (37).

In addition to target position, we anticipated that tlCRISPRi efficiency would depend on the mRNA abundance in the cell. We used a pTet- inducible mRFP reporter to measure tlCRISPRi while titrating aTc, effectively titrating the number of target mRNA copies in the cell. tlCRISPRi is most effective at 2 nM aTc. At higher aTc concentrations and correspondingly higher target mRNA abundance, repression may be limited by the amount of CRISPR complexes in the cell (Figure 2B) (38). At lower aTc concentrations

and correspondingly lower transcript abundance, the expression level approaches the lower limit of detection, which limits the detectable dynamic range of repression.

tlCRISPRi has smaller polar effects than txCRISPRi

While transcriptional control systems act on entire operons, RNA targeting with dCas13 may enable independent regulation of single genes in multi-gene operons. The use of dCas13 may greatly improve our ability to regulate bacterial functions, as a large fraction of genes are contained in multi-gene operons. To test whether dCas13 can independently repress one gene while leaving the other unperturbed, we built a synthetic reporter constitutively expressing sfGFP and mRFP in a single transcriptional unit. The dual reporter operon uses a strong constitutive promoter (J23110) to express the sfGFP ORF followed by the mRFP ORF, each translated from a separate synthetic Bujard RBS (39).

We first tested whether dCas9 repression of the dual reporter operon would produce strong polar effects, in which targeting one gene for repression leads to silencing of both genes (40). As expected, dCas9 repression exhibited strong polar effects when targeting either gene in the operon (Figure 3A). We quantified polar effects by comparing the ratio of repression between the target gene and the other gene in the operon. Targeting the downstream mRFP gene gave 32-fold repression of mRFP and 5.7-fold repression of sfGFP, an overall 5.6-fold greater repression of mRFP than sfGFP (Figure 3A). This reverse polar effect, in which targeting a downstream gene can affect an upstream gene, has been reported before for txCRISPRi in bacteria (40, 41). This effect is likely due to destabilization of the mRNA transcript that leads to improper

folding or reduced half life. Targeting the upstream sfGFP gene gave 15-fold repression of sfGFP and 34-fold repression of mRFP, representing a 2.3-fold weaker repression of sfGFP than the downstream mRFP. The weaker polar effect on sfGFP when downstream mRFP is targeted may be due to incomplete mRNA transcripts that still allow for some translation of the first gene in the operon.

We then used dCas13 to test tlCRISPRi on each gene in the dual reporter operon. dCas13 was expressed from the medium strength J23107 promoter to produce a large tlCRISPRi effect, according to our previous results (Figure 1D). Compared to transcriptional repression with dCas9, we observed reduced polar effects (Figure 3B). Targeting mRFP, the second ORF in the operon, gave 16.3-fold repression of mRFP, but only 1.7-fold repression of sfGFP (Figure 3B). There was a 10-fold greater repression of mRFP than sfGFP, showing that we can specifically target a downstream gene in an operon with modest effects on upstream genes. Targeting sfGFP gave 6.5-fold repression of sfGFP and 2.5-fold repression of mRFP, representing a 2.6-fold greater repression of sfGFP than mRFP. This result indicates that there is a larger polar effect when targeting the upstream gene compared to the downstream gene. Although repression from dCas9 was stronger overall, dCas13 showed 5.9- and 1.7-fold higher specificity repression for sfGFP and mRFP, respectively (Figure S5). For both genes, the observed polar effects with dCas13 were less than that of dCas9 repression.

Multiplexed dCas9/dCas13 tools improve independent regulation of multi-gene operons

We next tested whether we could activate a subset of genes in a multi-gene operon by activating transcription of the entire operon and selectively repressing

translation for undesired genes. As a proof of concept, we envisioned using a previously described dCas9 txCRISPRa system (21, 22) to activate a two-gene operon together with dCas13 as a translational repressor. Briefly, dCas9 txCRISPRa uses a modified scaffold RNA (scRNA) to recruit transcriptional activators upstream of a gene of interest and initiate transcription. We hypothesized dCas13 would repress one gene in the operon while leaving the other activated.

To investigate simultaneous txCRISPRa and tlCRISPRi, we modified the dual sfGFP/mRFP operon used in previous experiments. We replaced the strong constitutive promoter with a weak promoter that could be activated with dCas9. Using the txCRISPRa system alone with dCas9 targeting the promoter and no crRNA for dCas13 targeting, sfGFP and mRFP were activated 10 and 16-fold, respectively (Figure 4A). When a crRNA was introduced to target mRFP with dCas13, we observed that mRFP returned to basal levels while sfGFP remained activated. This result shows that we can specifically activate the first gene in a multi-gene operon by upregulating the entire operon, then using tlCRISPRi to repress downstream genes. However, when we tried to repress sfGFP with dCas13, we observed weak and equal repression for both sfGFP and mRFP (Figure 4A). This result was unexpected because we previously observed preferential sfGFP repression when targeting sfGFP in the constitutively-expressed dual reporter, although by a relatively modest factor of 2.6-fold (Figure 3B). Taken together, our results suggest that independently repressing a downstream gene is effective, but independently repressing an upstream gene remains challenging.

The strong polar effect leading to equal repression on the downstream mRFP gene when targeting the upstream sfGFP could be due to translational coupling of the

two genes (42, 43). There are many examples of translational coupling between genes in natural systems. Translational coupling can occur if ribosomes translating the upstream gene are recruited to the downstream gene, either through stop codon read-through or increased local concentration near the downstream RBS (42, 44, 45). Alternatively, ribosome-mediated mRNA unfolding can disrupt secondary structures that inhibit translation initiation at downstream genes (46–48). Thus, if translation of the downstream gene, mRFP in this case, depends wholly or in part on translation of the upstream gene, polar effects on downstream genes with tlCRISPRi may be difficult to avoid.

To test whether translational coupling could explain the observed polar effects, we constructed modified reporters where the RBS had been removed for either sfGFP or mRFP and repeated the multiplexed txCRISPRa/tlCRISPRi experiment. In the absence of tlCRISPRi, the reporter lacking a RBS for the upstream sfGFP showed greatly reduced mRFP levels (Figure S7) compared to the reporter with the sfGFP RBS (Figure 4A). This result suggests that a significant amount of the observed mRFP expression is from translational coupling to the upstream sfGFP. However, the mRFP decrease could also result from decreased mRNA transcript stability that can occur when the RBS is removed from the 5’ end (49). We found stronger evidence for translational coupling by removing the RBS from the downstream mRFP. With this reporter, we observed that mRFP was still upregulated 20-fold when the operon was activated with dCas9 (Figure S7), again suggesting that there is significant translational coupling from the upstream sfGFP to the downstream mRFP.

To test if independent repression of the upstream gene in an operon is possible, we needed a reporter with minimal translational coupling between genes. Ribosome read-through is thought to be dependent on the rate of translational initiation at the upstream gene (42, 43). We therefore tested an alternative *rc* RBS for the upstream sfGFP that has lower sfGFP expression when activated by txCRISPRa (Figure S6). We designed two crRNAs targeted to this RBS to repress sfGFP translation, *rc-1* and *rc-2*. With these two crRNAs, we observed 3.2- and 11-fold repression of sfGFP, while the effects on the downstream mRFP were 1.6- and 3.6-fold, respectively. These effects correspond to 2.0- or 3.1-fold preferential repression of sfGFP than mRFP (Figure 4B, left), in contrast to equivalent repression of sfGFP and mRFP seen with the original *buj* RBS at sfGFP (Figure 4A). Thus, it is possible to selectively repress the upstream gene, although the preferential effect is relatively modest. Notably, the preferential dCas13 repression effects with this txCRISPRa-activated reporter are similar to the 2.6-fold preferential repression we observed with the original constitutive reporter (Figure 3B). In the constitutive reporter, the initial sfGFP and mRFP levels were similar to the levels produced by txCRISPRa-mediated activation with the *rc* RBS (Figure 4B, left). Thus, the extent of translational coupling may be dependent on the absolute expression level (43), which may consequently affect the possibility of independently repressing the first gene in an operon.

We then assessed whether the sfGFP mRNA sequence context could be contributing to the observed translational coupling and polar effects. We replaced the sfGFP reporter with mBFP (mTagBFP2, blue fluorescent protein) under the same *rc* RBS. Qualitatively, we observed similar effects with either mBFP or sfGFP in the

upstream position. Targeting the downstream mRFP with dCas13 strongly silences mRFP by ~35-fold with only a 1.1-fold effect on mBFP (Figure 4B, right). Targeting the RBS of the upstream mBFP with the *rc-1* and *rc-2* crRNAs produced 2.4- and 6.7-fold repression of mBFP, while the effects on the downstream mRFP were 1.1- and 1.6-fold, respectively. Overall, mRFP remained activated 3.8- and 2.7-fold relative to the dCas9 non-targeting gRNA. These effects correspond to 2.2- and 4.2-fold higher repression of mBFP than the downstream mRFP, respectively (Figure 4B, right). These effects are qualitatively similar to those observed with sfGFP in the upstream position (2.0- and 3.1-fold, respectively) but the quantitative differences suggest that transcript sequence context can contribute to polar effects.

The experiments described above were performed with the broad specificity crRNA DR (Figure 1). We proceeded to test whether the cognate dCas13Rfx crRNA DR (Figure S1) performs similarly for individual gene targeting with txCRISPRa/tlCRISPRi. We observed qualitatively similar results with the cognate crRNA, but modestly improved individual gene targeting with reduced polar effects on the neighboring gene in the operon (Figure 4C).

In *E. coli*, roughly a third of all genes are contained in operons with overlapping ORFs (50). To test the ability of our system to independently regulate genes in overlapping operons, we created a modified synthetic reporter that contains the C terminus of *trpA* and N terminus of *trpB* fused between the ORFs for BFP and RFP. *trpA-trpB* is well-studied as a model system for translational coupling due to its overlapping stop and start codons (48, 51). We found that there were strong polar effects between the two genes when targeting either for repression (Figure S8),

indicating that it is difficult to target individual genes with dCas13 in tightly-coupled overlapping operons. Engineering approaches to overcome tight translational coupling with dCas13 will require further investigation.

One potential disadvantage of the combined txCRISPRa/tlCRISPRi approach is that there may be a growth burden associated with expressing two CRISPR systems. To test whether the cell could support the two systems operating simultaneously, we measured the growth kinetics of a cell expressing either the txCRISPRa, tlCRISPRi, or both txCRISPRa/tlCRISPRi machineries. We found that the expression of dCas13 and dCas9 together in the cell imposed no detectable additional burden to the cell compared to expressing the CRISPRa machinery alone (Figure S9). Both dCas9 and dCas13 were expressed from the medium-strength J23107 minimal promoter. We attempted to build constructs with Cas proteins under the strong J23119 promoter, but were unable to successfully clone these constructs, likely due to high metabolic burden. There is also a well established precedent that strong heterologous protein expression can be burdensome and deleterious to cell growth We therefore did not further pursue cloning of Cas proteins with strong promoters. Regardless, these results show that we can express dCas9 and dCas13 together at relevant concentrations for effective txCRISPRa/tlCRISPRi.

Multiplexed dCas9/dCas13 metabolic engineering improves LNT bioproduction

To test whether a multiplexed dCas9 and dCas13 system could be effective for metabolic engineering, we used bacterial lacto-*N*-tetraose (LNT) production as a model system. LNT is a human milk oligosaccharide (HMO) that is important for infant immune

development and microbiome health (57, 58). To produce LNT in *E. coli*, two heterologous enzymes, LgtA and *Cv*GalT, can be combined with overexpression of the LacY lactose permease (59, 60). LacY imports lactose into the cell, where LgtA, a β-1,3-*N*-acetylglucosaminyltransferase from *N. meningitidis,* produces the intermediate lacto-*N*-triose II (LNT II). *Cv*GalT is a β-1,3-galactosyltransferase from *C. violaceum* that converts LNT II to the final product LNT. Knocking out endogenous β-galactosidase activity (encoded by the *lacZ* gene) is also necessary to prevent cleavage of the lactose feedstock into its constituent monosaccharides glucose and galactose (60).

Because LNT production requires up- and down-regulation of multiple enzymes, CRISPR-Cas-mediated gene regulation can be useful for tuning and optimizing gene expression levels. We previously used a combinatorial txCRISPRa library to optimize pathway enzyme levels in a *lacZ* knockout strain (Figure 5A) (24). Here, we attempted to further optimize the system by using CRISPRi to repress endogenous genes that divert flux away from LNT. Specifically, we targeted the *ugd, wecB,* and *nagB* genes. The enzymes from these genes consume the metabolites UDP-*N*-acetylglucosamine and UDP-galactose, which are used by LgtA and *Cv*GalT during LNT production. Previous work has shown that knocking these genes out can improve LNT production by factors ranging from 20 - 350% (25, 26, 61). However, when we activated the LNT pathway enzymes with txCRISPRa and repressed *ugd, wecB,* or *nagB* using txCRISPRi, we observed that the knockdowns produced no increase in LNT production relative to the strain with txCRISPRa alone (Figure 5B). Furthermore, the nagB knockdown resulted in a significant decrease in LNT production *($p < .05$).* We then tested two additional guides per gene in case we were not targeting optimal positions

and therefore not able to effectively knockdown gene expression. Still, we found that txCRISPRi resulted in no differential LNT production relative to the txCRISPRa-only strain (Figure S10). One possible explanation for the lack of improved LNT production from these CRISPRi targets is that the expression of additional gRNAs leads to competitive effects, in which gRNAs with varying binding affinities are competing for a limited pool of dCas9 (62, 63). However, the LNT titers are similar between the conditions with no CRISPRi gRNA and the non-targeting gRNA. This result suggests that competition from additional gRNAs is not significantly weakening the activation from dCas9 in this heterologous pathway.

We next tested whether dCas13-mediated tlCRISPRi knockdowns could improve LNT production. We targeted *lacZ* for knockdown in an *E. coli* strain containing the native copy to determine if we could produce a sufficient knockdown for LNT production. We compared *lacZ* knockdowns from dCas13 and three different dCas9 targets to the *lacZ* knockout strain. We found that the dCas13 knockdown gave 4.5 ± 0.5 mM LNT compared to 1.8 ± 0.8 mM, 4.2 ± 0.5 mM, or 2.5 ± 0.2 mM from the dCas9 knockdowns (Figure 5C). These results show that both tlCRISPRi and txCRISPRi can sufficiently repress *lacZ* to enable LNT production. However, all knockdowns produced less LNT than the 8.6 ± 0.7 mM produced from the *lacZ* knockout (Figure 5B). This result could be due to the CRISPR perturbations not producing a strong enough knockdown, or from increased cellular burden from the operation of the CRISPR systems. To test whether the knockdown strength was limiting, we multiplexed dCas13 and dCas9 knockdowns of *lacZ* and observed 6.6 ± 1.3 mM of LNT production (Figure 5C). This result suggests

that the knockdowns from each system were not as strong as the knockout, and that greater knockdowns can be achieved by combining perturbations.

One potential advantage of CRISPR regulation is the opportunity for rapid combinatorial multiplexing. In the strain containing the dCas13 *lacZ* crRNA, we introduced a second crRNA targeting either *ugd, nagB,* or *wecB*. The second crRNA was expressed from a separate transcriptional unit on the same plasmid as the other gRNAs (Table S2). While dCas9 knockdowns of these targets all gave lower or similar levels of LNT, the dCas13 knockdown of *ugd* gave a 3-fold improvement in LNT production, from 4.5 ± 0.5 mM to 14 ± 2.2 mM (Figure 5D) ($p < .005$). The combined *lacZ* and *ugd* dCas13-mediated knockdown strain was the highest producing strain, yielding a 46% improvement over the *lacZ* knockout strain with no translational regulation. In contrast, the wecB knockdown significantly decreased LNT production ($p < .05$) and the nagB knockdown gave a small increase in LNT production that was not statistically significant ($p = .07$). These results highlight the utility of combinatorial multiplexing with programmable CRISPR-Cas systems, and the use of dCas13 enabled the identification of a gene expression program with substantial improvements in bioproduction.

Discussion

CRISPR-Cas technologies have enabled a range of applications in bacterial systems, including gene editing, biosensing, and bioproduction (3, 64–67). The ability to combine orthogonal CRISPR-Cas systems with different regulatory properties can expand the possibilities for precise gene regulation and information processing.

Additionally, combining orthogonal CRISPR-Cas systems may overcome limitations of gRNA competition to allow regulation of more genes in the genome simultaneously (38, 63, 68). The ability to increase the upper limit of regulatable gene targets in a single metabolic engineering program will be an important consideration for future efforts (69).

Here, we show that multiplexed DNA- and RNA-targeting dCas9 and dCas13 CRISPR-Cas systems can improve bioproduction in bacteria. These results emphasize that different methods of regulation can yield different and difficult to predict effects when regulating endogenous genes. The use of dCas13, either alone or in addition to dCas9, may offer several advantages for metabolic engineering. Previous HMO metabolic engineering efforts have shown *lacZ* knockout to be critical for effective production (25, 26, 61). Here, we showed that multiplexed knockdowns of *lacZ* with dCas9 and dCas13 improved LNT production relative to either dCas9 or dCas13 alone. Multiplexed knockdowns of additional genes can yield further improvements to LNT production to 14 mM (9.9 g/L). To our knowledge, this titer is the highest achieved to date in shake flask conditions (25, 26, 59). One possible explanation for these results is that perturbations with dCas13 may lead to larger changes in protein levels than with dCas9, especially for genes that are strongly regulated at the transcriptional level. For example, the expression of *ugd* is regulated by several independent transcriptional activators, which could lessen the effect that a transcriptional knockdown will have on protein levels (70).

Many targets for knockdown may not produce the desired effects when repressed with dCas9 due to tight regulation at the transcriptional level. For example, *nagB* is known to be highly regulated due to its relevance for sugar carbon utilization

(72, 73). When targeted with dCas9, the *nagB* knockdown resulted in a significantly lower LNT production ($p < .05$) than the other knockdown strains (Figure 5B). We then showed when *nagB* was targeted with dCas13 that LNT production increased from 4.5 to 5.7 mM (Figure 5D) ($p = .07$), suggesting that dCas13 achieved a more effective knockdown than dCas9. dCas13 may also provide an advantage over dCas9 when targeting genes in the same operon as an essential gene. In the case of the *nagBACD* operon, *nagC* encodes for a transcriptional repressor of the operon. It is possible that transcriptional repression of *nagB* leads to polar effects on *nagC* expression, resulting in de-repression of the whole operon. We also saw that repressing *wecB* did not improve LNT titers when targeted with either dCas9 or dCas13. This result may indicate that *wecB* is not an important target to control metabolic flux towards LNT, as suggested by others (26), or that there are other regulatory effects that compensate for the knockdown. Systematic comparisons of dCas9 and dCas13 knockdowns may uncover other contexts in which translational control is preferred over transcriptional regulation.

Operons are often organized with colocalized genes acting in different steps in a shared biosynthetic pathway. The ability to independently regulate genes in such operons could be particularly useful for bioproduction if the target product requires diverting a pathway intermediate. For example, regulation of the *aroF-tyrA* operon within *E. coli* has received significant attention for production of value-added chorismate derivatives and aromatics (46–49). *aroF* is involved in the chorismate biosynthesis pathway and its upregulation could improve flux towards high-value aromatic compounds. However, *tyrA* diverts chorismate flux towards biosynthesis of aromatic amino acids, making it a potential target for downregulation. The ability to activate *aroF*

while maintaining or repressing *tyrA* levels could improve bioproduction of chorismate derivatives. Future work will explore whether the use of dCas9 and dCas13 to tune individual genes in endogenous operons like *aroF-tyrA* can have significant effects on bioproduction.

We have also shown that dCas9 and dCas13 can be combined to overcome polar effects in multi-gene transcriptional operons. The ability to independently regulate individual genes in a multi-gene operon depends on sequence context, expression level, and target position within the operon, but we have shown that independent gene regulation can be improved through screening different crRNA targets (Figure 4B). Genome-wide tlCRISPRi screens could help to further elucidate sequence and context dependent rules that determine the polar effects at multi-gene operons.

The ability to target individual genes may also be affected by the genetic structure of the operon. Previous work has shown that the distances between the stop codon of the upstream gene and the start codon of the downstream gene affect the degree of translational coupling. Distances less than the footprint of a ribosome (25-30 nt) are thought to promote translational coupling. Our expectation is that as the degree of coupling decreases, the ability to independently repress genes with dCas13 would improve. Therefore, the current tlCRISPRi toolbox will be most effective at targeting single genes in multi-gene operons when there are larger spacings between the ORFs.

To date, there have been few approaches reported for decoupling translation in a multi-gene operon. The use of mutations to introduce premature stop codons or remove Shine-Dalgarno motifs in the upstream gene has been shown to decouple translation in *E. coli* (71). However, these perturbations typically resulted in the loss of expression of

both genes in the operon, limiting their applicability for metabolic engineering. Other work has attempted to reduce translational coupling by knocking down the ribosome recycling factor gene, but this intervention led to less overall translation from mRNA transcripts (45). Our work provides an alternative approach using dCas13 to independently regulate gene expression in multi-gene operons without extensive genome engineering.

Further investigation into Cas13 systems has the potential to improve our ability to precisely regulate individual genes at the translational level. In this work, we show that gene regulation with dCas13d provides novel functional properties with immediate applications in metabolic engineering. Several other Cas13 isoforms have been described that could also be effective for translational regulation (11, 74, 75). Dynamic control of dCas13 activity could enable improved biosynthesis, molecular recording, biosensing, and biotherapeutic applications (76–79). Such a system could be constructed by regulating dCas13 expression with an inducible or conditionally-responsive promoter. However, because there is weak basal repression from the crRNA alone (Figure 1B), tight control of crRNA expression may be necessary for maximal dynamic range.

The ability to activate gene expression using dCas13 could also be useful for a wide range of bacterial engineering applications. Our results here suggest that dCas13 mediated activation (tlCRISPRa) could offer advantages over DNA-targeting dCas9 (txCRISPRa) by enabling control of individual genes in multi-gene operons. txCRISPRa is also subject to relatively stringent target requirements that have limited its generalizability to arbitrary gene targets (22, 56, 80). If tlCRISPRa with dCas13 is

subject to less stringent or even simply different target requirements, it will expand the range of accessible gene targets in bacteria. Previous work in *E. coli* has shown that dCas13 can be fused to an effector domain for translational activation (tlCRISPRa) (16). Still, robust tlCRISPRa on arbitrary mRNA targets remains to be demonstrated (16, 19).

Moving forward, the ability to use dCas13 to independently regulate individual genes across bacterial genomes, along with multiplexed regulation using dCas9, may greatly expand the possibilities for basic biology research and applied biotechnology. The key advance in this work is that dCas9 and dCas13 can be used together to achieve functional effects that would be inaccessible using either system alone. We first demonstrate the advantage of combined dCas9/dCas13 through their multiplexed use to independently regulate individual genes in multi-gene operons. Many genes are subject to complex transcriptional regulation, and multi-gene operons are ubiquitous across nature. However, tools to study tightly regulated genes or individual genes within these operons are limited. Therefore, we hypothesize that regulating at the translational level may enable improved control of individual protein levels compared to transcriptional regulation.

Genome-wide dCas13 screens may reveal novel insights into phenotypes from individual ORF perturbations that would be missed from dCas9 knockdown screens that act at the operon level. dCas13 has already proven a useful tool for studying gene essentiality in mammalian and bacterial systems (12, 36). Improved control over multi-gene operons will expand our understanding of individual gene functions. In addition to essentiality studies, genome-wide tlCRISPRi screens can inform strain design for a variety of applications.

An outstanding challenge in synthetic biology is identifying the genetic perturbations that can optimize biosynthetic programs. By implementing dCas13-based translational regulation in bacteria and demonstrating the capabilities of combined translational and transcriptional control, we provide new opportunities to reprogram gene expression in bacteria. Systematic combinatorial perturbations with dCas9 and dCas13 should reveal effective strategies to regulate bacterial systems. These approaches may prove particularly useful for exploring the functional space available to medically- and industrially-relevant microbes.

Figures

Figure 1. tlCRISPRi represses gene expression in E. coli

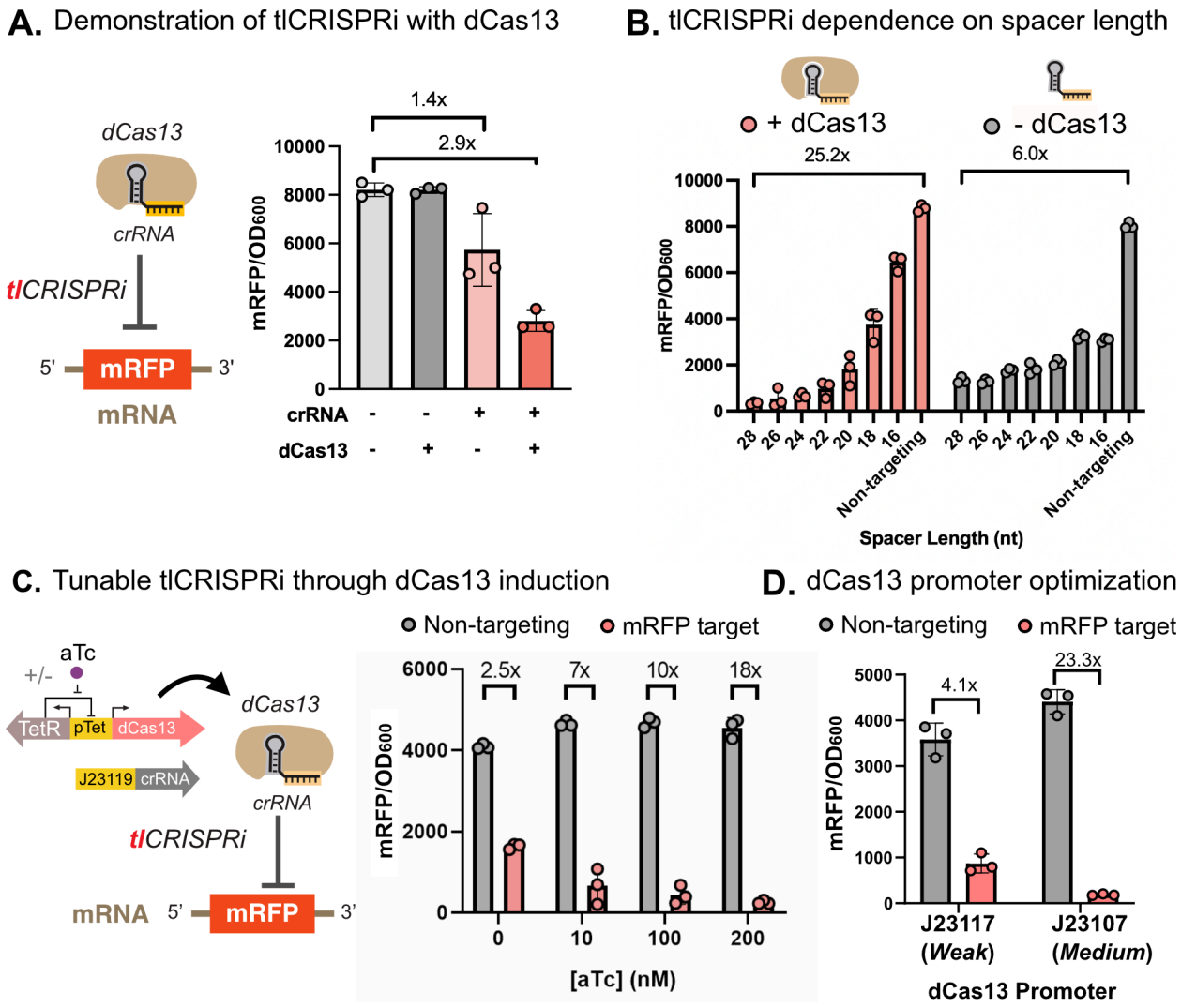


Figure 1. tlCRISPRi represses gene expression in *E. coli*.

A) Effect on reporter gene expression with dCas13 alone, crRNA alone, or both dCas13 and crRNA.The crRNA *rfp2* targets the ORF of the mRFP mRNA transcript (Table S3).

B) Reporter gene expression versus crRNA spacer length. The spacer is the region complementary to the target mRNA. Reporter gene expression levels were measured both in the presence and absence of dCas13. The *rfp2* crRNA was tested with spacer lengths from 16-28 by truncating 2 nt at a time from the 5' end.

C) Tuning tlCRISPRi activity by titrating dCas13 expression levels. dCas13 was expressed with an inducible pTet promoter and the crRNA was expressed from the constitutive J23119 promoter. Reporter gene expression levels were measured with 0, 10, 100 and 200 nM aTc concentrations.

D) tlCRISPRi with dCas13 expression from weak (J23117) or medium (J23107) strength constitutive promoters.

For all panels, values represent the mean ± standard deviation for at least three biological replicates.

Figure 2. tICRISPRi dependence on target position and abundance

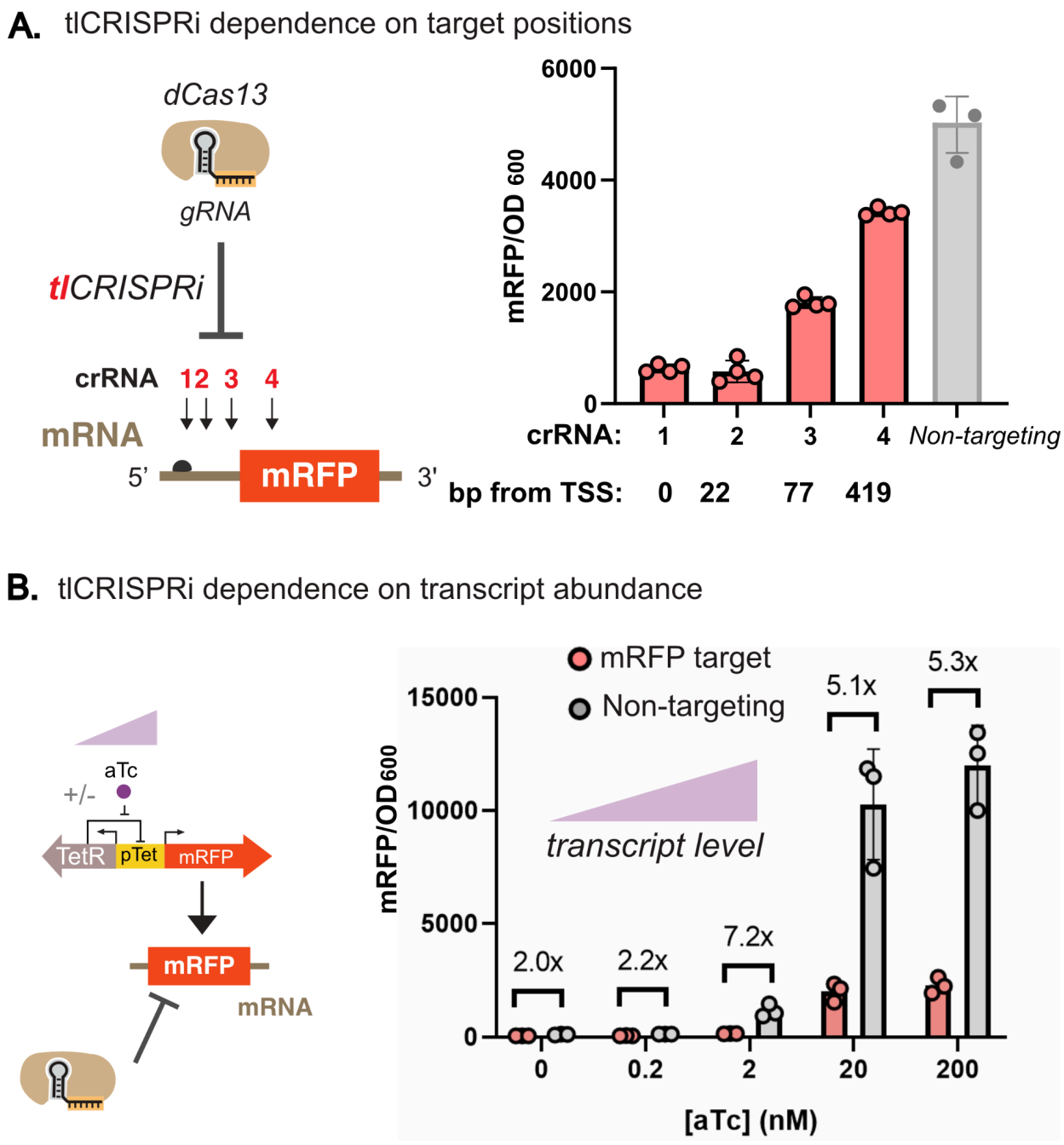


Figure 2. tICRISPRi dependence on mRNA target site position and transcript abundance.

A) Reporter gene expression versus tICRISPRi target position on the mRNA. *buj1* targets the unstructured RBS. *rfp2* targets the 3' end of the RBS and beginning of the

ORF. *rfp3* and *rfp4* target the middle of the ORF. The non-targeting control targets the template strand, which is not transcribed into mRNA.

B) tICRISPRi effects on reporter gene expression versus target mRNA transcript expression level. mRFP was expressed from the pTet promoter and induced with 0, 0.2, 2, 20, and 200 mM aTc. dCas13 and crRNA were constitutively expressed.

For all panels, values represent the mean ± standard deviation for at least three biological replicates.

Figure 3. CRISPRi polar effects for multi-gene operons

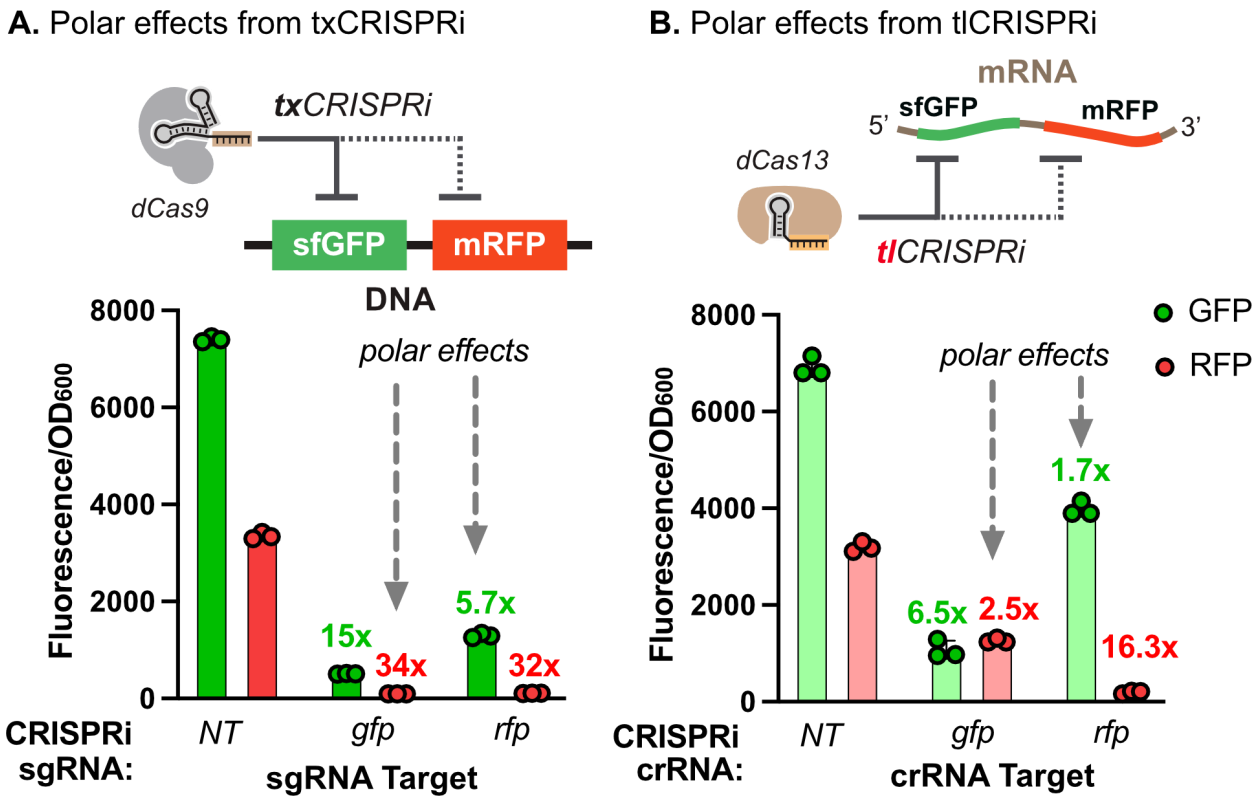


Figure 3. CRISPRi polar effects for multi-gene operons.

A) Polar effects from txCRISPRi with dCas9-mediated targeting of DNA to repress transcription. A synthetic multi-gene sfGFP-mRFP reporter transcript was expressed from the J3_J23110 promoter (see Methods). Each gene has a separate copy of the Bujard RBS. Polar effects on reporter gene expression were measured by comparing the ratio of repression between the target gene and the other gene in the operon. Fluorescent reporter levels were measured with dCas9 sgRNAs for sfGFP (*NT1)*, mRFP (*rr2)*, or a non-targeting (NT) (*AAV)* sequence (Table S3). dCas9 and dCas13 were both expressed from the J23107 promoter (Table S2).

B) Polar effects from tlCRISPRi with dCas13-mediated targeting of mRNA to repress translation. The reporter gene was the same as described above in (A). Fluorescent reporter levels were measured with dCas13 crRNAs for sfGFP (*gfp2*), mRFP (*rfp2)*, or a non-targeting (NT) (*AAV)* sequence (Table S3).

For all panels, values represent the mean ± standard deviation for at least three biological replicates.

Figure 4. Multiplexed txCRISPRa/tlCRISPRi in multi-gene operons

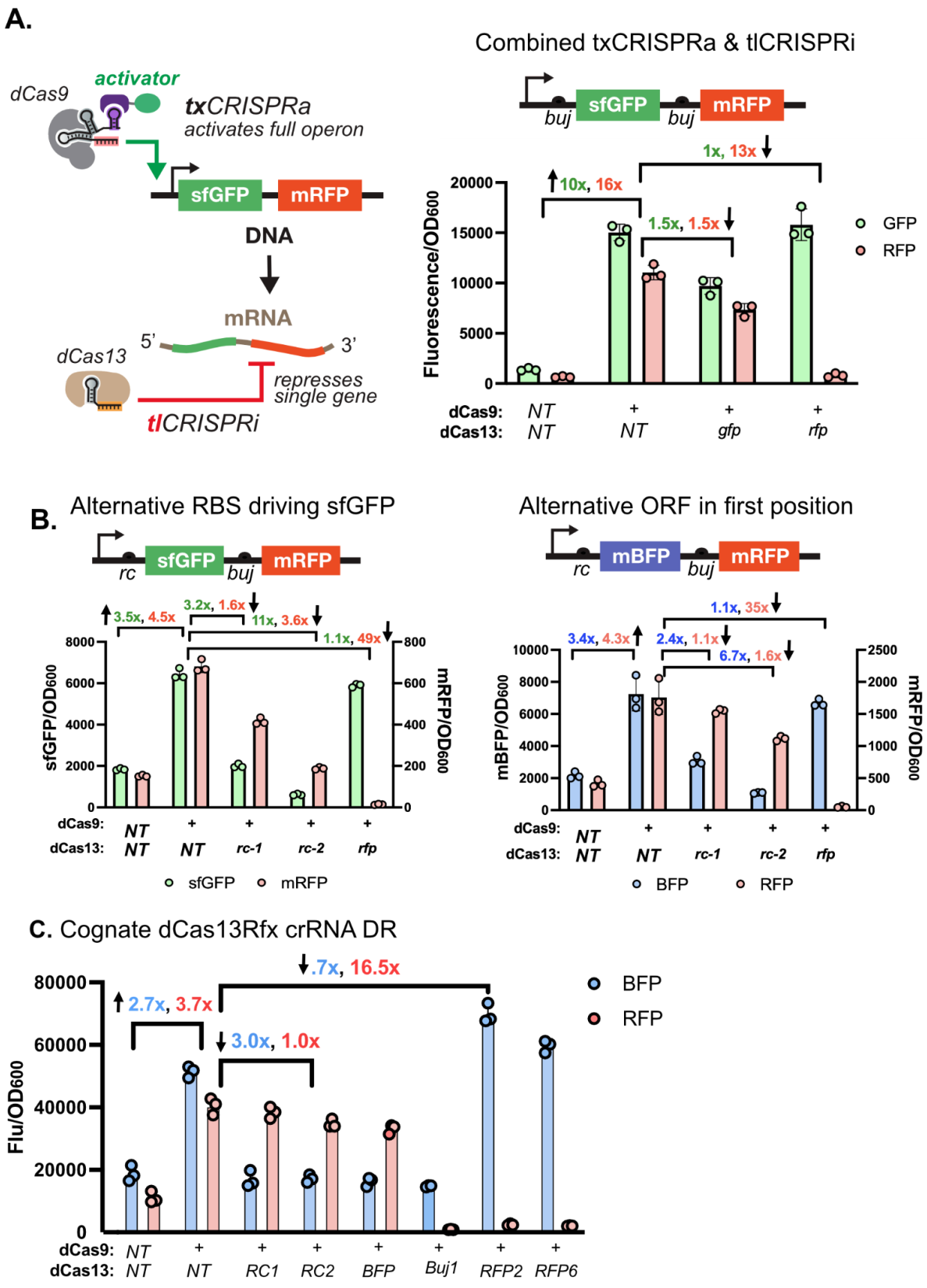


Figure 4. Multiplexed txCRISPRa/tlCRISPRi in multi-gene operons

A) Combined txCRISPRa and tlCRISPRi for independent gene activation in a multi-gene operon. For txCRISPRa, dCas9 targets the -81 position of the J3_J23117 promoter and recruits a transcriptional activator to activate the operon (22). For tlCRISPRi, dCas13 targets an individual gene with a crRNA targeting either sfGFP or mRFP. The dCas13 non-targeting (NT) sample indicates the fold activation from txCRISPRa without any tlCRISPRi repression. Fold repression from tlCRISPRi compares the dCas13 non-targeting and on-target crRNAs, always with the on-target J306 scRNA for txCRISPRa.

B) Combined txCRISPRa/tlCRISPRi as described in (A) using the *rc* RBS at the first gene in the operon instead of the Bujard RBS. The first gene in the operon is either sfGFP (left) or mBFP (right). Two new crRNA sequences were designed to target the *rc* RBS (Table S3).

C) Combined txCRISPRa/tlCRISPRi as described in (A) using the *Rfx*Cas13 cognate DR. The synthetic multi-gene BFP-RFP reporter is the same as described in Figure 4B (right). Two new crRNA sequences were designed to target the BFP and RFP ORFs (Table S3).

For all panels, values represent the mean ± standard deviation for at least three biological replicates.

Figure 5. Multiplexed txCRISPRa/tlCRISPRi for improved metabolic engineering

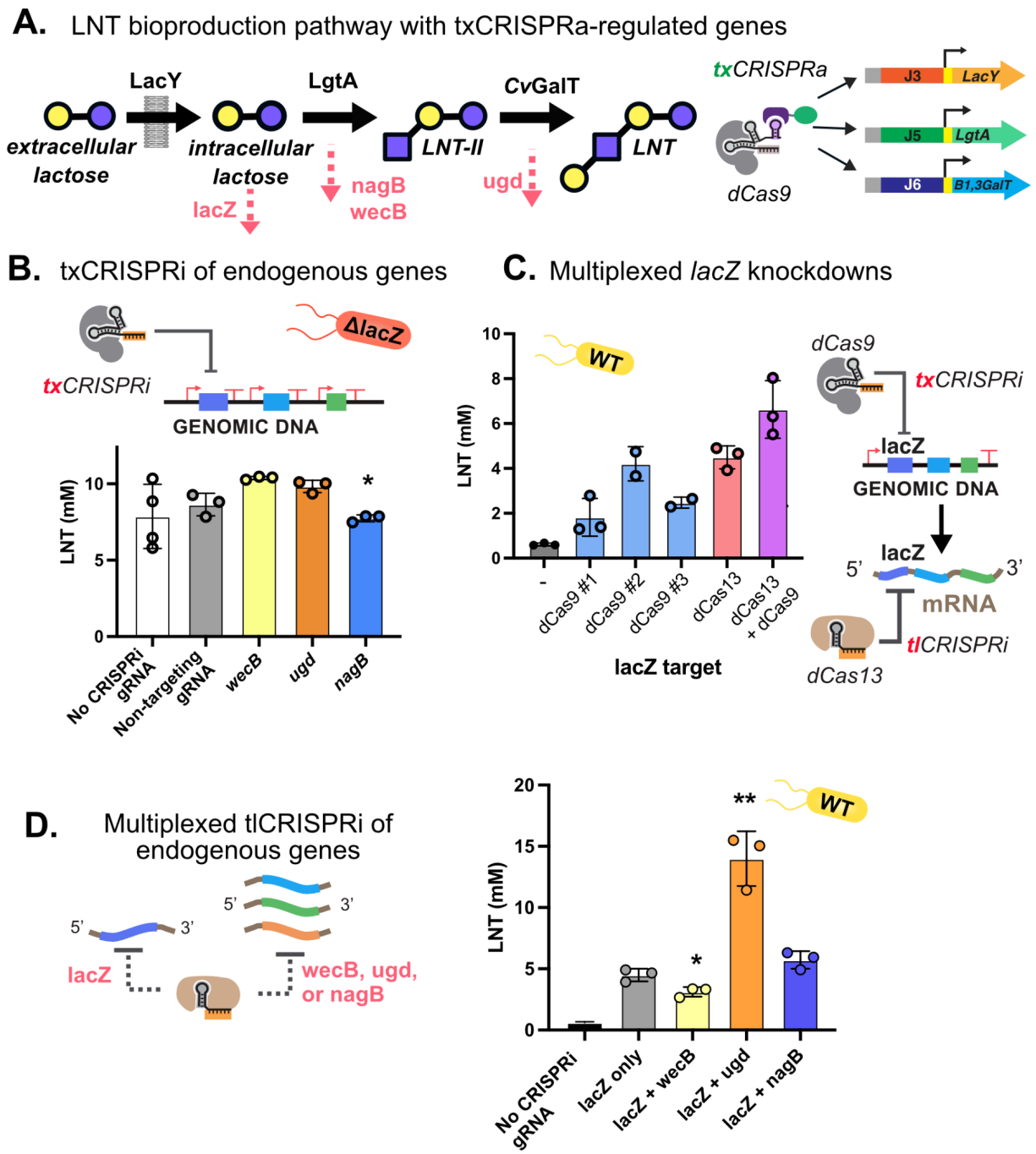


Figure 5. Multiplexed txCRISPRa/tlCRISPRi for improved metabolic engineering
A) Pathway schematic for LNT bioproduction in *E. coli* with CRISPRa-regulated genes. LNT is produced from lactose using three enzymes expressed from a two-plasmid system. The heterologous pathway is weakly expressed from synthetic promoters and activated using txCRISPRa. The pathway consists of *lacY* from *E. coli*, *lgtA* from

Neisseria meningitidis, and *B3GalT* from *Chromobacterium violaceum.* Endogenous genes that divert flux away from the LNT pathway, including *wecB, ugd,* and *nagB*, are shown in red. These genes are targets for repression to increase LNT flux.

B) LNT titers with different endogenous genes targeted for repression with txCRISPRi. All conditions contain the txCRISPRa program activating the heterologous LNT production pathway. Different txCRISPRi constructs were delivered to a parental starting strain containing a medium-*lacY*, medium-*lgtA*, high-*CvGalT* txCRISPRa program (JM109 cells containing the pJF234.21 and pIDF96C plasmids) as a starting point (24). The JM109 *E. coli* strain contains a *lacZ* deletion to prevent lactose metabolism. Asterisks in this panel indicate a statistically significant difference compared to no gRNA (∗: p-value < 0.05).

C) LNT titers with txCRISPRi and tlCRISPRi individual and multiplexed knockdowns of *lacZ.* The multiplexed knockdown uses the dCas9 #1 and dCas13 gRNAs (Table S3). Experiments were conducted in the CK024 *E. coli* strain, which contains wild type (WT), unmodified *lacZ* and genome integrated dCas9 (Table S1).

D) LNT titers for multiplexed dCas13 knockdowns of *lacZ* and other endogenous gene targets. Experiments were conducted in the CK024 *E. coli* strain (Table S1). Asterisks in this panel indicate a statistically significant difference compared to *lacZ* only (∗: p-value < 0.05, ∗∗: p-value < 0.005).

For all panels, values represent the mean ± standard deviation for at least two biological replicates.

Acknowledgements

We thank members of the Carothers and Zalatan groups, especially Cholpisit (Ice) Kiattisewee for advice, materials, and comments on the manuscript.

Competing interests

J.G.Z. and J.M.C are members of the Wayfinder Biosciences scientific advisory board.

Data sharing plans

All data for this work are available in the Supplementary Information file.

Funding information

This work was supported by US National Science Foundation (NSF) Award MCB 2225632 (J.M.C. and J.G.Z.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

1. 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.
2. 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.
3. 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.
4. Mateus, A., Shah, M., Hevler, J., Kurzawa, N., Bobonis, J., Typas, A. and Savitski, M.M. (2021) Transcriptional and Post-Transcriptional Polar Effects in Bacterial Gene Deletion Libraries. *mSystems*, **6**, e00813-21.
5. Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong, W.K., Paley, S., Subhraveti, P. and Karp, P.D. (2020) The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.*, **48**, D445–D453.
6. Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L., García-Sotelo, J.S., Alquicira-Hernández, K., Martínez-Flores, I., Pannier, L., Castro-Mondragón, J.A., *et al.* (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.*, **44**, D133-143.
7. Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A., *et al.* (2017) RNA targeting with CRISPR–Cas13. *Nature*, **550**, 280–284.
8. Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N. and Hsu, P.D. (2018) Transcriptome engineering with RNA-targeting Type VI-D CRISPR effectors. *Cell*, **173**, 665-676.e14.
9. Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J. and Zhang, F. (2017) RNA Editing with CRISPR-Cas13. *Science*, **358**,

1019–1027.

10. O'Connell, M.R. (2019) Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. *J. Mol. Biol.*, **431**, 66–87.
11. Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J., Verdine, V., Donghia, N., Daringer, N.M., Freije, C.A., *et al.* (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, **356**, 438–442.
12. Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z. and Sanjana, N.E. (2020) Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.*, **38**, 722–727.
13. Abudayyeh, O.O., Gootenberg, J.S., Franklin, B., Koob, J., Kellner, M.J., Ladha, A., Joung, J., Kirchgatterer, P., Cox, D.B.T. and Zhang, F. (2019) A cytosine deaminase for programmable single-base RNA editing. *Science*, **365**, 382–386.
14. Du, M., Jillette, N., Zhu, J.J., Li, S. and Cheng, A.W. (2020) CRISPR artificial splicing factors. *Nat. Commun.*, **11**, 2973.
15. Núñez-Álvarez, Y., Espie--Caullet, T. and Luco, R.F. (2022) "A CRISPR-dCas13 RNA-editing tool to study alternative splicing". 10.1101/2022.05.24.493209.
16. Otoupal, P.B., Cress, B.F., Doudna, J.A. and Schoeniger, J.S. (2022) CRISPR-RNAa: targeted activation of translation using dCas13 fusions to translation initiation factors. *Nucleic Acids Res.*, 10.1093/nar/gkac680.
17. Charles, E.J., Kim, S.E., Knott, G.J., Smock, D., Doudna, J. and Savage, D.F. (2021) Engineering improved Cas13 effectors for targeted post-transcriptional regulation of gene expression. 10.1101/2021.05.26.445687.
18. Zhang, K., Zhang, Z., Kang, J., Chen, J., Liu, J., Gao, N., Fan, L., Zheng, P., Wang, Y. and Sun, J. (2020) CRISPR/Cas13d-Mediated Microbial RNA Knockdown. *Front. Bioeng. Biotechnol.*, **8**.
19. Montagud-Martínez, R., Márquez-Costa, R. and Rodrigo, G. (2023) Programmable regulation of translation by harnessing the CRISPR-Cas13 system. *Chem. Commun.*, **59**, 2616–2619.
20. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene*, **33**, 103–119.

21. 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**, 2489.
22. 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.
23. Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de Las Heras, A., Páez-Espino, A.D., Durante-Rodríguez, G., Kim, J., Nikel, P.I., *et al.* (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.*, **41**, D666-675.
24. Fontana, J., Sparkman-Yager, D., Faulkner, I., Cardiff, R., Kiattisewee, C., Walls, A., Primo, T.G., Kinnunen, P.C., Martin, H.G., Zalatan, J.G., *et al.* (2023) Guide RNA structure design enables combinatorial CRISPRa programs for biosynthetic profiling. 10.1101/2023.11.17.567465.
25. Hu, M., Li, M., Miao, M. and Zhang, T. (2022) Engineering *Escherichia coli* for the High-Titer Biosynthesis of Lacto-N-tetraose. *J. Agric. Food Chem.*, **70**, 8704–8712.
26. Zhu, Y., Wan, L., Meng, J., Luo, G., Chen, G., Wu, H., Zhang, W. and Mu, W. (2021) Metabolic Engineering of *Escherichia coli* for Lacto-N-triose II Production with High Productivity. *J. Agric. Food Chem.*, **69**, 3702–3711.
27. Scherer, G.F., Walkinshaw, M.D., Arnott, S. and Morré, D.J. (1980) The ribosome binding sites recognized by *E. coli* ribosomes have regions with signal character in both the leader and protein coding segments. *Nucleic Acids Res.*, **8**, 3895–3907.
28. Kim, D.-H., Behlke, M.A., Rose, S.D., Chang, M.-S., Choi, S. and Rossi, J.J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.*, **23**, 222–226.
29. Kurreck, J. (2006) siRNA Efficiency: Structure or Sequence—That Is the Question. *J. Biomed. Biotechnol.*, **2006**, 83757.

30. Tambe, A., East-Seletsky, A., Knott, G.J., Doudna, J.A. and O'Connell, M.R. (2018) RNA-binding and HEPN-nuclease activation are decoupled in CRISPR-Cas13a. *Cell Rep.*, **24**, 1025–1036.
31. Xu, C., Zhou, Y., Xiao, Q., He, B., Geng, G., Wang, Z., Cao, B., Dong, X., Bai, W., Wang, Y., *et al.* (2021) Programmable RNA editing with compact CRISPR–Cas13 systems from uncultivated microbes. *Nat. Methods*, **18**, 499–506.
32. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., *et al.* (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
33. Zhang, S. and Voigt, C.A. (2018) Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Res.*, **46**, 11115–11125.
34. Cho, S., Choe, D., Lee, E., Kim, S.C., Palsson, B. and Cho, B.-K. (2018) High-Level dCas9 Expression Induces Abnormal Cell Morphology in *Escherichia coli*. *ACS Synth. Biol.*, **7**, 1085–1094.
35. Bot, J.F., van der Oost, J. and Geijsen, N. (2022) The double life of CRISPR–Cas13. *Curr. Opin. Biotechnol.*, **78**, 102789.
36. Adler, B.A., Hessler, T., Cress, B.F., Lahiri, A., Mutalik, V.K., Barrangou, R., Banfield, J. and Doudna, J.A. (2022) Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. *Nat. Microbiol.*, **7**, 1967–1979.
37. Radzisheuskaya, A., Shlyueva, D., Müller, I. and Helin, K. (2016) Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Res.*, **44**, e141.
38. 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 Synthetic Biology.
39. Lutz, R. and Bujard, H. (1997) Independent and Tight Regulation of Transcriptional Units in *Escherichia Coli* Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Res.*, **25**, 1203–1210.

40. Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.-M., Marta, E., *et al.* (2016) A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell*, **165**, 1493–1506.
41. Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V. and Bikard, D. (2018) A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.*, **9**, 1912.
42. Levin-Karp, A., Barenholz, U., Bareia, T., Dayagi, M., Zelcbuch, L., Antonovsky, N., Noor, E. and Milo, R. (2013) Quantifying Translational Coupling in *E. coli* Synthetic Operons Using RBS Modulation and Fluorescent Reporters. *ACS Synth. Biol.*, **2**, 327–336.
43. Tian, T. and Salis, H.M. (2015) A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Res.*, **43**, 7137–7151.
44. Schümperli, D., McKenney, K., Sobieski, D.A. and Rosenberg, M. (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell*, **30**, 865–871.
45. Saito, K., Green, R. and Buskirk, A.R. (2020) Ribosome recycling is not critical for translational coupling in *Escherichia coli*. *eLife*, **9**, e59974.
46. Rex, G., Surin, B., Besse, G., Schneppe, B. and McCarthy, J.E. (1994) The mechanism of translational coupling in *Escherichia coli.* Higher order structure in the atpHA mRNA acts as a conformational switch regulating the access of de novo initiating ribosomes. *J. Biol. Chem.*, **269**, 18118–18127.
47. Little, S., Hyde, S., Campbell, C.J., Lilley, R.J. and Robinson, M.K. (1989) Translational coupling in the threonine operon of *Escherichia coli* K-12. *J. Bacteriol.*, **171**, 3518–3522.
48. Oppenheim, D.S. and Yanofsky, C. (1980) Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics*, **95**, 785–795.
49. Iost, I. and Dreyfus, M. (1995) The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.*, **14**, 3252.
50. Johnson, Z.I. and Chisholm, S.W. (2004) Properties of overlapping genes are

conserved across microbial genomes. *Genome Res.*, **14**, 2268–2272.

51. Aksoy, S., Squires, C.L. and Squires, C. (1984) Translational coupling of the trpB and trpA genes in the *Escherichia coli* tryptophan operon. *J. Bacteriol.*, **157**, 363–367.
52. Masuo, S., Zhou, S., Kaneko, T. and Takaya, N. (2016) Bacterial fermentation platform for producing artificial aromatic amines. *Sci. Rep.*, **6**, 25764.
53. Juminaga, D., Baidoo, E.E.K., Redding-Johanson, A.M., Batth, T.S., Burd, H., Mukhopadhyay, A., Petzold, C.J. and Keasling, J.D. (2012) Modular Engineering of l-Tyrosine Production in *Escherichia coli*. *Appl. Environ. Microbiol.*, **78**, 89–98.
54. Juminaga, D. and Keasling, J.D. (2017) Metabolic engineering of the shikimate pathway.
55. Eudes, A., Juminaga, D., Baidoo, E.E.K., Collins, F.W., Keasling, J.D. and Loqué, D. (2013) Production of hydroxycinnamoyl anthranilates from glucose in *Escherichia coli*. *Microb. Cell Factories*, **12**, 62.
56. Kiattisewee, C., Karanjia, A.V., Legut, M., Daniloski, Z., Koplik, S.E., Nelson, J., Kleinstiver, B.P., Sanjana, N.E., Carothers, J.M. and Zalatan, J.G. (2022) Expanding the Scope of Bacterial CRISPR Activation with PAM-Flexible dCas9 Variants. *ACS Synth. Biol.*, **11**, 4103–4112.
57. Kulinich, A. and Liu, L. (2016) Human milk oligosaccharides: The role in the fine-tuning of innate immune responses. *Carbohydr. Res.*, **432**, 62–70.
58. Newburg, D.S. and Morelli, L. (2015) Human milk and infant intestinal mucosal glycans guide succession of the neonatal intestinal microbiota. *Pediatr. Res.*, **77**, 115–120.
59. Baumgärtner, F., Conrad, J., Sprenger, G.A. and Albermann, C. (2014) Synthesis of the human milk oligosaccharide lacto-N-tetraose in metabolically engineered, plasmid-free *E. coli*. *Chembiochem Eur. J. Chem. Biol.*, **15**, 1896–1900.
60. Priem, B., Gilbert, M., Wakarchuk, W.W., Heyraud, A. and Samain, E. (2002) A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. *Glycobiology*, **12**, 235–240.
61. Zhang, W., Liu, Z., Gong, M., Li, N., Lv, X., Dong, X., Liu, Y., Li, J., Du, G. and

Liu, L. (2021) Metabolic engineering of *Escherichia coli* for the production of Lacto-N-neotetraose (LNnT). *Syst. Microbiol. Biomanufacturing*, 10.1007/s43393-021-00023-1.

62. Barbier, I., Kusumawardhani, H., Chauhan, L., Harlapur, P.V., Jolly, M.K. and Schaerli, Y. (2023) Synthetic Gene Circuits Combining CRISPR Interference and CRISPR Activation in *E. coli*: Importance of Equal Guide RNA Binding Affinities to Avoid Context-Dependent Effects. *ACS Synth. Biol.*, **12**, 3064–3071.
63. Chen, P.-Y., Qian, Y. and Del Vecchio, D. (2018) A Model for Resource Competition in CRISPR-Mediated Gene Repression. In *2018 IEEE Conference on Decision and Control (CDC)*.pp. 4333–4338.
64. Tian, T. and Zhou, X. (2023) CRISPR-Based Biosensing Strategies: Technical Development and Application Prospects. *Annu. Rev. Anal. Chem.*, **16**, 311–332.
65. Kaminski, M.M., Abudayyeh, O.O., Gootenberg, J.S., Zhang, F. and Collins, J.J. (2021) CRISPR-based diagnostics. *Nat. Biomed. Eng.*, **5**, 643–656.
66. Anzalone, A.V., Koblan, L.W. and Liu, D.R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
67. Rottinghaus, A.G., Ferreiro, A., Fishbein, S.R.S., Dantas, G. and Moon, T.S. (2022) Genetically stable CRISPR-based kill switches for engineered microbes. *Nat. Commun.*, **13**, 672.
68. 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.
69. Shaw, W.M., Studená, L., Roy, K., Hapeta, P., McCarty, N.S., Graham, A.E., Ellis, T. and Ledesma-Amaro, R. (2022) Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. *Nat. Commun.*, **13**, 4984.
70. Mouslim, C., Latifi, T. and Groisman, E.A. (2003) Signal-dependent Requirement for the Co-activator Protein RcsA in Transcription of the RcsB-regulated ugd Gene. *J. Biol. Chem.*, **278**, 50588–50595.
71. Huber, M., Faure, G., Laass, S., Kolbe, E., Seitz, K., Wehrheim, C., Wolf, Y.I., Koonin, E.V. and Soppa, J. (2019) Translational coupling via

termination-reinitiation in archaea and bacteria. *Nat. Commun.*, **10**, 4006.

72. Pennetier, C., Domínguez-Ramírez, L. and Plumbridge, J. (2008) Different regions of Mlc and NagC, homologous transcriptional repressors controlling expression of the glucose and N-acetylglucosamine phosphotransferase systems in *Escherichia coli*, are required for inducer signal recognition. *Mol. Microbiol.*, **67**, 364–377.
73. Plumbridge, J.A. (1991) Repression and induction of the nag regulon of *Escherichia coli* K-12: the roles of nagC and nagA in maintenance of the uninduced state. *Mol. Microbiol.*, **5**, 2053–2062.
74. Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., *et al.* (2017) Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol. Cell*, **65**, 618-630.e7.
75. Colognori, D., Trinidad, M. and Doudna, J.A. (2023) Precise transcript targeting by CRISPR-Csm complexes. *Nat. Biotechnol.*, **41**, 1256–1264.
76. 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.
77. Shinoda, H., Taguchi, Y., Nakagawa, R., Makino, A., Okazaki, S., Nakano, M., Muramoto, Y., Takahashi, C., Takahashi, I., Ando, J., *et al.* (2021) Amplification-free RNA detection with CRISPR–Cas13. *Commun. Biol.*, **4**, 1–7.
78. Kellner, M.J., Koob, J.G., Gootenberg, J.S., Abudayyeh, O.O. and Zhang, F. (2019) SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Protoc.*, **14**, 2986–3012.
79. Ding, Y., Tous, C., Choi, J., Chen, J. and Wong, W.W. (2023) Orthogonal inducible control of Cas13 circuits enables programmable RNA regulation in mammalian cells. 10.1101/2023.03.20.533499.
80. Call, S.N. and Andrews, L.B. (2022) CRISPR-Based Approaches for Gene Regulation in Non-Model Bacteria. *Front. Genome Ed.*, **4**, 892304.

Supplementary Information

Supplementary Figures

Figure S1: Effect of tlCRISPRi on mRFP expression from crRNAs using either the *Ur* or cognate direct repeat (DR) sequence
Figure S2: Effect of tlCRISPRi on mRFP expression using dCas13d or dCas13a
Figure S3: Time-course growth curves for dCas13 and Cas13
Figure S4: Reporter gene expression for tlCRISPRi using nuclease-inactive dCas13 and nuclease-active Cas13
Figure S5: Fold repression ratio for two fluorescent reporters in an operon targeted by both dCas9 and dCas13
Figure S6: Gene expression for txCRISPRa of two reporter constructs
Figure S7: Gene expression for combined txCRISPRa/tlCRISPRi of a synthetic reporter lacking an RBS for sfGFP or mRFP
Figure S8: Gene expression for combined txCRISPRa/tlCRISPRi of a synthetic reporter with high translational coupling.
Figure S9: Growth kinetics measured from cells containing dual-CRISPR systems
Figure S10: LNT titers with multiple txCRISPRi gRNAs targeting different endogenous genes for repression.

Supplementary Tables

Table S1: Strains used in this study
Table S2: Plasmids used in this study
Table S3: gRNA sequences

Supplementary Methods

S1. DNA Sequences

Supplementary Figures

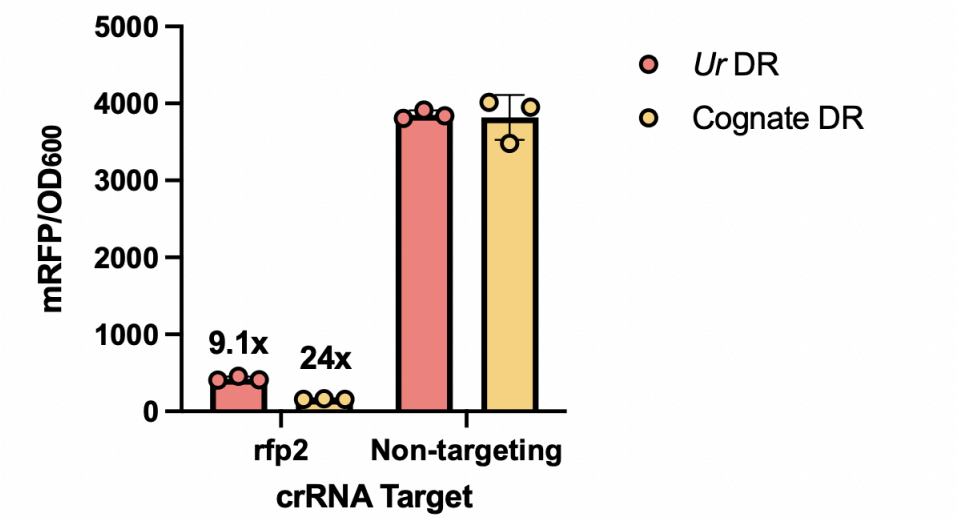


Figure S1. Effect of tICRISPRi on mRFP expression from crRNAs using either the *Ur* or the RfxCas13d cognate direct repeat (DR) sequence. The crRNAs *rfp2* and *AAV* were used as on- and non-targeting, respectively (Table S3). dCas13 was induced with 200 nM aTc. Values represent the mean ± standard deviation for at least three biological replicates. Fold-repression is shown relative to the respective non-targeting for each DR sequence.

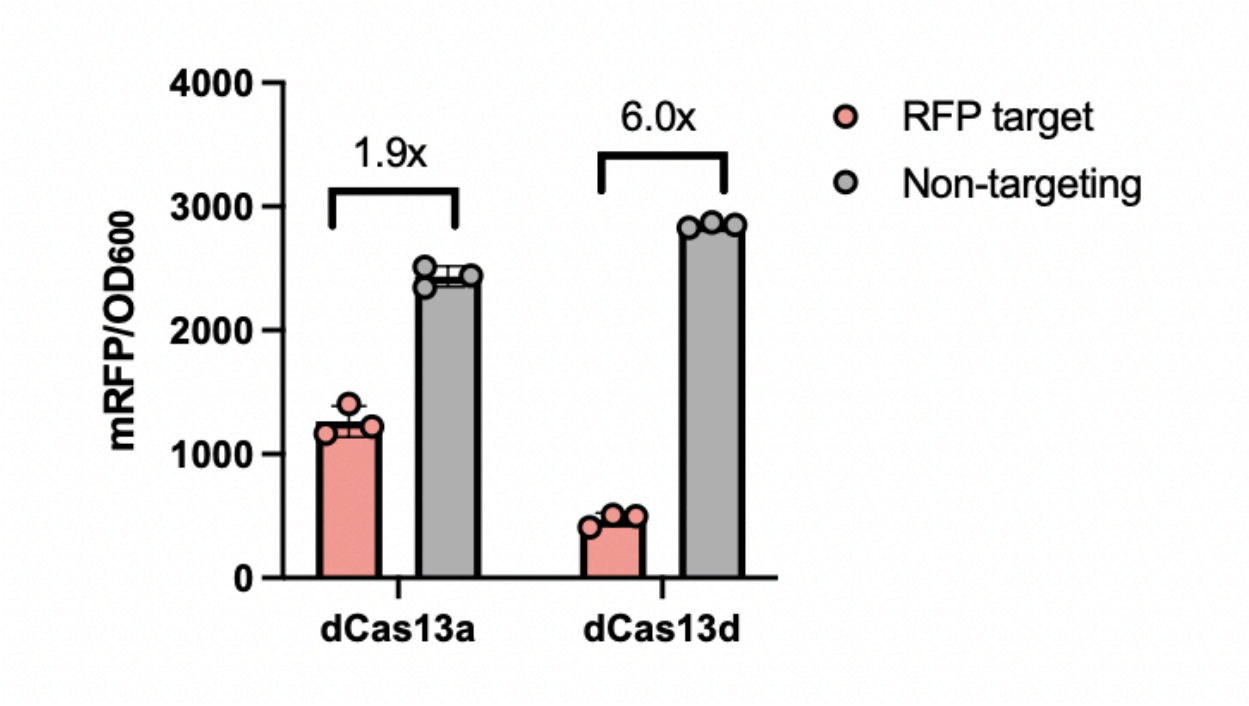


Figure S2. Effect of tlCRISPRi on mRFP expression using dCas13d or dCas13a. The crRNAs *rfp2* and *AAV* were used as on-target and non-targeting, respectively (Table S3). dCas13 variants were induced with 200 nM aTc. Values represent the mean ± standard deviation for at least three biological replicates.

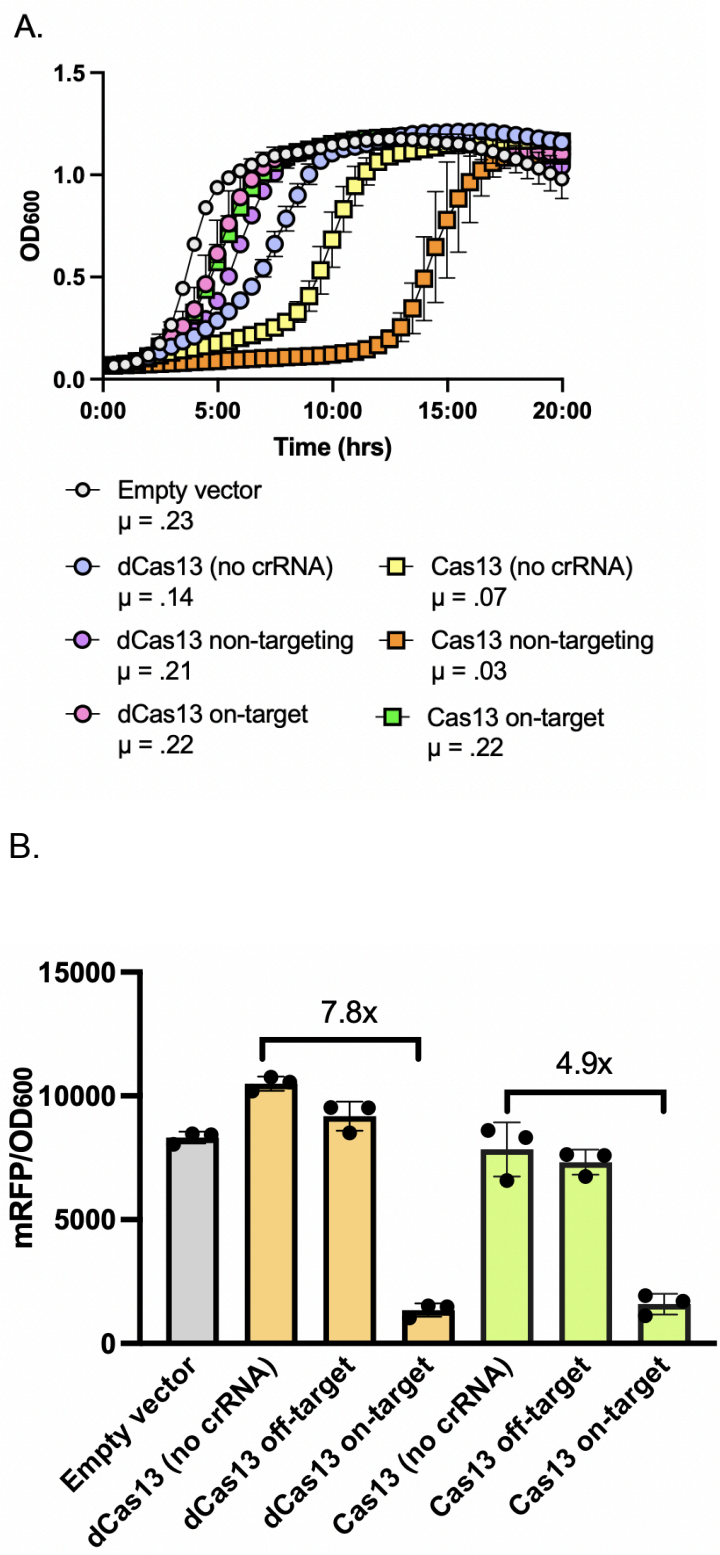


Figure S3. A) Time-course OD_{600} values for *E. coli* cultures were measured over 20 hours with dCas13 and Cas13 induced at 200 nM aTc 1:100 from overnight liquid

cultures that had also been grown with 200 nM aTc. Specific growth rates (μ) were calculated using the equation: $\mu = (\Delta \log OD_{600})/\Delta t$ between hours 2 and 6. B) Reporter gene expression after 20 hours of the strains shown in the top panel. The crRNAs *rfp2* and *AAV* were used as on-targets and non-targeting, respectively, targeting a mRFP reporter gene (Table S3). The mRFP reporter was constitutively expressed from the J3_J23110 promoter. Growth conditions are described in the Materials & Methods. Values represent the mean ± standard deviation for at least three biological replicates.

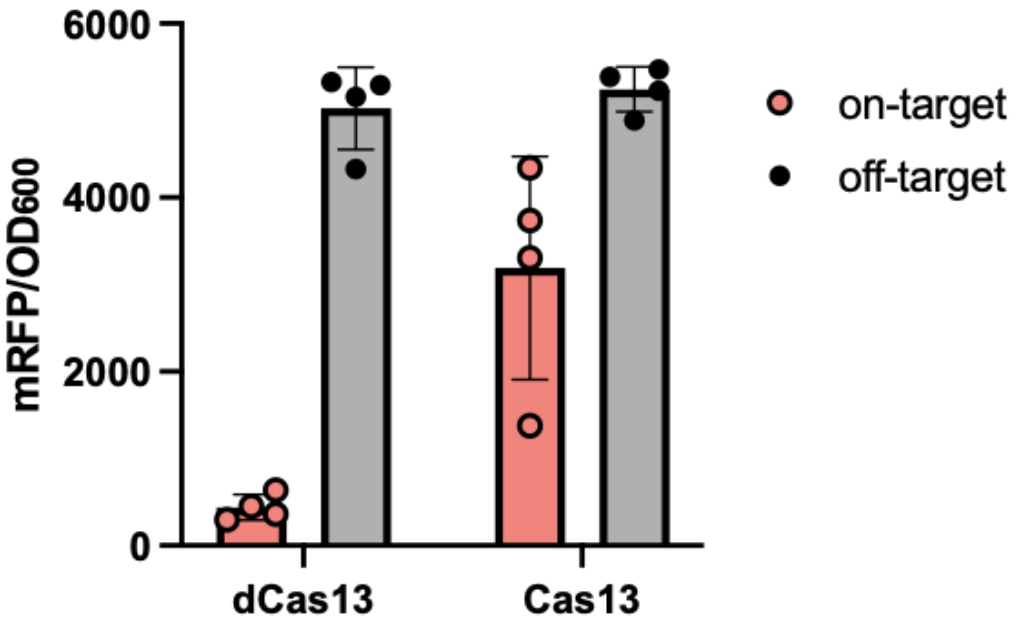


Figure S4. Reporter gene expression is shown for tICRISPRi comparing nuclease-inactive dCas13 and nuclease-active Cas13. These cultures were initiated into minimal media directly from single colonies on freshly-grown plates. In this context, we observed that dCas13d was 7-fold more effective at mRFP repression than Cas13d. This distinct behavior compared to Figure S3 may result from significant collateral Cas13d activity that disrupts the repression effect. This distinct behavior may result from the burden of expressing many heterologous constructs while at the same time switching from rich media plates to minimal media. We also cannot exclude the possibility that there are differences in protein expression levels for Cas13 and dCas13 when directly induced. The crRNAs *rfp2* and *AAV* are used as on- and off-targets, respectively (Table S3). dCas13 and Cas13 were induced with 200 nM aTc. The mRFP reporter was constitutively expressed from the J3_J23110 promoter. Values represent the mean ± standard deviation for at least three biological replicates.

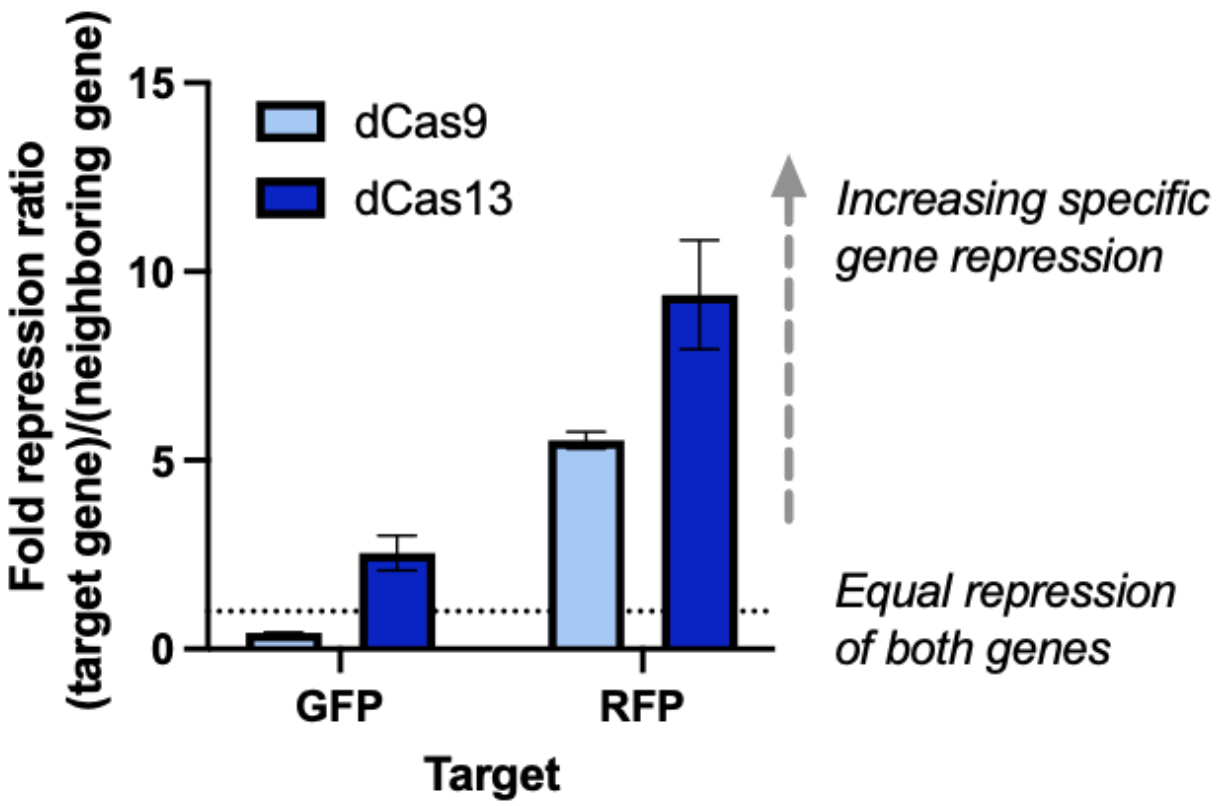


Figure S5. Fold repression ratio for two fluorescent reporters in an operon targeted by both dCas9 and dCas13. For each reporter, the ratio of fold repression for the target gene to fold repression of the neighboring operon gene is shown on the Y-axis as a measure of polar effects. CRISPRi polar effects are compared between dCas9 (light blue) and dCas13 (dark blue). The reporter operon consists of sfGFP followed by mRFP, expressed from a strong constitutive promoter (J3_J23110) and separate copies of the Bujard RBS. Accompanying raw data and fold repression are shown in Figure 3. For dCas9, the sgRNA targets *rr2* and *NT1* were used to target mRFP and sfGFP respectively, and fold repression values were calculated relative to cells expressing an non-targeting sgRNA for *AAV* (Table S3). For dCas13, the crRNA targets *rfp2* and *gfp2* were used to target mRFP and sfGFP, and fold repression values were calculated relative to cells expressing an non-targeting crRNA for *AAV* (Table S3).

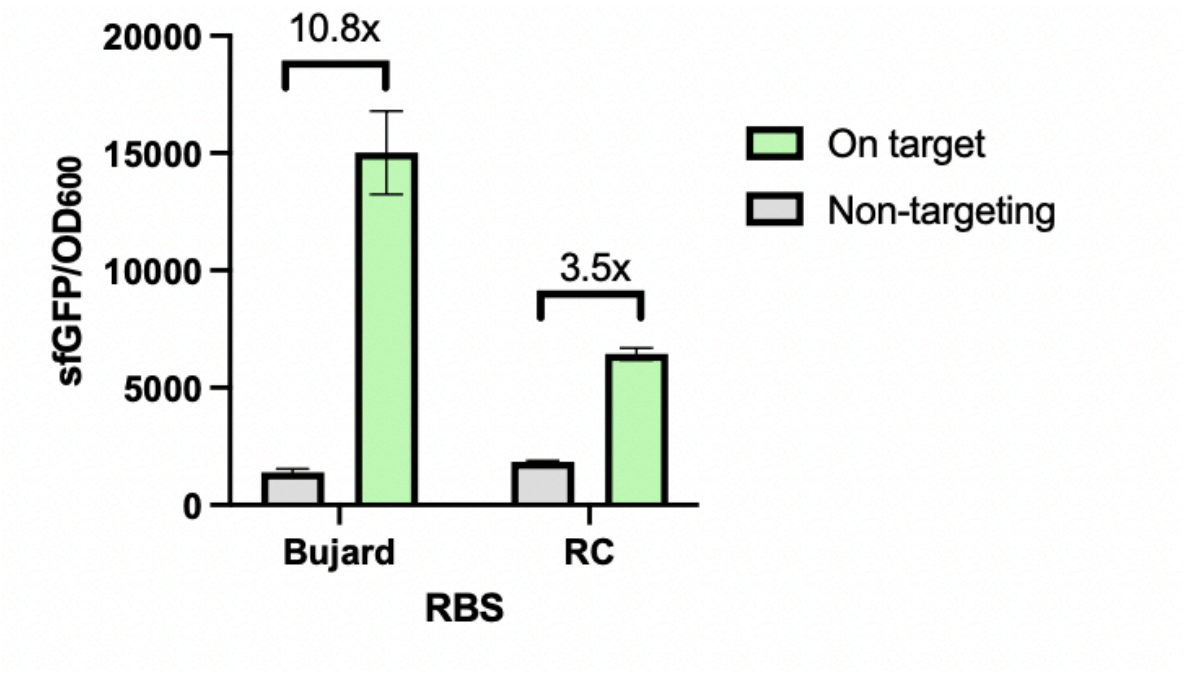


Figure S6. Gene expression for txCRISPRa of two reporter constructs. txCRISPRa was performed on multi-gene operons containing either the Bujard (left) and *rc* (right) synthetic RBSs driving expression of sfGFP, the first gene in the operon. The reporter uses the weak constitutive promoter J3_J23117, which allows for improved activation range over the previously used J23110 minimal promoter. The on-target gRNA used for dCas9 was J306 targeting -81 bp from the TSS, and the non-targeting gRNA is AAV (Table S3). Values represent the mean ± standard deviation for at least three biological replicates.

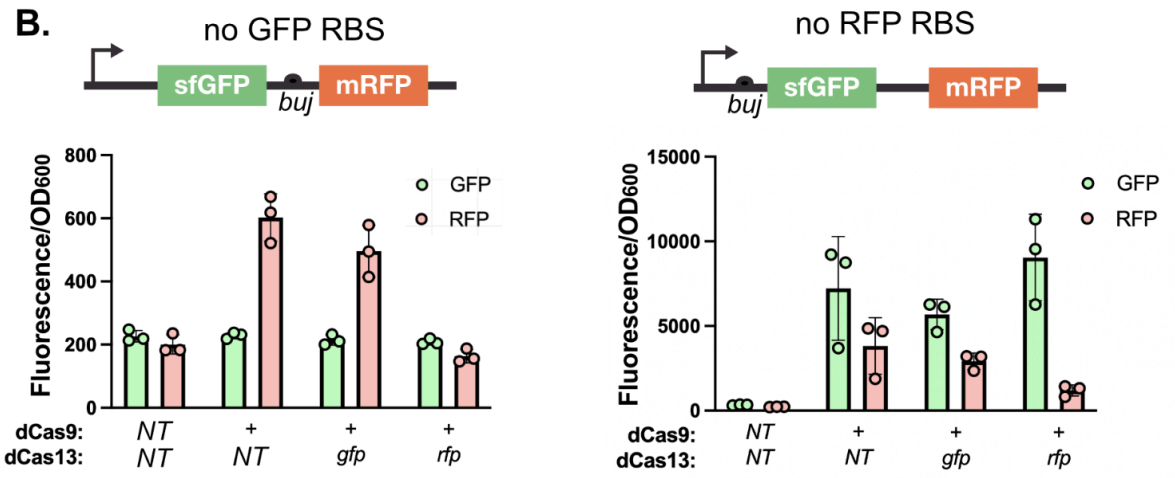


Figure S7. Combined txCRISPRa/tlCRISPRi as described in Figure 4A using a modified reporter lacking an RBS for either sfGFP (left) or mRFP (right) to investigate translational coupling. All strains contain dCas9 and dCas13 expressed from the J23107 promoter. For both dCas9 and dCas13, "NT" denotes expression of a non-targeting gRNA. For dCas9, "+" denotes the presence of an on-target, activating scRNA. crRNAs used for tlCRISPRi are labeled and sequences are given in Table S3. Values represent the mean ± standard deviation for at least three biological replicates.

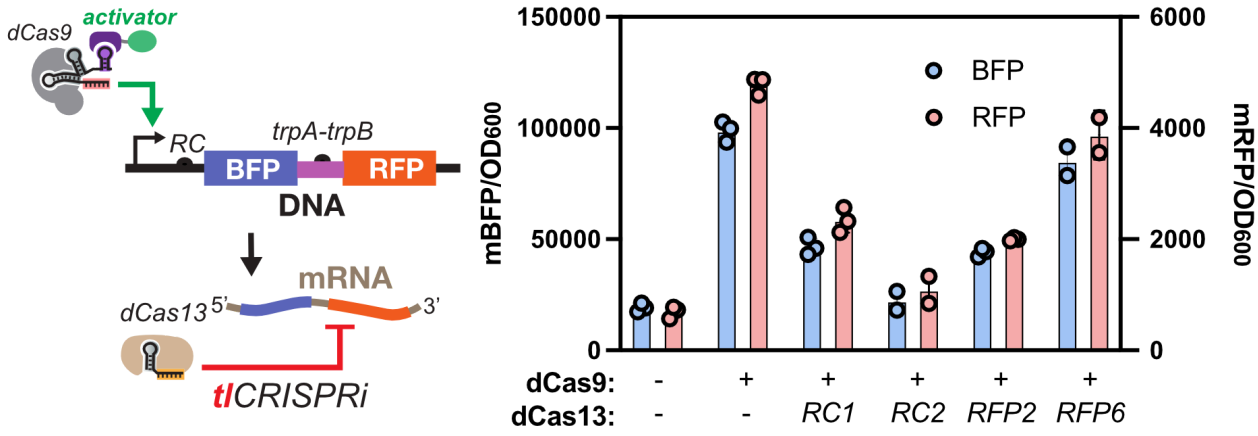


Figure S8. Gene expression for combined txCRISPRa/tlCRISPRi of a synthetic reporter with high translational coupling. The synthetic reporter used to model high translational coupling contains the last 60 nucleotides of the *trpA* gene and the first 60 nucleotides of the *trpB* gene, preceded by BFP and followed by RFP. All strains contain dCas9 and dCas13 expressed from the J23107 promoter. For both dCas9 and dCas13, "-" denotes expression of a non-targeting gRNA. "+" denotes the presence of an on-target, activating scRNA for dCas9. crRNAs used for tlCRISPRi are labeled and sequences are given in Table S3. Values represent the mean ± standard deviation for at least three biological replicates.

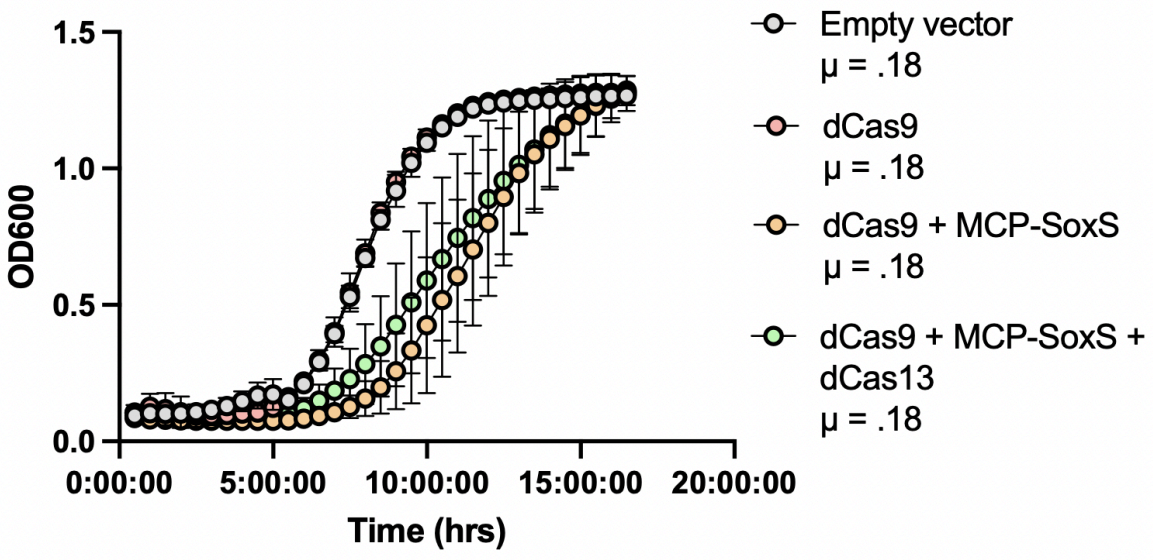


Figure S9. Growth time course measured from cells containing dual-CRISPR systems. For the cells containing empty vector and dCas9 + MCP-SoxS, the JM109 strain was used (Table S1) containing the plasmids pJF043 or pJF234.21, respectively (Table S2). For the cells containing dCas9 alone or dCas9, MCP-SoxS, and dCas13, the CK024 strain was used (Table S1). For the dCas9, MCP-SoxS, and dCas13 condition, the plasmid pRC172 was used (Table S2). Growth conditions are described in the Materials & Methods. Values represent the mean ± standard deviation for at least three biological replicates. Specific growth rates (μ) were calculated using the equation: $\mu = (\Delta \log OD600)/\Delta t$ between hours 6 and 10.

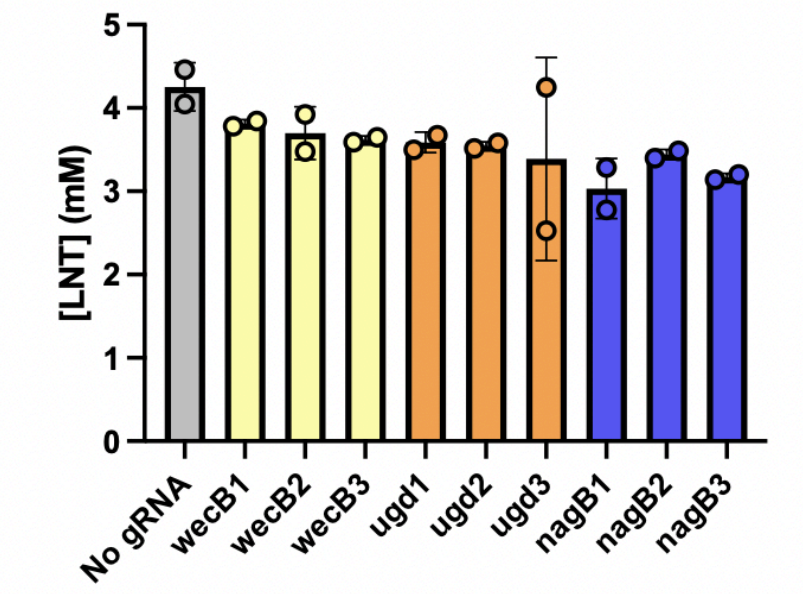


Figure S10. LNT titers with multiple txCRISPRi gRNAs targeting different endogenous genes for repression. Three sgRNAs are tested for each target. sgRNA #1 corresponds to the same targets shown in Figure 5B. The data reported here are from an independent experiment. Modest differences in the titers between the data shown here and in Figure 5B may be explained by differences in minimal media or LC/MS solvent preparation. Experiments were conducted in the JM109 *E. coli* strain, which contains a *lacZ* deletion to prevent lactose metabolism. All gRNA sequences are given in Table S3. Values represent the mean ± standard deviation for two biological replicates.

Supplementary Tables

Table S1. Strains used in this study

Strain	**Description**	**Genotype**
MG1655	Wildtype *E. coli* strain	F^- λ^- ilvG$^-$ rfb-50 rph-1
CK024	Integrated dCas9 used for combined dCas9 + dCas13 experiments	MG1655_arsB::Sp.pCas9-dCas9
JM109 (20)	*lacZ* knockout used for dCas9-only LNT experiments	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *lac*I^qZΔM15]

Table S2. Plasmids used in this study

Plasmid	Marker	Replicon	Promoter	RBS	Gene	Terminator	Reference
pJF143.J3_J23110	Amp	pSC101**	J3_J23110	buj	mRFP	dblTerm	(21), Figures 1 and 2
pRC055	Chlor	p15a	pTet	rc-2	dRxCas13d	dblTerm	This work, Figure 1
pRC056	Chlor	p15a	pTet	rc-2	RxCas13d	dblTerm	This work, Figure S1
pRC057.crRNA	Chlor	p15a	pTet J23119	rc-2 -	dRxCas13d crRNA	dblTerm ECK120033736	This work, Figure 1
pRC058.crRNA	Chlor	p15a	pTet J23119	rc-2 -	RxCas13d crRNA	dblTerm ECK120033736	This work, Figure S1
pRC055.117	Chlor	p15a	J23117	rc-2	dRxCas13d	dblTerm	This work, Figure 1
pRC055.107	Chlor	p15a	J23107	rc-2	dRxCas13d	dblTerm	This work, Figure 1
pRC065.crRNA	Chlor	p15a	J23107 J23119	rc-2 -	dRxCas13d crRNA	dblTerm ECK120033736	This work, Figure 1
pRC065d.crRNA	Chlor	p15a	J23107 J23119	rc-2 -	dRxCas13d crRNA	dblTerm ECK120033736	This work, Figure 1
pRC073.crRNA	Chlor	p15a	J23119	-	crRNA	ECK120033736	This work, Figure 1
pRC146	Kan	ColE1	J23119	-	crRNA	ECK120033736	This work, Figure 1
pRC072	Amp	pSC101**	J3_23110	buj, buj	sfGFP, mRFP	dblTerm	This work, Figure 3
pJB003	Amp	pSC101**	J3_J23117	buj, buj	sfGFP, mRFP	dblTerm	This work, Figure 4

pRC239	Amp	pSC101**	J3_J23117	rc, buj	sfGFP, mRFP	dblTerm	This work, Figure 4
pRC241	Amp	pSC101**	J3_J23117	rc, buj	mBFP, mRFP	dblTerm	This work, Figure 4
pJF083	Amp	pSC101**	pTet	buj	mRFP	dblTerm	This work, Figure 2
pCK667.crRNA	Chlor	pSC101**	pTet J23119	Rbs50k -	dCas13a crRNA	dblTerm	This work, Figure S1
pCK513.gRNA	Spec	pRSF1010	J23107, J23107, J23119	Buj, Buj, -	dCas9, MCP-SoxS, scRNA	dblTerm, BBa_B1002, ECK120033736	This work, Figure 3
pJF234.21	Chlor	p15a	Sp-Pdcas9, J23107 J23105, J23105, J23105	Native RBS, Buj, - - -	dCas9, MCP-SoxS, J306_17 J506_17 J606_20	dblTerm, BBa_B1002, ECK120033736 (x3)	This work, Figure 5
pRC280.gRNA	Chlor	p15a	Sp-Pdcas9, J23107 J23105, J23105, J23105, J23110	Native RBS, Buj, - - - -	dCas9, MCP-SoxS, J306, J506, J606, dCas9 gRNA	dblTerm, BBa_B1002, ECK120033736 (x3)	This work, Figure 5
pIDF96C	Amp	pSC101**	J3, J5, J6	RBS-A, RBS-B, RBS-C	lacY, lgtA, galT	ECK120033737, ECK120010818, ECK120015440	This work, Figure 5
pRC172	Chlor	p15a	J23107, J23107 J23105, J23105, J23105	rc2, buj, - - -	dCas13, MCP-SoxS, J306 J506 J606	dblTerm, BBa_B1002, ECK120033736 (x3)	This work, Figure 5

pRC182	Chlor	p15a	J23107, J23107 J23105, J23105, J23105, J23110	rc2, buj, - - - -	dCas13, MCP-SoxS , J306, J506, J606, dCas13 gRNA	dblTerm, BBa_B1002, ECK120033736 (x4)	This work, Figure 5
pRC188	Chlor	p15a	J23107 J23105, J23105, J23105	buj, - - -	MCP-SoxS , J306 J506 J606	BBa_B1002, ECK120033736 (x3)	This work, Figure 5
pRC189	Chlor	p15a	J23107 J23105, J23105, J23105, J23110	buj, - - - -	MCP-SoxS , J306, J506, J606, dCas13 gRNA	dblTerm, BBa_B1002, ECK120033736 (x4)	This work, Figure 5
pRC271	Chlor	p15a	J23107 J23105, J23105, J23105, J23110, J23110	rc2, buj, - - - - -	dCas13, MCP-SoxS , J306, J506, J606, crRNA 1, crRNA 2	dblTerm, BBa_B1002, ECK120033736 (x5)	This work, Figure 5
pJF043	Chlor	p15a	-	-	-	-	(20), Figure S6
pRC289	Amp	pSC101**	J23110	rc, endog. trpB RBS	mBFP, mRFP	dblTerm	This work, Figure S8

Table S3. gRNA sequences

gRNA	DNA Sequence	CRISPR System	Target
buj1	TGGTACCTTTCTCCTCTTTAATGAATTC	dCas13	RFP
rfp2	AgAACGTCTTCGCTACTCGCCATGGTAC	dCas13	RFP
rfp3	ctcgtgaccgttaacggaaccttccata	dCas13	RFP
rfp4	ttttttctgcataaccggaccgtcgga	dCas13	RFP
rfp5	tgcttacaaaaccgacatcaaactggac	dCas13	RFP
gfp2	ATGAGCAAAGGAGAAGAACTTTTCACTG	dCas13	GFP
rc1	GTGCTCCCTCGTGAGAACTTTTTCGAAC	dCas13	*rc* RBS (expressing either sfGFP or mBFP)
rc2	ATAAAACCTCCTTACTCTGTGCTCCCTC	dCas13	*rc* RBS (expressing either sfGFP or mBFP)
lacZ-9	TTCTCCGTGGGAACAAACGG	dCas9	lacZ
nagB-9	GTTCACGCAACCTTCAACGG	dCas9	nagB
ugd-9	TACATAGCCAGTACCGGAAA	dCas9	ugd
wecB-9	ACAAACGGTAAATACTCCTG	dCas9	wecB
lacZ-13	cggccagtgaatccgtaatcatggtcat	dCas13	lacZ
nagB-13	ggatcagtctcattattcacctcaataa	dCas13	nagB
ugd-13	agccagtaccggaaatggtgattttcat	dCas13	ugd
wecB-13	aatacagtcagtactttcacatcgattc	dCas13	wecB
J306	TTGTGTCCAGAACGCTCCGT	dCas9	-81bp from the TSS of synthetic J3 promoter
J506	AGCAGCATGAGCAGCATTGA	dCas9	-81bp from the TSS of synthetic J5 promoter

J606	GTCGCAGTCGCGCGAGCACT	dCas9	-81bp from the TSS of synthetic J6 promoter
AAV	GGGGCCACTAGGGACAGGAT	dCas13/dCas9	Off-target
rr2	TGGAACCGTACTGGAACTGC	dCas9	RFP
NT1	CATCTAATTCAACAAGAATT	dCas9	GFP
rfp6	CTTTGATAACGTCTTCGCTACTCGCCAT	dCas13	RFP ORF
bfp	gcatgttctccttaatcagctcgctcat	dCas13	BFP ORF
ugd.2	GACATAATCAGCATCCCGGT	dCas9	ugd
ugd.3	CGCCACGCGCATCGCCAGGT	dCas9	ugd
nagB.2	gtcggcaggcccagtacaaa	dCas9	nagB
nagB.3	ccccctgactaccgctgaac	dCas9	nagB
wecB.2	gcgccatcttgatggcttcc	dCas9	wecB
wecB.3	gcatcgggagatgctcgatc	dCas9	wecB
lacZ.2	gtcacgacgttgtaaaacga	dCas9	lacZ
lacZ.3	cgaaaggggatgtgctgca	dCas9	lacZ

Supplementary Methods

DNA sequences used in this study

Reporters

>J3_J23110_**buj**_sfGFP_**buj**_mRFP_dblTerm

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTA
CGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGG
AGCGTTCTGGACACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctag
gtacaatgctagc**GAATTCATTAAAGAGGAGAAAGGTACC**ATGAGCAAAGGAGAAGAACTTTTC
ACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCC
GTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGG
AAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCC
CGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTAC
AGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGA
AGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATT
CTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAA
AGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGC
AGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTAC
CTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTG
AGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAtaaggatcc**GAATT**
CATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATgcgtttc
aaagttcgtatggaaggttccgttaacggtcacgagttcgaaatcgaaggtgaaggtgaaggtc
gtccgtacgaaggtacccagaccgctaaactgaaagttaccaaaggtggtccgctgccgttcgc
ttgggacatcctgtccccgcagttccagtacggttccaaagcttacgttaaacacccggctgac
atcccggactacctgaaactgtccttcccggaaggtttcaaatgggaacgtgttatgaacttcg
aagacggtggtgttgttaccgttacccaggactcctccctgcaagacggtgagttcatctacaa
agttaaactgcgtggtaccaacttcccgtccgacggtccggttatgcagaaaaaaaccatgggt
tgggaagcttccaccgaacgtatgtacccggaagacggtgctctgaaaggtgaaatcaaaatgc
gtctgaaactgaaagacggtggtcactacgacgctgaagttaaaaccacctacatggctaaaaa
accggttcagctgccgggtgcttacaaaaccgacatcaaactggacatcacctcccacaacgaa
gactacaccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccggtgcttaaggat
ccaaactcgagtaaggatctCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCC
TTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGG
GTGGGCCTTTCTGCGTTTATA

>J3_J23117_**rc**_mBFP_**buj**_mRFP_dblTerm

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTA
CGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGG
AGCGTTCTGGACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctag

ggattgtgctagc**GTTCGAAAAAGTTCTCACGAGGGAGCACAGAGTAAGGAGGTTTTAT**atgag
cgagctgattaaggagaacatgcacatgaagctgtacatggagggcaccgtggacaaccatcac
ttcaagtgcacatccgagggcgaaggcaagccctacgagggcacccagaccatgagaatcaagg
tggtcgagggcggccctctccccttcgccttcgacatcctggctactagcttcctctacggcag
caagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggc
ttcacatgggagagagtcaccacatacgaagacgggggcgtgctgaccgctacccaggacacca
gcctccaggacggctgcctcatctacaacgtcaagatcagaggggtgaacttcacatccaacgg
ccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgtaccccgctgac
ggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaa
acgccaagaccacatatagatccaagaaacccgctaagaacctcaagatgcctggcgtctacta
tgtggactacagactggaaagaatcaaggaggccaacaacgagacctacgtcgagcagcacgag
gtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaattaaggatcc**G**
AATTCATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATgcg
tttcaaagttcgtatggaaggttccgttaacggtcacgagttcgaaatcgaaggtgaaggtgaa
ggtcgtccgtacgaaggtacccagaccgctaaactgaaagttaccaaaggtggtccgctgccgt
tcgcttgggacatcctgtccccgcagttccagtacggttccaaagcttacgttaaacacccggc
tgacatcccggactacctgaaactgtccttcccggaaggtttcaaatgggaacgtgttatgaac
ttcgaagacggtggtgttgttaccgttacccaggactcctccctgcaagacggtgagttcatct
acaaagttaaactgcgtggtaccaacttcccgtccgacggtccggttatgcagaaaaaaaccat
gggttgggaagcttccaccgaacgtatgtacccggaagacggtgctctgaaaggtgaaatcaaa
atgcgtctgaaactgaaagacggtggtcactacgacgctgaagttaaaaccacctacatggcta
aaaaaccggttcagctgccgggtgcttacaaaaccgacatcaaactggacatcacctcccacaa
cgaagactacaccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccggtgcttaa
ggatccaaactcgagtaaggatctCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTG
GGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCT
TCGGGTGGGCCTTTCTGCGTTTATA

dCas9

>Sp-PdCas9_dCas9_dblTerm

TTACGAAATCATCCTGTGGAGCTTAGTAGGTTTAGCAAGATGGCAGCGCCTAAATGTAGAATGA
TAAAAGGATTAAGAGATTAATTTCCCTAAAAATGATAAAACAAGCGTTTTGAAAGCGCTTGTTT
TTTTGGTTTGCAGTCAGAGTAGAATAGAAGTATCAAAAAAAGCACCGACTCGGTGCCACTTTTT
CAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATGCTGTTTTGAATGGTTCCAACAAGAT
TATTTTATAACTTTTATAACAAATAATCAAGGAGAAATTCAAAGAAATTTATCAGCCATAAAAC
AATACTTAATACTATAGAATGATAACAAAATAAACTACTTTTTAAAAGAATTTTGTGTTATAAT
CTATTTATTATTAAGTATTGGGTAATATTTTTTGAAGAGATATTTTGAAAAAGAAAAATTAAAG
CATATTAAACTAATTTCGGAGGTCATTAAAACTATTATTGAAATCATCAAACTCATTATGGATT
TAATTTAAACTTTTTATTTTAGGAGGCAAAAATGGATAAGAAATACTCAATAGGCTTAGcTATC
GGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCA
AGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGA
CAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG

AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTT
TCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTAT
TTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGA
AAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATA
TGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGA
CAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCA
AGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATC
TCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATT
GGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCA
AAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATT
TGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAC
TGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTG
ATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAA
ATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGT
GAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGG
GTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGA
GAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAAT
AGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAG
TTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCT
TCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAA
TTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGA
AGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGA
AGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTT
AATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATA
ATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGA
GATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTT
AAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATA
AGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTAT
GCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGA
CAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTA
TTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAA
TATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAG
CGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTG
TTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACAT
GTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATgcCATTGTTCCA
CAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTG
GTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACT
TCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGT
TTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTA
AGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTAT

TCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAA
TTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCG
TTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAA
AGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAA
TATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGA
TTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCG
AGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAA
GTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTG
CTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTC
AGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTA
CTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTA
AAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTT
AGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCT
CTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTC
CAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTAT
TGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTT
AGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTAT
TTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAA
ACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTT
TATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTAACtcgagtaaggatctccaggca
tcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtg
aacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata

MCP-SoxS

J23107_**Buj**_MCP-SoxS_BBa-B1002

tttacggctagctcagccctaggtattatgctagc**GAATTCATTAAAGAGGAGAAAGGTACC**at
ggggcccgcttctaactttactcagttcgttctcgtcgacaatggcggaactggcgacgtgact
gtcgccccaagcaacttcgctaacgggatcgctgaatggatcagctctaactcgcgttcacagg
cttacaaagtaacctgtagcgttcgtcagagctctgcgcagaatcgcaaatacaccatcaaagt
cgaggtgcctaaaggcgcctggcgttcgtacttaaatatggaactaaccattccaattttcgcc
acgaattccgactgcgagcttattgttaaggcaatgcaaggtctcctaaaagatggaaacccga
ttccctcagcaatcgcagcaaactccggcatctacGGTGGCGGAGGTAGCATGTCCCATCAGAA
AATTATTCAGGATCTTATCGCATGGATTGACGAGCATATTGACCAGCCGCTTAACATTGATGTA
GTCGCAAAAAAATCAGGCTATTCAAAGTGGTACTTGCAACGAATGTTCCGCACGGTGACGCATC
AGACGCTTGGCGATTACATTCGCCAACGCCGCCTGTTACTGGCCGCCGTTGAGTTGCGCACCAC
CGAGCGTCCGATTTTTGATATCGCAATGGACCTGGGTTATGTCTCGCAGCAGACCTTCTCCCGC
GTTTTCGCGCGGCAGTTTGATCGCACTCCCGCGGATTATCGCCACCGCCTGTAAGCGGCCGCca
cgcaaaaaaccccgcttcggcggggtttttcgc

dCas9 gRNA

>J23105_**spacer**-tracrRNA-MS2_TrrnB

tttacggctagctcagtcctaggtactatACTAGT**NNNNNNNNNNNNNNNNNNNN**GTTTTAGAG
CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACATGAGGAT
CACCCATGTGCTTTTTTTGAAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAGGATCTGAAT
AGCGCCGTCGACCATCATCATCATCATCATTGAGTTTAAACGGTCTCCAGCTTGGCTGTTTTGG
CGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAA
CAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGA
AACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATC
AAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAA
CT

dCas13

>J23107_**rc2-rbs**_dCas13_dblTerm

tttacggctagctcagccctaggtattatgctagc**GATATACTAAAAGAGGAGAAAACTGAT**AT
Gatcgaaaaaaaaaagtccttcgccaagggcatgggcgtgaagtccacactcgtgtccggctcc
aaagtgtacatgacaaccttcgccgaaggcagcgacgccaggctggaaaagatcgtggagggcg
acagcatcaggagcgtgaatgagggcgaggccttcagcgctgaaatggccgataaaaacgccgg
ctataagatcggcaacgccaaattcagccatcctaagggctacgccgtggtggctaacaaccct
ctgtatacaggacccgtccagcaggatatgctcggcctgaaggaaactctggaaaagaggtact
tcggcgagagcgctgatggcaatgacaatatttgtatccaggtgatccataacatcctggacat
tgaaaaaatcctcgccgaatacattaccaacgccgcctacgccgtcaacaatatctccggcctg
gataaggacattattggattcggcaagttctccacagtgtatacctacgacgaattcaaagacc
ccgagcaccatagggccgctttcaacaataacgataagctcatcaacgccatcaaggcccagta
tgacgagttcgacaacttcctcgataaccccagactcggctatttcggccaggccttttcagc
aaggagggcagaaattacatcatcaattacggcaacgaatgctatgacattctggccctcctga
gcggactggcgcactgggtggtcgctaacaacgaagaagagtccaggatctccaggacctggct
ctacaacctcgataagaacctcgacaacgaatacatctccaccctcaactacctctacgacagg
atcaccaatgagctgaccaactccttctccaagaactccgccgccaacgtgaactatattgccg
aaactctgggaatcaaccctgccgaattcgccgaacaatatttcagattcagcattatgaaaga
gcagaaaaacctcggattcaatatcaccaagctcagggaagtgatgctggacaggaaggatatg
tccgagatcaggaaaaatcataaggtgttcgactccatcaggaccaaggtctacaccatgatgg
actttgtgatttataggtattacatcgaagaggatgccaaggtggctgccgccaataagtccct
ccccgataatgagaagtccctgagcgagaaggatatctttgtgattaacctgaggggctccttc
aacgacgaccagaaggatgccctctactacgatgaagctaatagaatttggagaaagctcgaaa
atatcatgcacaacatcaaggaatttaggggaaacaagacaagagagtataagaagaaggacgc
ccctagactgcccagaatcctgcccgctggccgtgatgtttccgccttcagcaaactcatgtat
gccctgaccatgttcctggatggcaaggagatcaacgacctcctgaccaccctgattaataaat
tcgataacatccagagcttcctgaaggtgatgcctctcatcggagtcaacgctaagttcgtgga
ggaatacgccttttttcaaagactccgccaagatcgccgatgagctgaggctgatcaagtccttc
gctagaatgggagaacctattgccgatgccaggagggccatgtatatcgacgccatccgtattt
taggaaccaacctgtcctatgatgagctcaaggccctcgccgacacctttttccctggacgagaa
cggaaacaagctcaagaaaggcaagcacggcatgagaaatttcattattaataacgtgatcagc

aataaaaggttccactacctgatcagatacggtgatcctgcccacctccatgagatcgccaaaa
acgaggccgtggtgaagttcgtgctcggcaggatcgctgacatccagaaaaaacagggccagaa
cggcaagaaccagatcgacaggtactacgaaacttgtatcggaaaggataagggcaagagcgtg
agcgaaaaggtggacgctctcacaaagatcatcaccggaatgaactacgaccaattcgacaaga
aaaggagcgtcattgaggacaccggcagggaaaacgccgagagggagaagtttaaaaagatcat
cagcctgtacctcaccgtgatctaccacatcctcaagaatattgtcaatatcaacgccaggtac
gtcatcggattccattgcgtcgagcgtgatgctcaactgtacaaggagaaaggctacgacatca
atctcaagaaactggaagagaagggattcagctccgtcaccaagctctgcgctggcattgatga
aactgcccccgataagagaaaggacgtggaaaaggagatggctgaaagagccaaggagagcatt
gacagcctcgagagcgccaaccccaagctgtatgccaattacatcaaatacagcgacgagaaga
aagccgaggagttcaccaggcagattaacagggagaaggccaaaaccgccctgaacgcctacct
gaggaacaccaagtggaatgtgatcatcagggaggacctcctgagaattgacaacaagacatgt
accctgttcgcaaacaaggccgtcgccctggaagtggccaggtatgtccacgcctatatcaacg
acattgccgaggtcaattcctacttccaactgtaccattacatcatgcagagaattatcatgaa
tgagaggtacgagaaaagcagcggaaaggtgtccgagtacttcgacgctgtgaatgacgagaag
aagtacaacgataggctcctgaaactgctgtgtgtgcctttcggctactgtatccccaggttta
agaacctgagcatcgaggccctgttcgataggaacgaggccgccaagttcgacaaggagaaaaa
gaaggtgtccggcaattcctaactcgagtaaggatctccaggcatcaaataaaacgaaaggctc
agtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtca
cactggctcaccttcgggtgggcctttctgcgtttata

>TetR-Ptet_**rbs**_dCas13a_dblTerm
ttaagacccactttcacatttaagttgtttttctaatccgcatatgatcaattcaaggccgaat
aagaaggctggctctgcaccttggtgatcaaataattcgatagcttgtcgtaataatggcggca
tactatcagtagtaggtgtttccctttcttctttagcgacttgatgctcttgatcttccaatac
gcaacctaaagtaaaatgccccacagcgctgagtgcatataatgcattctctagtgaaaaacct
tgttggcataaaaaggctaattgattttcgagagtttcatactgtttttctgtaggccgtgtac
ctaaatgtacttttgctccatcgcgatgacttagtaaagcacatctaaaacttttagcgttatt
acgtaaaaaatcttgccagctttccccttctaaagggcaaaagtgagtatggtgcctatctaac
atctcaatggctaaggcgtcgagcaaagcccgcttatttttacatgccaatacaatgtaggct
gctctacacctagcttctgggcgagtttacgggttgttaaaccttcgattccgacctcattaag
cagctctaatgcgctgttaatcactttacttttatctaatctagacatcattaattcctaattt
ttgttgacactctatcgttgatagagttattttaccactccctatcagtgatagagaaaa**gaat**
tcaaaGCGGTTACACTATAAGGAGGTAcaTatgaaagtgacgaaggtaggaggcatttcgcata
agaagtacacgtccgaaggccgcttagtgaagtcagaatcggaagaaaatcgcacagacgaacg
tctgtcggcgttgcttaatatgcgccttgacatgtatatcaagaatcccagcagcacggaaacc
aaggaaaatcaaaaacgcattgggaaattaaagaaattcttctcaaacaaaatggtctatctta
aagacaataccttgagtttgaagaatgggaaaaaggagaacattgatcgtgagtattctgagac
tgacatccttgagagcgatgtccgtgacaagaaaaacttcgccgtgttgaaaaagatctatctg
aatgaaaacgtgaactcggaggaattggaagtttttcgtaacgacattaagaagaaactgaaca
aaatcaacagcctgaagtactcatttgaaaagaataaggcgaattatcaaaagattaatgagaa

taacatcgagaaggttgaaggtaagtcaaagcgtaacattatttacgattattatcgtgagtca
gcgaaacgtgacgcttatgtaagcaatgtgaaagaagcctttgataagctttacaaggaagagg
acattgcaaaacttgttcttgaaattgagaaccttacgaagttagagaaatacaagattcgcga
gttctaccacgaaattattggacgtaagaatgacaaggaaaactttgcaaaaatcatctacgaa
gaaatccagaatgttaataacatgaaagagttgatcgagaaggtaccggacatgagtgagttga
aaaagagccaagtattttacaagtattacttagacaaagaagagttgaacgacaagaacatcaa
atacgcgttttgtcatttcgtggaaatcgaaatgagtcagttgctgaagaactacgtatataag
cgcttaagtaatatctcgaatgacaaaattaagcgtatctttgaataccagaacttgaaaaaat
tgatcgaaaataagctgttaaacaaacttgacacgtacgtccgtaattgtggaaagtataatta
ttatttgcaagacggcgaaattgccacttcagatttcatcgcccgcaaccgtcagaatgaagcg
tttcttcgcaacatcattggggtgtcatctgtggcctacttttctcttcgcaacattcttgaaa
cggagaacgagaatgatattactgggcgtatgcgcggcaaaacagttaagaacaataaaggtga
agagaagtacgtgtccggagaagttgataagatctataatgaaaataagaagaacgaggttaag
gagaacttaaaaatgttctattcgtacgatttcaatatggacaacaagaatgaaatcgaagatt
tcttcgccaacatcgacgaggcgatttcttccatcgctcacggtattgtcgccttcaacttgga
attagaaggtaaggatatctttgcgttcaagaacattgcgccatccgaaatctcaaagaagatg
tttcagaatgagattaacgagaaaaaactgaaattgaagatctttcgtcaactgaactctgcca
acgtgttccgctatctcgaaaagtataaaattctgaattaccttaaacgtacacgcttcgagtt
tgtcaataaaaatatcccattcgtcccgtctttcaccaaattatattcgcgcattgatgacctg
aagaatagtcttgggatttactggaaaactccgaaaacaaacgacgacaataagactaaggaga
ttattgatgcccaaatctatttgcttaaaaacatctattacggggagttcctgaattatttcat
gtcgaacaatggtaatttctttgagatttctaaagaaatcatcgaattgaacaagaacgataaa
cgcaacttaaagactgggttttacaagctgcaaaagtttgaagacatccaggagaagattccaa
aggaatacttggcgaatatccagtccctgtacatgattaatgccggtaatcaggacgaagaaga
aaaggacacttatattgatttcattcaaaagatcttcttaaagggatttatgacgtatcttgct
aataacggtcgtttaagtctgatttacatcggctcggatgaagaaacaaatacgtcattagcag
aaaagaagcaagagtttgacaagttcttgaagaagtacgagcagaacaataatatcaagatccc
ctatgagatcaatgaattcctgcgtgagatcaaactgggaaacatcctgaagtatactgagcgt
ttaaacatgttctaccttatcttaaagcttttgaatcacaaggagctgacaaatctgaagggta
gtcttgaaaaatatcagtctgccaataaggaagaagcgttctctgaccaattggagttaattaa
cctgcttaaccttgacaacaaccgcgtgacggaagacttcgaattagaggccgacgagattgga
aaatttcttgatttcaatggcaacaaagttaaggataacaaggaactgaaaaagttcgatacaa
acaagatctactttgacggcgagaacattatcaaacaccgtgccttctacaatattaagaaata
tggcatgttaaacttactggagaaaattgccgacaaggctggatacaagatctcgatcgaagag
ctgaagaaatactccaataaaaagaatgagatcgagaagaaccataagatgcaggaaaatctgc
accgcaaatacgctcgtccccgtaaagacgagaagtttacagatgaggactatgaaagttacaa
gcaagctattgagaatattgaggagtacacccaccttaagaacaaggtagaattcaatgagctg
aatttactgcagggcctgttgctgcgcattttacatcgtttagtcggatatacctcaatttggg
aacgcgatctgcgcttccgccttaaaggtgagttcccagaaaaccaatacatcgaagagatctt
caactttgaaaataagaagaacgtgaagtacaaaggggtcagattgtagagaaatacattaaa
ttctacaaggaattacatcaaaatgatgaagttaagatcaacaagtacagttccgcgaatatca

aggtgttgaagcaagaaaagaaggacctttatattgctaattacatcgccgcattcaattatat
tcctcacgccgagatctcactgctggaagtccttgaaaatttgcgtaaattgctgtcctacgat
cgcaaactgaaaaatgccgtaatgaaatcagtagttgatatccttaaggagtatggttttgtag
ccacattcaaaatcggggcggacaagaagatcggtattcagacactggagagcgaaaaaatcgt
gcatcttaagaatcttaagaagaagaagttaatgactgaccgcaattccgaggaactttgcaaa
ttggtgaagattatgtttgaatacaaaatggaagagaaaaagtctgaaaacTAActcgagtaag
gatctccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgt
tgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcg
tttata

_Ur_Cas13 gRNA

>J23110_Direct-repeat_**spacer**_ECK120033736

tttacggctagctcagtcctaggtacaatgctagcCACTAGTGCGAATTTGCACTAGTCTAAAA
C**NNNNNNNNNNNNNNNNNNNNNNNNNNNN**AAAGCCCCCGGAAGATCACCTTCCGGGGGCTTTtt
tattgcgc

_Rfx_Cas13 gRNA

>J23110_Direct-repeat_**spacer**_ECK120033736

tttacggctagctcagtcctaggtacaatgctagcAACCCCTACCAACTGGTCGGGGTTTGAAA
C**NNNNNNNNNNNNNNNNNNNNNNNNNNNN**AAAGCCCCCGGAAGATCACCTTCCGGGGGCTTTtt
tattgcgc

LNT pathway

>J3_J23117_**RBS-A**_lacY

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTA
CGAGCTGTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGG
AGCGTTCTGGACACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctag
ggattgtgctagc**AAAGATCTTTTAAGAAGGAGATATACAT**atgtactatttaaaaaacacaaa
cttttggatgttcggtttattcttttttcttttacttttttatcatgggagcctacttcccgttt
ttcccgatttggctacatgacatcaaccatatcagcaaaagtgatacgggtattatttttgccg
ctatttctctgttctcgctattattccaaccgctgtttggtctgctttctgacaaactcgggct
gcgcaaatacctgctgtggattattaccggcatgttagtgatgtttgcgccgttctttatttt
atcttcgggccactgttacaatacaacattttagtaggatcgattgttggtggtatttatctag
gcttttgttttaacgccggtgcgccagcagtagaggcatttattgagaaagtcagccgtcgcag
taatttcgaatttggtcgcgcgcggatgtttggctgtgttggctgggcgctgtgtgcctcgatt
gtcggcatcatgttcaccatcaataatcagtttgttttctggctgggctctggctgtgcactca
tcctcgccgttttactctttttcgccaaaacggatgcgccctcttctgccacggttgccaatgc
ggtaggtgccaaccattcggcatttagccttaagctggcactggaactgttcagacagccaaaa
ctgtggttttgtcactgtatgttattggcgtttcctgcacctacgatgttttgaccaacagt

ttgctaatttctttacttcgttctttgctaccggtgaacagggtacgcgggtatttggctacgt
aacgacaatgggcgaattacttaacgcctcgattatgttctttgcgccactgatcattaatcgc
atcggtgggaaaaacgccctgctgctggctggcactattatgtctgtacgtattattggctcat
cgttcgccacctcagcgctggaagtggttattctgaaaacgctgcatatgtttgaagtaccgtt
cctgctggtgggctgctttaaatatattaccagccagtttgaagtgcgtttttcagcgacgatt
tatctggtctgtttctgcttctttaagcaactggcgatgattttatgtctgtactggcgggca
atatgtatgaaagcatcggtttccagggcgcttatctggtgctgggtctggtggcgctgggctt
caccttaatttccgtgttcacgcttagcggccccggcccgctttccctgctgcgtcgtcaggtg
aatgaagtcgcttaa

>J5_J23117_**RBS-B**_lgtA
TGGCAATTCCGACGTCTATACATCGCATCACTACACTATTGATTATCATTGTGTACGTAACGAG
CTTGCACAACGTGAAGTTCTTCGAGCACTTCAGCTCGCAACGTAAATGACAGTTGCTGTTAAGT
GACGTGAATCCTTCAATGCTGCTCATGCTGCTGTCGTAAATAAGTAAGTCACTCCCACttgaca
gctagctcagtcctagggattgtgctagc**AACTGGTAATTTGAGGAGGTAATTT**ATGCCGTCGG
AAGCCTTTCGTCGCCATCGTGCGTACCGTGAAAACAAACTTCAGCCTTTGGTGTCTGTTTTGAT
CTGCGCGTATAATGTAGAGAAGTATTTTGCACAGAGTCTGGCTGCGGTTGTGAACCAAACCTGG
CGTAACTTAGACATCTTAATCGTCGATGACGGAAGTACAGATGGCACGTTGGCGATCGCACAAC
GTTTCCAAGAGCAAGATGGTCGCATCCGTATCTTAGCCCAGCCACGCAATTCGGGACTTATCCC
ATCTTTGAACATTGGACTTGATGAACTGGCTAAAAGTGGCATGGGAGAATACATTGCGCGCACT
GATGCGGATGACATTGCAGCGCCCGACTGGATCGAGAAAATCGTGGGAGAGATGGAGAAAGATC
GCTCTATCATCGCTATGGGAGCATGGTTAGAGGTTCTGTCAGAGGAGAAGGACGGCAACCGTTT
AGCTCGTCACCATGAGCATGGAAAAATCTGGAAAAAACCCACCCGCCACGAAGACATCGCCACG
GTCTTCCCATTCGGTAACCCAATCCACAATAATACAATGATTATGCGTCGTAGCGTGATTGATG
GAGGTCTTCGCTACAACACTGAGCGCGATTGGGCCGAGGATTACCAGTTTTGGTATGACGTGAG
CAAGTTAGGCCGTCTGGCTTACTATCCAGAAGCGCTGGTGAAGTACCGTCTTCATGCGAACCAG
GTGTCAAGTAAATACAGTGTACGTCAACATGAAATCGCACAGGGTATCCAAAAAACAGCGCGCA
ATGATTTTTTGCAGAGCATGGGTTTTAAAACACGCTTCGACAGCTTGGAGTATCGCCAAATCAA
AGCCGTAGCCTATGAATTACTGGAAAAGCATCTGCCGGAGGAAGATTTTGAGCGTGCCCGCCGC
TTTTTATACCAATGTTTTAAACGCACCGACACCCCACCAGCTGGCGCCTGGCTTGATTTTGCCG
CCGACGGCCGTATGCGTCGCTTATTTACTTTGCGCCAATATTTCGGTATCTTACGCCGTCTGTT
GAAAAATCGCTGA

>J6_J23117_**RBS-C**_galT
cctgcacgagttcgctgtcgagacaagtctcttagcgacgtattacgaagatcacatagtcaga
tgaagctatagagcacgacgctaacgattacgtcacgcttgacacaacagtttcgctacctagt
gctcgcgcgactgcgacgttgtccttctagtcgcccatgactcttgacagctagctcagtccta
gggattgtgctagc**AATCTCATAAATCAAATATAGGGAGGATCAT**ATGGACACCATCATGATTA
AACGTCCGCTGGTTAGCGTTATTCTGCCGGTGAATAAAAACAATCCGCATCTGGAAGAAGCAAT
CCAGAGCATTAAAAACCAGACCTATAAAGAGCTGGAACTGATCATTATTGCCAACAACTGCGAG
GATAACTTTTATAGCCTGCTGCTGAAATATCAGGACCAGAAAACCAAAATTATCCGCACCAGCA

TCAAATATCTGCCGTTTAGCCTGAATCTGGGTGTTCATCTGAGCCAGGGTGAATATATTGCACG
TATGGATTCAGATGATATCAGCGTTCTGGATCGCATTGAAAAACAGGTTAAACGCTTTCTGAAT
ACACCGGAACTGAGCATTCTGGGTAGCAATGTTGAATATATCAATGAAGCCAGCGAAAGCATTG
GCTATAGCAACTATCCGCTGGATCATAGCAGCATTGTTAATAGCTTTCCGTTTCGTTGTAATCT
GGCACATCCGACCATTATGGTTAAAAAAGAAGTGATTACCACGCTTGGTGGCTATATGTATGGT
AGCCTGAGCGAAGATTATGATCTGTGGATTCGTGCAAGCCGTCATGGCAATTTCAAATTTAGCA
ATATTGATGAACCGCTGCTGAAGTACCGTATTCATAAAGGTCAGGCAACCAATAAAAGCAACGC
CTATAACATCTTTGCCTTTGATAGCAGCCTGAAAATCCGTGAATTTCTGCTGAATGGTAATGTG
CAGTATCTGCTGGGTGCAGCACGTGGTTTTTTTGCATTTCTGTATGTGCGCTTCATCAAAAAAT
GA

Acknowledgements

I am incredibly grateful to have been able to spend the last five years contributing to projects that satisfy both my scientific curiosity and my interest in societally impactful work, an opportunity I do not take for granted.

I would like to thank my advisors, James Carothers and Jesse Zalatan. From James, I learned the importance of storytelling and selling the “big picture”. Jesse taught me to be rigorous in my questioning and thoughtful in my experimental design. Both have greatly improved my scientific writing and thinking. While getting pulled in two different directions was difficult at times, it taught me to focus on the problems that I found important and strengthened my scientific judgement.

I am also especially grateful to Jorge Marchand, who has served as a mentor and an unofficial third advisor throughout my PhD. It was by Jorge’s LC/MS that I spent countless frustrated, but productive hours and cut my teeth in analytical chemistry. I appreciate your generosity with both your time and guidance.

There are several other scientific mentors who I would like to thank:

- Mrs. Boyes, my high school chemistry teacher, who encouraged me to apply for a research internship program at the local state college, and helped set me on this path.
- Professor Jonathan Long, my first PI at Stanford, who gave me my first real research experience, and gave me more independence than I probably deserved at that time as an undergrad who had just learned how to pipette.
- Gabe Reder, who began as a TA and became an important advocate and mentor to me, who encouraged me to pursue graduate school and gave me the confidence that I could be successful there.
- Savannah Braden, and the entire Apeel Sciences team, including Cody Vild, Matthew Kahlscheuer, and Wade Ingram, among others, who taught me the importance of crafting a story in science, and that scientists could be cool and fun.
- Orlando de Lange, who taught me the value of being intentional with teaching and mentorship. Thank you for the excellent advice you’ve given me throughout my PhD about lab culture and mentorship. Thank you for encouraging me to pursue more opportunities in education and outreach through SoundBio, which became a defining aspect of my PhD.

To the many labmates and friends that I have had the opportunity to work with and get to know over the years, thank you for creating an incredibly supportive and collaborative scientific environment, you have all made just showing up to work every day more fun. This includes, Michael Guzman, Ava Karanjia, Widi Sugianto, Jake Brandner, Janis Shin, Tommy Primo, Brian Darst, Kira Olander, Yejun Kim, Yujia Huang, Amanda Robert, Stephen Fedak, Noel Jameson, Jess Caruso, Nidhi Mehta, Brianne King, Morgan Bean, Hayden Henry, Karl Anderson, and many others.

I want to give special thanks to Diego Alba, who I'm happy to share one brain with and am excited to work with for what is hopefully many years in the future. Thank you to Ben Tickman, who challenged me to deepen my understanding of the science i was working on, to keep asking questions, and try to have a new idea every day. Most of the work in this thesis was born out of conversations with him. Thank you to Ian Faulkner and Ice Kiattisewee, who have laid the foundations for much of this work and have answered one billion stupid science questions from me over the last several years.

I've had the privilege of mentoring and collaborating with some incredible students, including Stella Anastasakis, Sam Dyer, Maggie Cook, Kieran Heiberg, and Juliana Beall, among others. Working with you has been among the most rewarding parts of graduate school and has helped me grow as both a scientist and a teacher.

Thank you to the SoundBio Lab community. I would encourage all graduate students, and scientists at any stage in their career, to challenge themselves to do more to invest in their community and the next generation of scientists. Conducting research is such a small part of the job - communicating science is a big part of it, but so is training the next generation. Science is a slow process and is often built on the idea that incremental change over time amounts to meaningful impact. This whole thing only works if we invest in the next generation who will keep moving the ball forward.

Thank you to my amazing friends for their support. One of the most important lessons I've learned during my PhD, and am still learning, is how to balance work with all the other joys of life. The friends I have made in Seattle have consistently given me a good excuse to leave work early or skip Fridays, for which I am extremely grateful. Also, a huge shout out to all of my roommates, Brandon, John, Luciano, with "real jobs" who have subsidized countless meals (and my rent) over the years. Your sacrifices have not gone unnoticed. A massive, sincere thank you to Marco who has been with me on this journey since the beginning.

Thank you Alice, not just for accompanying me on countless visits to the lab on weekends and evenings, but for always being so incredibly supportive and interested in my work. Countless times over the last three years, you have patiently listened to me

complain about how the mass spec isn't working again, and how it is broken but in a different way than last time, enough times that you could probably fix one yourself at this point. Thank you for all of your love and support and patience.

Thank you to my moms, Amy Levine and Suzanne Cardiff, for always being so supportive of me and doing everything in your power to help me reach my goals. For fostering an environment that encouraged questioning conventional beliefs and asking challenging questions. I love you very much and couldn't be here without you. Thank you to the rest of my family: my grandfather Herbert Levine, my late grandmother Linda Levine who we miss everyday, David Levine, and my brothers, Thomas and Joshua Kline, and their beautiful families.