Competition-based CRISPR-dCas9 transcriptional control mechanisms and application of dCas9 biosensors for high-throughput, cell-based protease inhibitor screens

by

Daniel Allen Anderson

B.A. Molecular and Cell Biology

University of California, Berkeley, 2016

Submitted to the Department of Biological Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Signature of author……

Daniel A. Anderson

Department of Biological Engineering

June 28, 2021

Certified by……

Christopher A. Voigt

Professor of Biological Engineering

Thesis Supervisor

Accepted by……

Katharina Ribbeck

Professor of Biological Engineering

Chair of Graduate Program, Department of Biological Engineering

Competition-based CRISPR-dCas9 transcriptional control mechanisms and application of a dCas9 biosensor for high-throughput, cell-based protease inhibitor screens

by

Daniel Allen Anderson

Submitted to the Department of Biological Engineering on June 28, 2021 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Engineering

Abstract

Catalytically-dead Cas9 (dCas9) is a programmable transcription factor that can be targeted to promoters through the design of small guide RNAs (sgRNAs), where it can function as an activator or repressor. In Chapter 1 of this thesis, I outline the multitude of tools and applications that have been developed for dCas9 circuits. I then discuss the limitations and advantages of these systems and and outline some of the most promising opportunities for dCas9-based genetic circuits.

In Chapter 2, I devise, model, and implement a new-to-nature transcriptional control mechanism using dCas9. Natural promoters use overlapping binding sites as a mechanism for signal integration, where the binding of one transcription factor can augment the activity of another. Here, I implement this strategy in *Escherichia coli* using pairs of sgRNAs designed to repress and then derepress transcription through competitive binding. I demonstrate that this mechanism can control both transcriptional initiation and transcriptional elongation with over 30-fold dynamic range. This work characterizes and demonstrates a new genetic control modality that could be used to build analog circuits or to implement cis-regulatory logic on CRISPRi-targeted native genes.

In the final chapter of this thesis, I use a dCas9 genetic circuit to create an *in vivo* selection system for protease inhibitors. By leveraging a previously-described dCas9 toolkit, I create a synthetic genetic circuit that responds to SARS-CoV-2 viral protease activity. Using this circuit as an *in vivo* biosensor, I integrate it with a RiPP-based molecular library and an *in vivo* selection system to screen for inhibitors of the SARS-CoV-2 Papain-like protease (PLpro). With this integrated system, I screened tens of millions of RiPPs and identified DAA680, a 13-AA modified peptide with PLpro inhibitory activity. However, follow-up studies showed that this peptide also inhibits another SARS-CoV-2 viral protease, CLpro, indicating a non-specific mechanism of inhibition. Nonetheless, these results validate our system's ability to identify and isolate RiPP-based protease inhibitors from large libraries. Additionally, our extensive characterization of the selection system should be generalizable to any biosensor with a transcriptional output. This should enable the rapid deployment of novel cell-based selection methods that can identify molecules with diverse bioactivities.

Thesis Supervisor: Christopher A. Voigt

Title: Daniel I.C. Wang Professor of Advanced Biotechnology in the MIT Department of Biological Engineering

Contents

List of Figures

1 Chapter 1: Perspectives and Opportunities for dCas9-based Genetic Circuits

1.1 Introduction to CRISPR-Cas Systems

CRISPR-Cas systems are bacterial viral defense mechanisms that have been coopted by humans to create a variety of new, revolutionary biotechnologies. Most of the fundamental work in CRISPR-Cas systems have focused on the Type II CRISPR-Cas system from *Streptococcus pyogenes*, which (unless otherwise noted) will be the primary CRISPR system discussed here. While originally a 3-component system comprising two RNA species and a protein, early engineering efforts on CRISPR-Cas systems, fused the two RNA species into a small guide RNA (sgRNA) [1]. This created a two-component system where a sgRNA can be programmed to direct a protein (Cas9) to cleave any DNA sequence with a nearby NGG protospacer adjacent motif (PAM) sequence [1]. While this programmable cutting mechanism was clearly enabling for gene editing efforts, an entirely new class of applications originated from the creation of a catalytically-dead double mutant (D10A, H841A) version, dCas9 [2, 3]. This double mutant retains binding activity, but no longer cleaves DNA [2, 3]. This turned Cas9 into a programmable, DNA-binding system which can now interface with transcription, a fundamental signal type in biological systems.

1.2 Transcriptional control with dCas9

1.2.1 CRISPRi

The fundamental mechanisms of dCas9-based transcriptional control can be thought of in two conceptual bins: transcriptional inhibition (CRISPRi) and transcriptional activation (CRISPRa)[2, 4]. In CRISPRi, dCas9 binding reduces an otherwise-high transcriptional signal. In bacteria, this can be accomplished in one of two ways: either by directing dCas9 to the core of the promoter, thus preventing RNA polymerase (RNAP) binding (transcriptional initiation control); or by directing it downstream of a transcriptional initiation site, thus

dislodging elongating RNAPs (transcriptional elongation control) [2, 4]. Notably, elongation control has a polarity-dependence, where termination of RNAP elongation is greatly enhanced when RNAP collides with the PAM-proximal side of a DNA-dCas9 complex [2, 4, 5]. The underlying mechanism for this phenomenon is poorly-understood, but well-replicated [6-9].

Target-sgRNA mismatches have been shown to have a negative effect on transcriptional initiation and elongation inhibition [2, 5]. Biophysical characterization of these mismatch-induced effects point to multiple different mechanisms for dCas9's reduced binding capacity, depending on the mismatch location [10]. In elongation control, mismatches have been proposed to increase the probability of dCas9 dislodgement by RNAP in *E. coli* [5]. This ability to titrate the repressive effect of sgRNAs with simple mismatches has also been harnessed in high-throughput CRISPRi-based screens to elucidate dose-responses of genes in both bacterial and mammalian systems [11, 12].

In yeast cells, dCas9 binding alone is enough for transcriptional repression, however the fusion of a chromatin-remodeling protein Mxi1 results in greater silencing effects [13]. Lastly, mammalian cells also show efficient CRISPRi, but this requires the fusion of the repressive chromatin modifier protein, KRAB, to dCas9 [13].

1.2.2 CRISPRa

Due to the modular nature of transcription in eukaryotes, CRISPRa is largely a solved problem. Eukaryotic CRISPRa is usually accomplished through targeting of a dCas9-VP64 fusion to promoter regions [14]. However, in bacterial systems, CRISPRa is inherently more difficult to execute than CRISPRi. Initial efforts fused the ω-subunit of RNAP (RpoZ) to dCas9 and directed it to the upstream region of a promoter. However, this system required endogenous *rpoZ* knockouts and weak promoters for good fold-activation [4]. More recently, different activating domains, AsiA and SoxS, have been shown to be capable of large fold-activations without requiring endogenous gene knockouts [15, 16]. RpoZ, AsiA, and SoxS activation domains generally only increase transcriptional flux through sigma-70 promoters, which make up the majority of endogenous *E. coli* promoters [17]. However, a recent effort has shown that recruitment of different activation domain, PspF, by dCas9 enables efficient activation of sigma-54 promoters in *E. coli* and *Klebsiella oxytoca*, thus expanding the

capabilities of the bacterial CRISPRa toolkit [16].

Many of the advances in bacterial CRISPRa have been due to the elucidation of strict spacing requirements for activation in bacteria [14, 16, 18, 19]. And, while dCas9 is highly programmable, it can only be targeted to sequences adjacent to an NGG PAM site. While synthetic promoters can be designed with specific NGG locations, this PAM constraint can prevent efficient activation of endogenous promoters with RpoZ, SoxS, and PspF (AsiA doesn't seem to have such strict distance constraints) [15, 16, 18, 20]. There have been several different solutions for this problem. Both SoxS and PspF CRISPRa systems have been adapted to use dxCas9(3.7), an engineered dCas9 with reduced PAM constraints [18, 19, 21]. Additionally, linker length adjustments and the use of circularly-permuted dCas9 to reorient fusion proteins have been shown to successful in getting around these constraints [20, 22]. Both of these distance-constraint solutions highlight one of the strengths of building tools on top of the CRISPR platform: that you can often leverage the new advancements that are being made across diverse biological engineering disciplines.

1.3 Diverse signal integration methods of CRISPRi/a systems

If the aforementioned CRISPRi/a mechanisms are thought of as controlling a transcriptional output signals, then there are a variety of signal types that can act as inputs into CRISPRi/a systems. The most obvious inputs that can feed into CRISPRi/a systems are those of sgRNA and dCas9 transcription. In exponentially-growing *E. coli*, the titration of sgRNA or dCas9 (and presumably the dCas9-sgRNA complex) shows a non-saturating, log-linear output promoter response, indicating a non-cooperative binding mechanism with a low dissociation constant [23, 24]. Beyond simple expression-level control, other signal types can be integrated through various mechanisms that modify the activity of CRISPRi/a systems. These different inputs can be thought of as interfacing at one of three distinct points: controlling sgRNA activity, controlling dCas9 activity, and modulating effector recruitment to bound dCas9.

1.3.1 sgRNA-based control

RNA is a versatile biopolymer that has been exploited by both nature and engineers to integrate signals into biological systems [25]. Given that sgRNAs are just another RNA species, many of those mechanisms can also be harnessed to act as inputs into CRISPRi/a systems. Small molecule input integration has been enabled by RNA ribozymes and aptamers that self-cleave or fold into an inaccessible states in response to the presence / absence of ligands [26, 27]. Additionally, arbitrary oligonucleotide inputs can be sensed through toehold-mediated and Small Transcription Activating RNA (STAR) control of sgRNA activity and expression [28, 29]. Lastly, sequence-specific antisense RNAs can be used to target sgRNAs and derepress the effects of CRISPRi in *E. coli* [30].

1.3.2 dCas9-based control

Another location for signal integration is the dCas9 protein itself. The protein is amenable to split-protein methods of inducible control and different dimerization domains have been fused to enable light [31] and rapamycin-inducible [32] CRISPRa systems in mammalian cells. The introduction of protease-cleavable, constrained linkers inhibiting dCas9 activity has enabled the sensing of protease activity in both bacterial and mammalian systems [22]. Lastly, the recent discovery of AntiCRISPR proteins, a viral counter-countermeasure, has created new methods of modulating dCas9 functionality through a variety of different inhibition mechanisms [33]. One of the most potent AntiCRISPRs, AcrIIA4, has also been adapted into a photoactivatable form, creating another handle for light input into CRISPRi/a systems [34].

1.3.3 Effector recruitment

The final method for signal integration is at the stage of transcriptional effector recruitment. As previously mentioned, CRISPRi/a transcriptional regulation is dependent on the recruitment of protein effectors in all cases except bacterial transcriptional repression. While some CRISPRi/a designs simply fuse these effectors directly to dCas9 [13, 16, 19, 35],

other designs make the recruitment of these effectors dependent on other signal. Fusing ligand-dependent heterodimerization domains to both dCas9 and the effector protein enables sensing and even simple logic computation for small molecules [36, 37]. Small molecule inputs have also be integrated through the attachment of ligand-stabilized degron domains to the transcriptional effectors [38].

1.4 CRISPRi/a-based synthetic genetic circuits

1.4.1 Feedforward logic circuits

Through composition of the aforementioned tools and mechanisms, engineers have been able to create complex signal processing circuits. Early CRISPRi circuits utilized feed-forward designs and demonstrated the generation of 2-input logic functions through the composition of Boolean-complete NOT and NOR gates in *E. coli* [24]. This approach was also transferred to a non-model bacterium, *Bacteroides thetaiotaomicron* [39]. As this method was scaled up in *E. coli*, signal attenuation occurred with more than few sgRNAs or cascade layers, due to a lack of cooperativity and retroactivity from a shared dCas9 resource [23]. This effect was also observed in yeast-based CRISPRi circuits, but to a lesser extent, where it was shown that signal propagation was possible through a seven-layer circuit with seven different sgRNA species (the largest yeast-based circuit to date) [40]. Despite the limitations in signal propagation for CRISPRi circuits, the largest circuits to date were recently demonstrated in *E. coli*: a digital MUX / DEMUX genetic circuit pair [41]. This MUX / DEMUX circuit pair, with 10 and 7 sgRNAs, respectively, was split between "sender" (MUX) and "receiver" (DEMUX) strains. The logic of the circuits allowed the sender strain to communicate one of its two input signals to the receiver strain along a single "chemical wire" (AHL produced by LuxI). The receiver strain would then "demultiplex" the signal by directing it to the appropriate one of its two outputs. An additional, shared "selector" signal would indicate to both the sender and receiver cells which of its two inputs the sender should transmit. Although a 2:1 MUX / DEMUX system doesn't truly decrease the number of chemical signals required to communicate two multiplexed signals (since a selector chemical signal must be used alongside the transmitted signal), this system may be necessary when there is only one inter-system signal available. Notably, the longest path through their MUX

/ DEMUX system was 8 layers (across 2 cell types and an AHL-based chemical transmission), which by some metrics is the "deepest" synthetic genetic circuit demonstrated to date.

1.4.2 Feedback and dynamical circuits

In addition to standard feedforward architectures, CRISPRi mechanisms have recently been used to recreate classical transcription factor (TF)-based circuits such as the genetic toggle switch and the 3-ring oscillator in *E. coli* [42-44]. Although previously thought to be impossible due to dCas9's lack of cooperativity, recent modeling efforts [45], and—most-convincingly—experimental demonstrations, have shown that CRISPRi can be harnessed for these types of circuits. An interesting comparison arises from the fact that three independent groups have each attempted to create a CRISPRi-based oscillator circuit [42-44]. Schaerli *et al.* and Silver *et al.* both utilized a 3-sgRNA ring-repression topology, and, despite having similar designs, the Schaerli *et al.* circuit provided long-term oscillations while the Silver *et al.* circuit did not. This points to the apparent fragility of CRISPRi-based oscillation circuits, which has also been indicated theoretically [45]. The third oscillator was designed by Simmel *et al.*, and integrated a single CRISPRi node into the standard 3-node repressilator topology [46]. Their circuit demonstrated robust, periodic oscillations. However, there were also periodic growth fluctuations likely due to the fact that they chose to control dCas9 expression rather than sgRNA levels within the circuit [44].

In addition to oscillators, CRISPRi/a circuits have also been implemented to create incoherent feedforward loops (iFFLs) in both bacterial and mammalian systems. This has allowed the generation of interesting functionalities like pattern formation and pulse-generation [28, 33, 47]. One of the most-impressive recent demonstrations was the simultaneous use of two iFFL systems, 6 sgRNAs in all, to create two independent concentration bandpass filters within a single *E. coli* strain [42].

1.4.3 Mixed-modality circuits

A unique aspect to CRISPRi/a systems is the fact that the same dCas9 protein can act as both a repressor and an activator within a given cell. This enables the use of multiple

computational modalities with little additional genetic material. This multi-modal approach has been demonstrated in bacteria using both the AsiA-based and SoxS-based CRISPRa systems. In both of these examples, simultaneous activation and repression of different genes was conducted by targeting dCas9 to an upstream region and the core region of a promoter, respectively [15, 16]. The multi-modal approach must be modified for eukaryotic systems, where transcriptional activation and repression are both dependent on effector proteins being recruited by the dCas9. In this case, mixed-modality circuits can be accomplished by encoding the transcriptional effector protein type (either activator or repressor) via orthogonal RNA aptamers in the sgRNAs. This has been shown to allow simultaneous activation and repression of genes within mammalian cells [14].

Another recent example of mixed computation comes in the form of multiple sgRNAs targeted to overlapping regions of a single promoter in *E. coli* [48]. Using the RpoZ-based CRISPRa system, Tan *et al.* harnessed a library of different repression sgRNA/activation sgRNA pairs. For each pair, they surveyed different combinations of: binding location and sgRNA expression. Across library members, they observed differential mean expression from the promoter (about 10-fold overall) and differential noise values (30-50% change for each mean expression level). From this they derived a set of empirical rules they used to predict different combinations of sgRNAs that can produce the same mean expression but different noise and vice versa. Due to the empirical nature of their derived rules, the broad applicability of these results remain uncertain.

Unlike CRISPRi-based systems, the design principles for composing mixed-modality circuits into large synthetic gene systems is not obvious. Ultimately, these systems may serve as smaller subcircuits, providing specialized computational abilities in resource-efficient manners through the reuse of dCas9.

1.5 Limitations and Possibilities for CRISPRi/a circuits

When using CRISPRi/a to build synthetic genetic circuits, comparisons are often made to TF- based circuits designed with protein effectors such as LacI, TetR, and cI [46, 49, 50]. CRISPRi/a circuits differ from these circuits in several ways, which will be discussed below.

1.5.1 Dynamics

The binding and unbinding dynamics for CRISPRi/a are generally much slower than TFs like LacI. Single molecule *in vivo* studies in *E. coli* have parameterized dCas9's association rate at k_a = 2.7x10^-3 min^-1 (molecule / cell)^-1, which is 100-fold slower than LacI's observed association rate of k_a = 0.3 min^-1 (molecule / cell)^-1 [51, 52]. However, this slow binding rate is offset by dCas9's complete lack of dissociation, which seems to only occur due to dislodgement from DNA replication [52, 53]. In exponentially-growing *E. coli* with a doubling time of 30 minutes, this leads to an effective dissociation rate for dCas9 of k_d = 0.03 min^-1, which is roughly 50-fold slower than LacI's dissociation rate of k_d = 1.32 min^-1 [51, 52]. Notably, if one uses these parameters to calculate the dissociation constant (half-maximal repression), they come to similar values of 11 and 4 (molecules / cell)^-1 for dCas9 and LacI, respectively.

Even though the binding/unbinding dynamics of dCas9 are much slower than standard TFs, the effect of this is minimal under most circumstances. This is due to the fact that in exponentially-growing *E. coli* cells TF-based circuits also converge to their steady-state values based on the cellular division rate. This is because the half-lifes of the effector TF proteins are much longer than the time scale of cellular division. From the perspective of a single cell, this makes dilution the primary degradation mechanism for both TF and dCas9-based circuits [54]. Indeed, the recent demonstration of CRISPRi reproducing the TF-based repressilator circuit indicates the slow dynamics of dCas9 does not preclude its use for highly-dynamical circuits [42, 55]. It should be noted that while many genetic circuit studies utilize exponentially-growing bacterial cells for simplicity, dCas9's non-Markovian binding nature should be acknowledged and modeled appropriately for any CRISPRi/a applications with variable / non-existent growth conditions such as stationary phase circuits [56] or mammalian systems with slow growth rates. Additionally, circuits with faster dynamics like RNA-based or phosphor-relay circuits may require further dynamic consideration.

1.5.2 Scalability

One of the initial purported advantages of CRISPRi/a circuits was the ability to easily create thousands of orthogonal parts, which would allow the scaling of previously part-

limited TF-based genetic circuits. While the creation of thousands of orthogonal parts was borne out, simultaneous use of these parts was confounded by the fact that they all share a common dCas9 resource. It has been shown both mathematically and experimentally that the co-expression of multiple sgRNAs titrates a shared dCas9 resource away from the co-expressed sgRNAs, a phenomenon known as retroactivity [23, 57, 58]. Using maximum non-toxic dCas9 expression levels and a threshold of minimum 10-fold repression per sgRNA, the number of simultaneously-expressed sgRNAs was shown to be limited to seven and 14 gRNAs for a standard dCas9 and a specially-engineered non-toxic dCas9, respectively [23]. Recently, a control theoretical approach had some success in reducing retroactivity effects through integration of a dCas9 regulator alongside a sgRNA-based circuit [57]. In this circuit, the dCas9 regulator is a sgRNA-based negative feedback loop on dCas9 expression, so that as the free pool of dCas9 is titrated away by sgRNA expression, dCas9 expression increases accordingly [57]. This allowed the creating of a two-sgRNA (plus one for the dCas9 controller) circuit that did not show retroactivity. Additionally, theoretical efforts have identified certain CRISPRi circuit motifs that are less susceptible to retroactivity than others [58]. These motifs could be integrated into current circuit design methods to possibly increase the size of CRISPRi circuits. However, the exponentially-scaling nature of the CRISPRi retroactivity problem means that further optimization will likely not make much progress.

Interestingly, a theoretical paper recently proposed that CRISPRa mechanisms are much less susceptible to retroactivity effects than CRISPRi [59]. In this work, Clamons and Murray used the same 10-fold expression criteria mentioned above and found that CRISPRa systems should be able to simultaneously handle dozens to hundreds of co-expressed sgRNAs. Despite the computational limitations of activation-only genetic circuits, the prospect of achieving the original scale-up promise of CRISPRi/a systems encourages deeper consideration for large CRISPRa systems.

1.5.3 Cooperativity

Cooperativity is a sought after property in genetic circuits due to theoretical results indicating that it is a prerequisite for complex functionality like toggle switches, oscillations, and digital cascades [49, 60, 61]. While there have been some CRISPRi/a systems engineered to exhibit cooperativity through either fused multimerization domains or fused chromatin

remodeling proteins [23, 40], standard dCas9 exhibits a non-cooperative input / output response due to its single-component binding mechanism [23, 24]. Due to this, many thought that standard CRISPRi/a circuits would not be capable of reproducing these interesting circuit functionalities. However, recently both toggle switches and oscillators have been produced using exclusively CRISPRi components [42, 43]. Additionally, the single-gate assessment during the building of the MUX / DEMUX circuits showed cooperative responses varying from $n = 0.69 - 4.3$ [41]. Several explanations have been proposed for CRISPRi's ability to exhibit some type of hidden cooperativity including: non-specific DNA binding, bound-repressor degradation due to cellular growth, sgRNA titration effects, and deviation from Hill-like performance [42, 43, 45]. These phenomena point to a theoretical and experimental blind spot in our understanding of both CRISPRi/a mechanisms and the requirements for cooperativity, which merits further exploration.

It should also be noted that the apparent non-cooperativity of CRISPRi/a systems is not always a negative attribute. Its graded, non-saturating response makes CRISPRi a great candidate for feedback control, where it can be used without concern of controller saturation [57, 62, 63]. Similar requirements for graded control have had to utilize TALE proteins, which much be specifically tailored for their binding sequence [64].

1.5.4 Transferability

The fact that CRISPR-Cas systems seem to work readily across every domain of life coupled with their highly programmable nature makes them promising candidates for transferring synthetic genetic circuit functionalities into non-model organisms [65-69]. In an ideal scenario, one would be able to develop a synthetic genetic circuit for a given organism and then be able to easily transfer that circuit to other organisms sharing similar transcriptional regulation methods (e.g. within gram-negative bacteria). With standard TF-based circuits this usually requires rigorous elucidation of promoter structures and constraints involving time-intensive operator placement scanning for each gate [70-72]. CRISPRi/a approaches also require similar scanning efforts, but this is readily multiplexed through sgRNA libraries [11, 12, 16, 18, 19, 73]. Although building complex synthetic transcriptional networks requires consideration of many other factors, including genetic context insulation and gate composability, the ease of transcriptional control gained through

CRISPRi/a is fundamental starting point for the transfer of synthetic genetic circuits [72, 74].

One of the benefits of CRISPRi/a for transferrable synthetic circuits is that once a suitable synthetic promoter architecture is found, it can be readily scaled up by simply diversifying the sgRNA target sequence [18, 19, 23, 24, 39, 40]. With deliberate sgRNA sequence design, these promoters are almost assured to be orthogonal. While the simple biophysical mechanisms of bacterial CRISPRi make it readily transferrable across bacterial species, the complexities of CRISPRa have caused its transfer to lag behind. However, recent work has demonstrated various CRISPRa systems being used across diverse bacterial species including: *Escherichia coli, Pseudomonas putida, Salmonella enterica, Klebsiella oxytoca, Lysobacter enzymogenes, Myxococcus xanthus, and Bacillus subtilus* [4, 16, 18, 75-77]. This points to the potential of both CRISPRi and CRISPRa mechanisms for transfer to non-model organisms.

1.6 Future Perspectives

Recent results indicate that at 10 sgRNAs, we're likely near the limit for scaling CRISPRi-based circuits [41]. Furthermore, the exponentially-scaling nature of retroactivity makes it unlikely that further optimization will provide many advances [23, 41, 57]. We propose that, instead of thinking of CRISPRi/a as a substitute for TF repressors in large synthetic genetic circuits, it may be better to think of harnessing CRISPRi/a in subcircuits that exploit their strengths. One option is to create hybrid synthetic genetic circuits where the primary computation is done through a scalable transcriptional control mechanism and CRISPRi/a is limited to a final actuation layer that interfaces with endogenous systems. This approach exploits the programmability of CRISPRi/a, and has already been used to great effect in several circuits [24, 44, 78, 79]. Another option is to implement CRISPRi/a when a graded, non-saturating response is needed, such as in feedback control [57, 62].

In addition to the ones discussed here, there are many other emerging CRISPRi/a technologies whose potential still need to be assessed and explored. Recent advances in bacterial CRISPRa showing large fold-changes across multiple species [15, 16, 18], along with theoretical results showing that CRISPRa can possibly scale to hundreds of co-expressed sgRNAs [59], points to a possible route for large-scale CRISPRa-based circuits. However, since much of the foundational work for scaling genetic circuits depends on signal inversion

by repressors [50, 80], new frameworks and approaches will need to developed for activator-based systems. Additionally, homologous CRISPR-Cas proteins like dCas12a are beginning to be applied in synthetic genetic circuits [43, 81]. The advantages and disadvantages of dCas12a relative to dCas9 have not been fully elucidated. However, its unique properties have been used to create an RNA-sensitive CRISPRi circuit [82], and its use in an oscillator circuit proved notably better than an identical circuit made with dCas9 [43].

There are many different, interesting avenues for exploration and expansion of CRISPRi/a genetic circuits. And while CRISPR-Cas may not have immediately revolutionized synthetic genetic circuits like it did the field of genetic editing, we anticipate that it will ultimately find a substantial and useful niche.

2 Chapter 2: Competitive dCas9 binding as a mechanism for transcriptional control

2.1 Main Text

2.1.1 Introduction

Regulatory networks integrate environmental and cellular signals to ensure genes are expressed under the correct conditions. Integration can occur at individual promoters through the arrangement of DNA operators to which regulatory proteins bind and either recruit or interfere with transcriptional machinery [83, 84]. Within a promoter, the binding of one regulator can also affect the binding of others through positive or negative interactions; for example, through DNA looping or overlapping operators [85-95]. This collectively generates a "cis-regulatory logic," which dictates the conditions for gene expression based on the combination of regulators that are active [84, 96-103]. For synthetic genetic circuits, the placement of multiple operators within a promoter has been used to implement combinatorial logic, for example to create an X AND (NOT Y) gate for an edge detector by placing an operator for an activator (X) and repressor (Y) within a promoter [104-114]. As a design principle, combining operators is useful for "compressing" large logic operations to reduce the resource burden [115]. It is challenging to insert an operator into a promoter because the change in sequence can affect the promoter strength, an effect that can be mitigated with insulators and computational predictions [70, 116-120]. Prokaryotic promoters are small, thus making it difficult to fit multiple operators and it is difficult or impossible to design operators that bind to multiple regulatory proteins [70, 121-123].

dCas9 can be programmed to bind to different DNA sequences by changing the targeting sequence of its bound sgRNA [2-4, 124]. DNA binding requires 5-10 bp sgRNA-DNA complementarity adjacent to an NGG PAM sequence [10], although dCas9 can be engineered to remove the PAM sequence requirements [21, 125-128]. When bound to DNA, dCas9 covers 30 bp and melts the strands to form a bubble that results in very slow unbinding rates [52, 53]. Thus, it can function as a repressor by sterically blocking the binding of RNA polymerase (RNAP) to a promoter or by blocking RNAP progression through a gene [2]. When targeting constitutive *E. coli* σ70 promoters, repression is strongest when the sgRNA is targeted within

the -10 to -35 promoter core and does not depend on which strand is targeted [2, 4, 24]. Targeting dCas9 downstream of a promoter represses transcriptional elongation, with much stronger repression observed when the sgRNA is targeted to the non-template strand [2, 4]. dCas9 can also serve as an activator by fusing it to an activating domain that recruits RNAP [4, 14-16, 129]. There are often tight spacing requirements for activation, particularly for prokaryotic promoters. Lastly, dCas9 can also be directed to bind to the operators of protein repressors, activators or enhancers, thus blocking their impact on expression [90, 130, 131].

An advantage of using dCas9 is that its regulatory effect can be directed to a promoter without having to insert an operator sequence. This has been used to control the regulation of native genes, for example, controlling enzymes at branch points to redirect flux through a metabolic pathway [11, 63, 78, 132-134]. It also simplifies the integration of multiple signals at a single promoter by designing sgRNAs that target it to different positions. For instance, NOR gates have been constructed using two sgRNAs that target different positions in the output promoter, either of which leads to repression [40, 135]. When multiple sgRNAs target overlapping regions, this leads to mutually exclusive binding [2]. Tan and co-workers harnessed this effect to control the strength and noise of an *E. coli* constitutive promoter by co-transcribing different ratios of two sgRNAs that direct dCas9 (fused to an activating domain) to overlapping positions that either recruit or block RNAP [48].

Once expressed, the regulatory effects of dCas9 only end when the protein or sgRNA degrade or are diluted by cell division. Several approaches have been taken to control the activity of either dCas9 or the sgRNAs after they are expressed. One is to express antiCRISPR proteins derived from phage genomes that bind to and inactivate dCas9 [33, 34]. This leads to the complete inactivation of dCas9, thus eliminating its ability to implement any sgRNA-mediated regulation in the cell. Different sets of genes can be controlled by expressing orthogonal dCas9 variants, each of which binds a different set of sgRNAs [35, 36]. These can be changed dynamically from being repressors to activators by expressing the corresponding domains as separate proteins that bind to dCas9 using modular protein-protein interaction domains [36]. Another approach is to design an RNA to bind to and augment the activity of a specific sgRNA. Antisense RNAs will inactivate sgRNAs by targeting them for degradation via the native bacterial Hfq system [30]. The sgRNA can also be designed to fold into an inactive hairpin, thus requiring the co-expression of toehold RNAs to unfold and bind to dCas9 [29, 136, 137]. Both of these techniques require modifying the

sgRNA to have additional sequences such that it can be bound by the modulating RNA.

While there are many natural examples of different repressors binding to the same operator or a repressor displacing an activator, to our knowledge there are no examples of a repressor's action being negated by a second protein binding to an overlapping operator. To this end, we have developed a mechanism for sgRNA-specific derepression through competitive dCas9 binding to overlapping regions. Transcription is blocked when a repressing sgRNA (sgR) directs dCas9 to the first position and a second derepressing sgRNA (sgD) directs dCas9 to a mutually-exclusive second position that does not impact transcription (Figure 2.1-1a). We find that the two regions are competitive so long as they are within 14 bp and the PAM sequences are between the target sites. The repression/derepression switch is implemented in two ways (Figure 2.1-1b). The first is to design the sgR to overlap the -35 σ70-binding region of promoter and the sgD to bind just upstream. The second approach is to design repression/derepression sites within a gene by exploiting the strand-dependence of RNAP elongation inhibition by dCas9. These results demonstrate a new mechanism to control the activity of dCas9-directed regulation that could be used for efficient genetic circuit design to integrate signals in a promoter or to derepress subsets of native genes to subregulate genome-encoded functions.

2.1.2 Results

2.1.2.1 Promoter control through repression/derepression

dCas9 can repress a promoter by sterically blocking the binding of RNAP (Figures 2.1-1b, 2.1-2a) [2, 4]. An *E. coli* constitutive promoter can be repressed by targeting the repressing sgRNA (sgR) to overlap the -10 or -35 σ70-binding sites [2, 4, 24]. Our design for derepression is based on using a second sgRNA (sgD) to recruit dCas9 to an upstream site that blocks its ability to bind to the repressing position. Importantly, sgD cannot interfere with RNAP binding or else it will also lead to repression. To quantify the impact of dCas9:sgRNA binding on mRNA transcription rates, we developed a parameter α, defined as the ratio of maximal mRNA production rates (completely unbound DNA) to the production rate when all DNA is bound by dCas9:sgRNA (saturated DNA). The ideal repression / derepression sgRNA pair

location would have a large α value for sgR (α >>1) and α→1 for sgD, indicating strong repression by sgR and no repressive effect by sgD.

A system was designed to evaluate how targeting sgR and sgD to different positions in the promoter impacts the effectiveness of repression and derepression. Two inducible promoters were used to independently transcribe sgR (IPTG-inducible P_{Tac}) and sgD (Vanillic acid-inducible P_{Van}) encoded on a p15A plasmid. The targeted constitutive promoter (P_{C1}) was built based on the -35 to +1 core of P_{Gate08} [23] to which a randomly-generated 100 bp sequence was added upstream (Methods). Positions in P_{C1} were selected to be targeted for sgR/sgD by exploiting NGG PAM sites in the promoter sequence (Figure 2.1-2b). The sgR and sgD sequences were designed based on the same scaffold [23], with mutations the 20 bp spacer sequence corresponding to the target region of the promoter. To measure promoter activity, P_{C1} was placed upstream of a ribozyme [74] and gene encoding red fluorescent protein (*rfp*). From the same plasmid, dCas9 was expressed using the aTc-inducible P_{Tet} promoter. A concentration of 1.25 nM aTc was used for all experiments and leads to approximately 500 dCas9 molecules per cell during exponential growth [23].

Experiments were performed to measure the promoter activity that results when dCas9 is targeted to different positions. The plasmids were transformed into *E. coli*. The cells were grown in defined media (EZ Rich) and induced with aTc and either IPTG or Vanillic acid (Van) for 5.5 hours (Methods). Fluorescence was then measured using flow cytometry. First, we compared the repression obtained by targeting the -10 (sgR_{P1}) or -35 (sgR_{P2}) positions of the promoter (Figure 2.1-2b). Upon maximum sgR expression (1 mM IPTG), both sgRs are able to repress the promoter by >30-fold (Figure 2.1-2c). This result is consistent with previously observed fold-repressions at these locations [23, 24].

A model was derived to capture the repression of a promoter by dCas9:sgR,

$$P + S_R \overset{K_R}{\Leftrightarrow} P_R \tag{1}$$

where P and P_R are the concentrations of promoters in the unbound and bound state, S_R is the concentration of dCas9:sgR and K_R is the association constant. The production of mRNA transcripts m from the promoter is described by

$$\frac{dm}{dt} = \beta_m \frac{1+\alpha^{-1}K_R S_R}{1+K_R S_R} + \beta - \delta_m m \quad (2)$$

where β_m is the maximum transcription rate, β is the leaky transcription rate and δ_m is the degradation rate. Solving for steady-state yields

$$m = \left(\frac{\beta_m}{\delta_m}\right) \frac{1+\alpha^{-1}K_R S_R}{1+K_R S_R} + \frac{\beta}{\delta_m} \quad (3)$$

This equation can be further converted to the activity of the output promoter P_{C1},

$$y = (y_{max} - y_{min}) \frac{1+\alpha^{-1}\kappa_R x_R}{1+\kappa_R x_R} + y_{min} \quad (4)$$

where y is in arbitrary units (au) of RFP fluorescence and y_{max}/y_{min} are the maximum/minimum measured values. The parameter x_R is the strength of the promoter driving the expression of sgR, measured as au of RFP fluorescence and κ_R is rescaled to be in the same units. This assumes that the binding of sgR to dCas9 to form S_R is in the linear (unsaturated) regime. Equation 4 was fit to the sgR_{P1} and sgR_{P2} induction curves (Figure 2.1-2c), and the parameters were extracted (Table 2.1-1).

We then designed experiments to determine the constraints of targeting dCas9:sgD to the promoter without evoking repression (Figure 2.1-2d). Seven positions were selected between the -28 and -66 region of the promoter, including orientations that target both strands (Figure 2.1-2b). The promoter activity was measured for each position when sgD is expressed (100 μM Van) and unexpressed (0 μM Van) and these data were used to estimate α as the ratio of these values (Figure 2.1-2d, e). We observe that α continuously decreases as the sgD targets regions farther from the promoter core, indicating weaker repressive abilities (Figure 2.1-2e). When sgD targets regions upstream of -60, negligible repression is observed.

It has been previously shown that dCas9:sgRNAs will compete for binding when their

target regions overlap [2, 48]. However, the distance constraints we measured showed that it would not be possible to obtain repression/derepression by targeting overlapping regions. Even if we used the upper-bound on TSS distance for the repressing sgRNA (30 bp) and the lower-bound on distance for the derepressing sgRNA (60 bp), the sgRNA target sequences would still have to be 10 bp away from each other. We hypothesized that while we could not utilize sgRNA target overlap, we may still be able to harness steric hindrance effects from overlap of the dCas9 protein footprint on DNA. To investigate this, we examined a crystal structure of the dCas9-sgRNA-DNA complex [138], which shows that dCas9 has an overhang of ~9 bp on the PAM-proximal binding side, but only a ~1 bp overhang on the PAM-distal side of the sgRNA target. This indicated that we might be able to obtain competitive dCas9 binding without overlapping the sgRNA binding regions if we oriented two sgRNAs with their PAM sequences facing each other.

To test this hypothesis, we redesigned the system to express sgR_{P2} from a choline (Chol)–inducible promoter and sgD_{P5} from a Van-inducible promoter (Figure 2.1-2f). These promoters were chosen because they are not predicted to append disruptive 5′ sequences onto the sgRNAs [2, 139]. In accordance with the crystal structure estimation of dCas9's DNA footprint, we designed sgD_{P5} to be 14 bp from sgR_{P2}. To bind sgD_{P5} at this distance, we had to introduce a 3 bp mutation from -60 to -62 of P_{C1} to create a PAM site for sgD_{P5}, resulting in P_{C2}. We screened for derepression by fully inducing sgR_{P2} with 9 mM Chol and titrating the expression of sgD_{P5} from 0 to 100 μM Van. Derepression occurs in a graded manner as more sgRNA is expressed, ultimately returning the promoter activity to its unrepressed state (Figure 2.1-2g).

We then determined the promoter response when different amounts of sgR and sgD are expressed. This response can be viewed as a cis-regulatory logic operation [100], where the signals from these regulators are integrated by the promoter. To obtain this function, 60 combinations of sgR_{P2} and sgD_{P5} induction levels were measured. These data points are shown in Figure 2.1-2h as circles, where their colors are the output values of the circuit and each data point is positioned at the respective promoter activity values for sgR and sgD (error bars for each measurement are provided in Supplementary Figure 2.2-4). Output from the promoter increased with sgD induction and decreased with sgR expression.

It has been shown that the co-expression of multiple sgRNAs titrates a shared dCas9

resource away from the co-expressed sgRNAs [23, 57]. This could complicate the interpretation of the derepression data, where the expression of sgD could deplete the dCas9 pool, thus indirectly reducing the concentration of dCas9:sgR in the cell. To test for this effect, we conducted a control experiment where we maximally expressed a sgR and then titrated in either a non-targeting sgRNA, sgN, (does not bind to the promoter or genome) or an off-target sgRNA, sgO, (binds to an inert region of the plasmid) (Supplementary Figure 2.2-2). Neither of these sgRNAs showed derepression, indicating that the derepression we observed was not due to dCas9 titration.

Our model was then expanded to include derepression. Equation 1 can be modified to include the competitive reaction of dCas9:sgD (S_D) binding to the promoter to form P_D,

$$P_D + S_R \overset{K_D}{\Leftrightarrow} P + S_R + S_D \overset{K_R}{\Leftrightarrow} P_R + S_D \tag{5}$$

where K_D is the association constant between S_D and P. The production of mRNA transcripts m from the promoter is described by

$$\frac{dm}{dt} = \beta_m \frac{1+\alpha_R{}^{-1}K_R S_R+\alpha_D{}^{-1}K_D S_D}{1+K_R S_R+K_D S_D} + \beta - \delta_m m \tag{6}$$

where β_m/α_R and β_m/α_D are the maximum transcription rates when either S_R or S_D are bound to the promoter. Solving for steady-state yields

$$m = \left(\frac{\beta_m}{\delta_m}\right) \frac{1+\alpha_R{}^{-1}K_R S_R+\alpha_D{}^{-1}K_D S_D}{1+K_R S_R+K_D S_D} + \frac{\beta}{\delta_m} \tag{7}$$

As described for Equation 4, this can be further converted to the activity of the output promoter P_{C2},

$$y = (y_{max} - y_{min}) \frac{1+\alpha_R{}^{-1}\kappa_R x_R + \alpha_D{}^{-1}\kappa_D x_D}{1+\kappa_R x_R + \kappa_D x_D} + y_{min} \quad . \quad (8)$$

This equation was then fit to the 2-dimensional response of the promoter to the expression of sgR and sgD (Figure 2.1-2h). This fit was performed while keeping the parameters previously fit to Equation 4 for sgR_{P2} (κ_R, α_R, y_{min}, and y_{max}) constant. The fit is shown as the heatmap coloration in Figure 2.1-2h, the parameters for which are provided in Table 2.1-1. The newly-fit association constant κ_D for sgD is similar to κ_R indicating that sgR_{P2} and sgD_{P5} have similar apparent binding strengths. Additionally, the observed α_D of 1 is consistent with sgD_{P5} having little repressive effect on its own.

2.1.2.2 Control of transcriptional elongation through repression/derepression

dCas9 can block transcription when it is directed to bind internally to a gene by physically interfering with the progression of RNAP (Figure 2.1-3a). However, being located in a gene complicates the design of a derepressing sgRNA position, which must bind to disrupt the repressing sgRNA without itself blocking elongation. To this end, we exploited the strand-dependence of dCas9-based elongation repression, where it has been found that RNAP elongation is more likely to terminate when it collides with the PAM-proximal side of a dCas9-DNA complex (Figure 2.1-3b) [2, 4, 5]. Similar to the promoter repression / derepression mechanism, these constraints on sgR and sgD repressive effects can be captured empirically with the parameter α, which is the maximum fold-repression of mRNA production rates under saturating conditions.

A genetic system was constructed to test this design. dCas9 produces stronger repression when it is directed to the 5'-end of the gene [2, 4], so we selected a position at +232 to be targeted by a repressing sgRNA (sgR_{O1}) (Figure 2.1-3c). This location was selected because it has two adjacent PAM sites that could be used for targeting sgD, one of which has the same 14 bp spacing as was found to be optimal when derepressing a promoter (sgD_{O2}). sgR_{O1} was placed under the control of a Chol-inducible P_{BetI} promoter. It has been observed that elongation repression can be weaker with stronger promoters, presumably due to dislodgement of dCas9 by RNAP [5]. To evaluate this effect, we used an IPTG-inducible

promoter (P_{Tac}) to drive the transcription of the *rfp* reporter. When the promoter is strong (1 mM IPTG), we observed a 7-fold repression of RFP upon the maximum induction of sgR_{O1} (10 mM Chol) (Figure 2.1-3d). This is lower than values reported in the literature for similar distances from the TSS (20- to 100-fold) [2, 4]. However, if one calculates the fold-repression by normalizing to RFP expression of a P_{Tac}-only plasmid (pDAA040) without any sgRNAs, we then observe 34-fold repression. This discrepancy from fold-repression values derived from Chol induction is likely due to the high basal rate of transcription from an uninduced P_{BetI} promoter (Supplementary Figure 2.2-1), which results in leaky expression of sgR_{O1} in the 0 mM Chol condition. The repression data for sgR_{O1} was fit to Equation 4 and the parameters are provided in Table 2.1-1. While we initially developed the models in Equations 4 and 9 for the promoter repression / derepression mechanism, these models can be generalized to the elongation repression / derepression mechanism if the promoter states P, P_R, and P_D are instead considered to be general DNA states. After parameter fitting, we observed that sgR_{O1} had a α of 27, which aligns with the putative 34-fold repression possible when factoring in P_{BetI} leaky expression. However, the α for sgR_{O1} is 4-fold lower than that from promoter repression, indicating a limitation of elongation repressive control at this location.

The dependence of repression on the P_{Tac} promoter strength was then measured by varying IPTG (Figure 2.1-3e). Due to leaky expression of sgR_{O1} from P_{BetI}, we chose to look at repression levels relative to a strain with just the P_{Tac} promoter alone (pDAA040). For each level of P_{Tac} induction, the relative expression was calculated as the ratio of the RFP fluorescence from the P_{Tac}-only plasmid to RFP fluorescence from the circuit plasmid with maximal sgR_{O1} expression (10 mM Choline). These data show that the repressive effects of sgR_{O1} initially increase with promoter strength, but stabilize to roughly 30-fold repression over the top 10-fold range of P_{Tac} expression (Figure 2.1-3e). This result differs from previous work showing that elongation repression is promoter-strength invariant as long as dCas9 levels are saturating [5]. However, that work only examined elongation repression of very strong promoters with strengths varying over a 12-fold range. Our constant relative repression levels in the high P_{Tac} activity range are consistent with those results.

Two sgDs were designed to target positions up- (sgD_{O2}) and down-stream (sgD_{O1}) of the sgR_{O1} position (Figure 2.1-3c). These were placed under the control of a Van-inducible promoter, as before. We initially attempted to calculate an α value for these derepressing sgRNAs, however induction of sgD_{O2} resulted in higher, rather than lower, RFP expression,

likely due to derepression of leaky sgR_{O1} expression from P_{BetI} (Figure 2.1-3f). Nonetheless, we continued on with testing the derepression capabilities of sgD_{O1} and sgD_{O2} (Figure 2.1-3g). To do this, sgR_{O1} was maximally expressed (10 mM Chol) with the strongest induction of the *rfp* reporter (1 mM IPTG). Maximum expression of sgD_{O2} (100 μM Van) showed 4-fold derepression (Figure 2.1-3g). This derepression was confirmed to not be a result of the titration of dCas9 (Supplementary Figure 2.2-2). In contrast, sgD_{O1} did not show any derepression and instead its expression slightly reduces the expression of the reporter, implying it may have its own repressing effect.

The regulatory logic for the co-expression of sgR_{O1} and sgD_{O2} was then determined (Figure 2.1-3h). Similar to the promoter repression / derepression system, the elongation mechanism exhibited increasing expression with sgD induction and decreasing expression with sgR induction (error bars for each measurement are provided in Supplementary Figure 2.2-4). Equation 8 was then fit to the 60-point sampling of the system, with x_R corresponding to P_{BetI} activity and x_D corresponding to P_{Van} activity values.

The fit parameters imply that the derepressing sgRNA association constant κ_D was 2.5-fold lower than the repressing sgRNA association constant κ_R (Table 2.1-1). The association rates of dCas9 to DNA are likely to be similar whether binding to the promoter or gene. However, the dissociation rate of elongation-blocking sgRNA bound to the non-template strand is expected to be larger because dCas9 can be displaced by RNAP collisions [5]. Furthermore, the dissociation rate for template-bound sgRNAs in a gene would be expected to be even higher since this sgRNA orientation is less repressive to elongation. The asymmetric effect of elongating RNAP on sgRNA displacement can presumably be seen when comparing the fit κ_R and κ_D values between the two repression / derepression mechanisms (Table 2.1-1). In the promoter-based mechanism, κ_R and κ_D are similar, however, in the elongation control mechanism they are 2.5-fold different. This is presumably due to elongating RNAP preferentially increasing the off-rate for sgD_{O2} relative to sgR_{O1}, thus decreasing its effective association constant.

The dependence of derepression on the P_{Tac} promoter strength was then measured by varying IPTG (Figure 2.1-3e). Similar to the promoter-strength test for relative repression, we normalized the fluorescence to that of the P_{Tac}-only plasmid (pDAA040). For each level of P_{Tac} induction, the relative expression during derepression was calculated as the ratio of the

RFP fluorescence from the P_{Tac}-only plasmid to the RFP fluorescence from the circuit plasmid with maximal sgR_{O1} and sgD_{O2} induction (10 mM Chol and 100 μM Van). From this experiment, we observed that derepression is invariant over a wider range of transcription rates than that observed for repression.

2.1.2.3 Ratiometric performance of the repression / derepression

Ratiometric signal processing is where a system exhibits a response to the relative value of two inputs, ignoring their absolute magnitude. Naturally-occurring ratiometric responses have been observed in ATP/ADP management [140], X vs. autosomal chromosome levels [141], circadian clock determination [142], cancer cell clinical resistance [143], and sugar source utilization in yeast [144]. To examine the ratiometric performance of our repression / derepression mechanisms, we looked at the circuit outputs relative to the ratio of promoter activities producing sgD and sgR (P_{Van} / P_{BetI}). From this, we observed that the promoter repression / derepression circuit effectively responded to over 2 orders of magnitude of sgD / sgR ratios with a ~50-fold dynamic range (Figure 2.1-4a). The elongation mechanism had a smaller overall dynamic range (~10-fold) and the circuit did not perform predictably across all ratios, presumably due to the apparent asymmetry in the effective association constants for sgR_{O1} and sgD_{O2} (Figure 2.1-4b).

For the promoter repression / derepression mechanism, the ratiometric response between repressing and derepressing sgRNA saturates at the maximal output when sgD promoter activity values are greater than sgR promoter activities (Figure 2.1-4a). This is consistent with the notion that saturation of the promoter by derepressing sgRNA will completely prevent repression by sgR. This also makes the ratiometric response of our circuit one-sided, in that it can only process ratios where the magnitude of one signal is larger than the other.

2.1.3 Discussion

This work introduces a new mode of regulatory control, where the binding of a repressor is displaced by the binding of a nearby derepressor. This motif, where both repressor and

derepressor bind to the same operator, has not been observed in natural systems. This may be due to the difficulty of designing a common operator that binds to multiple proteins and a derepressing protein that blocks the repressor, but not RNAP. The programmability of CRISPRi makes the design of multiple sgRNAs binding nearby sites almost trivial and the size of dCas9, and its melting of DNA, simplifies its use to sterically inhibit the binding or elongation of RNAP. Note that the off-rate of dCas9 binding DNA is very slow, effectively irreversible, until it is displaced by DNAP or RNAP [5, 52, 53]. In our repression / derepression mechanism this makes it unlikely that dCas9:sgD can displace dCas9:sgR when it is already bound. Rather, it depends on which binds first, with DNAP replication (and to a lesser-extent RNAP elongation) "resetting" the DNA state at the rate of plasmid replication (cellular growth rate) and RNAP elongation (promoter strength) [5, 52]. Thus, the kinetics of the system reaching steady-state are expected to be dependent on growth phase and may impact their use in circuits with components with faster dynamics [47, 145-147].

Repression/derepression could be applied to design synthetic circuits that perform new signal integration functions. As simple cis-regulatory logic, the operation performed by these circuits is A IMPLY B where sgD (B) overrides the regulation imposed by sgR (A) (Figures 2.1-2g and 2.1-3g). If instead the displaced regulator were an activator (based on dCas9 recruiting RNAP [15, 16, 18, 20]), this logic would be A NIMPLY B.

However, this digital combinatorial logic does not capture the potential for new signal integration performed by these mechanisms. The signal integration follows a graded, rather than switch-like, transition (Figures 2.1-2h and 2.1-3h) and this can be harnessed to build useful analog functions. The circuit based on the repression/derepression of a promoter (sgR_{P2} and sgD_{P5}) is capable of responding to transcriptional input ratios over two orders of magnitude (Figure 2.1-4a). Ratiometric signal processing is common in natural regulatory systems and is usually achieved by a motif where two species compete for a third component and only one of the bound species results in the output [140-144, 148-153]. For example, this competition can occur at a DNA binding site where an activating transcription factor competes with an inactive or repressive transcription factor [154, 155]. Ratiometric circuits can be used to determine which of two continuously-variable signals is larger. The sgR_{P2} / sgD_{P5} circuit in log-space is performing a (sgD – sgR) calculation, which if connected to a cooperative switch-like output could act as a single neuron in a neural network [156].

Various approaches have been developed to place native genes under synthetic

regulatory control through the insertion of synthetic promoters (*e.g.*, that respond to T7 RNAP) or other regulatory motifs [157-159]. CRISPRi has been rapidly adopted because it can exert regulatory control without having to mutate the target; however, it only imparts a single on/off signal. To this end, depression/derepression can be used to inactivate the effects of CRISPRi in the control of native genes [24, 44, 78, 79]. For example, CRISPRi/a has been used to dynamically repress or activate enzymes to control carbon flux in metabolic engineering applications [75, 76, 78, 133, 160]. Derepression allows for the disruption of a subset of the genes being influenced by an activating or repressing sgRNA; in effect, this would introduce cis-regulatory logic into the native genes without needing to insert operators. In this work, we had to deal with the challenge of identifying or designing PAM sites so that the sgR and sgD binding sites are appropriately positioned to satisfy the distance constraints. This restricts their utility in the control of genome-encoded genes, where the likelihood of PAM sites being appropriately positioned in a promoter is small or restricts where the regulation can occur within a gene. This is likely why the effect that we observed for the blockage and release of elongation is small. However, recent efforts to engineer dCas9 to not require a PAM site have made progress [21, 125-128]. Collectively, these advances will allow the programming of cis-regulatory logic to be applied to any gene in the genome without having to insert or modify genomic DNA.

2.1.4 Materials and Methods

2.1.4.1 Strains, plasmids, media, and chemicals.

E. coli NEB 10-beta (C3019I, New England BioLabs, Ipswich, MA, USA) was used for all routine cloning. All genetic circuit measurements were done using *E. coli* K-12 MG1655 * [F- λ- ilvG- rfb-50 rph-1 Δ(araCBAD) Δ(LacI)] [24, 161]. Cells were grown in in MOPS EZ Rich Defined Medium (Teknova, M2105) with 0.2% glucose (Teknova, G0520). Kanamycin (50 μg/ml, GoldBio, K-120-5) was used to maintain plasmids. Chemical inducers used: vanillic acid (Van) (Millipore Sigma, 94770); isopropyl ß-d-1-thiogalactopyranoside (IPTG) (GoldBio, I2481C); anhydrotetracycline (aTc) (Millipore Sigma, 37919); and choline chloride (Chol) (Millipore Sigma, C7017). DNA oligos and genes were ordered from Integrated DNA Technologies (Coralville, IA) and Twist Biosciences (San Francisco, CA). All plasmids were

constructed from the parental pDAA038 backbone using TypeIIS assembly to insert circuit components between BsaI sites. A table of plasmids is provided in Supplementary Table 2.2-1. Key plasmid maps are shown in Supplementary Figure 2.2-5. Monomeric RFP was used and is referred to as "RFP" throughout the paper.

2.1.4.2 Computational methods.

The random 100 bp sequence within P_{C1} was generated using the online Random DNA Sequence Generator (http://www.faculty.ucr.edu/~mmaduro/random.htm) with the GC content set to 50%. Non-linear fitting was completed with the Python scipy.optimize.curve_fit() function.

2.1.4.3 Induction assays.

All growth was performed in 96-well V-bottom plates (Roskilde, Denmark, #249952) at 1000 rpm in a microplate shaker (ELMI, #DTS-4). The day before, individual colonies were inoculated into 150 µL EZ Rich media and Kan for overnight (16 hr) growth at 37 °C. The next day, cultures were diluted 200-fold by adding 0.75 µL of overnight culture into 150 µL of EZ Rich media and Kan. After 2 hr growth, cultures were diluted 1000-fold into inducing conditions by adding 6 µL of culture into 198 µL fresh media and Kan and then 5 µL of that dilution into media with inducers and Kan. To induce dCas9, aTc was added to a final concentration of 1.25 nM in all experiments. Growth was performed for 5.5 hours, after which samples were prepared for flow cytometry.

2.1.4.4 Flow cytometry analysis.

Fluorescence characterization was performed using a BD LSR Fortessa flow cytometer with the HTS attachment (BD, Franklin Lakes, NJ). Samples were prepared by aliquoting 40 µL of cell culture into 160 µL of PBS containing 200 µg/mL Kan. All samples were run in standard mode at a flow rate of 2 µL/s. RFP fluorescence measurements were made using the

green (561 nm) laser and all data was derived from the PE-Texas Red-A channel (PMT voltage of 700 V). The FSC and SSC voltages were 650 V and 270 V, respectively. At least 10,000 events were collected for each sample and the Cytoflow Python package (https://github.com/cytoflow/cytoflow) was used for analysis, including gating. The geometric mean fluorescence is calculated for all cytometry distributions.

2.1.4.5 Promoter input calculations.

The following procedure was followed to convert inducer concentrations (e.g., [IPTG]) to the activities of the inducible promoter (e.g., P_{Tac}) reported in RFP fluorescence (au). This approach has been described previously [23, 162]. Using a separate plasmid based on pDAA038, the response function of the inducible system was measured separately by measuring RFP fluorescence using cytometry. The RFP fluorescence values in au for the inducer concentrations were then plotted as "Promoter Activity." The response functions used are shown in Supplementary Figure 2.2-1.

2.1.5 Tables

Table 2.1-1: Model Parameters

Parameters							
Mechanism [a]	**sgRNAs**	y_{min}	y_{max}	κ_R	κ_D	α_R	α_D
Promoter	sgR_{P1}	1.2	7700	0.10		150	
Promoter	sgR_{P2}	16	6800	0.043		120	
Promoter	sgR_{P2}+sgD_{P5}	16	6800	0.043	0.063	120	1.0
Elongation	sgR_{O1}	37	4200	0.019		27	
Elongation	sgR_{O1}+sgD_{O2}	37	4200	0.019	0.0077	27	1.0

a. Single sgRNA experiments were fit to Equation 4 and dual-sgRNA experiments were fit to Equation 8.

2.1.6 Figures

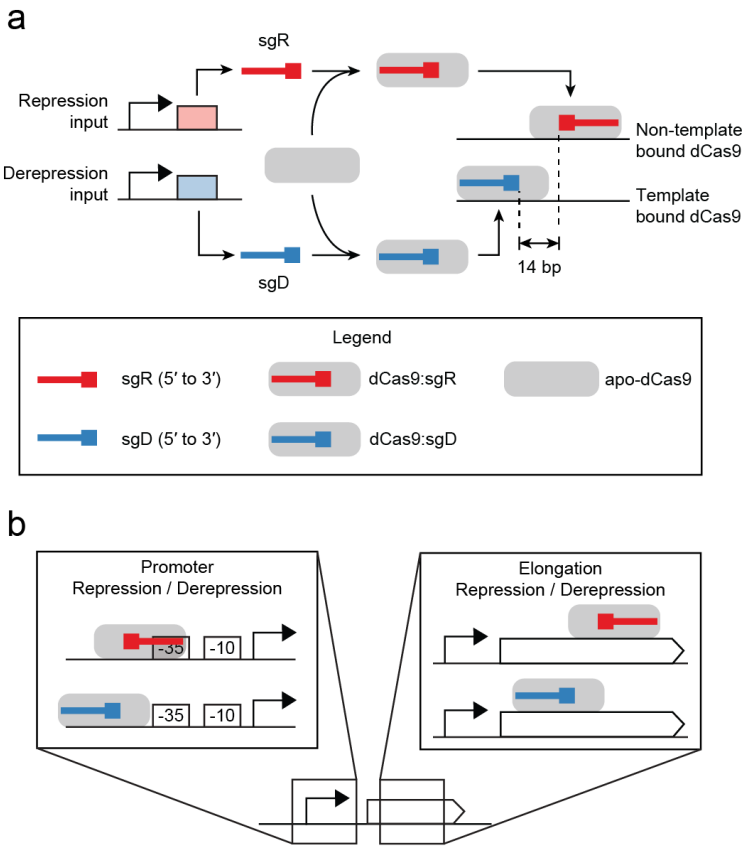


Figure 2.1-1: The mechanisms of dCas9 repression and derepression

(a) The footprint of dCas9:sgD blocks the binding of dCas9:sgR without requiring that the sgRNA-binding regions overlap. The PAM-adjacent regions of sgRNAs are shown as boxed ends. **(b)** The sgR/sgD sequences can be targeted to the promoter or internal to a gene.

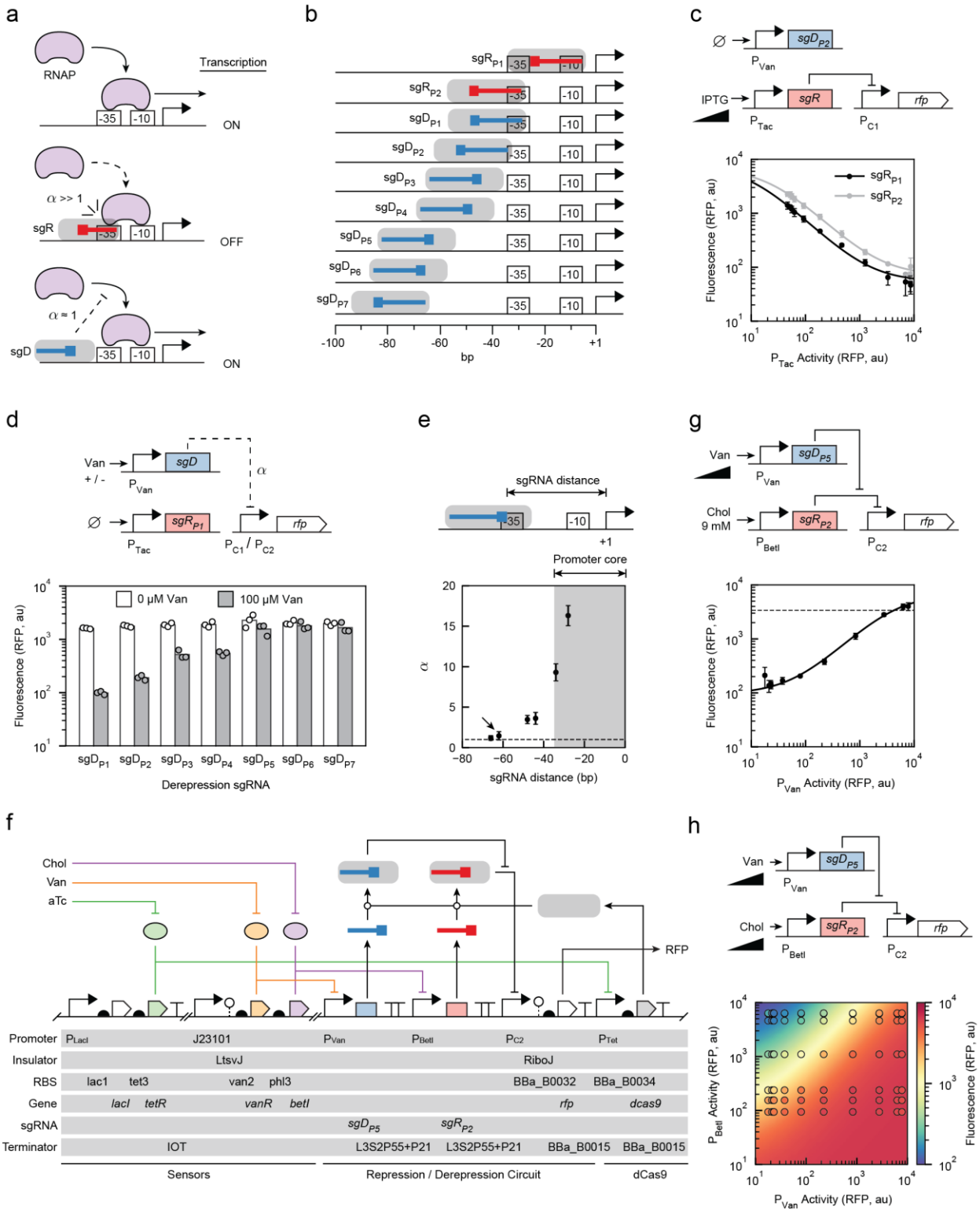
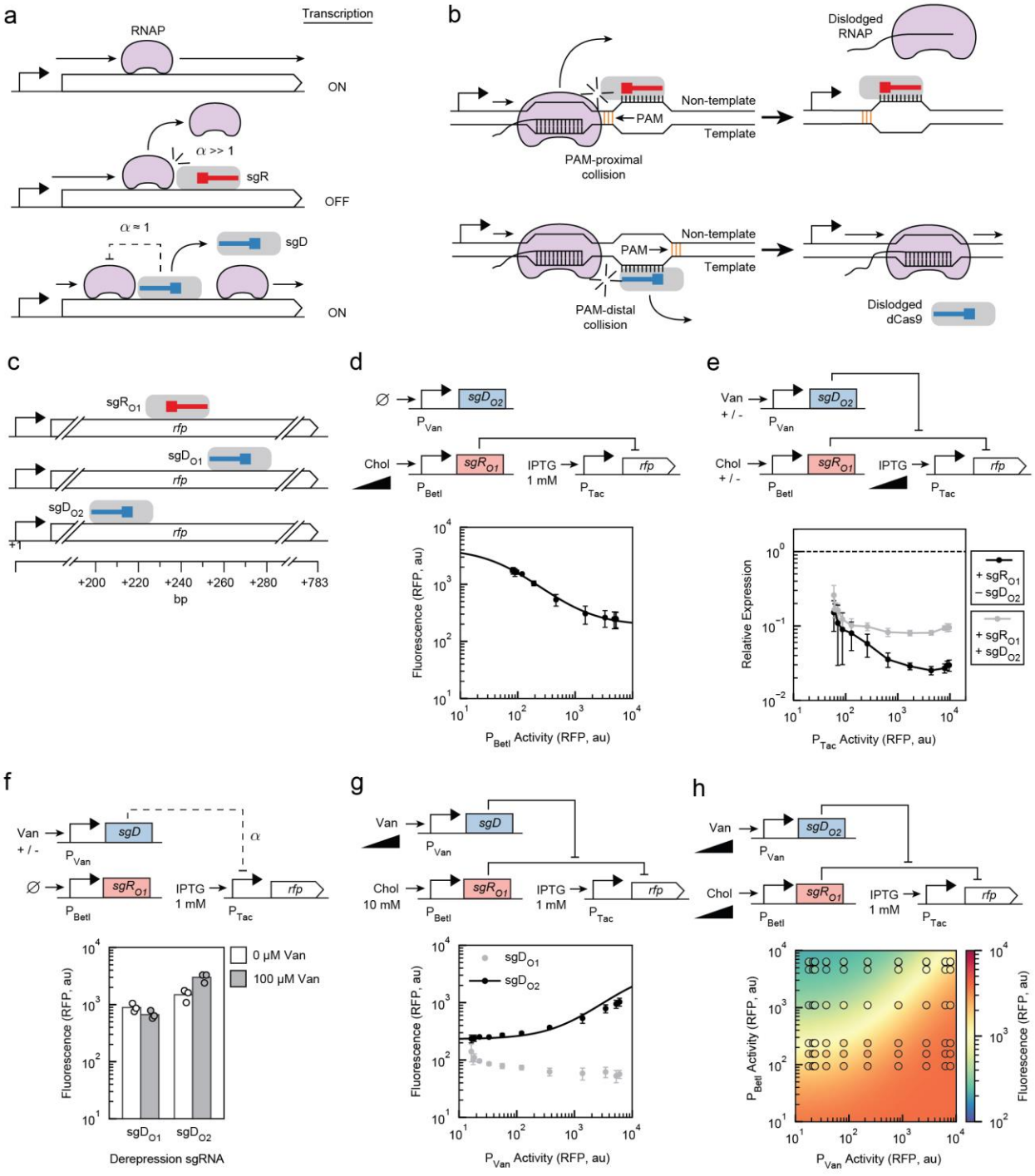


Figure 2.1-2: Repression and derepression targeted to a promoter

(a) The promoter states for sgR, sgD and RNAP binding to a promoter are shown,

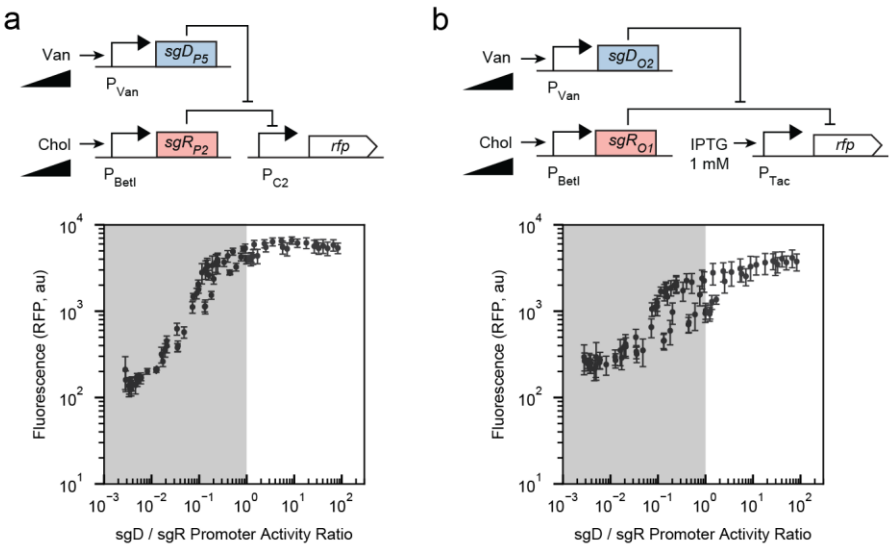
highlighting the -10 and -35 sites to which σ70 binds. Ideal α values for each sgRNA type are indicated (see text). **(b)** Positions of P_{C1} tested for targeting sgR and sgD. The boxes on the sgRNAs mark the PAM-adjacent regions, the orientation of which determines which strand is targeted. **(c)** Repression by the expression of sgRs. The experiments are performed with plasmids pDAA043 and pDAA656 for sgR_{P1} and sgR_{P2}, respectively. Both plasmids contain sgD_{P2} (uninduced for these experiments). The lines are the fit to Equation 4, the parameters of which are in Table 2.1-1 (Methods). The x-axis is converted from IPTG to promoter activity (Methods). Inducer concentrations used to generate the induction curve were IPTG (μM) = [0, 0.977, 1.95, 3.91, 7.81, 15.6, 31.3, 62.5, 125, 250, 500, 1000]. **(d)** The position dependence of repression. The experiments are performed with plasmids pDAA042, 043, 050, 051, 654, 052, and 044 for sgD_{P1}-sgD_{P7} (Supplementary Table 2.2-1) The circuits also contain sgR_{P1} (uninduced for these experiments). The pDAA654 plasmid testing sgD_{P5} used the promoter P_{C2}, all other sgDs used P_{C1}. **(e)** Values of α are calculated for the sgDs tested in part d. The horizontal dashed line is drawn at α = 1 (ideal α value). The arrow marks sgD_{P5}, which was selected to build the repression/derepression circuit. **(f)** Genetic circuit schematic of the circuit based on sgR_{P2} and sgD_{P5} circuit, encoded on a single plasmid (pDAA107). **(g)** Derepression induction curve for sgD_{P5} with full sgR_{P2} induction. The line is a fit to Equation 8 with parameters in Table 2.1-1. A dashed line indicates the uninduced expression value (0 mM Chol, 0 μM Van). The x-axis is converted from Van to promoter activity (Methods). Inducer concentrations used to generate the induction curve were Van (μM) = [0, 0.0977, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]. **(h)** Two-dimensional response function for the induction of sgR_{P2} and sgD_{P5}. The circles are experimental measurements colored by the mean fluorescence values of three replicates (standard deviations in Supplementary Figure 2.2-4). The continuous color in the background is the model prediction from Equation 8, R^2=0.87 (Supplementary Figure 2.2-4). The x-axis and y-axis are converted from Van and Chol, respectively, to promoter activity (Methods). Inducer concentrations used to generate the induction curves were Van (μM) = [0, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]; Chol (μM) = [0, 37.0, 111, 333, 1000, 3000, 9000]. Representative cytometry distributions for all parts are shown in Supplementary Figure 2.2-3. Data points are the means of three replicates performed on different days and the error bars are the standard deviations of these measurements. The error bars in part e are calculated as $\sigma = |A/B|((\sigma_A/A)^2+(\sigma_B/B)^2]^{1/2}$, where A and B are the means with 0 and 100 μM Van sgD induction, respectively, and σ_A and σ_B are the standard deviations of these measurements.



<u>Figure 2.1-3</u>: Repression/derepression by blocking RNAP progression within a gene

(a) RNAP transcribes the full length transcript in the absence of dCas9:sgR. When dCas9:sgR is bound, this blocks RNAP (large α), causing it to terminate prematurely (note that RNAP can also dislodge dCas9, not shown). Ideally, dCas9:sgD does not block RNAP

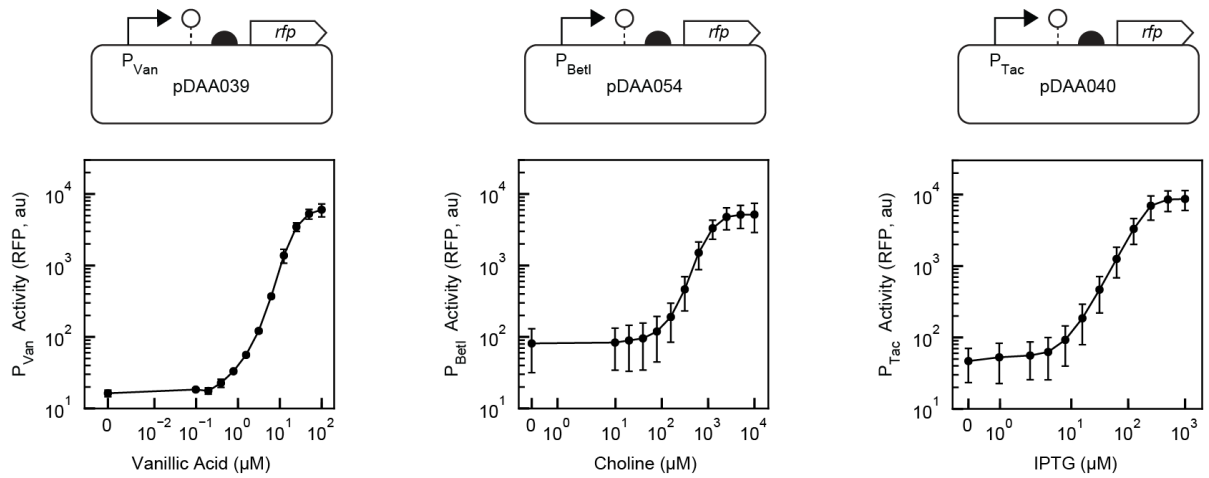
progression, providing an $\alpha = 1$. **(b)** When dCas9 is bound to the non-template strand, the PAM is proximal to RNAP, and dCas9 tends to stay bound after collision. When dCas9 is bound to the template strand, the PAM is distal to RNAP collision and RNAP can continue to elongate. **(c)** Repressing and derepressing sgRNA binding sites tested relative to *rfp*. **(d)** Repression of transcriptional elongation by sgR_{O1}. The circuit is encoded on plasmid pDAA056, which also includes sgD_{O2} (uninduced in these experiments). The curve is a fit to Equation 4 with parameters in Table 2.1-1. The x-axis is converted from Chol to promoter activity (Methods). Inducer concentrations used to generate the induction curve were Chol (μM) = [0, 9.77, 19.5, 39.1, 78.1, 156, 313, 625, 1250, 2500, 5000, 10000]. **(e)** The dependence of sgR_{O1} repression and sgD_{O2} derepression on the strength of the promoter controlling the repressed gene (P_{Tac}). Relative expression values for the repression test (+sgR_{O1} /-sgD_{O2} condition) were calculated by dividing the mean fluorescence of the P_{Tac}-only plasmid pDAA040 by the mean fluorescence of the circuit plasmid with sgR_{O1} induction (10 mM Chol). Relative expression values for the derepression test (+sgR_{O1} /+sgD_{O2} condition) were calculated by dividing the mean fluorescence of the P_{Tac}-only plasmid (pDAA040) by the mean fluorescence of the circuit plasmid with sgR_{O1} and sgD_{O2} induction (10 mM Chol and 100 μM Van). The circuit is encoded on plasmid pDAA056. Connecting lines are a visual guide. The x-axis is converted from IPTG to promoter activity (Methods). Inducer concentrations used to generate the induction curve were IPTG (μM) = [0, 0.977, 1.95, 3.91, 7.81, 15.6, 31.3, 62.5, 125, 250, 500, 1000]. **(f)** Fully-induced P_{Tac} Promoter activity with (100 μM Van) and without (0 μM Van) induction of sgDs. Plasmids pDAA057 and pDAA056 were used for the sgD_{O1} and sgD_{O2} experiments, respectively. sgR_{O1} was present but uninduced. **(g)** Derepression induction curves with full sgR_{O1} induction. The line is a fit to Equation 8, with fit parameters in Table 2.1-1. The x-axis is converted from Van to promoter activity (Methods). Inducer concentrations used to generate the induction curve were Van (μM) = [0, 0.0977, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]. **(h)** Two-dimensional response function for the induction of sgR_{O1} and sgD_{O2}. The circles are experimental measurements colored by the mean fluorescence values of three replicates (standard deviations in Supplementary Figure 2.2-4). The continuous color in the background is the model prediction from Equation 8, R^2=0.87 (Supplementary Figure 2.2-4). The x-axis and y-axis are converted from Van and Chol to promoter activity, respectively (Methods). Inducer concentrations used to generate the induction curves were Van (μM) = [0, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]; Chol (μM) = [0, 37.0, 111, 333, 1000, 3000, 9000]. Representative cytometry distributions for all parts are shown in Supplementary Figure 2.2-3. Data points are the means of three replicates performed on different days and the error bars are the standard deviations of these measurements. The error bars in parts e were calculated as $\sigma = |A/B|((\sigma_A/A)^2+(\sigma_B/B)^2]^{1/2}$, where A is the mean of the P_{Tac}-only expression, B is the mean of the circuit with either sgR_{O1} or sgR_{O1} + sgD_{O2} induction, and σ_A and σ_B are the standard deviations of their measurements.



<u>Figure 2.1-4</u>: Ratiometric performance of repression / derepression circuits

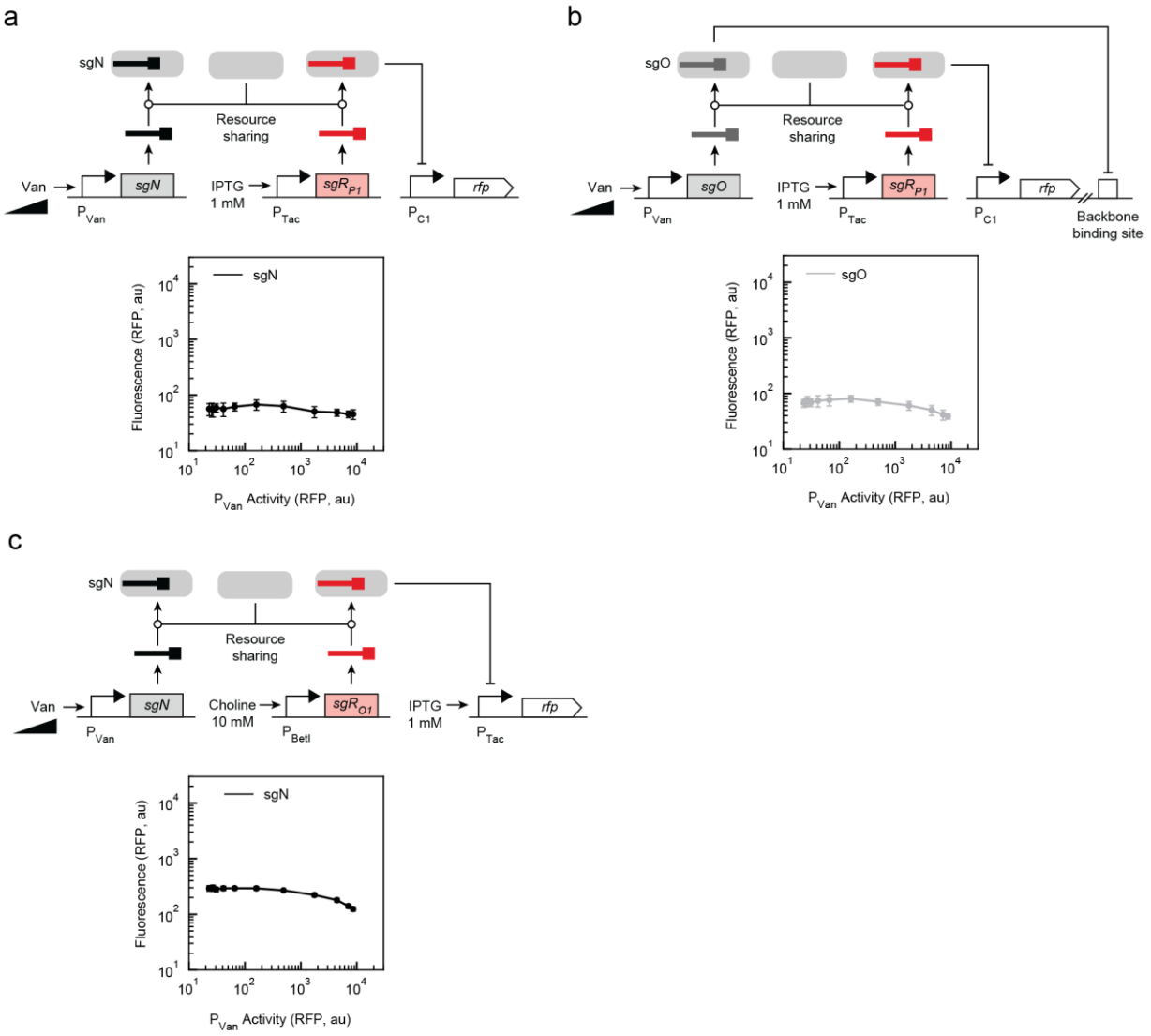
(a) Promoter repression / derepression circuit experimental results from Figure 2.1-2h replotted as the ratio of the two inputs. **(b)** Elongation repression / derepression circuit experimental results from Figure 2.1-3h replotted as the ratio of the two inputs. For both parts, the x-axis values are derived from converting Chol (sgR) and Van (sgD) values to promoter activity (Methods) and then dividing sgD promoter activity by sgR promoter activity. Inducer concentrations used to generate the induction curves were Van (μM) = [0, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]; Chol (μM) = [0, 37.0, 111, 333, 1000, 3000, 9000]. sgD / sgR ratios below 1 are indicated in gray. Data points are the means of three replicates performed on different days and the error bars are the standard deviations of these measurements.

2.2 Supplementary



<u>Figure 2.2-1</u>: Response functions for inducible promoters

The plasmids are based on the pDAA038 backbone. These data were used to convert inducer concentrations into promoter activities for the response functions shown in the main text figures and are used for the xR and xD values in equation fitting. Inducer concentrations used for each induction curve were: Van (μM) = [0, 0.0977, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]; Chol (μM) = [0, 9.77, 19.5, 39.1, 78.1, 156, 313, 625, 1250, 2500, 5000, 10000]; IPTG (μM) = [0, 0.977, 1.95, 3.91, 7.81, 15.6, 31.3, 62.5, 125, 250, 500, 1000]. The points are means of three replicates performed on different days and the error bars are the standard deviations of those measurements.



<u>Figure 2.2-2</u>: Controls to determine the impact of dCas9 depletion

The expression of one sgRNA can titrate dCas9 from a second sgRNA, thus creating the illusion of derepression activity. **(a)** The genetic system to assess the impact of the expression of a non-targeting sgRNA (sgN) that does not target the plasmid or genome is shown. The data show the impact on repression by sgR_{P1} expression (1mM IPTG) as sgN is expressed by increasing vanillic acid concentration. **(b)** The genetic system for the expression of an off-target sgRNA (sgO) that targets a location in the plasmid that has no functional effect and its impact on repression by sgR_{P1} expression (1mM IPTG). **(c)** The genetic system to evaluate the impact of sgN expression (Van) on elongation repression by sgR_{O1} (10mM Chol). The plasmids used to generate these data are described in Supplementary Table 2.2-1. Inducer concentrations used for each induction curve were: Van (μM) = [0, 0.0977, 0.195, 0.391, 0.781, 1.56, 3.13, 6.25, 12.5, 25, 50, 100]. The points are means of three replicates performed on different days and the error bars are the standard deviations of these measurements.

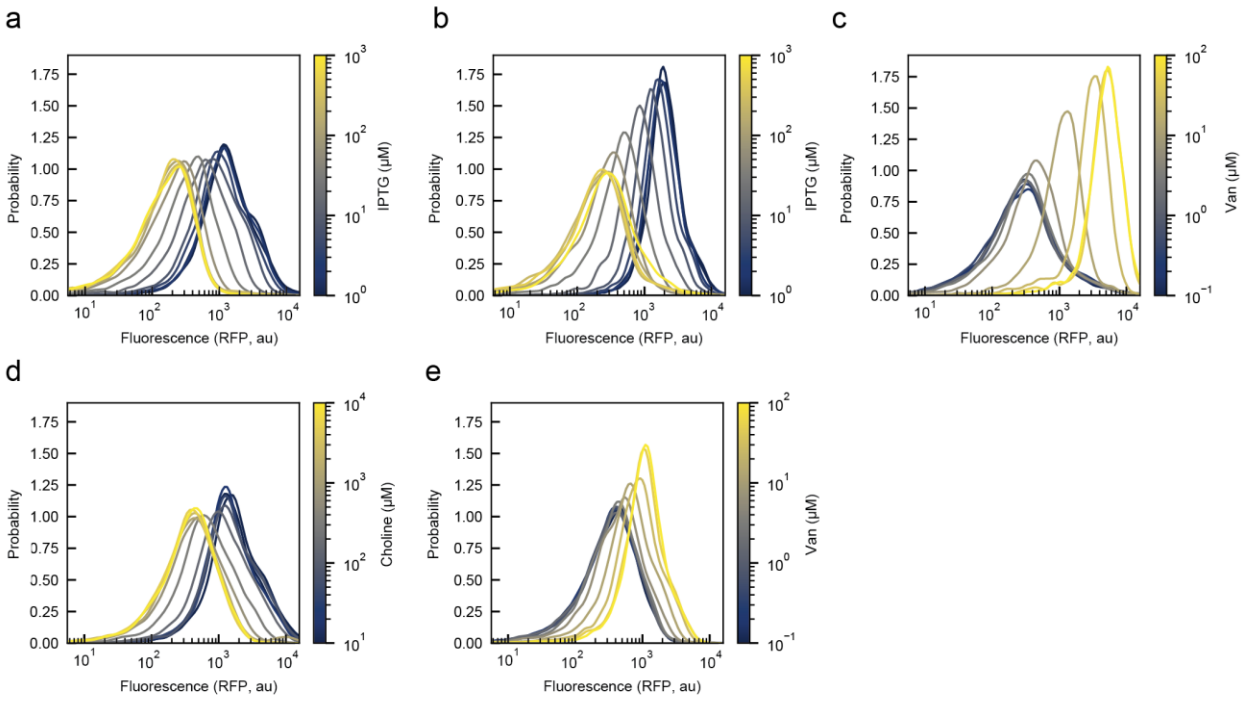
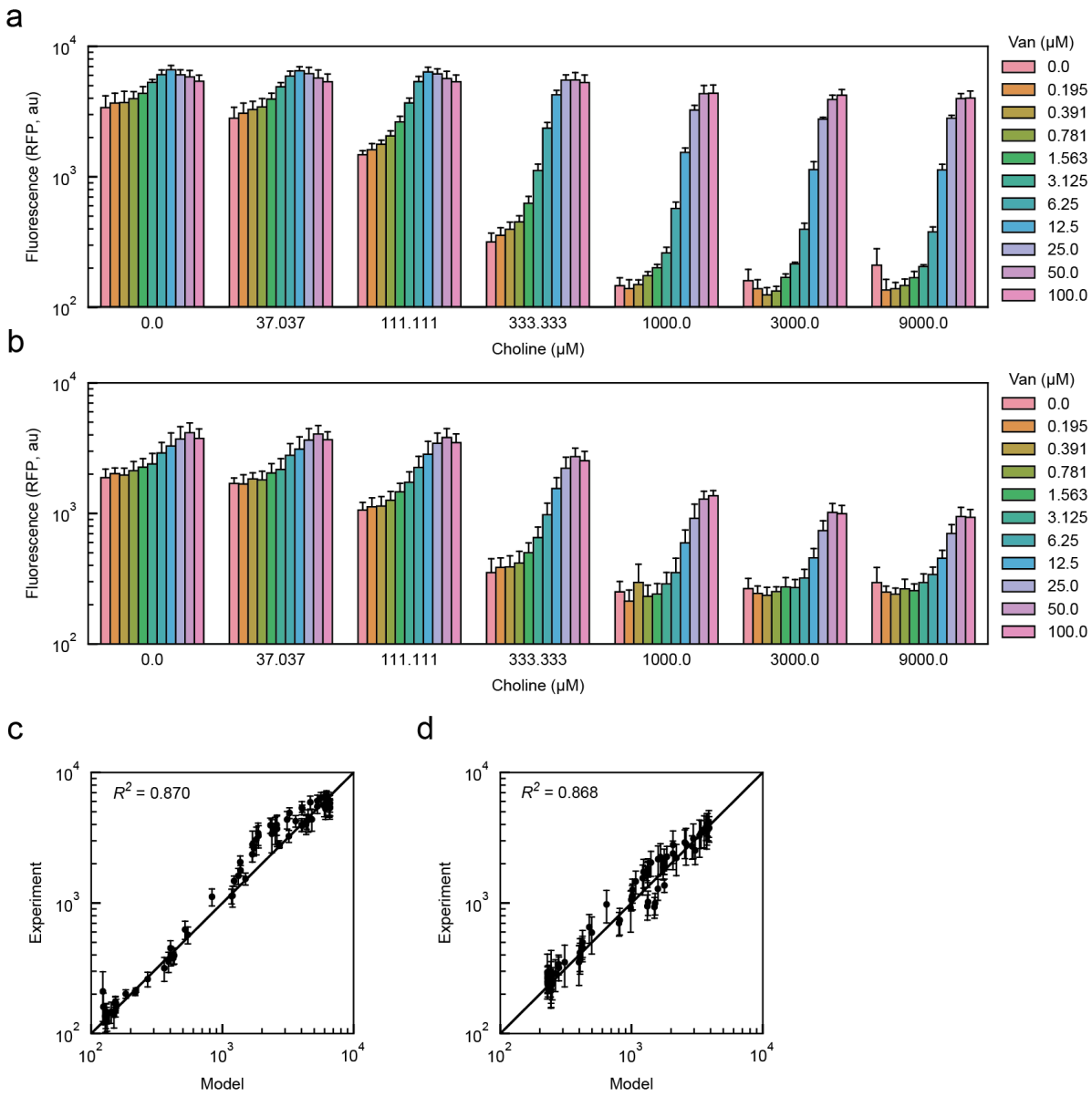


Figure 2.2-3: Representative cytometry distributions

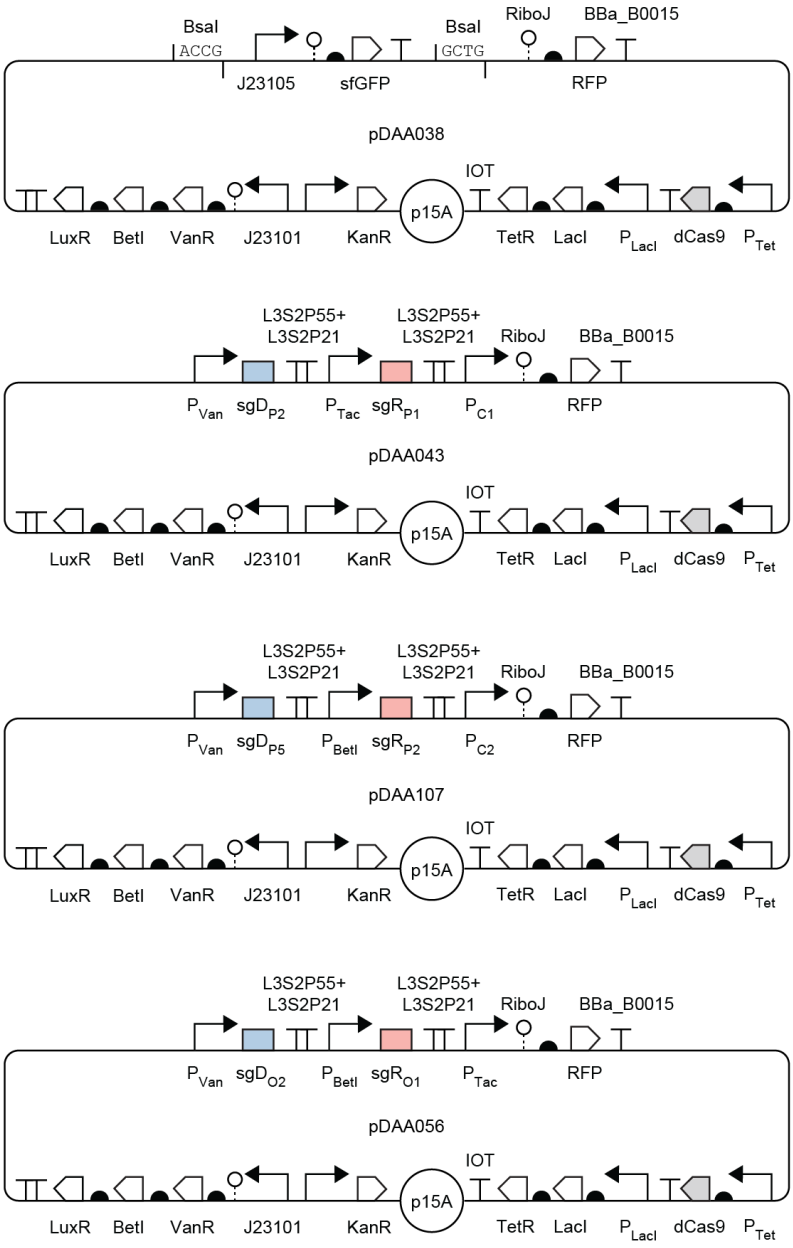
Distributions are shown for: **(a)** Figure 2.1-2c (sgR_{P1}) **(b)** Figure 2.1-2c (sgR_{P2}) **(c)** Figure 2.1-2g, **(d)** Figure 2.1-3d, **(e)** Figure 2.1-3g. All distributions are representative of three experiments performed on different days.



<u>Figure 2.2-4</u>: Standard deviations of two-dimensional response functions and model fits

(a) Output of the promoter repression / derepression circuit (pDAA107) for all 60 experimental conditions. The bars represent the mean values of three replicates performed on different days and correspond to the colored circles in Figure 2.1-2h. **(b)** Output of the elongation repression / derepression circuit (pDAA056) for all 60 experimentally-sampled conditions. The bars represent the mean values of three replicates performed on different days and correspond to the colored circles in Figure 2.1-3h. **(c)** Evaluation of goodness-of-fit for Equation 9 to the sgR_{P2}+sgD_{P5} promoter repression / derepression circuit with parameters shown in Table 2.1-1. A reference line is drawn at y = x. **(d)** Evaluation of goodness-of-fit for Equation 9 to the sgR_{O1}+sgD_{O2} elongation repression / derepression circuit with parameters shown in Table 2.1-1. A reference line is drawn at y = x. The data are the means of three

replicates performed on different days and the error bars are the standard deviations of these measurements.



<u>Figure 2.2-5</u>: Key plasmid maps

Key plasmid maps used in this study. pDAA038: plasmid backbone used to build all circuits. pDAA043: initial promoter repression and derepression α testing construct. pDAA107: final promoter repression / derepression circuit. pDAA056: final elongation repression / derepression circuit.

Table 2.2-1: Plasmids used in this study

Name	Mechanism	sgR	sgD	sgR Inducer	sgD Inducer	Promoter [a]
			Circuits			
pDAA042	Initiation	sgR_{P1}	sgD_{P1}	IPTG	Van	P_{C1}
pDAA043	Initiation	sgR_{P1}	sgD_{P2}	IPTG	Van	P_{C1}
pDAA050	Initiation	sgR_{P1}	sgD_{P3}	IPTG	Van	P_{C1}
pDAA051	Initiation	sgR_{P1}	sgD_{P4}	IPTG	Van	P_{C1}
pDAA654	Initiation	sgR_{P1}	sgD_{P5}	IPTG	Van	P_{C2}
pDAA052	Initiation	sgR_{P1}	sgD_{P6}	IPTG	Van	P_{C1}
pDAA044	Initiation	sgR_{P1}	sgD_{P7}	IPTG	Van	P_{C1}
pDAA656	Initiation	sgR_{P2}	sgD_{P2}	IPTG	Van	P_{C1}
pDAA107	Initiation	sgR_{P2}	sgD_{P5}	Chol	Van	P_{C2}
pDAA057	Elongation	sgR_{O1}	sgD_{O1}	Chol	Van	P_{Tac}
pDAA056	Elongation	sgR_{O1}	sgD_{O2}	Chol	Van	P_{Tac}
pDAA055	Elongation	sgR_{O1}	sgD_{O3}	Chol	Van	P_{Tac}
pDAA047	Initiation	sgR_{P1}	sgO	IPTG	Van	P_{C1}
pDAA048	Initiation	sgR_{P1}	sgN	IPTG	Van	P_{C1}
pDAA080	Elongation	sgR_{O1}	sgN	Chol	Van	P_{Tac}
			Promoter Activity Profiling			
pDAA039	N/A	N/A	N/A	N/A	N/A	P_{Van}
pDAA040	N/A	N/A	N/A	N/A	N/A	P_{Tac}
pDAA054	N/A	N/A	N/A	N/A	N/A	P_{BetI}

a. P_{C2} is the P_{C1} promoter with a 3 base pair mutation from -62 to -60.

3 Chapter 3: A genetic-circuit-enabled dual selection system for SARS-CoV-2 protease inhibitors

3.1 Main Text

3.1.1 Results

3.1.1.1 Development of a protease-activity sensing synthetic genetic circuit

Our approach was to first develop a protease-activity responsive synthetic genetic circuit and then connect the output of that circuit into a dual-selection system. To develop the protease-activity detection circuit, we started with the previously-described protease-sensitive circularly-permuted (CP)-dCas9 protein [22]. The protease sensitivity of the CP-dCas9 system is derived from the introduction of a short linker connecting the original N and C termini of a CP dCas9 protein, which, when present, "cages" the dCas9 protein into an inactive state thus preventing DNA binding. If a protease cleavage site is introduced into the linker and its cognate protease is present, the linker is cleaved thus returning dCas9 to an active, "uncaged" state (Figure 3.1-1b).

We adapted the CP-dCas9 system to sense the activity of the two SARS-CoV-2 viral proteases, PLpro and CLpro, by introducing their cognate cut sites, TLKGGAPT and SVKLQSGF, into the CP-dCas9 linker. These sequences were chosen after testing a series cut site sequences and finding these to be the most performant (data not shown). Guided by previous characterization of the CP-dCas9 system [22], we also chose to shorten the N-terminal end of the linker (the original dCas9 C-terminus) by 6 amino acids to decrease leaky caged dCas9 activity (Methods). The CP-dCas9 mechanism was then developed into a synthetic genetic circuit activated by protease inhibition by providing a small guide RNA (sgRNA) targeting the -35 region of a constitutively active σ70 promoter, P_8. The repressing action of uncaged dCas9 effectively inverts the protease activity signal, providing high transcriptional output in the presence of an inhibited protease (Figure 3.1-1b).

A system was then designed to integrate the protease-sensitive CP-dCas9 circuit with a selection system. The P_8 promoter was placed upstream of a bicistronic selection system

containing GFP-chloramphenicol acetyltransferase (GFP-CAT) and Herpes Simplex Virus thymidine kinase-RFP (HSVtk-RFP). The fluorescence from the fused fluorescent proteins was used for readout of the system. The protease-sensitive CP-dCas9, SARS-CoV2 protease, and the p8 sgRNA were all constitutively expressed on a ColE1* plasmid mutated to have similar copy numbers as p15A [64].

Experiments were performed to measure the effect of protease inhibition on selection protein expression. Plasmids were transformed into *E. coli* and cells were grown to stationary phase overnight in rich media (TB) before measurement with flow cytometry (Methods). Perfect protease inhibition was simulated by introducing Cys→Ala mutations at the catalytic cysteine residue of each of the proteases (C111A for PLpro and C145A for CLpro) [163, 164]. Active / dead protease fold-changes of GFP-CAT were 130 and 270-fold for PLpro and CLpro, respectively (Figure 3.1-1c). The downstream protein in the bicistron, HSVtk-RFP, exhibited smaller fold-change values of 45 and 75-fold (Figure 3.1-2d), likely due to transcriptional attenuation [165-167]. Our fold-changes are notably higher than the 10-fold changes previously reported with protease-sensitive CP-dCas9 systems [22], although we observed that there is still some room for improvement indicated by the 1080 and 176-fold changes observed for GFP-CAT and HSVtk-RFP with non-targeting sgRNA (NT) and normal dCas9 controls (Supplementary Figure 3.2-1). Nevertheless, we deemed our fold-changes to be high enough to carry on with characterization of the connected selection system.

3.1.1.2 Characterization of the GFP-CAT + HSVtk-RFP dual selection system

Our bicistronic selection cassette comprised GFP-CAT and HSVtk-RFP for positive and negative selections, respectively. CAT is an antibiotic resistance protein that modifies chloramphenicol (Cm), a translation-inhibiting molecule (Figure 3.1-1e). CAT expression rescues growth rates in a graded manner making it ideal for selection stringency tuning [168]. Conversely, HSVtk is viral thymidine kinase that phosphorylates the synthetic nucleoside dP such that it can be incorporated into endogenous DNA synthesis processes. High expression of HSVtk leads to cellular death in the presence of dP due to high mutational loads from the synthetic nucleotide (Figure 3.1-1f) [169-171].

Relative growth assays were completed to determine the relationship between selection protein (CAT and HSVtk) expressions and selection molecule (Cm and dP) concentrations. The goal being to define the ideal concentrations of Cm and dP that preferentially enrich for inhibited proteases. To this end, we redesigned the CP-dCas9 system by removing the constitutively-produced protease, sgRNA, and CP-dCas9, We then inserted different constitutive promoters upstream of the GFP-CAT : HSVtk-RFP selection cassette. Six constitutive promoters of different strengths were taken from the Anderson promoter series and one terminator were chosen to create a series of different selection cassette expression levels (Supplementary Figure 3.2-2). We then looked at the relationship between selection protein and selection molecule via growth curves for each of the constitutive constructs with different Cm and dP concentrations. Normalized growth rates were derived by dividing the growth rate of each well by the growth rate of a terminator-only construct ran in the same plate (Methods). From this, we observed that we would expect to see maximum growth rates with fully-inhibited PLpro for any value of Cm above 125 μM, with the highest selection stringency possible being 1000 μM Cm (Figure 3.1-1g). In the case of HSVtk, 0.625 μM of dP was sufficient to provide a growth rate advantage for dPLpro over PLpro (Figure 3.1-1h). Mild growth-rate hits were also observed at 0.625 μM even when cells were expressing minimal amounts of HSVtk, indicating some non-specific growth rate inhibition by dP at that concentration.

While we expected to be able to use the GFP and RFP fusions as a proxy for selection protein expression levels, we sought to validate this assumption with the previously-constructed active / dead protease controls. If our fluorescent protein fusions were appropriate expression indicators, we would expect to be able to predict relative growth rates from the active / dead PLpro control fluorescence mapped onto the constitutive promoter survey (dashed lines in Figures 3.1-1g and 3.1-1h). To this end, we subjected the active / dead controls for PLpro to a range of Cm and dP concentrations. For CAT, the active protease control showed reduced growth above 63 μM Cm, while the catalytically-dead protease did not show a growth hit until 1000 μM Cm (Figure 3.1-1i). These results are consistent with the relative growth rates estimated from the constitutive promoter survey (Figure 3.1-1g). We also conducted a similar experiment for HSVTk-induced growth rate inhibition, which, as a negative selection, should follow the opposite trend. Concentrations up to 0.625 μM dP showed little growth rate inhibition of the active protease control and significant inhibition

of the dead protease control (Figure 3.1-1j). As previously observed with the constitutive promoters, concentrations of dP above 0.625 μM also showed some non-specific inhibition.

At a functional level, any cell-based selection method is dependent on relative growth rate differences between cells containing "hits" and cells lacking "hits". Ideally, a selection stringency would be chosen such that the normalized growth rates of the hits and non-hits (dead and active proteases in our case) would be 1 and 0, respectively. Using the previously-described selection system characterization, we now have a fine-grained mapping of selection stringency to Cm and dP concentrations.

3.1.1.3 Selection campaign for protease inhibitors identifying modification-dependent inhibition of PLpro

To generate the chemical diversity for protease inhibitor selections, we chose to base our library off the trunkamide pathway in the cyanobactin RiPP class [172]. We chose this pathway because the order of post-translational modification (Figure 3.1-2a) is well-understood [173] and the core sequences have been shown to be highly tolerant to mutation [174, 175] allowing for highly-diverse libraries. A 9-AA macrocyclic peptide library was designed by randomizing positions 1-6, and 8 with NNK degenerate codons (Figure 3.1-2a). The Pro at position 6 was kept because tolerance to its mutation has not been well-explored [174]. Additionally a Cys was fixed at position 9 because heterocyclization by LynD has been shown to be necessary precursor to macrocyclization by TruG [176]. Previously-described leader and follower sequences containing LynD, TruA, and TruG recognition sites [175] were appended to the N and C terminus of the core. Expression was stabilized via the N-terminal appendage of a His6-SUMO-TEV sequence. Ideal post-translational modification would result in heterocyclization by LynD, N-terminal cleavage by TruA, and lastly macrocylization by TruG to form a 9-membered ring (Figure 3.1-2a).

For *in vivo* expression, the library was placed on separate plasmid from the circuit, using a pSC101(var2) backbone. A P_{T5LacO} promoter (IPTG inducible) and a strong RBS was used to drive it. In addition to the protease detection circuit plasmid and library plasmids, a separate plasmid with the three modifying enzymes was also constructed (pEG06_047). The modifying enzymes were each controlled by their own inducible system: LynD (aTc inducible),

TruA (OC14-AHL inducible), and TruG (cumate inducible) and placed on a p15A backbone. This resulted in a three-plasmid system with protease detection, modifying enzymes, and library members each being on their own compatible backbone.

The initial peptide library was generated via TypeIIs cloning. The library plasmid (pEG02_217) was grown up, miniprepped and then transformed into a Marionette-Clo [139] (DH10B derivative) *E. coli* strain containing the PLpro detector circuit and modifying enzyme plasmids, resulting in the Round 0 selection library.

We then conducted rounds of selection to enrich for modification-dependent inhibitors of PLpro (Figure 3.1-2b). Fluorescence of GFP-CAT was monitored via flow cytometry (Methods) to assess the progression of the selection campaign. Between each round of selection, library plasmids were isolated via miniprep and then treated with restriction enzymes to remove the modifying enzyme and circuit plasmids. This step was designed to prevent the propagation of cheaters. The first selection was a positive selection round conducted at 125 μM Cm--the minimum Cm concentration known to completely inhibit growth of fully-active PLpro protease (Figure 3.1-2b and 3.1-1g). After Round 1, we observed a distinct shift in the protease activity circuit output towards inhibited proteases (Figure 3.1-2b). We lowered expression of the peptide by reducing IPTG and then continued with two more positive selection rounds, Round 2 and 3, increasing the stringency with 400 and 600 μM Cm (Figure 3.1-2b). After Round 3, we had a clear population of cells that seemed to contain inhibited protease. We then wanted to ensure that any hits we derived had modification enzyme dependent bioactivity. To select for this, we isolated the library and transformed it into a strain lacking the modifying enzyme plasmid. The peptide was induced as before and a negative selection round was conducted with 0.625 μM dP to deplete inhibitors that worked without the modifying enzymes present, resulting in Round 4. Following the negative selection, we isolated and retransformed the library into a strain with the modifying enzymes and conducted another positive selection, resulting in Round 5. Round 5 showed a high-expressing (inhibited protease) population, but also a low-expressing population despite the stringent positive selection round (Figure 3.1-2b). Considering that the library plasmid was carried through all of the selections and that dP has a high mutagenesis rate, we chose to PCR-isolate just the core region of the library and then reclone it into the library backbone. Transforming this recloned library into a strain with modifying

enzymes and the PLpro protease activity circuit showed that nearly the entire population contained inhibited PLpro protease (Figure 3.1-2b).

Individual colonies of the Recloned library were isolated then tested for bioactivity and modification dependence using the protease activity circuit. Three conditions were tested for each isolated colony: no peptide and no modifying enzymes; peptide and no modifying enzymes (62.5 μM IPTG); and peptide plus modifying enzymes (62.5 μM IPTG, 0.2 μM aTc, 10 μM OC14-AHL, and 100 μM Cuma). Out of the 24 clones tested, one hit showed a distinct modification enzyme dependence for its putative inhibitory activity (Supplementary Figure 3.2-5). This hit was named DAA680 and was carried through for further characterization.

3.1.1.4 Characterization of DAA680 implies a non-specific inhibition mechanism

While it's clear that DAA680 required modification enzyme expression for its bioactivity, the full trunkamide pathway has three distinct modification enzymes (Figure 3.1-2a), any combination of which may be responsible for the activity of DAA680. To further understand the specific post-translational modifications required for inhibitory activity, we re-ran the protease activity cytometry assay with the peptide induced (62.5 μM IPTG) and all eight different combinations of modifying enzyme induction (Figure 3.1-2c). From this, we observed the most inhibitory activity in the LynD + TruG condition. The level of GFP-CAT expression observed from the protease activity circuit was consistent with nearly-complete inhibition of PLpro (Figure 3.1-1c and 3.1-2c). The condition with all three enzymes expressed also showed significant, albeit reduced inhibition. The slow kinetics of macrocyclization by TruG [173, 176, 177] and the bimodality observed in the cytometry (data not shown) imply that LynD + TruA is likely the active state. Mapping the modifications by LynD and TruA onto the peptide sequence of DAA680 implied the active structure is a 13-AA peptide with a heterocycle (from LynD) at position 9 and N-terminal cleavage (from TruA). Notably, since the non-dependent enzyme TruG is responsible for both cleavage of the follower and macrocyclization, the predicted active from still contains the 4-AA TruG follower sequence (Figure 3.1-2d).

Next, DAA680 was assessed for its inhibitory specificity by transforming it into the CLpro detection circuit and monitoring its activity via cytometry. DAA680 was tested under

two conditions: no peptide with LynD + TruA (0 μM IPTG, 0.2 μM aTc, and 10 μM OC14-AHL); and peptide + LynD + TruA (62.5 μM IPTG, 0.2 μM aTc, and 10 μM OC14-AHL). From this, we observed that DAA680 had significant inhibitory activity against CLpro (Figure 3.1-2e). While the apparent inhibitor activity of DAA680 against CLpro seems to be smaller than that of PLpro, this could simply be due to lower relative expression of PLpro in the protease activity circuit. In fact, an expression level calculator predict 2-fold lower expression of PLpro relative CLpro in our protease detection circuits [178, 179]. While non-specific protease inhibition is the most obvious mechanism explaining the inhibitory results for both PLpro and CLpro, it is possible that the putative non-specific activity of DAA680 could originate from antiCRISPR activity against uncaged dCas9 used for readout of the protease activity circuits. We ruled out this mechanism by testing DAA680 in a dCas9-only circuit and did not observe any significant inhibition of repression (Supplementary Figure 3.2-6).

To further validate the protease inhibitor selections via *in vitro* activity assays, we conducted a large-scale production and isolation of DAA680. Since cleavage by TruA removes the N-terminal His-tag (Figure 3.1-2a) thus complicating isolation from cellular lysate, we chose to only conduct LynD modification during *in vivo* expression. This allowed for easy isolation of the LynD-modified peptide via its His-tag. N-terminal cleavage was subsequently conducted *in vitro* with the LynA enzyme (Methods) (Supplementary Figure 3.2-7). Following cleavage, the peptide was isolated via a methanol crash to remove the protease, solid-phase extraction, and HPLC (Methods). LCMS verification of the isolated peptide showed a high-purity peak with a mass consistent with heterocyclization (Supplementary Figure 3.2-7).

After isolation, DAA680 inhibitory activity was tested via *in vitro* protease activity assays for PLpro and CLpro (Methods). While we did not test high enough concentrations to get full inhibition of PLpro, we observed full inhibition of CLpro (Figure 3.1-2f). Hill function fits to the initial velocity V_0 values indicated half-inhibitory concentrations K_i of 85 μM and 14 μM for PLpro and CLpro, respectively (Table 1). These micromolar inhibitory concentrations are well-within the range of concentrations that various non-specific inhibition mechanisms have been observed to occur [180-183]. Additionally, the large Hill coefficient observed for CLpro (PLpro could not be determined) is a bellwether for non-specific artifacts often found in high-throughput screening methods (Table 1) [182, 184].

3.1.2 Discussion

This work demonstrates the integration of RiPP engineering for molecular diversity generation, synthetic genetic circuits for bioactivity readout, and cell-based selection methods for high-throughput screening. This approach allowed us to screen tens of millions of RiPPs for protease inhibition activity in a single test tube, which is orders of magnitude more than what can be accomplished through complex, plate-based screens [185-187].

Our use of a dual selection system enabled us to select for RiPPs with modification dependence. To our knowledge, this is the first explicit selection for modification enzyme dependence on bioactivity performed with RiPP-based libraries. Peptide modifications by RiPP enzymes are diverse [188], and many of these modifications can provide favorable medicinal properties such as increased membrane permeability [189], conformational constrainment [189, 190], proteolytic stability [191], and improved pharmacokinetics [192]. Interestingly, in our DAA680 hit, we seemed to have selected for modification dependence of only two of the three modifying enzymes (Figures 3.1-2a and 3.1-2c). Due to the non-specific nature of DAA680's inhibition activity, it is not certain that the heterocycle at position 9 is necessary for bioactivity. One possibility is that LynD heterocyclization may simply reduce the degradation rate of DAA680, increasing its effective concentration. The effect of modification enzyme dependent expression levels should be considered whenever modification-dependent bioactivity is observed via intracellular biosensors.

Non-specific activity is a universal issue that has plagued high-throughput screening campaigns for decades [184, 193-196]. While the highly-concentrated and crowded nature of the *E. coli* cytoplasm has quite different properties from the relatively dilute solutions used in high-throughput screens [197-199], we can look towards the extensive literature describing non-specific hits in high-throughput screens for insight into the mechanisms of DAA680 non-specificity. There are a variety of chemical mechanisms that can lead to non-specific bioactivity, however colloidal aggregation is the most common, occurring at a rate of 1 to 3% in high-throughput compound libraries [200]. While we did not explicitly explore to inhibitory mechanisms of DAA680, one of the hallmarks of aggregation-based inhibition is a large inhibition curve Hill coefficient [182, 184], which we observed in our *in vitro* assay for CLpro activity (Figure 3.1-2f). Additionally, DAA680 has an amphipathic structure with a highly hydrophobic N-terminal sequence and a hydrophilic C-term (Figure 3.1-2d). We observed

similarly hydrophobic sequences in many of the other active hits pulled from the Recloned library (data not shown). Small, amphipathic peptides are known to exhibit colloidal aggregation [201-204], providing one possible route for our library members to exhibit non-specific inhibition.

Whether or not aggregation is the primary mechanism of non-specific inhibition in our system, there are several possible approaches to reducing the number of non-specific hits coming out of our screens. Earlier interventions in the selection campaign are obviously more time and cost-efficient. The earliest possible change is to alter the protease activity sensor selection conditions to encourage the enrichment of better hits. The P_{T5LacO} promoter and RBS used to drive the production of our peptides in the selections was co-opted from our previous RiPP engineering efforts, which favored high expression levels for downstream isolation and mass spectrometry characterization. Even with the reduction of peptide expression after selection Round 1 (Figure 3.1-2b), we expect expression from P_{T5LacO} to still be high at the IPTG concentrations used [139]. Micromolar concentrations of compounds have been shown to be sufficient for colloidal aggregation [180, 193]. This concentration is readily achieved in *E. coli* heterologous expression systems with strong promoters [205]. Furthermore, macromolecular crowding effects in the *E. coli* cytoplasm can encourage protein aggregation [198]. Like has been done in plate-based high-throughput screening methods [193], lowering the intracellular concentration of our library members via further IPTG reduction may increase the likelihood of specific hits.

Another possible method of reducing non-specific inhibitors is to leverage the negative selection system developed and validated in this work (Figure 3.1-1h) with off-target proteases. Negative selections against non-specific hits are a fundamental step for many *in vitro* and cell-based selection technologies [139, 206-208]. In our selection campaign, one negative selection round was conducted without modifying enzymes to remove hits with modification-independent bioactivity (Figure 3.1-2b). In a similar way, negative selection rounds could be conducted against CLpro to enrich for variants that specifically inhibited PLpro. We note that while the HSVtk-based negative selection was relatively effective in our case, later experimentation determined that a PheS-RFP negative selection protein was a superior negative selection marker. This was due to the fact that the negative selection molecule for HSVtk, dP, was highly mutagenic and allowed many cells to escape from negative selection conditions (data not shown).

One of the primary contributions of this work is the exhaustive and generalizable characterization of CAT and HSVtk-based selections (Figure 3.1-1g and 3.1-1h). In the field of biotechnology, selections have been incredibly successful in producing bioactive compounds, however many selection-based approaches are *ad hoc*, difficult to monitor progression, and require extensive tuning for each new application. Since we fused fluorescent proteins to both CAT and HSVtk, we are able to explicitly define the relationships between intracellular protein concentrations, Cm/dP concentrations, and relative growth rates. Using basic active / dead protease controls (Figures 3.1-1c and 3.1-1d), we were able to estimate the ideal Cm/dP concentrations to use during selections. We validated this approach by explicitly looking at growth-rate changes in those active / dead controls (Figures 3.1-1g and 3.1-1h). Biosensors are a common input into cell-based selection methods [125, 139, 209-213] and this same approach can be easily generalized to selection with any biosensor that has a transcriptional output. With our fluorescence-based characterization data, it is no longer necessary to optimize both the biosensor input and the selection system performance for every new target. One can simply tune / optimize the biosensor to provide an output in the valid expression ranges defined by our characterization. This should enable rapid deployment of novel cell-based selections for molecules with diverse bioactivities.

3.1.3 Materials and Methods

3.1.3.1 Strains, media, and chemicals.

E. coli NEB 10-beta (C3019I, New England BioLabs, Ipswich, MA, USA) was used for all routine cloning. *E. coli* Marionette-Clo (with chloramphenicol resistance marker removed using the FRT sites) was used for all selection experiments [214]. *E. coli* Marionette-X, a Marionette-compatible derivative of NEB Express (C2523I, New England BioLabs, Ipswich, MA, USA) was used for large-scale peptide expression experiments. TB (T0311, Teknova, Hollister, CA, USA) supplemented with 0.4% glycerol (BDH1172-4LP, VWR, OH, USA) was used for all selection experiments and peptide expression. 2xYT liquid media (B244020, BD, Franklin Lakes, NJ, USA) and 2xYT + 2% agar (B214010, BD, Franklin Lakes, NJ, USA) plates were used for routine cloning and strain maintenance. SOB liquid media (S0210, Teknova, Hollister, CA, USA) was used for making competent cells. SOC liquid media (B9020S, New England BioLabs, Iwsich, MA, USA) was used for outgrowth. Unless noted

otherwise, cells were induced with the following chemicals: cuminic acid (268402, Millipore Sigma, Saint Louis, MO, USA) added as 1000X stock (200 mM) in EtOH or DMSO; 3-oxohexanoyl-homoserine lactone (3OC6-AHL) (K3007, Millipore Sigma, Saint Louis, MO, USA) added as a 1000X stock (1 mM) in DMSO; anhydrotetracycline (aTc) (37919, Millipore Sigma, Saint Louis, MO, USA) added as a 1000X stock (100 µM) in DMSO; isopropyl ß-D-1-thiogalactopyranoside (IPTG) (I2481C, Gold Biotechnology, Saint Louis, MO, USA) added as 1000X stock (1 M) in water. Cells were selected with the following antibiotics: carbenicillin (carb, C-103-5, Gold Biotechnology, Saint Louis, MO, USA) added as 1000X stock (100 mg/mL in H2O); kanamycin (kan, K-120-10, Gold Biotechnology, Saint Louis, MO, USA) as 1000X stock (50 mg/mL in H2O); spectinomycin (spec, S-140-5, Gold Biotechnology, Saint Louis, MO, USA); and chloramphenicol (Cm, C-105-5, Gold Biotechnology, Saint Louis, MO, USA). Liquid chromatography was performed with Optima Acetonitrile (A996-4, Thermo Fisher Scientific, MA, USA) and water (Milli-Q Advantage A10, Millipore Sigma, Saint Louis, MO, USA) supplemented with LCMS Grade Formic Acid (85178, Thermo Fisher Scientific). The following solvents/chemicals were also used: Ethanol (V1001, Decon Labs, King of Prussia, PA, USA), Methanol (3016-16, Avantor, Center Valley, PA, USA), Ammonium bicarbonate (A6141 Millipore Sigma, Saint Louis, MO, USA), dimethyl sulfoxide (DMSO) (32434, Alfa Aesar, Ward Hill, MA, USA), Imidazole (IX0005, Millipore Sigma, Saint Louis, MO, USA), sodium chloride (X190, VWR, OH, USA), sodium phosphate monobasic monohydrate (20233, USB Corporation, Cleveland, OH, USA), sodium phosphate dibasic anhydrous (204855000, Acros, NJ, USA), guanidine hydrochloride (50950, Millipore Sigma, Saint Louis, MO, USA), tris (75825, Affymetrix, Cleveland, OH, USA), TCEP (51805-45-9, Gold Biotechnology, Saint Louis, MO, USA), and EDTA (0.5M stock, 15694, USB Corporation, Cleveland, OH, USA). DNA oligos and gBlocks were ordered from Integrated DNA Technologies (IDT) (San Francisco, CA, USA).

3.1.3.2 Plasmids and genes.

All protease detection circuit plasmids were generated via BsaI TypeIIs insertion into a pDAA352 backbone, which contains a ColE1* (p15 copy number) origin and Kan resistance marker. The modifying enzyme plasmid used during selections, pEG06_047 was constructed from previously-described modifying enzymes. For all protease-sensitive CP-dCas9 variants,

we started with the dCas9-CP[199] variant described in [22] and used the -6AA C-term truncation. Inserting the cleavage sites for PLpro and CLpro resulted in linkers RIDLTLKGGAPTDKKY and RIDLSVKLQSGFDKKY, respectively, where underlined portions are the cleavable linker and flanking regions are the original dCas9 N-term and C-term. Superfolder GFP (GFP) [215] and mScarlet-I (RFP) [216] were fused to the N-term and C-term of CAT and HSVTK with a GG and AEAAAKEAAAKAAAAKA (H-linker [217]) linkers, respectively.

3.1.3.3 Fluorescence readout assays.

Experiments were initiated by picking single colonies into 1mL TB + antibiotics. Overnight growth (~20 hr) was then performed at 30 in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) at 900 r.p.m. in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). If induction was not necessary (as for constitutive systems), samples were then prepared for cytometry analysis as outlined in "flow cytometry analysis". If induction was necessary, cultures were then diluted 1:1000 into inducing conditions by adding 1 μL of saturated overnight culture into 1mL TB + antibiotic(s) + inducer(s). This was then grown overnight (~20 hr) in the same conditions and prepared for cytometry analysis the next day.

3.1.3.4 Flow cytometry analysis.

Fluorescence characterization was performed on a BD LSR Fortessa flow cytometer with the HTS attachment (BD, Franklin Lakes, NJ, USA). Samples were prepared by diluting saturated overnight cultures 1:600 by adding 0.5 μL of cell culture into 300 μL of PBS containing 0.2 mg/mL Kan. All samples were run in high throughput mode at a flow rate of 0.5 μL/s. PBS-only wells were ran between sample wells to prevent cross-sample carryover. Fluorescence measurements were made using the blue (488 nm) and green (532 nm) lasers for GFP and RFP, respectively. GFP fluorescence readings were derived from the FITC-A channel (PMT voltage of 400 V) and RFP / mScarlet-I readings were derived from the PE-Texas Red-A channel (PMT voltage of 650 V). The FSC and SSC voltages were 650 V and 270

V, respectively. At least 20,000 events were collected for each sample and the Cytoflow Python package (https://github.com/cytoflow/cytoflow) was used for downstream analysis. Gating was completed by fitting a 2D Gaussian function to the FSC-A and SSC-A distributions and excluding all events greater than three standard deviations from the mean. When a single summary statistic is derived, the median fluorescence value of the distribution is used.

3.1.3.5 Relative growth assays.

Saturated overnight cultures were created as described for cytometry induction assays. Cells were brought into exponential growth by diluting 6.5 μL of overnight culture into 1mL TB + antibiotics (OD ~0.08). These cultures were grown for 2 hours using the same growth conditions until an OD ~0.2. Cultures were then OD-matched and then diluted to OD = 0.01 in varying amounts of Cm, dP. 200 μL of these cultures were grown in BioTek Synergy H1 plate reader with shaking for at 30C. Continuous OD600 measurements were taken for 16 hours. Single-sample growth rates were derived by log-transforming the OD600 measurements, and extracting the linear portion of the curve with the maximum slope with custom Python scripts. This slope corresponds to the exponential growth rate constant for that given sample. See Supplementary Figures 3.2-3 and 3.2-4 for representative examples of growth curves and their fits. A normalization strain (sDAA626 with plasmid, pDAA424: Terminator-sfGFP:CAT-hsvtk:mScarlet) was also inoculated across 3 different locations in each plate. Single-sample growth rates were normalized by dividing each samples' growth rate by the mean maximum growth rate of the normalization strain ran within the same plate.

3.1.3.6 Constitutive promoter expression calculations.

A series of six different constitutive promoters (and one terminator) were placed in front of the sfGFP-CAT-HSVtk-RFP cassette to generate a range of expression levels. The expression level of CAT and HSVtk was assessed by the fluorescence of their fused fluorescent proteins. Following the protocols in described in *Induction assays for cytometry* and *Flow cytometry*

analysis, fluorescence values were derived from three different experiments performed on three different days. The mean fluorescence value of the three replicates is used to define the expression level of the promoter in fluorescence units. See Supplementary Figure 3.2-2a and 2b for all replicate values.

3.1.3.7 DAA680 expression and purification.

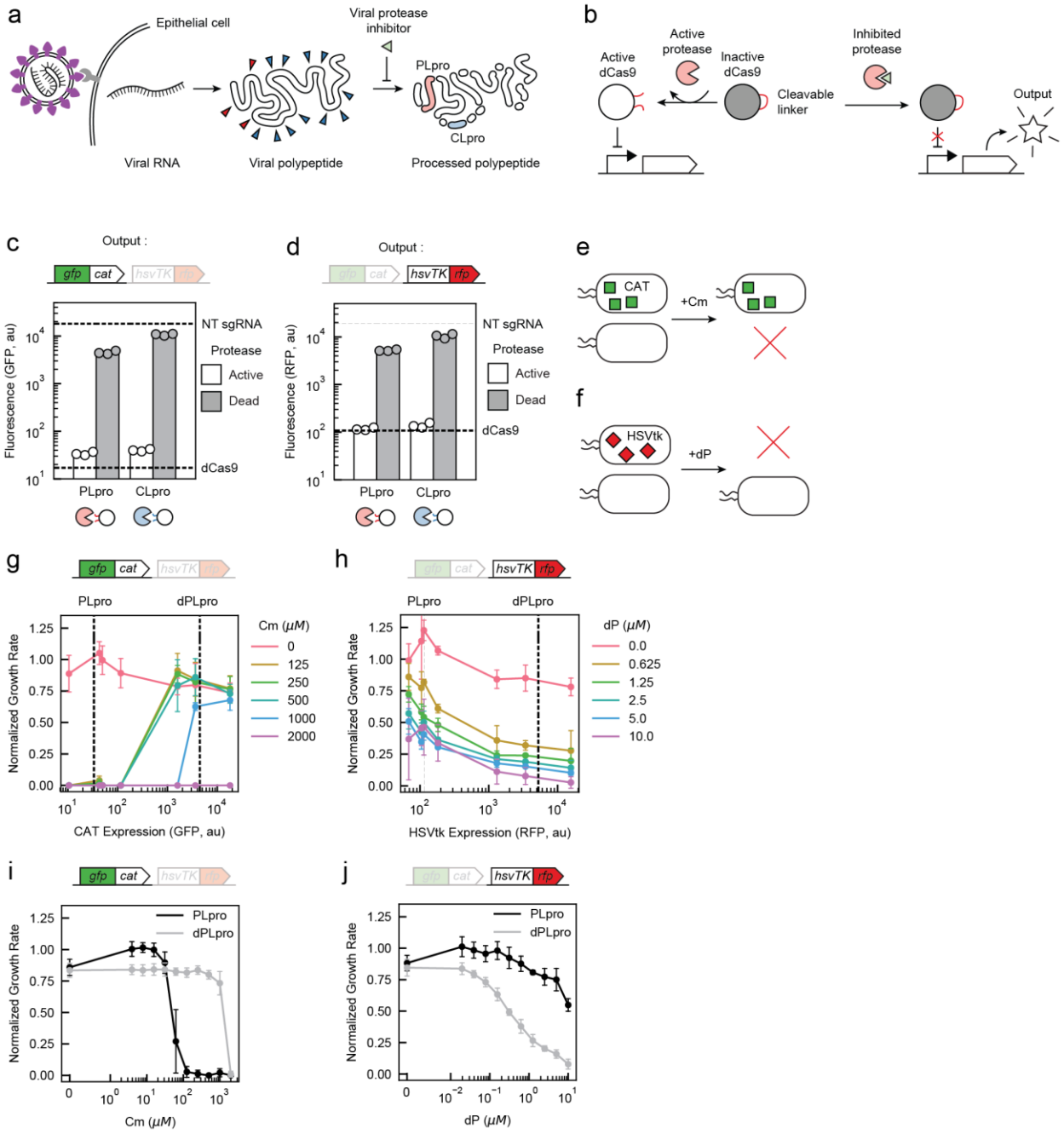
Peptide (pDAA646) and LynD-only (pEG07_160) plasmids were transformed into *E. coli* Marionette-X and plated on 2xYT agar + Kan + Carb. The next day, single colonies were picked into 30 mL LB + antibiotics and grown overnight at 30 C, 250 r.p.m. 10 mL of saturated overnight culture was then subcultured into 0.5 L of TB + antibiotics and incubated at 30 C, 250 r.p.m. After 3 hours (OD ~1) 1 mM IPTG and 100 μM Cuma were added to induce the peptide and modifying enzyme. The cultures were then transferred to 18 C growth at 250 r.p.m. for 20 hours. These cultures were then spun down and freeze-thawed to promoter lysis. Lysed cells were then centrifuged to remove cellular debris and the supernatant was kept. Nickel columns were used for His-tag purification. After His-tag pulldowns, the peptide was eluted and then dialyzed overnight to remove salts. LynA cleavage was then performed for 24 hours at 37C before a methanol crash. Solid phase extraction was used to isolate the cleaved peptide and HPLC and subsequent lyophilization lead to highly-pure compound.

3.1.3.8 In vitro *protease activity assays.*

All protease assays were conducted in a BioTek Synergy H1 plate reader without shaking for 1 hour at 37C. BPS Bioscience protease assay kits were used for both PLpro and CLpro assays (#79995-1 and #78015, respectively). Wavelengths of 360 nm / 460 nm were used for excitation / emission. Assuming valid Michaelis-Menten conditions, initial velocity V_0 values were extracted by fitting a linear function to the 460 nm emission with respect to time between 20 and 40 minutes. Normalized V_0 for each protease was calculated by dividing each reading by the V_0 of the protease-only well. Hill function fits of protease inhibition curves used Python's scipy.optimize.curve_fit() function to fit $y = y_{max}/(1+(x/K_I)\text{^}n)$, where y is V_0, x

is the inhibitor concentration, K_I is the IC50 for the inhibitor, n is the Hill coefficient, and y_{max} is the maximum V_0 observed.

3.1.4 Figures



<u>Figure 3.1-1</u>: An in vivo selection system for protease inhibitors

(a) Schematic of the viral lifecycle intervention point for viral protease inhibitors. Viral PLpro and CLpro autocatalytic cleavage sites are shown with red and blue triangles, respectively. Released PLpro and CLpro proteases are shown in red and blue, respectively. **(b)** Schematic of a caged-dCas9 protease inhibitor detection synthetic genetic circuit. "Caged" (inactive) dCas9 is colored gray and the "uncaged" (active) form is shown in white. The

constraining linker containing the protease cut site is indicated in red. **(c)** sfGFP-CAT expression levels for PLpro and CLpro protease inhibitor detection systems. Expression from circuits with both active and dead versions of the proteases are shown in white and gray, respectively. Dashed reference lines are the mean fluorescence values for normal dCas9 and dCas9 with a non-targeting sgRNA (NT sgRNA) (see Supplementary Figure 3.2-1). **(d)** Same as for (c), but showing HSVtk-RFP expression levels (see Supplementary Figure 3.2-1 for dCas9 control reference value determination). **(e)** Schematic showing positive selection for CAT-producing cells with chloramphenicol (Cm). **(f)** Schematic showing negative selection against HSVtk-producing cells with dP. **(g)** Relative growth values for the positive selection system across different Cm concentrations. CAT expression values on the x-axis were determined from flow cytometry measurements of the fused sfGFP protein (see Methods and Supplementary Figure 3.2-2). Dashed reference lines are the mean sfGFP-CAT fluorescence values for the active and catalytically-dead PLpro systems shown in (c). Representative growth curve fits of a single replicate are shown in Supplementary Figure 3.2-3. **(h)** Relative growth values for the negative selection system across different dP concentrations. HSVtk expression values on the x-axis were determined from flow cytometry measurements of the fused RFP protein (see Methods and Supplementary Figure 3.2-2). Reference lines are the mean HSVtk-RFP fluorescence values for the active and catalytically-dead PLpro systems shown in (d). Representative growth curve fits of a single replicate are shown in Supplementary Figure 3.2-3. **(i)** Relative growth rates for active and dead protease controls in the positive selection system across a range of Cm concentrations. Representative growth curve fits of a single replicate are shown in Supplementary Figure 3.2-4. Values of Cm (μM) = [0, 3.91, 7.81, 15.6, 31.3, 62.5, 125, 250, 500, 1000, 2000]. **(j)** Relative growth rates for active and dead protease controls in the negative selection systems across a range of dP concentrations. Representative growth curve fits of a single replicate are shown in Supplementary Figure 3.2-4. Values of dP (nM) = [0, 19.5, 39.1, 78.1, 156, 313, 625, 1250, 2500, 5000, 10000] For subfigures (c), (d), three experimental replicates were measured on three separate days and are shown with a bar drawn at their mean value. For subfigures (g), (h), (i), and (j), all growth values are normalized to the sDAA626 strain, which was ran in the same plate on the same day (Methods). For subfigures (g), (h), (i), and (j), single points are the mean of three replicates performed on three separate days. Standard deviations of each measurement are indicated with error bars.

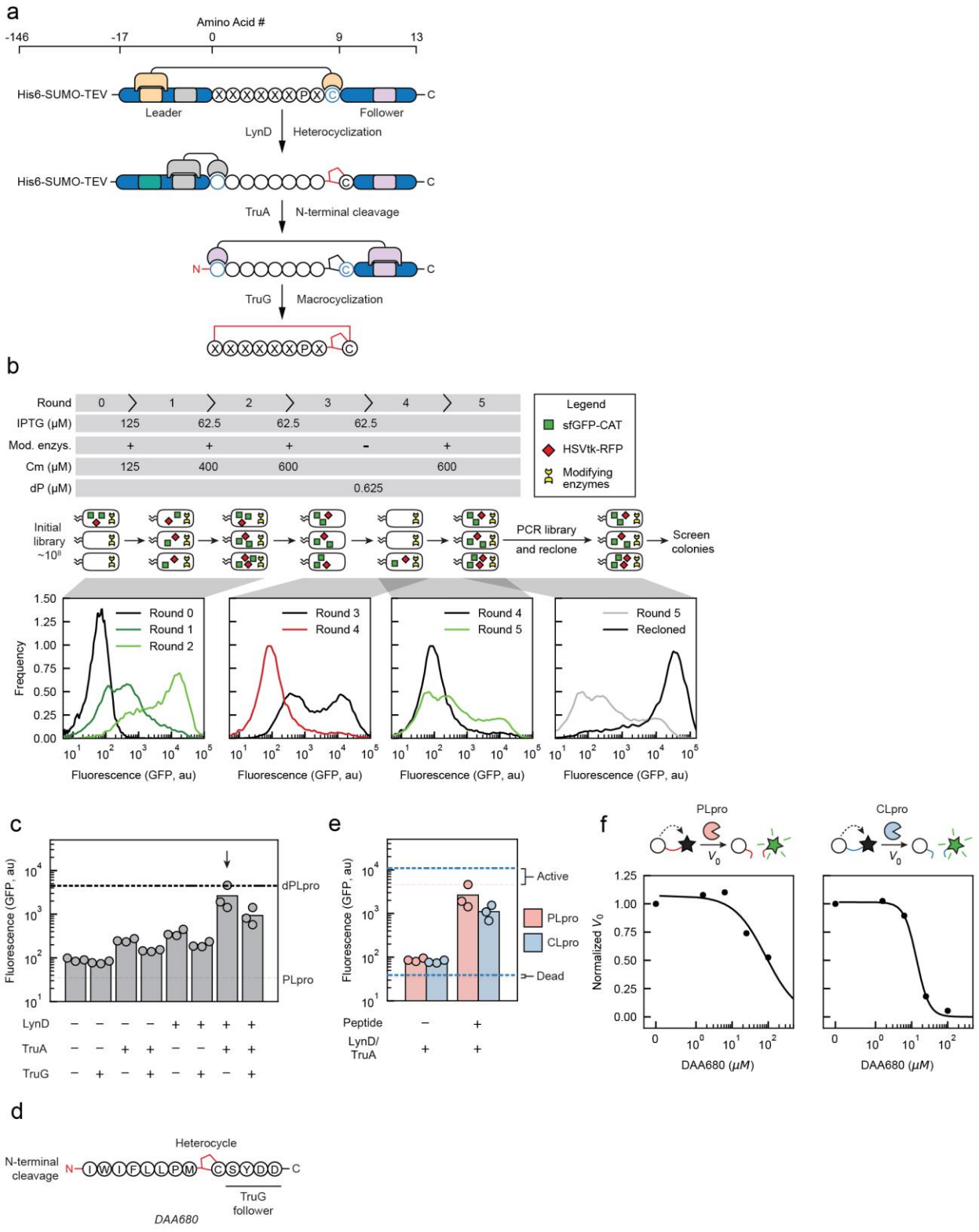


Figure 3.1-2: PLpro protease inhibitor selection campaign and result

(a) Intended post-translational modification pattern for the trunkamide-derived RiPP library members. Three enzymatic modifications are shown. Residues that are being modified in a

given step are colored blue and red indicates that they were modified in the previous step. NNK-randomized amino acids are designated with X. An amino acid scale is shown relative to the start of the diversified library core. **(b)** Selection campaign used to derive DAA680. For Rounds 1, 2, 3, and 5, cytometry distributions are from stationary-phase, post-selection cultures with all inducer conditions derived from the previous selection round. Round 4 cytometry was derived from a fresh outgrowth of post-negative selection culture without dP and with modifying enzyme induction (Methods). The "Recloned" cytometry distribution was derived from the final recloned library. All distributions are the result of a single experimental replicate collected during the selection campaign. **(c)** PLpro inhibitor detection circuit readout showing the putative modifying-enzyme-dependence of the isolated hit, DAA680. In all cases, the peptide was induced with 62.5 μM IPTG. Eight different modifying enzyme combinations were derived from combinations of 0 / 0.2 μM aTc (LynD), 0 / 10 μM OC14-AHL (TruA), and 0 / 100 μM Cuma (TruG). Dashed lines show the circuit's output with catalytically-active (PLpro) and catalytically-dead (dPLpro) protease (see Figure 3.1-1c). An arrow indicates the modifying enzyme combination with the most inhibition. **(d)** The modification state with the most activity as suggested by synthetic genetic circuit assessment in (c) and isolated via large-scale production and purification (see Supplementary Figure 3.2-6 for isolation procedure and mass spectrometry structural verification). **(e)** Non-specific protease inhibition of DAA680 observed with the PLpro and CLpro inhibition detection circuits. Peptide induction conditions are 0 / 62.5 μM IPTG, and LynD / TruA were induced together with 0 / 0.2 μM aTc (LynD) and 0 / 10 μM OC14-AHL (TruA). Red and blue dashed lines show the fluorescence from the active and dead protease controls for PLpro and CLpro systems, respectively (see Figure 3.1-1c). **(f)** *In vitro* assessment of PLpro and CLPro inhibition by DAA680. Initial velocity (V_0) values were derived from the time-dependent change in fluorescence of the probe and were normalized to the 0 μM DAA680 condition. Data points are derived from a single replicate. Lines are model fits of the Hill function to all replicates (Methods) with fit parameters in Table 1. For subfigures (c), (d), three experimental replicates were measured on three separate days and are shown with a bar drawn at their mean value.

3.1.5 Tables

Table 3.1-1: Hill function parameter fits for inhibition by DAA680 within in vitro protease assays

Protease	y_{max}	K_i (μM)	n [a]
PLpro	1.02	85.2	-
CLpro	1.07	13.8	2.5

a) Hill coefficient n could not be accurately determined for PLpro since concentrations were not high enough to reach full inhibition.

3.2 Supplementary

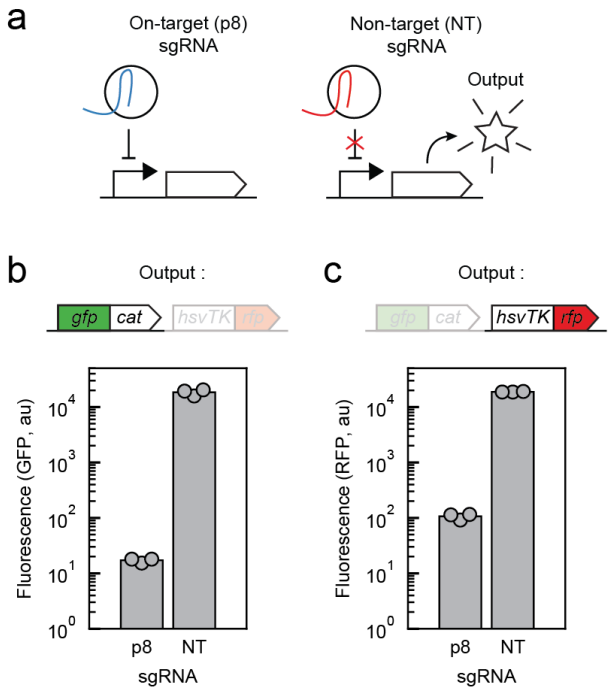


Figure 3.2-1: dCas9-based circuit controls

This is a companion figure to Figures 3.1-1c and 3.1-1d. **(a)** Schematic of the expected performance of the standard dCas9 circuit with promoter-targeting (p8) and non-targeting (NT) sgRNAs. **(b)** Fluorescence output of sfGFP-CAT for on-target (p8) and non-targeting (NT) sgRNAs. **(c)** Fluorescence output of HSVtk-RFP with dCas9 controls for on-target (p8) and non-targeting (NT) sgRNAs. For subfigures (b) and (c), three experimental replicates were measured on three separate days and are shown with a bar drawn at their mean value. These mean values are shown in Figure 3.1-1c, and Figure 3.1-1d.

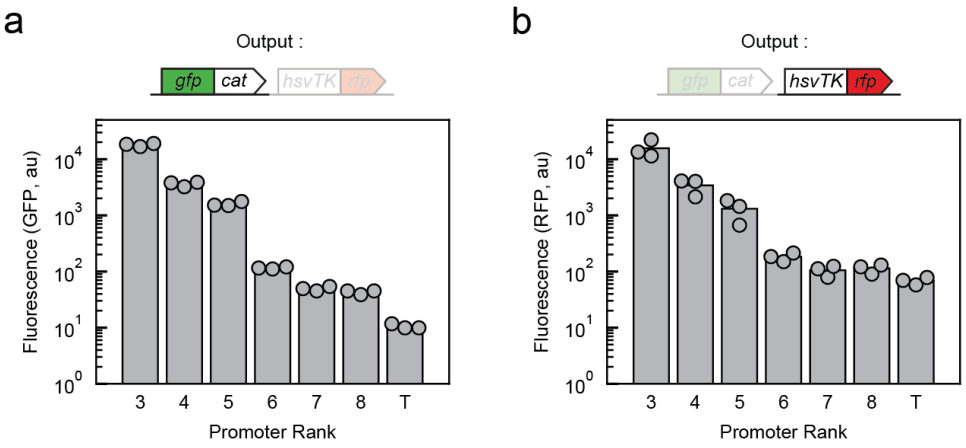


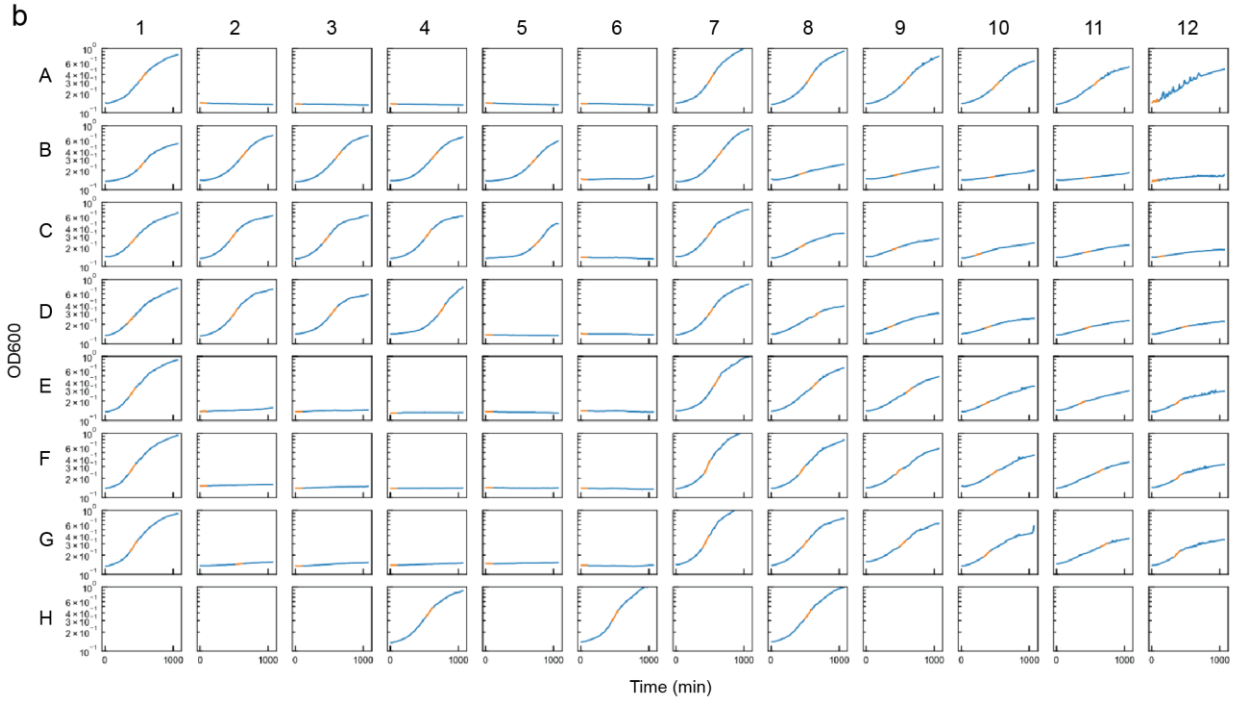
Figure 3.2-2: Constitutive promoter strengths for relative growth rate experiments

This is a companion figure to Figures 3.1-1g and 3.1-1h. **(a)** Fluorescence values of sfGFP-CAT for the constitutive promoters (and terminator, T) used to test positive selection growth rates at different expression levels. Mean fluorescence values are used to define the x-axis values in Figure 3.1-1g. **(b)** Fluorescence values of HSVTK-RFP for the constitutive promoters (and terminator, T) used to test positive selection growth rates at different expression levels. Mean fluorescence values are used to define the x-axis values in Figure 3.1-1h.

a

	1	2	3	4	5	6	7	8	9	10	11	12	
A	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	sDAA626	
B	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	sDAA627	
C	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	sDAA628	
D	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	sDAA629	
E	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	sDAA630	
F	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	sDAA631	
G	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	sDAA632	
H	blank	blank	blank	sDAA626	blank	sDAA626	blank	sDAA626	blank	blank	blank	blank	
	0	0	0	0	0	0	0	625	1250	2500	5000	10000	dP (nM)
	0	125	250	500	1000	2000	0	0	0	0	0	0	Cm (uM)

b



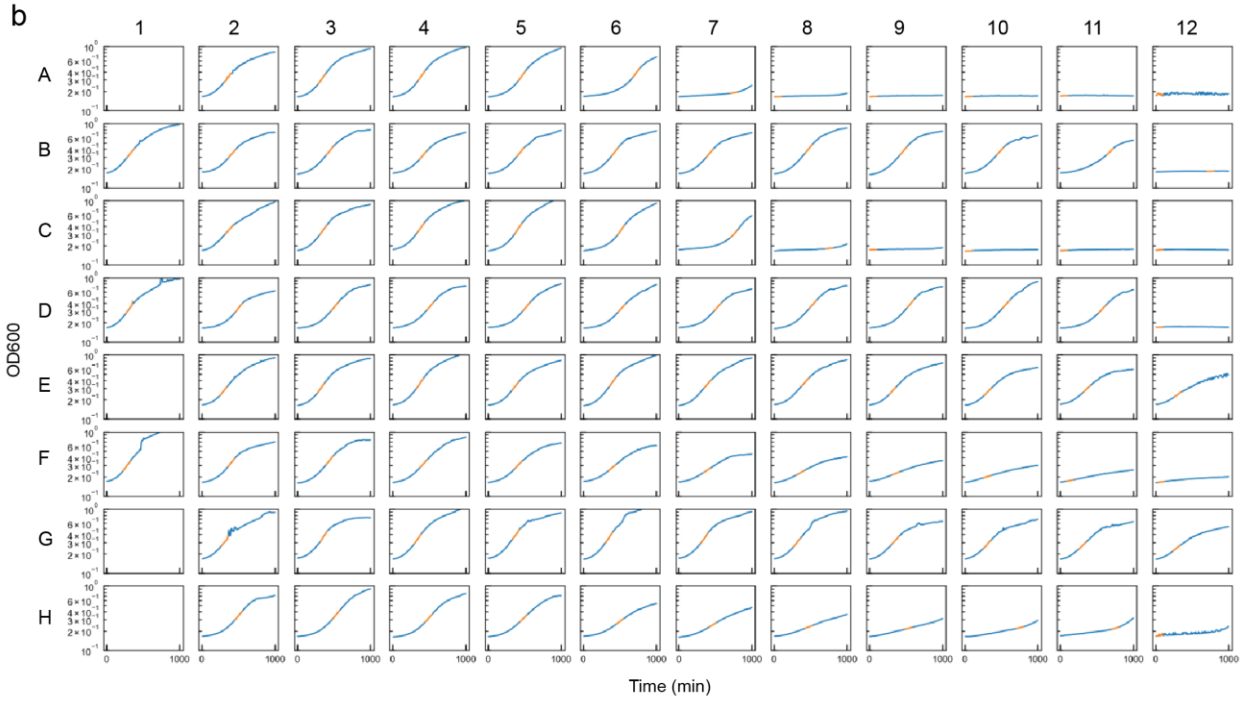
<u>Figure 3.2-3</u>: Raw growth curve data and fits for positive and negative selection system testing with constitutive promoters

This is a companion figure to Figures 3.1-1g and 3.1-1h. **(a)** Experimental plate setup for the relative growth assay. Positive and negative selection wells are colored green and red, respectively. Strains in each well are indicated. Location of normalization strain sDAA626 is also shown. **(b)** Raw growth curves for a single representative replicate of the experiment. For all subplots, x-axis values are the time in minutes and y-axis values are OD600 readings. The portion of the growth curve that was algorithmically chosen to define each well's growth rate is highlighted in orange (see Methods).

a

	1	2	3	4	5	6	7	8	9	10	11	12	
A	blank	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	
B	sDAA626	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	
C	blank	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	
D	sDAA626	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	
E	blank	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	sDAA641	
F	sDAA626	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	sDAA643	
G	blank	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	sDAA1110	
H	blank	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	sDAA1111	
	0	0	3.90625	7.8125	15.625	31.25	62.5	125	250	500	1000	2000	Cm (uM)
	0	0	19.53125	39.0625	78.125	156.25	312.5	625	1250	2500	5000	10000	dP (nM)

b



<u>Figure 3.2-4</u>: Raw growth curve data and fits for active / dead protease controls

This is a companion figure to Figures 3.1-1i and 3.1-1j. **(a)** Experimental plate setup for the relative growth assay. Positive and negative selection wells are colored green and red, respectively. Cm and dP values within each well correspond to the values shown below the plate that match their color (no wells have both Cm and dP). Strains in each well are indicated. Location of normalization strain sDAA626 is also shown. **(b)** Raw growth curves for a single representative replicate of the experiment. For all subplots, x-axis values are the time in minutes and y-axis values are OD600 readings. The portion of the growth curve that was algorithmically chosen to define each well's growth rate is highlighted in orange (see *Relative growth assays* in Methods).

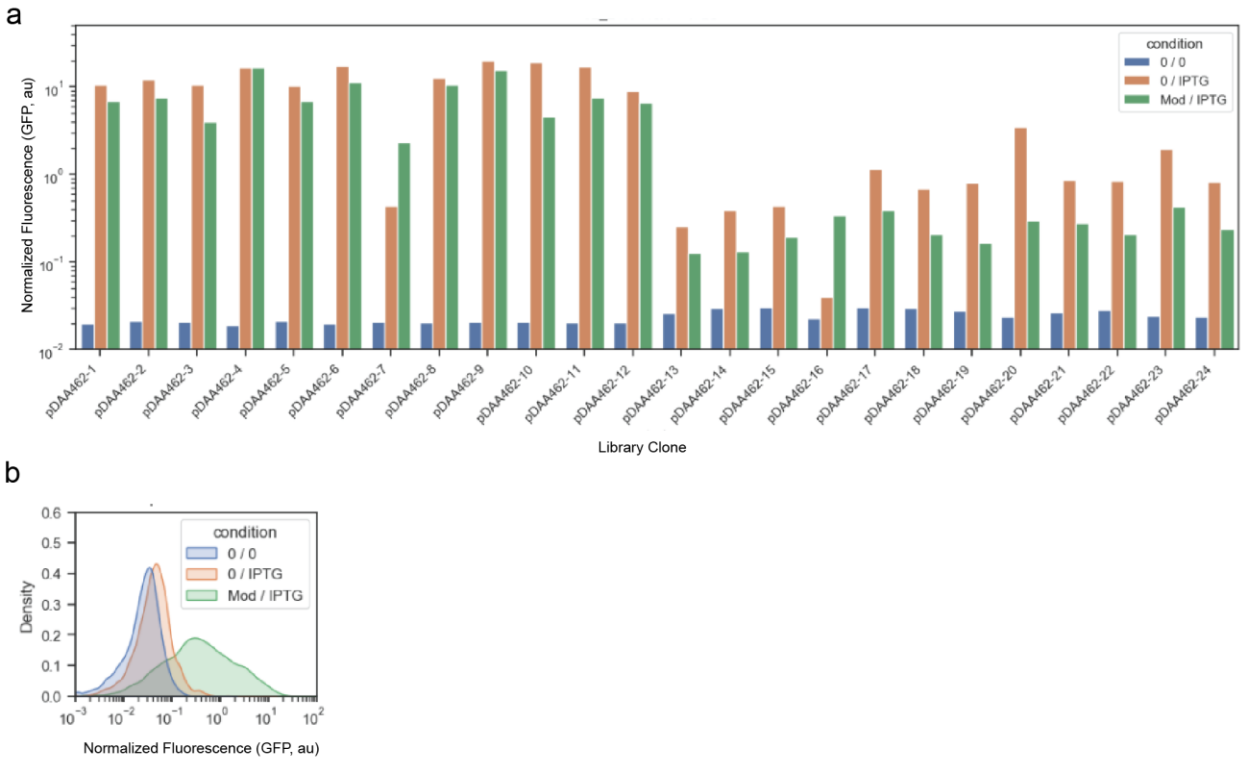


Figure 3.2-5: Activity and modification enzyme dependence of isolated PLpro inhibitors

(a) Fluorescence measurements of the 24 colonies isolated from the final recloned library. Each clone was tested under three conditions: no peptide and no modifying enzyme induction (0/0), no modifying enzyme with peptide induction (0/IPTG), and modifying enzymes with peptide induction (Mod / IPTG). When induced, modifying enzymes were induced with 0.2 µM aTc, 10 µM OC14-AHL, and 100 µM Cuma. When induced, peptide was induced with 62.5 µM IPTG. **(b)** Raw cytometry profile of the modification-dependent hit found in the screen shown in part (a). All data points are derived from a single replicate.

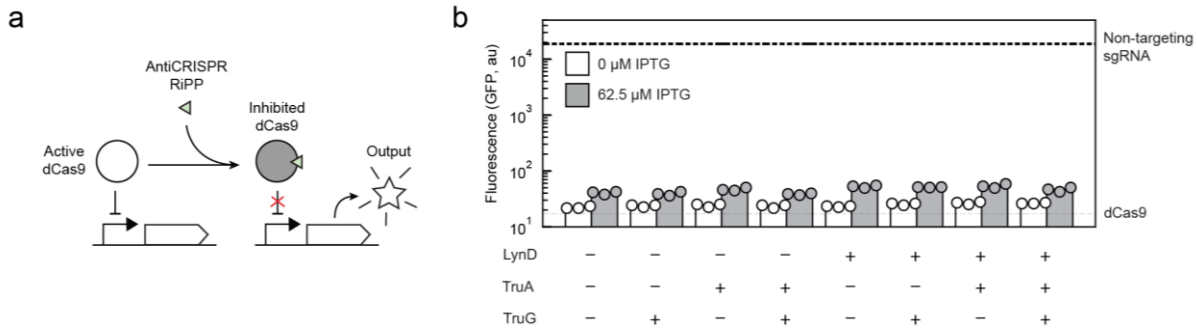


Figure 3.2-6: Testing DAA680 for AntiCRISPR activity

(a) Schematic representation of how the dCas9 circuit would be expected to perform if we had selected for an AntiCRISPR RiPP. Active dCas9 is indicated in white and the inhibited dCas9 is indicated in gray. **(b)** Standard dCas9 repression circuit readout showing with expression of DAA680 and eight different modifying enzyme expression combinations. Unmodified peptide induced with 0 / 62.5 μM IPTG. Eight different modifying enzyme combinations were derived from 0 / 0.2 μM aTc (LynD), 0 / 10 μM OC14-AHL (TruA), and 0 / 100 μM Cuma (TruG) inducer concentrations. Reference dashed lines are the mean fluorescence values for normal dCas9 and dCas9 with a non-targeting sgRNA (NT sgRNA) (see Supplementary Figure 3.2-1b for dCas9 control replicates). Three experimental replicates were measured on three separate days and are shown with a bar drawn at their mean value.

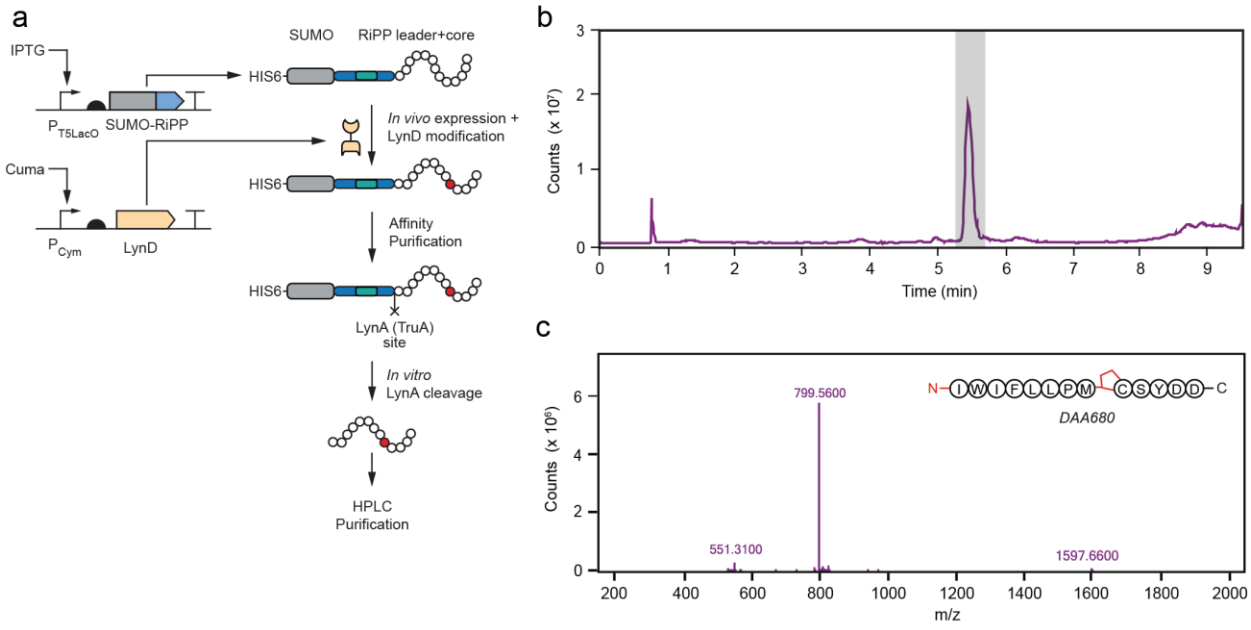


Figure 3.2-7: Production and purification of DAA680

(a) Schematic representation of the peptide production and purification process for DAA680. For large-scale production and purification, the TruA homolog, LynA, is used to conduct *in vitro* N-terminal cleavage rather than TruA. **(b)** The total ion chromatogram (TIC) of the final HPLC-purified DAA680 peptide. The time slice used to generate the mass spectrum in (c) is shown in gray. **(c)** Mass spectrum of the time slice shown in (b) indicating that the singly-charged DAA680 peptide is the primary constituent.

References

[1] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science,* vol. 337, pp. 816-821, 2012.

[2] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin*, et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression," *Cell,* vol. 152, pp. 1173-1183, 2013.

[3] K. M. Esvelt, P. Mali, J. L. Braff, M. Moosburner, S. J. Yaung, and G. M. Church, "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing," *Nature Methods,* vol. 10, pp. 1116-1123, 2013.

[4] D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, and L. A. Marraffini, "Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system," *Nucleic Acids Research,* vol. 41, pp. 7429-7437, 2013.

[5] A. Vigouroux, E. Oldewurtel, L. Cui, D. Bikard, and S. van Teeffelen, "Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes," *Molecular Systems Biology,* vol. 14, pp. e7899-e7899, 2018.

[6] R. Clarke, R. Heler, M. S. MacDougall, N. C. Yeo, A. Chavez, M. Regan*, et al.*, "Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks," *Molecular Cell,* vol. 71, pp. 42-55.e8, 2018.

[7] J. R. Widom, V. Rai, C. E. Rohlman, and N. G. Walter, "Versatile transcription control based on reversible dCas9 binding," *Rna,* vol. 25, pp. 1457-1469, 2019.

[8] Q. Zhang, F. Wen, S. Zhang, J. Jin, L. Bi, Y. Lu*, et al.*, "The post-PAM interaction of RNA-guided spCas9 with DNA dictates its target binding and dissociation," *Science Advances,* vol. 5, pp. 1-11, 2019.

[9] K. S. Whinn, G. Kaur, J. S. Lewis, G. D. Schauer, S. H. Mueller, S. Jergic*, et al.*, "Nuclease dead Cas9 is a programmable roadblock for DNA replication," *Scientific Reports,* vol. 9, pp. 1-9, 2019.

[10] E. A. Boyle, J. O. L. Andreasson, L. M. Chircus, S. H. Sternberg, M. J. Wu, C. K. Guegler*, et al.*, "High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding," *Proceedings of the National Academy of Sciences,* vol. 114, pp. 5461-5466, 2017.

[11] J. S. Hawkins, M. R. Silvis, B. M. Koo, J. M. Peters, H. Osadnik, M. Jost*, et al.*, "Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in Escherichia coli and Bacillus subtilis," *Cell Systems,* vol. 11, pp. 523-535.e9, 2020.

[12] M. Jost, D. A. Santos, R. A. Saunders, M. A. Horlbeck, J. S. Hawkins, S. M. Scaria*, et al.*, "Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs," *Nature Biotechnology,* vol. 38, pp. 355-364, 2020.

[13] A. Gilbert, Luke, H. Larson, Matthew, L. Morsut, Z. Liu, A. Brar, Gloria, E. Torres, Sandra*, et al.*, "CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes," *Cell,* vol. 154, pp. 442-451, 2013-07-01 2013.

[14] J. G. Zalatan, M. E. Lee, R. Almeida, L. A. Gilbert, E. H. Whitehead, M. La Russa*, et al.*, "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds," *Cell,* vol. 160, pp. 339-350, 2015.

[15] C. Dong, J. Fontana, A. Patel, J. M. Carothers, and J. G. Zalatan, "Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria," *Nature Communications,* vol. 9, 2018.

[16] H.-i. Ho, J. R. Fang, J. Cheung, and H. H. Wang, "Programmable CRISPR-Cas transcriptional activation in bacteria," *Molecular Systems Biology,* vol. 16, pp. 1-12, 2020.

[17] R. K. Shultzaberger, Z. Chen, K. A. Lewis, and T. D. Schneider, "Anatomy of Escherichia coli s 70 promoters," vol. 35, 2007.

[18] J. Fontana, C. Dong, C. Kiattisewee, V. P. Chavali, B. I. Tickman, J. M. Carothers*, et al.*, "Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements," *Nature Communications,* vol. 11, pp. 1-11, 2020.

[19] Y. Liu, X. Wan, and B. Wang, "Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria," *Nature Communications,* vol. 10, pp. 3693-3693, 2019.

[20] 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 Research,* pp. 1-10, 2021.

[21] J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun*, et al.*, "Evolved Cas9 variants with broad PAM compatibility and high DNA specificity," *Nature,* pp. 1-24, 2018.

[22] B. L. Oakes, C. Fellmann, H. Rishi, K. L. Taylor, S. M. Ren, D. C. Nadler*, et al.*, "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," *Cell,* vol. 176, pp. 254-267.e16, 2019.

[23] S. Zhang and C. A. Voigt, "Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design," *Nucleic Acids Research,* pp. 1-11, 2018.

[24] A. A. K. Nielsen and C. a. Voigt, "Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks," *Molecular systems biology,* vol. 10, pp. 763-763, 2014.

[25] A. Serganov and D. J. Patel, "Ribozymes, riboswitches and beyond: regulation of gene expression without proteins," *Nature reviews. Genetics,* vol. 8, pp. 776-790, 2007.

[26] W. Tang, J. H. Hu, and D. R. Liu, "Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation," *Nature Communications,* vol. 8, pp. 1-8, 2017.

[27] K. Kundert, J. E. Lucas, K. E. Watters, C. Fellmann, A. H. Ng, B. M. Heineike*, et al.*, "Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs," *Nature Communications,* vol. 10, 2019-12-01 2019.

[28] J. Chappell, A. Westbrook, M. Verosloff, and J. B. Lucks, "Computational design of small transcription activating RNAs for versatile and dynamic gene regulation," *Nature Communications,* vol. 8, pp. 1-11, 2017.

[29] K. H. Siu and W. Chen, "Riboregulated toehold-gated gRNA for programmable CRISPR–Cas9 function," *Nature Chemical Biology,* vol. 15, pp. 217-220, 2019.

[30] Y. J. Lee, A. Hoynes-O'Connor, M. C. Leong, and T. S. Moon, "Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system," *Nucleic Acids Research,* vol. 44, pp. 2462-2473, 2016.

[31] Y. Nihongaki, F. Kawano, T. Nakajima, and M. Sato, "Photoactivatable CRISPR-Cas9 for optogenetic genome editing," *Nature Biotechnology,* vol. 33, pp. 755-760, 2015.
[32] B. Zetsche, S. E. Volz, and F. Zhang, "A split-Cas9 architecture for inducible genome editing and transcription modulation," *Nature Biotechnology,* vol. 33, pp. 139-142, 2015-02-01 2015.
[33] M. Nakamura, P. Srinivasan, M. Chavez, M. A. Carter, A. A. Dominguez, M. La Russa*, et al.*, "Anti-CRISPR-mediated control of gene editing and synthetic circuits in eukaryotic cells," *Nature Communications,* vol. 10, 2019-12-01 2019.
[34] F. Bubeck, M. D. Hoffmann, Z. Harteveld, S. Aschenbrenner, A. Bietz, M. C. Waldhauer*, et al.*, "Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9," *Nature Methods,* vol. 15, pp. 924-927, 2018.
[35] H. Kim, D. Bojar, and M. Fussenegger, "A CRISPR/Cas9-based central processing unit to program complex logic computation in human cells," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 116, pp. 7214-7219, 2019.
[36] 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," *Nature Methods,* vol. 13, pp. 1043-1049, 2016.
[37] Z. Bao, S. Jain, V. Jaroenpuntaruk, and H. Zhao, "Orthogonal Genetic Regulation in Human Cells Using Chemically Induced CRISPR/Cas9 Activators," 2017.
[38] B. Maji, C. L. Moore, B. Zetsche, S. E. Volz, F. Zhang, M. D. Shoulders*, et al.*, "Multidimensional chemical control of CRISPR-Cas9," *Nature Chemical Biology,* vol. 13, pp. 9-11, 2017.
[39] M. Taketani, J. Zhang, S. Zhang, A. J. Triassi, Y. J. Huang, L. G. Griffith*, et al.*, "Genetic circuit design automation for the gut resident species Bacteroides thetaiotaomicron," *Nature Biotechnology,* vol. 38, pp. 962-969, 2020.
[40] 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," *Nature Communications,* vol. 8, pp. 15459-15459, 2017.
[41] J. T. Sexton and J. J. Tabor, "Multiplexing cell-cell communication," *Molecular Systems Biology,* vol. 16, 2020-07-01 2020.
[42] J. Santos-Moreno, E. Tasiudi, J. Stelling, and Y. Schaerli, "Multistable and dynamic CRISPRi-based synthetic circuits," *Nature Communications,* vol. 11, pp. 1-8, 2020.
[43] J. Kuo, R. Yuan, S. Carlos, J. Paulsson, and P. A. Silver, "Toward a translationally independent RNA-based synthetic oscillator using deactivated CRISPR-Cas," *Nucleic Acids Research,* pp. 1-13, 2020.
[44] J. Henningsen, M. Schwarz-Schilling, A. Leibl, J. Guttierez, S. Sagredo, F. C. Simmel*, et al.*, "Single Cell Characterization of a Synthetic Bacterial Clock with a Hybrid Feedback Loop Containing dCas9-sgRNA," *ACS Synthetic Biology,* vol. 9, pp. 3377-3387, 2020.
[45] S. E. Clamons and R. M. Murray, "Modeling Dynamic Transcriptional Circuits with CRISPRi," *bioRxiv,* 2017.
[46] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature,* vol. 403, pp. 335-338, 2000.
[47] A. Westbrook, X. Tang, R. Marshall, C. S. Maxwell, J. Chappell, D. K. Agrawal*, et al.*, "Distinct timescales of RNA regulators enable the construction of a genetic pulse generator," *Biotechnology and Bioengineering,* vol. 9, pp. 377572-377572, 2019.
[48] F. Wu, J. Shim, and C. Tan, "Orthogonal tuning of gene expression noise using CRISPR – Cas," pp. 1-10, 2020.

[49] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature,* vol. 403, pp. 339-342, 2000.
[50] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski*, et al.*, "Genetic circuit design automation," *Submitted,* vol. 352, pp. aac7341-aac7341, 2016.
[51] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, "The lac Repressor Displays Facilitated Diffusion in Living Cells," *Science,* vol. 336, pp. 1595-1598, 2012-06-22 2012.
[52] D. L. Jones, P. Leroy, C. Unoson, D. Fange, V. Ćurić, M. J. Lawson*, et al.*, "Kinetics of dCas9 target search in Escherichia coli," *Science,* vol. 357, pp. 1420-1424, 2017.
[53] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, and J. A. Doudna, "DNA interrogation by the CRISPR RNA-guided endonuclease Cas9," *Nature,* vol. 507, pp. 62-67, 2014.
[54] J. Shin, S. Zhang, B. S. Der, A. A. K. Nielsen, and C. A. Voigt, "Programming Escherichia coli to function as a digital display," pp. 1-12, 2020.
[55] L. P.-t. Nathan, D. L. Glenn, and V. Johan, "Synchronous long-term oscillations in a synthetic gene circuit," *Nature,* vol. 538, pp. 1-4, 2016.
[56] J. Fontana, C. Dong, J. Y. Ham, J. G. Zalatan, and J. M. Carothers, "Regulated Expression of sgRNAs Tunes CRISPRi in E. coli," *Biotechnology Journal,* vol. 1800069, 2018.
[57] H. H. Huang, M. Bellato, Y. Qian, P. Cárdenas, L. Pasotti, P. Magni*, et al.*, "dCas9 regulator to neutralize competition in CRISPRi circuits," *Nature Communications,* vol. 12, pp. 1-7, 2021.
[58] P.-y. Chen, Y. Qian, and D. D. Vecchio, "A Model for Resource Competition in CRISPR-Mediated Gene Repression," *57th IEEE Conference on Decision and Control,* 2018.
[59] S. Clamons and R. Murray, "Modeling predicts that CRISPR-based activators , unlike CRISPR-based repressors , scale well with increasing gRNA competition and dCas9 bottlenecking," *bioRxiv,* 2019.
[60] Q. Zhang, S. Bhattacharya, and M. E. Andersen, "Ultrasensitive response motifs: basic amplifiers in molecular signalling networks," *Open Biology,* vol. 3, pp. 130031-130031, 2013.
[61] J. E. Ferrell and S. H. Ha, "Ultrasensitivity part III: cascades, bistable switches, and oscillators," *Trends in Biochemical Sciences,* vol. 39, pp. 612-618, 2014-12-01 2014.
[62] F. Ceroni, A. Boo, S. Furini, T. E. Gorochowski, O. Borkowski, Y. N. Ladak*, et al.*, "Burden-driven feedback control of gene expression," *Nature Methods,* vol. 15, pp. 387-393, 2018.
[63] Y. Wu, T. Chen, Y. Liu, R. Tian, X. Lv, J. Li*, et al.*, "Design of a programmable biosensor-CRISPRi genetic circuits for dynamic and autonomous dual-control of metabolic flux in Bacillus subtilis," vol. 48, pp. 1-14, 2019.
[64] T. H. Segall-Shapiro, E. D. Sontag, and C. A. Voigt, "Engineered promoters enable constant gene expression at any copy number in bacteria," *Nature Biotechnology,* 2018.
[65] L. G. Lowder, D. Zhang, N. J. Baltes, J. W. Paul, X. Tang, X. Zheng*, et al.*, "A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation," *Plant Physiology,* vol. 169, pp. 971-985, 2015-10-01 2015.
[66] J. C. Wagner, R. J. Platt, S. J. Goldfless, F. Zhang, and J. C. Niles, "Efficient CRISPR-Cas9–mediated genome editing in Plasmodium falciparum," *Nature Methods,* vol. 11, pp. 1-6, 2014.

[67] E. Semenova, M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner*, et al.*, "Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence," *Proceedings of the National Academy of Sciences,* vol. 108, pp. 10098-10103, 2011.

[68] Lei S. Qi, B. Chen, B. Huang, J. Schnitzbauer, G.-W. Li, Elizabeth H. Blackburn*, et al.*, "Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System," *Cell,* vol. 155, pp. 1479-1491, 2013.

[69] H. Deng, R. Gao, X. Liao, and Y. Cai, "CRISPR system in filamentous fungi: Current achievements and future directions," *Gene,* vol. 627, pp. 212-221, 2017-09-01 2017.

[70] B. C. Stanton, A. a. K. Nielsen, A. Tamsir, K. Clancy, T. Peterson, and C. a. Voigt, "Genomic mining of prokaryotic repressors for orthogonal logic gates," *Nature chemical biology,* vol. 10, pp. 99-105, 2014.

[71] B. C. Stanton, V. Siciliano, A. Ghodasara, L. Wroblewska, K. Clancy, A. C. Trefzer*, et al.*, "Systematic transfer of prokaryotic sensors and circuits to mammalian cells," *ACS Synthetic Biology,* vol. 3, pp. 880-891, 2014.

[72] Y. Chen, S. Zhang, E. M. Young, T. S. Jones, D. Densmore, and C. A. Voigt, "Genetic circuit design automation for yeast," *Nature Microbiology,* vol. 5, pp. 1349-1360, 2020-11-01 2020.

[73] S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena*, et al.*, "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex," *Nature,* vol. 517, pp. 583-588, 2015.

[74] C. Lou, B. Stanton, Y.-J. Chen, B. Munsky, and C. a. Voigt, "Ribozyme-based insulator parts buffer synthetic circuits from genetic context," *Nature Biotechnology,* vol. 30, pp. 1137-1142, 2012.

[75] Z. Lu, S. Yang, X. Yuan, Y. Shi, L. Ouyang, S. Jiang*, et al.*, "CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in Bacillus subtilis," *Nucleic Acids Research,* vol. 47, 2019.

[76] R. Peng, Y. Wang, W.-W. Feng, X.-J. Yue, J.-H. Chen, X.-Z. Hu*, et al.*, "CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in Myxococcus xanthus," *Microbial Cell Factories,* vol. 17, 2018-12-01 2018.

[77] L. Yu, W. Su, P. D. Fey, F. Liu, and L. Du, "Yield Improvement of the Anti-MRSA Antibiotics WAP-8294A by CRISPR/dCas9 Combined with Refactoring Self-Protection Genes inLysobacter enzymogenesOH11," *ACS Synthetic Biology,* vol. 7, pp. 258-266, 2018-01-19 2018.

[78] F. Moser, A. E. Borujeni, A. N. Ghodasara, E. Cameron, Y. Park, and C. A. Voigt, "Dynamic control of endogenous metabolism with combinatorial logic circuits," *Mol Syst Biol,* vol. 14, pp. 8605-8605, 2018.

[79] B. H. Weinberg, N. T. H. Pham, L. D. Caraballo, T. Lozanoski, A. Engel, S. Bhatia*, et al.*, "Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells," *Nature Biotechnology,* 2017.

[80] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, "Genetic programs constructed from layered logic gates in single cells," *Nature,* vol. 491, pp. 249-253, 2012.

[81] C. Schilling, M. A. G. Koffas, V. Sieber, and J. Schmid, "A novel prokaryotic CRISPR-Cas12a based tool for programmable transcriptional activation and repression," *bioRxiv,* pp. 2020.08.05.232744-2020.08.05.232744, 2020.

[82] S. P. Collins, W. Rostain, C. Liao, and C. L. Beisel, "Sequence-independent RNA

sensing and DNA targeting by a split domain CRISPR–Cas12a gRNA switch," *Nucleic Acids Research,* vol. 49, pp. 2985-2999, 2021.
[83] S. Weingarten-Gabbay and E. Segal, "The grammar of transcriptional regulation," *Human Genetics,* vol. 133, pp. 701-711, 2014-06-01 2014.
[84] S. A. F. T. Van Hijum, M. H. Medema, and O. P. Kuipers, "Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation," *Microbiology and Molecular Biology Reviews,* vol. 73, pp. 481-509, 2009-09-01 2009.
[85] H. G. Garcia, A. Sanchez, T. Kuhlman, J. Kondev, and R. Phillips, "Transcription by the numbers redux: experiments and calculations that surprise," *Trends Cell Biol,* vol. 20, pp. 723-33, Dec 2010.
[86] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev*, et al.*, "Transcriptional regulation by the numbers: models," *Curr Opin Genet Dev,* vol. 15, pp. 116-24, Apr 2005.
[87] N. N. Kreamer, R. Phillips, D. K. Newman, and J. Q. Boedicker, "Predicting the impact of promoter variability on regulatory outputs," *Sci Rep,* vol. 5, p. 18238, Dec 17 2015.
[88] M. K. Dahl, J. Degenkolb, and W. Hillen, "Transcription of the xyl operon is controlled in Bacillus subtilis by tandem overlapping operators spaced by four base-pairs," *J Mol Biol,* vol. 243, pp. 413-24, Oct 28 1994.
[89] F. Santos-Beneit, A. Rodriguez-Garcia, and J. F. Martin, "Complex transcriptional control of the antibiotic regulator afsS in Streptomyces: PhoP and AfsR are overlapping, competitive activators," *J Bacteriol,* vol. 193, pp. 2242-51, May 2011.
[90] N. Liu, S. Xu, Q. Yao, Q. Zhu, Y. Kai, J. Y. Hsu*, et al.*, "Transcription factor competition at the gamma-globin promoters controls hemoglobin switching," *Nat Genet,* vol. 53, pp. 511-520, Apr 2021.
[91] R. Maurer, B. J. Meyer, and M. Ptashne, "Gene regulation at the right operator (OR) of bacteriophage λ," *Journal of Molecular Biology,* vol. 139, pp. 147-161, 1980.
[92] G. Jubelin, A. Vianney, C. Beloin, J. M. Ghigo, J. C. Lazzaroni, P. Lejeune*, et al.*, "CpxR/OmpR interplay regulates curli gene expression in response to osmolarity in Escherichia coli," *J Bacteriol,* vol. 187, pp. 2038-49, Mar 2005.
[93] T. J. Lie, E. L. Hendrickson, U. M. Niess, B. C. Moore, A. K. Haydock, and J. A. Leigh, "Overlapping repressor binding sites regulate expression of the Methanococcus maripaludis glnK(1) operon," *Mol Microbiol,* vol. 75, pp. 755-62, Feb 2010.
[94] L. V. Wray, Jr. and S. H. Fisher, "Bacillus subtilis CodY operators contain overlapping CodY binding sites," *J Bacteriol,* vol. 193, pp. 4841-8, Sep 2011.
[95] 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," *Nature Communications,* vol. 11, pp. 1-15, 2020.
[96] J. Gertz, E. D. Siggia, and B. A. Cohen, "Analysis of combinatorial cis-regulation in synthetic and genomic promoters," *Nature,* vol. 457, pp. 215-8, Jan 8 2009.
[97] S. Kaplan, A. Bren, A. Zaslaver, E. Dekel, and U. Alon, "Diverse two-dimensional input functions control bacterial sugar genes," *Mol Cell,* vol. 29, pp. 786-92, Mar 28 2008.
[98] N. E. Buchler, U. Gerland, and T. Hwa, "On schemes of combinatorial transcription logic," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, pp. 5136-41, 2003.
[99] P. Schulthess, A. Loffler, S. Vetter, L. Kreft, M. Schwarz, A. Braeuning*, et al.*, "Signal integration by the CYP1A1 promoter--a quantitative study," *Nucleic Acids Res,* vol. 43, pp. 5318-30, Jun 23 2015.

[100] A. E. Mayo, Y. Setty, S. Shavit, A. Zaslaver, and U. Alon, "Plasticity of the cis-regulatory input function of a gene," *PLoS Biol,* vol. 4, p. e45, Apr 2006.
[101] S. Aerts, G. Thijs, B. Coessens, M. Staes, Y. Moreau, and B. De Moor, "Toucan: deciphering the cis-regulatory logic of coregulated genes," *Nucleic Acids Res,* vol. 31, pp. 1753-64, Mar 15 2003.
[102] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nat Genet,* vol. 31, pp. 64-8, May 2002.
[103] R. Hermsen, S. Tans, and P. R. ten Wolde, "Transcriptional regulation by competing transcription factor modules," *PLoS Comput Biol,* vol. 2, p. e164, Dec 1 2006.
[104] A. Rantasalo, J. Kuivanen, M. Penttila, J. Jantti, and D. Mojzita, "Synthetic Toolkit for Complex Genetic Circuit Engineering in Saccharomyces cerevisiae," *ACS Synth Biol,* vol. 7, pp. 1573-1587, Jun 15 2018.
[105] T. C. Yu, W. L. Liu, M. S. Brinck, J. E. Davis, J. Shek, G. Bower*, et al.*, "Multiplexed characterization of rationally designed promoter architectures deconstructs combinatorial logic for IPTG-inducible systems," *Nature Communications,* vol. 12, pp. 325-325, 2021.
[106] J. J. Tabor, H. M. Salis, Z. B. Simpson, A. A. Chevalier, A. Levskaya, E. M. Marcotte*, et al.*, "A Synthetic Genetic Edge Detection Program," *Cell,* vol. 137, pp. 1272-1281, 2009.
[107] W. Bacchus, W. Weber, and M. Fussenegger, "Increasing the dynamic control space of mammalian transcription devices by combinatorial assembly of homologous regulatory elements from different bacterial species," *Metab Eng,* vol. 15, pp. 144-50, Jan 2013.
[108] R. S. Cox, 3rd, M. G. Surette, and M. B. Elowitz, "Programming gene expression with combinatorial promoters," *Mol Syst Biol,* vol. 3, p. 145, 2007.
[109] T. Ellis, X. Wang, and J. J. Collins, "Diversity-based, model-guided construction of synthetic gene networks with predicted functions," *Nat Biotechnol,* vol. 27, pp. 465-71, May 2009.
[110] E. Sharon, Y. Kalma, A. Sharp, T. Raveh-Sadka, M. Levo, D. Zeevi*, et al.*, "Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters," *Nat Biotechnol,* vol. 30, pp. 521-30, May 20 2012.
[111] L. M. O. Monteiro, A. Sanches-Medeiros, C. A. Westmann, and R. Silva-Rocha, "Unraveling the Complex Interplay of Fis and IHF Through Synthetic Promoter Engineering," *Front Bioeng Biotechnol,* vol. 8, p. 510, 2020.
[112] K. F. Murphy, G. Balazsi, and J. J. Collins, "Combinatorial promoter design for engineering noisy gene expression," *Proc Natl Acad Sci U S A,* vol. 104, pp. 12726-31, Jul 31 2007.
[113] I. Mogno, J. C. Kwasnieski, and B. A. Cohen, "Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants," *Genome Res,* vol. 23, pp. 1908-15, Nov 2013.
[114] P. Perez-Pinera, D. G. Ousterout, J. M. Brunger, A. M. Farin, K. A. Glass, F. Guilak*, et al.*, "Synergistic and tunable human gene activation by combinations of synthetic transcription factors," *Nat Methods,* vol. 10, pp. 239-42, Mar 2013.
[115] R. E. Rondon, T. M. Groseclose, A. E. Short, and C. J. Wilson, "Transcriptional programming using engineered systems of transcription factors and genetic architectures," *Nat Commun,* vol. 10, p. 4784, Oct 21 2019.
[116] 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, p. e1002811, 2012.
[117] Y. Zong, H. M. Zhang, C. Lyu, X. Ji, J. Hou, X. Guo*, et al.*, "Insulated transcriptional elements enable precise design of genetic circuits," *Nature Communications,* vol. 8, pp. 52-52, 2017.
[118] D. M. Zong, S. Cinar, D. L. Shis, K. Josić, W. Ott, and M. R. Bennett, "Predicting Transcriptional Output of Synthetic Multi-input Promoters," *ACS Synthetic Biology,* pp. acssynbio.8b00165-acssynbio.8b00165, 2018.
[119] M. E. Hoque, G. Mustafa, S. Basu, and H. Balci, "Encounters between Cas9/dCas9 and G-Quadruplexes: Implications for Transcription Regulation and Cas9-Mediated DNA Cleavage," *ACS Synth Biol,* May 10 2021.
[120] W. Poole, A. Pandey, A. Shur, Z. A. Tuza, and R. M. Murray, "BioCRNpyler: Compiling Chemical Reaction Networks from Biomolecular Parts in Diverse Contexts," *bioRxiv,* pp. 2020.08.02.233478-2020.08.02.233478, 2020.
[121] L. M. O. Monteiro, L. M. Arruda, and R. Silva-Rocha, "Emergent Properties in Complex Synthetic Bacterial Promoters," *ACS Synth Biol,* vol. 7, pp. 602-612, Feb 16 2018.
[122] G. R. Amores, M. E. Guazzaroni, and R. Silva-Rocha, "Engineering Synthetic cis-Regulatory Elements for Simultaneous Recognition of Three Transcriptional Factors in Bacteria," *ACS Synth Biol,* vol. 4, pp. 1287-94, Dec 18 2015.
[123] M. E. Guazzaroni and R. Silva-Rocha, "Expanding the logic of bacterial promoters using engineered overlapping operators for global regulators," *ACS Synth Biol,* vol. 3, pp. 666-75, Sep 19 2014.
[124] F. Jiang and J. A. Doudna, "CRISPR – Cas9 Structures and Mechanisms," pp. 505-531, 2017.
[125] H. Nishimasu, X. Shi, S. Ishiguro, L. Gao, S. Hirano, S. Okazaki*, et al.*, "Engineered CRISPR-Cas9 nuclease with expanded targeting space," *Science,* vol. 361, pp. 1259-1262, Sep 21 2018.
[126] R. T. Walton, K. A. Christie, M. N. Whittaker, and B. P. Kleinstiver, "Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants," *Science,* vol. 368, pp. 290-296, Apr 17 2020.
[127] B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. Zheng*, et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities," *Nature,* vol. 523, pp. 481-485, 2015.
[128] D. Collias and C. L. Beisel, "CRISPR technologies and the search for the PAM-free nuclease," *Nature Communications,* vol. 12, pp. 555-555, 2021.
[129] C. Kiattisewee, C. Dong, J. Fontana, W. Sugianto, P. Peralta-Yahya, J. M. Carothers*, et al.*, "Portable bacterial CRISPR transcriptional activation enables metabolic engineering in Pseudomonas putida," *Metab Eng,* Apr 27 2021.
[130] A. Shur and R. M. Murray, "Repressing Integrase attachment site operation with CRISPR-Cas9 in E. coli," *bioRxiv,* vol. 9, pp. 8-13, 2017.
[131] S. A. Shariati, A. Dominguez, S. Xie, M. Wernig, L. S. Qi, and J. M. Skotheim, "Reversible Disruption of Specific Transcription Factor-DNA Interactions Using CRISPR/Cas9," *Mol Cell,* vol. 74, pp. 622-633 e4, May 2 2019.
[132] S. K. Kim, G. H. Han, W. Seong, H. Kim, S. W. Kim, D. H. Lee*, et al.*, "CRISPR interference-guided balancing of a biosynthetic mevalonate pathway increases terpenoid production," *Metab Eng,* vol. 38, pp. 228-240, Nov 2016.
[133] T. Tian, J. W. Kang, A. Kang, and T. S. Lee, "Redirecting Metabolic Flux via

Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in Escherichia coli," *ACS Synth Biol,* vol. 8, pp. 391-402, Feb 15 2019.
[134] J. van Gestel, J. S. Hawkins, H. Todor, and C. A. Gross, "Computational pipeline for designing guide RNAs for mismatch-CRISPRi," *STAR Protoc,* vol. 2, p. 100521, Jun 18 2021.
[135] T. Lebar and R. Jerala, "Benchmarking of TALE- and CRISPR/dCas9-Based Transcriptional Regulators in Mammalian Cells for the Construction of Synthetic Genetic Circuits," *ACS Synth Biol,* vol. 5, pp. 1050-1058, Oct 21 2016.
[136] L. Oesinghaus and F. C. Simmel, "Switching the activity of Cas12a using guide RNA strand displacement circuits," *Nat Commun,* vol. 10, p. 2092, May 7 2019.
[137] L. M. Hochrein, H. Li, and N. A. Pierce, "High-Performance Allosteric Conditional Guide RNAs for Mammalian Cell-Selective Regulation of CRISPR/Cas," *ACS Synth Biol,* Apr 30 2021.
[138] H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae*, et al.*, "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell,* vol. 156, pp. 935-949, 2014.
[139] A. J. Meyer, T. H. Segall-Shapiro, E. Glassey, J. Zhang, and C. A. Voigt, "Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors," *Nat Chem Biol,* vol. 15, pp. 196-204, Feb 2019.
[140] D. E. Atkinson, "The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers," *Biochemistry,* vol. 7, pp. 4030-4, Nov 1968.
[141] J. E. Madl and R. K. Herman, "Polyploids and sex determination in Caenorhabditis elegans," *Genetics,* vol. 93, pp. 393-402, Oct 1979.
[142] Y. Li, W. Xiong, and E. E. Zhang, "The ratio of intracellular CRY proteins determines the clock period length," *Biochem Biophys Res Commun,* vol. 472, pp. 531-8, Apr 8 2016.
[143] M. Raisova, A. M. Hossini, J. Eberle, C. Riebeling, T. Wieder, I. Sturm*, et al.*, "The Bax/Bcl-2 ratio determines the susceptibility of human melanoma cells to CD95/Fas-mediated apoptosis," *J Invest Dermatol,* vol. 117, pp. 333-40, Aug 2001.
[144] R. Escalante-Chong, Y. Savir, S. M. Carroll, J. B. Ingraham, J. Wang, C. J. Marx*, et al.*, "Galactose metabolic genes in yeast respond to a ratio of galactose and glucose," *Proceedings of the National Academy of Sciences,* vol. 112, pp. 1636-1641, 2015.
[145] D. Mishra, P. M. Rivera, A. Lin, D. Del Vecchio, and R. Weiss, "A load driver device for engineering modularity in biological networks," *Nature Biotechnology,* vol. 32, pp. 1268-1275, 2014.
[146] M. K. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal*, et al.*, "Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems," *ACS Synth Biol,* vol. 4, pp. 503-15, May 15 2015.
[147] M. Ali Al-Radhawi, D. Del Vecchio, and E. D. Sontag, "Multi-modality in gene regulatory networks with slow promoter kinetics," *PLoS Comput Biol,* vol. 15, p. e1006784, Feb 2019.
[148] J. Berg, Y. P. Hung, and G. Yellen, "A genetically encoded fluorescent reporter of ATP:ADP ratio," *Nat Methods,* vol. 6, pp. 161-6, Feb 2009.
[149] Y. Liu, R. Landick, and S. Raman, "A Regulatory NADH/NAD+ Redox Biosensor for Bacteria," *ACS Synthetic Biology,* vol. 8, pp. 264-273, 2019.
[150] Y. E. Antebi, J. M. Linton, H. Klumpe, B. Bintu, M. Gong, C. Su*, et al.*, "Combinatorial Signal Perception in the BMP Pathway," *Cell,* vol. 170, pp. 1184-1196.e24, 2017.

[151] R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu, "Synthetic analog computation in living cells," *Nature,* vol. 497, pp. 619-623, 2013.
[152] K. M. Cherry and L. Qian, "Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks," *Nature,* 2018.
[153] R. Lopez, R. Wang, and G. Seelig, "A molecular multi-gene classifier for disease diagnostics," *Nature Chemistry,* 2018.
[154] S. D. Perli and T. K. Lu, "Ratiometric logic in living cells via competitive binding of synthetic transcription factors," *Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication - NanoCom '17,* pp. 1-4, 2017.
[155] J. Zeng, A. Banerjee, J. Kim, Y. Deng, T. W. Chapman, and R. Daniel, "A Novel Bioelectronic Reporter System in Living Cells Tested with a Synthetic Biological Comparator," *Scientific Reports,* pp. 1-7, 2019.
[156] X. Li, L. Rizik, V. Kravchik, M. Khoury, N. Korin, and R. Daniel, "Synthetic neural-like computing in microbial consortia for pattern recognition," *Nat Commun,* vol. 12, p. 3139, May 25 2021.
[157] H. H. Wang, H. Kim, L. Cong, J. Jeong, D. Bang, and G. M. Church, "Genome-scale promoter engineering by coselection MAGE," *Nat Methods,* vol. 9, pp. 591-3, Jun 2012.
[158] D. Na, S. M. Yoo, H. Chung, H. Park, J. H. Park, and S. Y. Lee, "Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs," *Nat Biotechnol,* vol. 31, pp. 170-4, Feb 2013.
[159] J. R. Warner, P. J. Reeder, A. Karimpour-Fard, L. B. Woodruff, and R. T. Gill, "Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides," *Nat Biotechnol,* vol. 28, pp. 856-62, Aug 2010.
[160] 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," *Current Opinion in Biotechnology,* vol. 64, pp. 190-198, 2020.
[161] F. R. Blattner, "The Complete Genome Sequence of Escherichia coli K-12," *Science,* vol. 277, pp. 1453-1462, 1997-09-05 1997.
[162] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski*, et al.*, "Genetic circuit design automation," *Science,* vol. 352, pp. aac7341-aac7341, 2016.
[163] D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya*, et al.*, "Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity," *Nature,* 2020.
[164] W. Dai, W. Dai, B. Zhang, X.-m. Jiang, H. Su, J. Li*, et al.*, "Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease," vol. 4489, pp. 1-10, 2020.
[165] B. K. Cho, K. Zengler, Y. Qiu, Y. S. Park, E. M. Knight, C. L. Barrett*, et al.*, "The transcription unit architecture of the Escherichia coli genome," *Nature Biotechnology,* vol. 27, pp. 1043-1049, 2009.
[166] M. Naville and D. Gautheret, "Transcription attenuation in bacteria: theme and variations," *Brief Funct Genomics,* vol. 9, pp. 178-89, Mar 2010.
[167] J. M. Peters, A. D. Vangeloff, and R. Landick, "Bacterial transcription terminators: the RNA 3'-end chronicles," *J Mol Biol,* vol. 412, pp. 793-813, Oct 7 2011.
[168] J. B. Deris, M. Kim, Z. Zhang, H. Okano, R. Hermsen, A. Groisman*, et al.*, "The Innate Growth Bistability and Resistant Bacteria," vol. 342, 2013.
[169] Y. Tashiro, H. Fukutomi, K. Terakubo, K. Saito, and D. Umeno, "A nucleoside kinase as a dual selector for genetic switches and circuits," *Nucleic Acids Research,* vol. 39,

2011.
[170] M. Tominaga, K. Ike, S. Kawai-noma, and K. Saito, "Rapid and Liquid-Based Selection of Genetic Switches Using Nucleoside Kinase Fused with Aminoglycoside Phosphotransferase," pp. 1-11, 2015.
[171] S. Kawai-noma, K. Saeki, T. Yumoto, K. Minakata, and K. Saito, "Improvement of the dP-nucleoside-mediated herpes simplex virus thymidine kinase negative-selection system by manipulating dP metabolism genes," *Journal of Bioscience and Bioengineering,* vol. xxx, pp. 1-7, 2020.
[172] M. S. Donia, J. Ravel, and E. W. Schmidt, "A global assembly line for cyanobactins," *Nature Chemical Biology,* vol. 4, pp. 341-343, 2008.
[173] J. A. McIntosh and E. W. Schmidt, "Marine molecular machines: heterocyclization in cyanobactin biosynthesis," *Chembiochem,* vol. 11, pp. 1413-21, Jul 5 2010.
[174] D. E. Ruffner, E. W. Schmidt, and J. R. Heemstra, "Assessing the combinatorial potential of the RiPP cyanobactin tru pathway," *ACS Synthetic Biology,* vol. 4, pp. 482-492, 2015.
[175] D. Sardar, Z. Lin, and E. W. Schmidt, "Modularity of RiPP Enzymes Enables Designed Synthesis of Decorated Peptides," *Chemistry and Biology,* vol. 22, pp. 907-916, 2015.
[176] J. A. McIntosh, C. R. Robertson, V. Agarwal, S. K. Nair, G. W. Bulaj, and E. W. Schmidt, "Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst," *J Am Chem Soc,* vol. 132, pp. 15499-501, Nov 10 2010.
[177] V. Agarwal, E. Pierce, J. McIntosh, E. W. Schmidt, and S. K. Nair, "Structures of cyanobactin maturation enzymes define a family of transamidating proteases," *Chem Biol,* vol. 19, pp. 1411-22, Nov 21 2012.
[178] A. Espah Borujeni, D. Cetnar, I. Farasat, A. Smith, N. Lundgren, and H. M. Salis, "Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences," *Nucleic Acids Res,* vol. 45, pp. 5437-5448, May 19 2017.
[179] H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology,* vol. 27, pp. 946-U112, 2009.
[180] K. E. D. Coan and B. K. Shoichet, "Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors," *Journal of the American Chemical Society,* vol. 130, pp. 9606-9612, 2008.
[181] C. K. McLaughlin, D. Duan, A. N. Ganesh, H. Torosyan, B. K. Shoichet, and M. S. Shoichet, "Stable Colloidal Drug Aggregates Catch and Release Active Enzymes," *ACS Chemical Biology,* vol. 11, pp. 992-1000, 2016.
[182] B. K. Shoichet, "Interpreting steep dose-response curves in early inhibitor discovery," *Journal of Medicinal Chemistry,* vol. 49, pp. 7274-7277, 2006.
[183] S. L. McGovern, B. T. Helfand, B. Feng, and B. K. Shoichet, "A specific mechanism of nonspecific inhibition," *Journal of Medicinal Chemistry,* vol. 46, pp. 4265-4272, 2003.
[184] W. P. Walters and M. Namchuk, "Designing screens: how to make your hits a hit," *Nat Rev Drug Discov,* vol. 2, pp. 259-66, Apr 2003.
[185] J. E. Blanchard, N. H. Elowe, C. Huitema, P. D. Fortin, J. D. Cechetto, L. D. Eltis*, et al.*, "High-Throughput Screening Identifies Inhibitors of the SARS Coronavirus Main Proteinase," *Chemistry & Biology,* vol. 11, pp. 1445-1453, 2004.
[186] E. Smith, M. E. Davis-Gardner, R. D. Garcia-Ordonez, T. T. Nguyen, M. Hull, E. Chen*, et al.*, "High-Throughput Screening for Drugs That Inhibit Papain-Like Protease in SARS-CoV-2," *SLAS Discov,* vol. 25, pp. 1152-1161, Dec 2020.

[187] W. Zhu, M. Xu, C. Z. Chen, H. Guo, M. Shen, X. Hu, *et al.*, "Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative High-Throughput Screening," *ACS Pharmacol Transl Sci,* vol. 3, pp. 1008-1016, Oct 9 2020.
[188] M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. van Heel, J. H. Viel, *et al.*, "New developments in RiPP discovery, enzymology and engineering," *Natural Product Reports,* 2020.
[189] E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. Porco, D. Kozakov, S. Vajda, *et al.*, "How proteins bind macrocycles," *Nature Chemical Biology,* vol. 10, pp. 723-731, 2014.
[190] E. Marsault and M. L. Peterson, "Macrocycles are great cycles: Applications, opportunities, and challenges of synthetic macrocycles in drug discovery," *Journal of Medicinal Chemistry,* vol. 54, pp. 1961-2004, 2011.
[191] A. C. L. Lee, J. L. Harris, K. K. Khanna, and J. H. Hong, "A comprehensive review on current advances in peptide drug development and design," *International Journal of Molecular Sciences,* vol. 20, pp. 1-21, 2019.
[192] G. Bennett, A. Brown, G. Mudd, P. Huxley, K. Van Rietschoten, S. Pavan, *et al.*, "MMAE Delivery Using the Bicycle Toxin Conjugate BT5528," *Mol Cancer Ther,* vol. 19, pp. 1385-1394, Jul 2020.
[193] B. K. Shoichet, "Screening in a spirit haunted world," *Drug Discovery Today,* vol. 11, pp. 607-615, 2006.
[194] J. J. Irwin, D. Duan, H. Torosyan, A. K. Doak, K. T. Ziebart, T. Sterling, *et al.*, "An Aggregation Advisor for Ligand Discovery," *Journal of Medicinal Chemistry,* vol. 58, pp. 7076-7087, 2015.
[195] J. Bisson, J. B. McAlpine, J. B. Friesen, S. N. Chen, J. Graham, and G. F. Pauli, "Can Invalid Bioactives Undermine Natural Product-Based Drug Discovery?," *Journal of Medicinal Chemistry,* vol. 59, pp. 1671-1690, 2016.
[196] C. Aldrich, C. Bertozzi, G. I. Georg, L. Kiessling, C. Lindsley, D. Liotta, *et al.*, "The Ecstasy and Agony of Assay Interference Compounds," *ACS Central Science,* vol. 3, pp. 143-147, 2017.
[197] S. B. Zimmerman and S. O. Trach, "Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli," *J Mol Biol,* vol. 222, pp. 599-620, Dec 5 1991.
[198] R. J. Ellis, "Macromolecular crowding: obvious but underappreciated," *Trends Biochem Sci,* vol. 26, pp. 597-604, Oct 2001.
[199] B. R. Parry, I. V. Surovtsev, M. T. Cabeen, C. S. O'Hern, E. R. Dufresne, and C. Jacobs-Wagner, "The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity," *Cell,* vol. 156, pp. 183-94, Jan 16 2014.
[200] B. Y. Feng, A. Simeonov, A. Jadhav, K. Babaoglu, J. Inglese, B. K. Shoichet, *et al.*, "A high-throughput screen for aggregation-based inhibition in a large compound library," *J Med Chem,* vol. 50, pp. 2385-90, May 17 2007.
[201] S. Y. Fung, H. Yang, and P. Chen, "Formation of colloidal suspension of hydrophobic compounds with an amphiphilic self-assembling peptide," *Colloids and Surfaces B: Biointerfaces,* vol. 55, pp. 200-211, 2007.
[202] B. Zhou, L. Xing, W. Wu, X.-e. Zhang, and Z. Lin, "Small surfactant-like peptides can drive soluble proteins into active aggregates," pp. 2-9, 2012.
[203] H. Hong, A. Akbari, and J. Wu, "Small amphipathic peptides are responsible for the assembly of cruciferin nanoparticles," *Scientific Reports,* vol. 7, pp. 1-13, 2017.
[204] C. Hüttl, C. Hettrich, R. Miller, B. R. Paulke, P. Henklein, H. Rawel, *et al.*, "Self-assembled peptide amphiphiles function as multivalent binder with increased

hemagglutinin affinity," *BMC Biotechnology,* vol. 13, pp. 1-10, 2013.
[205] G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in Escherichia coli: advances and challenges," *Front Microbiol,* vol. 5, p. 172, 2014.
[206] L. Ledsgaard, M. Kilstrup, A. Karatt-Vellatt, J. McCafferty, and A. H. Laustsen, "Basics of Antibody Phage Display Technology," *Toxins (Basel),* vol. 10, Jun 9 2018.
[207] G. M. Cherf and J. R. Cochran, "Applications of Yeast Surface Display for Protein Engineering," *Methods Mol Biol,* vol. 1319, pp. 155-75, 2015.
[208] J. C. Carlson, A. H. Badran, D. A. Guggiana-Nilo, and D. R. Liu, "Negative selection and stringency modulation in phage-assisted continuous evolution," *Nat Chem Biol,* vol. 10, pp. 216-22, Mar 2014.
[209] A. M. Davis, A. T. Plowright, and E. Valeur, "Directing evolution: The next revolution in drug discovery?," *Nature Reviews Drug Discovery,* vol. 16, pp. 681-698, 2017.
[210] M. S. Packer, H. A. Rees, and D. R. Liu, "Phage-assisted continuous evolution of proteases with altered substrate specificity," *Nature Communications,* vol. 8, 2017.
[211] J. Pu, J. Zinkus-Boltz, and B. C. Dickinson, "Evolution of a split RNA polymerase as a versatile biosensor platform," *Nature Chemical Biology,* vol. 13, pp. 432-438, 2017.
[212] S. M. Miller, T. Wang, and D. R. Liu, "Phage-assisted continuous and non-continuous evolution," *Nature Protocols,* vol. 15, 2020.
[213] X. Yang, K. R. Lennard, C. He, M. C. Walker, A. T. Ball, C. Doigneaux*, et al.*, "A lanthipeptide library used to identify a protein–protein interaction inhibitor," *Nature Chemical Biology,* vol. 14, 2018.
[214] A. J. Meyer, T. H. Segall-shapiro, E. Glassey, J. Zhang, and C. A. Voigt, "Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors," *Nature Chemical Biology*.
[215] J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger, and G. S. Waldo, "Engineering and characterization of a superfolder green fluorescent protein," *Nat Biotechnol,* vol. 24, pp. 79-88, Jan 2006.
[216] D. S. Bindels, L. Haarbosch, L. van Weeren, M. Postma, K. E. Wiese, M. Mastop*, et al.*, "mScarlet: a bright monomeric red fluorescent protein for cellular imaging," *Nat Methods,* vol. 14, pp. 53-56, Jan 2017.
[217] R. Arai, H. Ueda, A. Kitayama, N. Kamiya, and T. Nagamune, "Design of the linkers which effectively separate domains of a bifunctional fusion protein," vol. 14, pp. 529-532, 2001.