Development and Scaling of Differentiation Circuit Architectures for Improving the Evolutionary Stability of Burdensome Functions in *E. coli*

Thesis by
Rory L. Williams

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Bioengineering

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended December 15th 2021

Rory L. Williams
ORCID: 0000-0003-2605-5790

ACKNOWLEDGEMENTS

In reflecting on my time in graduate school, I am grateful for the opportunities I have had in Caltech Bioengineering, and for the people whose support and friendship have made completing my PhD possible, and the time spent doing so as enjoyable and fulfilling as it has been. I first and foremost would like to thank Richard Murray, not only for giving me the good news on April 16th 2016 that I made it off the wait list, but for being a great mentor throughout graduate school. Richard provided abundant opportunities to develop, pursue, and iterate ideas with a good deal of independence, while providing the right amount of feedback and guidance along the way to ensure I did not stray too far off course. I am grateful for this mentorship and environment, as well as the impressive example he continuously provides. I would also like to thank my committee members Justin Bois, Lea Goentoro, Rustem Ismagilov, and Jared Leadbetter for their support and feedback throughout graduate school. I had the opportunity to take several classes and TA with Justin, and the skills and perspectives in data analysis, Python, and thinking about biology from these experiences proved invaluable during my graduate school research. I was also fortunate to work with Rustem and his lab members (particularly Tami Khazaei and Said Bogatyrev) in an extended rotation, and am grateful for this opportunity and their guidance.

Most of my time in graduate school was spent in the Murray Lab, which was made easier by it being a special place filled with group of unique, talented, intelligent individuals that impress me in many ways. I will always have fond memories of Ernie's after group meetings with all the big dogs (aka every Murray Lab member), game nights, and the daily conversations had while distracting one-another from lab work. The Murray Lab has seen more than a few people come and go since I started, and I am grateful for the times we shared, and the friends I have made. I cannot possibly be exhaustive, but I will share some brief highlights. Andy Halleran and John Marken consistently impressed me in their ability to communicate ideas in clear and engaging ways, as well as to make beautiful figures and slides. Their examples have certainly spurred my endeavor to step up my presentation game over the course of graduate school. I am particularly grateful for the time I shared with Andy Halleran, who although an incredibly kind person I have to admit intimidated me for a time with his breadth of knowledge, intelligence, and eloquence. The direction my research took in graduate school was spurred by our common interest in evolutionary

stability, and benefited greatly from our conversations (and from 3G). Their fellow William and Mary transplant Sam Clamons has consistently impressed me with his curiosity and appetite for knowledge. Anandh Swaminathan is an impressive and impressively modest boss who inside and outside of the lab helped get graduate school off to a good start. Throughout these years, Zoila Jurado has been one of the most generous people I know, both with time in and out of lab, and baked goods. Andrey Shur it seems can build or fix just about anything, and I'm not sure how the Murray Lab will get along without him. As well, his expertise with integrases and cloning were invaluable. For the past several years, I've had the pleasure of being bench mates with Reed McCardell, whose curiosity and positive attitude in and out of the lab have been a welcome example, and whose belief that I can keep my bench clean will stay with me. While many have come and gone, Miki Yun and Mark Prator have been the rocks of the Murray lab that have been essential in keeping it running and making it a great place. Miki, in addition to being a great lab manager, has been a combination of friend and lab mom. Mark Prator – DM, friend, and unofficial/offical lab social coordinator – has been a critical member of the lab, and I thank him for his dedication to the COVID Zoom happy hours. This list is not exhaustive, and there many others who have and continue to make the Murray Lab the great place that it is.

Outside of lab, I have been fortunate to have the support of family and friends throughout graduate school. The overlapping group of friends and roommates – Tom O'Connell, Taylor Stevens, Sho Harvey, Said Bogatyrev, Charlie Dorn, Jeremy Bernstein, Tom Naragon, Sharan Prakash, among others – I've had at Caltech have shared the ups and downs of graduate school, and without whom graduate school would have been more difficult, and less enjoyable. Friends from before graduate – high school, college, and the Largaespada Lab – have remained important throughout, and the cherished inspirational quote of my friend and Larg Lab mate Bryant Keller (Dream it, Believe it, Achieve it...) has frequently come to mind. In particular, I must thank my good friend Ian Goodwin and his fiance Sally, not only for their friendship, but for generous use of their guestroom in Pasadena after I moved to Irvine. Finally, my family – my parents, brothers, nieces, and grandparents – have been incredibly supportive throughout, and provided welcomed breaks from graduate school. Seu Sim, my wife and the newest member of my family, has become my best friend, the best COVID quarantine buddy one could ask for, and impresses and inspires me in many ways. I cannot thank her enough for her support and belief in me, nor can I imagine making it through graduate school without her.

In addition, I must thank her as Professor Sim, as well as UC Irvine, for welcoming me this past year in her lab to complete my graduate work.

ABSTRACT

With advances in the sequencing and synthesis of DNA, automation, and computation, we are increasingly able to rapidly and reliably program functions into cells. However, because the functions we engineer cells to perform are often both unnecessary for the cell's survival and burdensome to cell growth, mutation and natural selection can rapidly lead to loss of function. Though numerous strategies have made headway, improving the evolutionary stability of engineered functions remains a goal of the synthetic biology community. To address this problem generally, we developed a strategy relying on integrase-mediated recombination which allows non-producing progenitor cells to differentiate at a tunable rate, thereby continuously replenishing producer cells expressing the orthogonal T7 RNAP. While this strategy removes selective pressure for mutations inactivating the function of interest in the progenitor cell population, a strategy of terminal differentiation – in which the capacity of differentiated cells to grow is limited – was necessary to prevent the expansion of such mutations in the differentiated cell population. To experimentally implement terminal differentiation, we co-opted the R6K plasmid system, using differentiation to simultaneously activate expression of T7 RNAP, and inactivate expression of π protein (an essential factor for R6K plasmid replication), thereby allowing limitation of differentiated cell growth through antibiotic selection. Critically, we demonstrated computationally that terminal differentiation endows the circuit with robustness to mutations which disrupt T7 RNAP driven expression, and to plasmid instability effects that result in decreased expression. Intuitively and computationally identifying the category of mutations which disrupt the differentiation process as the Achilles's heel of terminal differentiation, we developed a redundant architecture using a novel split-π protein system which required 2 mutations to break the circuit. We experimentally demonstrated a trade-off between rate of production and duration of function as the differentiation rate is tuned, an increased benefit of terminal differentiation with higher-burden expression, and that redundancy improves the evolutionary stability of the terminal differentiation architecture. Specifically we achieve a maximum of ~2.8x (single-cassette terminal differentiation) and ~4.2x (redundant terminal differentiation) the total fluorescent protein production achieved from comparable high-burden naive expression in which all cells inducibly express T7 RNAP. We further demonstrate differentiation can enable the expression of even toxic functions, and develop a terminal differentiation

circuit architecture which will allow the degree of redundancy and therefore the evolutionary stability of the architecture to be scaled to arbitrary degrees.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. Williams, R. L. & Murray, R. M. Tunable integrase-mediated differentiation facilitates improved output of burdensome functions in E. coli. *bioRxiv*. `http://dx.doi.org/10.1101/614529` (2019).
   RW developed the project with guidance from RMM. RW planned and performed all experimental work, analyzed data, developed the computational models, executed the simulations, and wrote the manuscript with input and guidance from RM.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

As synthetic biology aims to engineer cells with the capacity to regulate and execute increasingly complex and burdensome functions, strategies which address the evolutionary potential of biology will only become more essential. The same force of Darwinian evolution which has provided incredible biological diversity does not discriminate between natural and engineered DNA, and consequently engineered functions may be readily lost in a population. It has long been observed that cell fitness negatively correlates with heterologous gene expression level [1], and increased burden results in a shorter evolutionary half-life of engineered functions [2, 3].

Efforts to improve evolutionary stability of engineered functions have taken a variety of forms, including the most straightforward goals of reducing mutation rate and minimizing burden. Strategies to reduce the rate of mutations have focused both on sequence design and host-genome engineering. At the level of sequence design, minimizing repeated sequences and parts diminishes mutations due to homologous recombination (HR) and improves circuit half-life [2], and researchers may evaluate sequences in silico for such HR and repeat-mediated mutations with the EFM calculator [4]. Alternatively, strains have been engineered to globally reduce mutation rates by disrupting the cell's capacity for HR with recA knockout, knocking out error-prone polymerases to reduce point mutations, and removing selfish transposon elements that otherwise may insert themselves and disrupt circuit function [5–7].

Though such strategies may delay the acquisition of destructive mutations, other approaches are necessary to impact the rate at which mutations are selected in the population. The simplest solution to delay the selection of these mutations is to reduce the expression level of circuit components and therefore the fitness difference between functional and non-functional cells [1, 2, 8]. Alternatively, rather than constitutively reducing expression, gene expression level may be dynamically regulated by co-opting transcriptional changes which occur during cell stress to drive negative feedback [9].

Additional strategies for improving evolutionary stability, rather than directly addressing rates of mutations and burden of functions, have sought to alter the con-

sequence of these mutations. A conceptually straightforward approach is to have multiple redundant copies of the synthetic construct, whereby multiple independent mutations are required to destroy function. The chemically inducible chromosomal evolution (CIChE) system was used to evolve strains with ~40 tandem copies of a circuit before deleting recA, resulting in expression of a polyhydroxybutyrate biosynthetic pathway for >100 generations compared to ~10 when expressed from a plasmid [10]. Importantly, this strategy removes random plasmid portioning as a mechanism for accelerating mutation propagation [11]. Alternatively, selection of destructive mutations could be limited by utilizing components whose mutation would inactivate not only the expression of the synthetic construct, but also an essential gene or selectable marker. Producing GFP with KanR on a single bicistronic transcript or as a fusion protein marginally improved evolutionary half-life when selecting with kanamycin, and using a bi-directional promoter to drive their expression separately increased half-life 4-10 fold [2, 12].

Strategies discussed so far have been limited to cell-level functions in uniform populations, and tactics which incorporate specialization and division of labor at a population level have not been addressed. With inspiration from microbial communities exhibiting metabolic division of labor and syntrophic interactions, there have been numerous successful implementations of metabolic division of labor for production of biomolecules of interest [13–15]. This design motif has numerous advantages, including reducing the number of genes and associated metabolic load in each specialized cell type, allowing independent optimization of separate pathways, and spatially separating potentially incompatible functions. While these benefits may be realized by combining in co-culture independently engineered strains or species, additional attractive properties become apparent with dynamically regulated division of labor in a population of genetically related or identical organisms. Such metabolic and reproductive division of labor is a recurring motif in microbiology [16, 17], but is underutilized in synthetic biology, particularly for addressing evolutionary constraints.

Examining specific instances of division of labor in bacteria gives insight into how we might use this motif in synthetic biology. In the cyanobacteria *Anabaena*, nitrogen deprivation induces a division of labor in which individual cells in a large filament terminally differentiate into heterocyst cells which are specialized for nitrogen fixation and are incapable of reproduction [18]. This reproductive and metabolic division of labor allows the collective to realize an inclusive fitness benefit

from a costly metabolic process; a process encoded by all cells, but expressed only in a fraction. If we imagine instead all cells expressing this function – nitrogen fixation – cells which mutate this function would certainly gain a fitness benefit, and would proliferate more quickly (assuming sufficient nitrogen). By instead having cells which have the genetic potential for nitrogen fixation, but do not express it, there is no selective pressure for mutations which would inactivate expression of genes necessary for this process. Though this function is indeed essential for *Anabaena* survival under nitrogen deprivation conditions, we could imagine using this same strategy in synthetic biology for the expression of burdensome functions, essential and non-essential alike.

To adopt this reproductive and metabolic division of labor into a synthetic context, we propose a circuit architecture much like that seen in *Anabaena*. The simplest form of this architecture consists of two cell types, with the first being specialized for the faithful replication of an encoded function in the absence of circuit burden, and the second – generated upon differentiation of the former – for the execution of the encoded function. Though largely orthogonal and complementary to previous approaches for improving the evolutionary stability of engineered functions, this strategy may be particularly suited for certain types of applications. Functions for which a subset of cells expressing a function is sufficient are ideal, as are functions which could be divided between cells of distinct phenotypes. Importantly, functions which are highly burdensome, toxic, or incompatible with cell proliferation enter the realm of possibility.

In Chapter 2, we propose differentiation combined with limiting the capacity of differentiated cells to grow (terminal differeniation) as a strategy to address evolutionary stability. We first develop a integrase-mediated differentiation architecture which allows the rate of differentiation to be tuned, the number of divisions a differentiated can undergo limited, and the fraction of the population in the differentiated state to be tuned by a combination of both factors. We then apply this to a circuit in which the orthogonal T7 RNAP is activated by integrase-mediated recombination, and demonstrate the capacity of this circuit to improve the evolutionary stability of burdensome functions. In Chapter 3, we address weaknesses of this initial demonstration of differentiation and terminal differentiation, and implement strategic redundancy in the differentiation architecture. The benefit of redundancy in the context of naive, differentiation, and terminal differentiation architectures is explored computationally, and the unique robustness of terminal differentiation to

mutations or other factors disrupting engineered functions examined. We further demonstrate that differentiation can enable the expression of toxic proteins in a manner that could potentially enable continuous bioproduction. In Chapter 4, we consider scaling the application of terminal differentiation to longer times and larger population sizes by increasing the degree of redundancy, investigate this computationally, and experimentally demonstrate a proof of concept circuit design that would enable the realization of scalable redundancy with repeated genomic integration of identical cassettes.

Chapter 2

DEVELOPMENT OF A TUNABLE INTEGRASE MEDIATED DIFFERENTIATION ARCHITECTURE FOR FRACTIONAL CONTROL AND T7 RNAP-DRIVEN BURDENSOME EXPRESSION

2.1 Rational for differentiation

In order to increase the duration of circuit life-time, either the rate of mutations which inactivate circuit function must be drastically decreased, or alternatively the opportunity for these mutations to be selected in a population of cells must be limited. To accomplish this, we reasoned that having a population of cells that encode the circuit function, but do not express it, would allow the genetic circuit to be replicated in the absence of selective pressure for inactivating mutations. By inducing these progenitor cells to differentiate at some rate into cells expressing the function, producer cells would be continuously replenished (Figure 2.1B). However, these producer cells are still susceptible to mutations which inactivate circuit function, and the opportunity for these mutants to be selected would need to be eliminated in order to prevent circuit failure. To accomplish this, we considered an architecture that would limit the number of divisions a cell could undergo following differentiation (Fig. 2.1C). In this way, mutations which inactivate circuit function would have negligible opportunity to be selected.

2.2 Experimental considerations for differentiation architectures

To experimentally implement this differentiation circuit as depicted in Figure 2.1, we required means of maintaining the function of interest fully off in the progenitor cell population, and an irreversible mechanism that would activate the function of interest at a tunable rate. To accomplish this, we turn to bacteriophage serine integrases, a class of proteins capable of unidirectional DNA recombination between specific sequences of DNA [19]. With strategic placement of integrase attachment sites on the genome, a single integrase-mediated recombination event can simultaneously activate and inactivate the expression of desired genes [20–22]. In order to tune the rate of differentiation, we then rely on the inherent stochasticity of the recombination process at low intracellular concentrations of integrase proteins, with higher expression increasing the probability that any given cell in the population

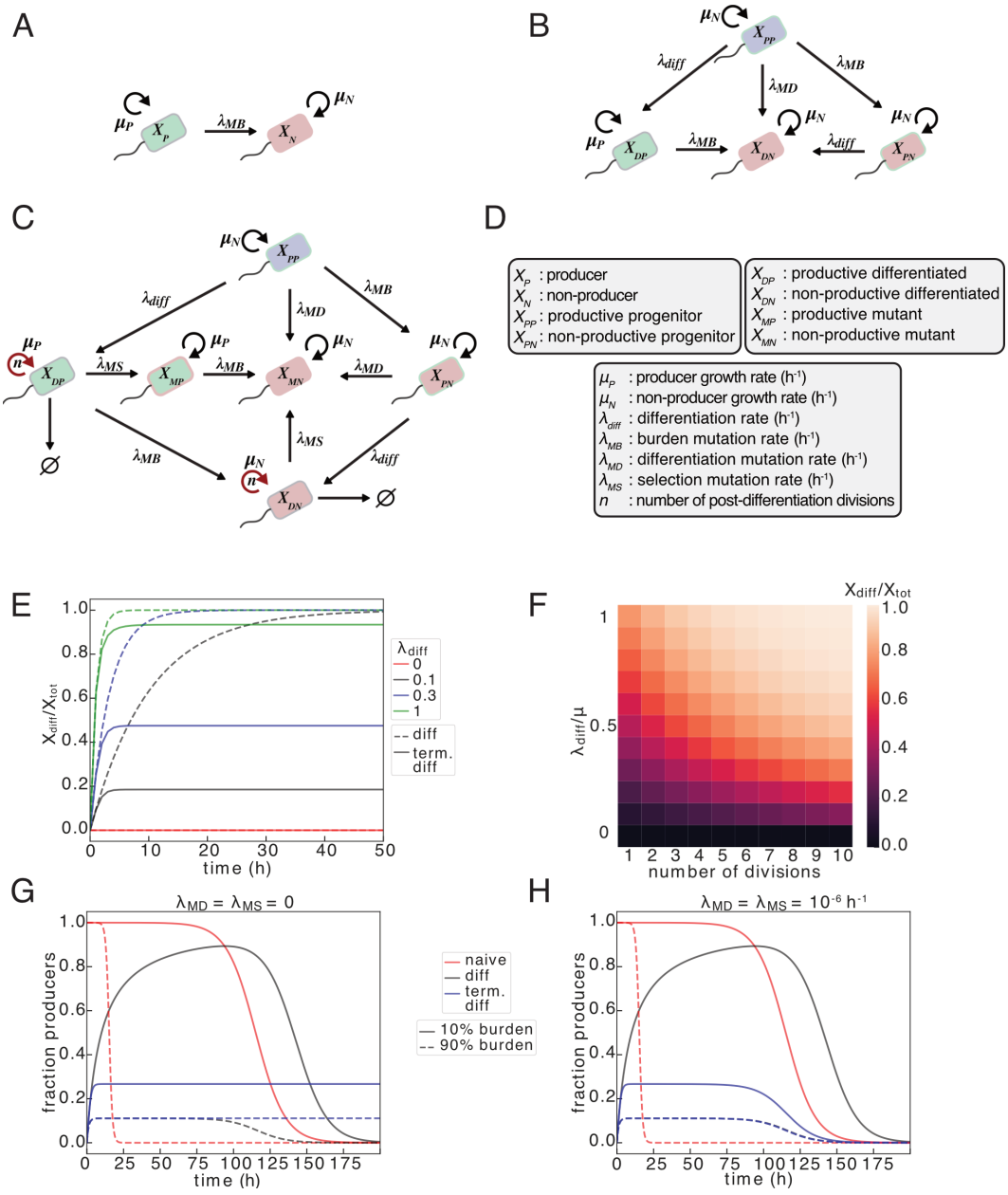

Figure 2.1: Architectures for implementation of differentiation circuits. (A-D) Schematics for a naive expression (A), differentiation-activated expression (differentiation) (B), and differentiation-activated expression in which the number of cell divisions following differentiation is limited (terminal differentiation) (C). (E) Deterministic ODE modeling in exponential growth conditions of differentiation circuits with and without selection ($\mu_N = \mu_P = 1$ h^{-1}; $\lambda_{MB} = \lambda_{MD} = \lambda_{MS} = 0$). (F) Modeling as in (E) for differentiation with selection varying differentiation rate and number of divisions. (G-H) Comparison of naive circuit with differentiation and terminal differentiation (G: $\lambda_{MB} = 10^{-6}$ h^{-1}, $\lambda_{MD} = \lambda_{MS} = 0$; H: $\lambda_{MB} = \lambda_{MD} = \lambda_{MS} = 10^{-6}$ h^{-1}).

will undergo the recombination event. In order to allow differentiation to control diverse expression programs, the genes regulated by the recombination event should encode proteins which control the expression of numerous genes, as in the case of transcription factors, sigma factors, or orthogonal RNA polymerases. Further, for the case of terminal differentiation we may limit the proliferation of differentiated cells by using this recombination to inactivate the expression of an essential, or conditionally essential gene. The specific circuit implementation will be discussed later, but it is sufficient here to recognize that the duration of growth or number of cell divisions possible after terminating the expression of an essential gene will depend on the role of the gene and its expression level, and can be viewed as a tunable variable.

2.3 Deterministic modeling of integrase-mediated differentiation

To gain intuition regarding the behavior of these proposed differentiation architectures, and specifically if and when differentiation-based circuits would be advantageous for improving the duration of circuit lifetime and/or total output achieved by an engineered function, we modeled the behavior deterministically using systems of ordinary differential equations. For comparison, we model a naive expression circuit in which all cells constitutively express the function (Figure 2.1A).

In the absence of any burden difference between the progenitor and differentiated cells, the differentiation architecture with unrestricted cell division in the differentiated cells results in all cells in population being differentiated if the differentiation rate is non-zero (Figure 2.1E). However, when the number of cell divisions is limited, the population achieves a steady-state fraction of differentiated cells that can be tuned by both the differentiation rate and the number of cell divisions allowed by differentiated cells (Figure 2.1E-F).

Considering the case when differentiated cells are performing some burdensome function and have a slower growth rate than the progenitor cells, we observe populations performing differentiation with and without limited cell divisions approaching or achieving a steady state fraction of differentiated producer cells (Figure 2.1G). However, as differentiated cells are able to incur mutations inactivating the production and relieving the associated burden, populations are eventually dominated by non-productive differentiated cells at a rate that increases with burden. Conversely, when the number of divisions of differentiated cells is limited, non-productive differentiated cells do not have the opportunity to expand in the population, and a

steady-state fraction of producers is achieved that will be disrupted only by the incredibly slow accumulation of mutations in the progenitor population (Figure 2.1G).

While differentiation with selection appears to permit indefinite circuit function, we have not yet considered additional classes of mutation that become possible when implementing these differentiation circuit architectures. In the case of differentiation alone, we posit a class of mutations in progenitor cells which would destroy the ability of the cell to undergo differentiation (Figure 2.1B-C). Additionally, when limiting the number of divisions a differentiated cell can undergo, a second new class of mutations becomes apparent which would restore the ability of a differentiated cell to proliferate indefinitely (Figure 2.1C). When we account for these two additional classes of mutations, we observe circuit failure even in the case where differentiated cells have limited capacity for cell division (Figure 2.1H). Though all circuit architectures – naive, differentiation, and terminal differentiation – are imperfect and ultimately fail due to mutation and natural selection, we reasoned that the best architecture would depend on a variety of factors, including the burden imposed by the engineered function, relative mutation rates (burden, differentiation, and selection mutations), as well as the needs of the specific application, such as maximizing total production or duration of circuit function.

To understand when each of these architectures would be best suited, we modeled each across varying production burdens, and mutation rates specific to burden, differentiation, and selection mutations. As the behavior of the two architectures involving differentiation are also impacted by the differentiation rate, we optimized this parameter in each case to maximize either the total production (Figure 2.2A) or the duration of population function (Figure 2.2B). In order for modeling results to be qualitatively comparable to our experiments, we modelled logistic growth with repeated 50X dilutions when the population reached 95 percent of the carrying capacity. Simulations were terminated when the production from an individual growth cycle was below a low threshold (Figure 2.2A), or when fewer than 10 percent of cells were producers (Figure 2.2B). Due to the mechanism of differentiation implemented experimentally – integrase-mediated recombination – differentiation is not coupled to cell division, and is modeled with a first-order rate constant coupled to time. As well, though the differentiation rate may indeed vary with the growth state of cells (i.e. exponential vs. stationary phase), we neglected this and assumed a constant differentiation rate regardless of growth phase.

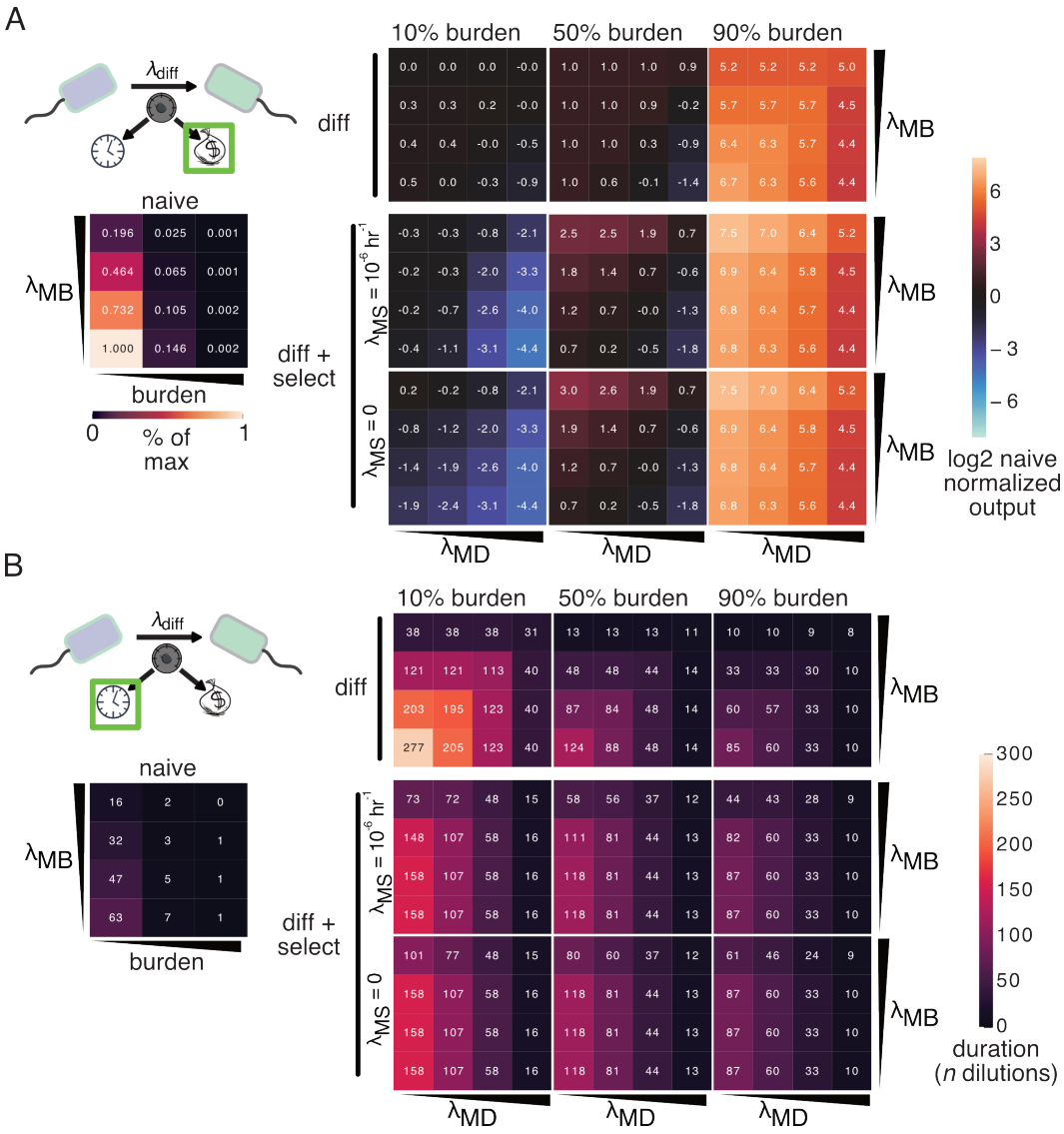

Figure 2.2: Differentiation architectures improve duration and output for high burden circuits. Deterministic modeling of naive and differentiation architectures with varying burden levels (10, 50, and 90 percent) and mutation rates (λ_{MB}, λ_{MD}: 10^{12}, 10^{-9}, 10^{-6}, and 10^{-3} h^{-1}). Simulations are of repeated 50X dilutions with logistic carrying-capacity limited growth, with dilutions occurring when the population reaches 95 percent of the carrying capacity (K). For differentiation and differentiation with selection, the differentiation rate was optimized to maximize total production (A) and duration (B). For all simulations, $K = 10^9$ cells, $n = 4$, $\mu_P = 1.5$ h^{-1}. (A) Production rate is modeled as proportional to growth rate and varies over time, and production rate is equal across burden levels. Heatmaps present (A) total production and (B) number of consecutive growths in which the ending fraction of producer cells is >10 percent of the population. (A) Heatmap (left) shows total production normalized maximum case ($\lambda_{MB} = 10^{-12}$ h^{-1} , 10 percent burden). Heatmaps (right) show the log2 of normalized total production (normalized to naive production with equivalent burden and λ_{MB}).

Modeling the naive implementation of production reveals matching intuition that both total production and circuit lifetime decrease with increased burden and burden mutation rate, with burden being the dominant factor. The relative benefit of differentiation for total production and circuit lifetime depends both on the burden, as well as on the relative rates of burden and differentiation mutations. For both total production and circuit lifetime, this benefit increases with increased burden, but decreases as the differentiation mutation rate increases relative to the burden mutation rate.

In the case of differentiation with selection, we see largely the same trends as with differentiation alone, however with a few key differences. While there is an increased benefit for total production relative to the other two architectures as burden increases, this strategy is counterproductive at low burdens, and decreases production particularly with higher differentiation mutation rates. Further, selection allows the population to be less susceptible to the burden mutation, demonstrated by increased production and duration relative to both differentiation and naive implementations as the burden mutation rate increases. Finally, comparing the second and third rows of Figure 2.2A and 2.2B reveals that the impact of the rate of the selection mutation is only revealed with low rates of the differentiation mutation, and becomes more apparent with increased burden mutation rates.

2.4 Integrase mediated differentiation allows tuning of population distribution

To experimentally investigate these qualitative predictions from our modeling, we first implement and characterize a differentiation architecture which allows tuning of the rate or probability of differentiation, selection against differentiated cells, and tuning of the duration of differentiated cell proliferation (Figure 2.3A). In this circuit, the salicylate-inducible promoter P_{SalTTC} and its cognate transcription factor $NahR^{AM}$ [23] control the expression Bxb1 integrase. To reduce Bxb1 concentration and differentiation in the absence of salicylate induction, we used a strong SsrA degradation tag [24] on the C-terminus (Bxb1-LAA). To allow tuning of the duration of differentiated cell proliferation, we take advantage of the reliance of R6K plasmid replication on the π protein encoded by *pir* [25]. We used the 3OC12-HSL (Las-AHL) inducible promoter P_{LasAM} and its cognate transcription factor $LasR^{AM}$ (personal communication with Adam Meyer) to control the expression of π protein, and placed its expression cassette such that the recombination event results in its excision (Figure 2.3A). The π protein abundance and R6K plasmid copy number at

the time of differentiation can therefore be tuned with Las-AHL (Figure 2.3C). As the R6K plasmid encodes the sole source of chloramphenicol resistance (CmR) and will be lost as a result of dilution through cell division when π protein is absent, the induction level of π protein sets the limit on number of divisions possible upon differentiation when chloramphenicol selection is applied (Figure 2.3A). In this initial characterization of the features of this system, Bxb1-LAA is expressed from a p15a plasmid (~20-30 copies per cell) with ampicillin resistance (AmpR), and catalyzes a genomic recombination event which activates the expression of the red fluorescent protein mScarletI, and excises the region encoding the π protein. The abundance of differentiated cells is inferred through mScarletI fluorescence, and the relative abundance of the R6K plasmid is inferred through the expression of its encoded constitutively produced green fluorescent protein sfGFP.

We characterized the long-term behavior of this differentiation circuit with varying differentiation rates and π protein expression, using flow cytometry to determine the population fraction of progenitor and differentiated cells. Progenitor cells are identified as the sfGFP-positive/mScarletI-negative population, while differentiated cells are identified as sfGFP-negative/mScarletI-positive population, with the activation of mScarletI expression occurring before loss of sfGFP (Figure 2.3E). Across all concentrations of Las-AHL, the population proceeds towards 100 percent differentiated mScarletI-positive cells in the absence of chloramphenicol when integrase is induced, with this occurring more quickly at higher induction of the integrase (Figure 2.3F). However, when selecting with chloramphenicol, the population appears to approach a steady state population distribution containing both progenitor and differentiated cells, with the relative abundance depending on the induction level of both integrase and π protein (Fig 3F). With 5 μM salicylate, differentiated cells comprise 59.2 +/- 0.03 percent (mean +/- SD of two replicates), 69.9 +/- 0.004 percent, and 70.3 +/- 0.05 percent of the population after four plate generations with 0.3 μM, 1 μM, and 3 μM Las-AHL, respectively. With 7.5 μM salicylate differentiated cells comprise 88.9 +/- 0.02 percent, 94.6 +/- 0.02 percent, and 96.4 +/- 0.01 percent of the population after four plate generations with 0.3 μM, 1μM, and 3μM Las-AHL, respectively. These distributions qualitatively align with our deterministic modeling, namely that a higher differentiation rate and a larger number of divisions allowed by differentiated cells both increase the steady state fraction of differentiated cells.

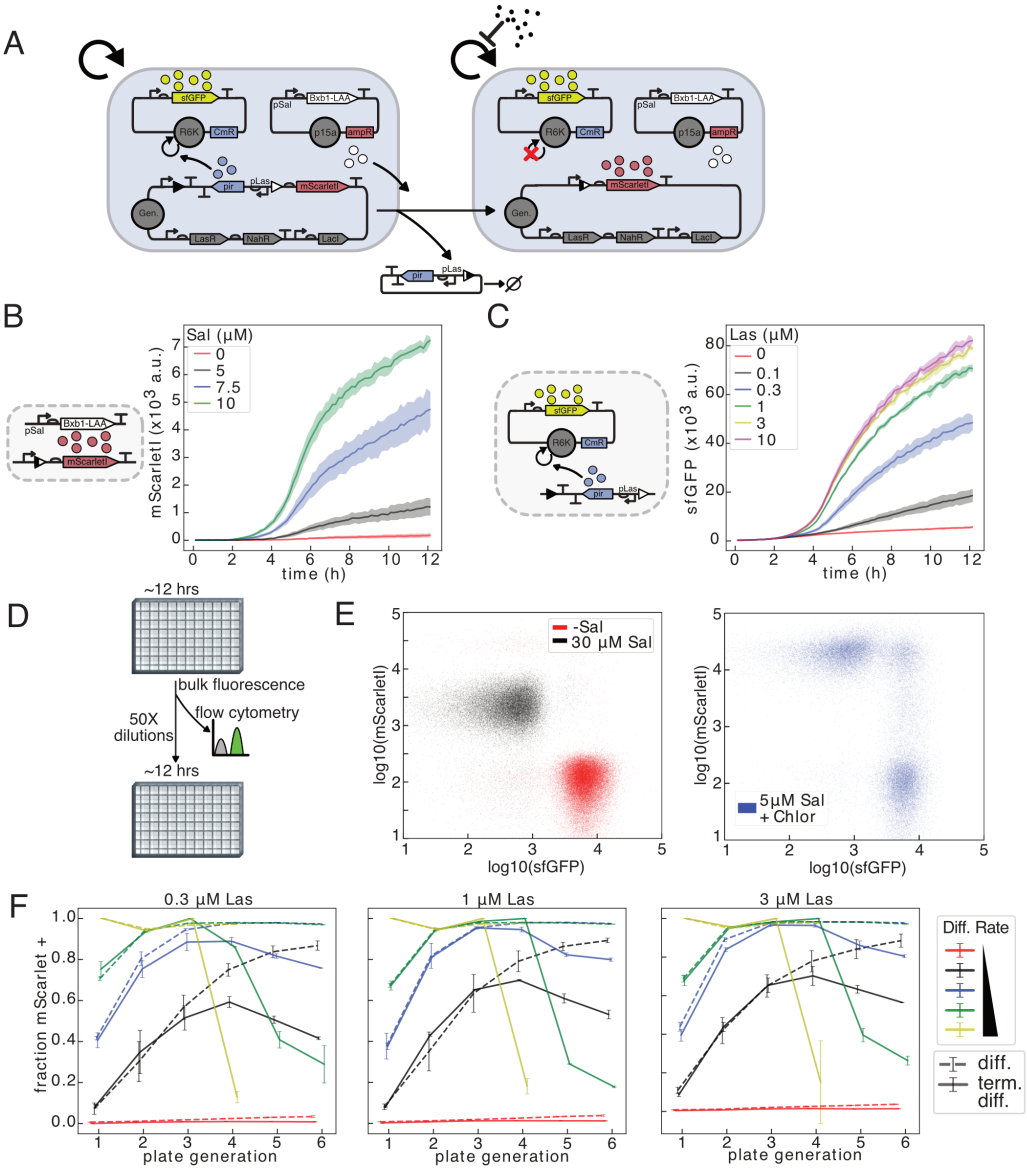

Figure 2.3: Implementation of a tunable integrase-mediated differentiation circuit. (A) Schematic of a tunable integrase differentiation circuit. Las-AHL and salicylate (sal) induce expression of *pir* encoded π protein and degradation-tagged Bxb1, respectively. (B-C) Batch culture experiments of JS006 with circuit depicted in (A) grown in M9CA media. (B) mScarletI fluorescence with varying induction levels of sal in M9CA + carb. (C) sfGFP fluorescence with varying induction levels of Las-AHL in M9CA + carb/chlor. (B-C) Mean +/- standard deviation of three replicates. (D-F) Cells are grown in 300 μL M9CA media with varying inducer concentrations, and diluted 50X every 12 hours. Samples are taken for flow cytometry after each growth. (E) Flow cytometry results after the third 12-hour growth for cells grown in 0.3 μM Las-AHL with carb +/- 30 μM sal(left), and cells grown media with carb/chlor, 0.3 μM Las-AHL, 5 μM sal(right). (F) Results of flow cytometry analysis of cells grown for six consecutive 12-hour growths in varying inducer concentrations. Average fraction mScarletI positive for two replicate wells +/- standard deviation.

However, at higher inductions of salicylate, we clearly see circuit failure when selecting with chloramphenicol. This is revealed at all Las-AHL concentrations in flow-cytometry of the fifth plate generation for 10 μM salicylate, and in the fourth plate generation for 15 μM salicylate (Figure 2.3F). Here, instead of achieving a steady state distribution comprised largely of differentiated cells, a population which is sfGFP-positive/mScarletI-negative comes to dominate the population. Though these resemble progenitor cells in gene-expression, they are no longer able to undergo differentiation, and have likely incurred a mutation analogous to the differentiation mutation we proposed in our model.

2.5 Differentiation-activated T7 RNAP expression improves burdensome function performance

Both to allow any arbitrary function to be expressed and to prevent leaky expression of the function in progenitor cells, we selected T7 RNAP, an orthogonal RNA polymerase broadly used in synthetic biology and bioproduction [26], to be activated by this recombination. To allow the expression level and burden to be tuned, the evolved IPTG inducible promoter P_{Tac} and associated transcription factor $LacI^{AM}$ [23] were used to control the expression of T7 RNAP. Recombination-activated T7 RNAP was integrated in a single copy on the *E. coli* genome, and a high copy Cole1 AmpR plasmid with T7 RNAP-driven sfGFP was used to report T7 RNAP expression. To assess T7 RNAP expression upon recombination, a variant of this Cole1 plasmid additionally encoding salicylate-inducible Bxb1 was used, while the Cole1 plasmid lacking Bxb1 was used to assess leaky T7 RNAP expression without recombination. As the positioning of the ribosomal binding site (RBS) with respect to the coding sequence is critical [27], we initially constructed the circuit such that recombination between the Bxb1 attB and attP sites would excise the intervening sequence, bringing the P_{Tac} promoter in proximity to the RBS and T7 RNAP coding sequence (Figure 2.4A). This design, however, resulted in a level of leaky T7 RNAP expression in the absence of any Bxb1 integrase as evidenced by higher T7 RNAP-driven GFP expression than the control with no T7 RNAP present (Figure 2.4D). As incorporating a terminator upstream of the T7 RNAP RBS did not improve this (Figure 2.4B), to fully eliminate leaky expression we relied on previous studies splitting T7 RNAP into functional domains to rationally choose a split site [28]. With this strategy, there is no potential for leaky expression of functional T7 RNAP prior to differentiation, and the full-length coding sequence that is generated upon recombination contains a 17 amino acid insertion from the

attL site and additional bases inserted to conserve the reading frame (Figure 2.4C). In the absence of Bxb1 integrase, GFP production was equivalent to the control without T7 RNAP presence, with induction of Bxb1 allowing high level T7 RNAP-driven expression (Figure 2.4E) that is tunable with IPTG (Figure 2.5B). This is also observed in the growth phenotype, with significant growth defect occurring only with addition of both salicylate and 30 or 100 μM IPTG (Figure 2.5B).

Using this system, we can directly compare the total output and duration of production for the two differentiation circuit architectures with naive inducible production at varying burden levels. The differentiation circuit for these experiments has a very similar design to the initial circuit described in Figure 2.3, with the distinctions that T7 RNAP instead of mScarletI is activated by differentiation, and Bxb1 is expressed from the R6K plasmid. The differentiation architectures with and without restricted cell division differ only in the presence of chloramphenicol, while the naive architecture is identical to a differentiated cell lacking the R6K plasmid, and is grown in the absence of chloramphenicol (Figure 2.5A-B). To read out T7 RNAP activity and to serve as the burdensome function being assessed, the high copy ColE1 AmpR plasmid with T7 RNAP-driven sfYFP was used for both differentiation and naive circuits. As with our mScarletI/sfGFP differentiation circuit, we can tune the differentiation rate with salicylate, and the population distribution of producers and non-producers over consecutive dilutions can be measured by flow cytometry. Further, we can compare the total production using end-point bulk-fluorescence measurements monitored during growth in a plate reader. For reference, JS006 cells with the naive circuit have experimentally determined growth rates of ~1.04 +/- 0.02 h^{-1} (mean/SD of two independent colonies with six total replicates), ~1.03 +/- 0.02 h^{-1}, ~0.84 +/- 0.09 h^{-1}, and ~0.48 +/- 0.06 h^{-1} when grown in M9CA + carbenicillin with 0, 10, 30, and 100 μM IPTG, respectively (Figure 2.6). This equates to burdens of ~1, ~19, and ~54 percent with 10, 30, and 100 μM IPTG relative to the uninduced case.

At relatively low-burden (10 μM IPTG), total production increases linearly in the naive case for the first ~6 plate generations before cells no longer producing sfYFP emerge and total production flattens to 44980 +/- 958 a.u. (mean +/- SD of four replicates; Figure 2.5E left panel). With higher burden (30 or 100 μM IPTG), nearly all production occurs in the first growth, and non-producers dominate the population by the end of the second growth (total production 11167 +/- 414; 5756 +/- 2061 a.u.). In the case of differentiation without selection, total production approaches

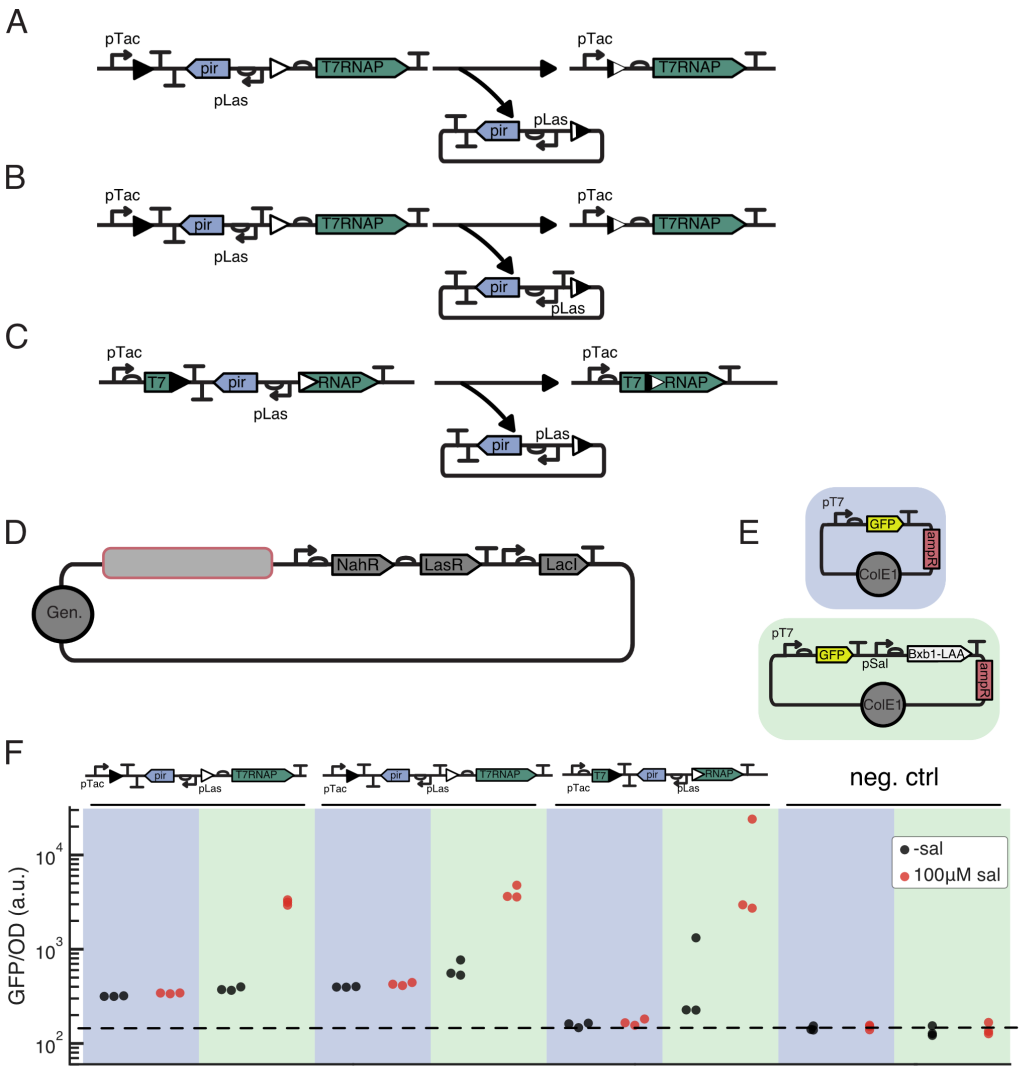

Figure 2.4: Integrase mediated activation allows leak-free expression of T7 RNAP. (A-D) Circuit designs for Bxb1 integrase activated expression of genomically integrated T7 RNAP. (A-B) Integrase recombination joins P_{Tac} promoter with RBS and full length CDS with (B) or without (A) a terminator in front of the RBS. (C) Integrase recombination joins promoter, RBS, and left half of CDS with right half of CDS. (F) JS006 negative control and genomically integrated strains were transformed with a ColE1 plasmid encoding T7 RNAP-driven GFP alone (blue shaded, F: top) or with salicylate inducible Bxb1-LAA (green shaded, F: bottom). OD normalized fluorescence recorded after 12 h of growth in LB + carb with (red) or without (black) 100 μM salicylate.

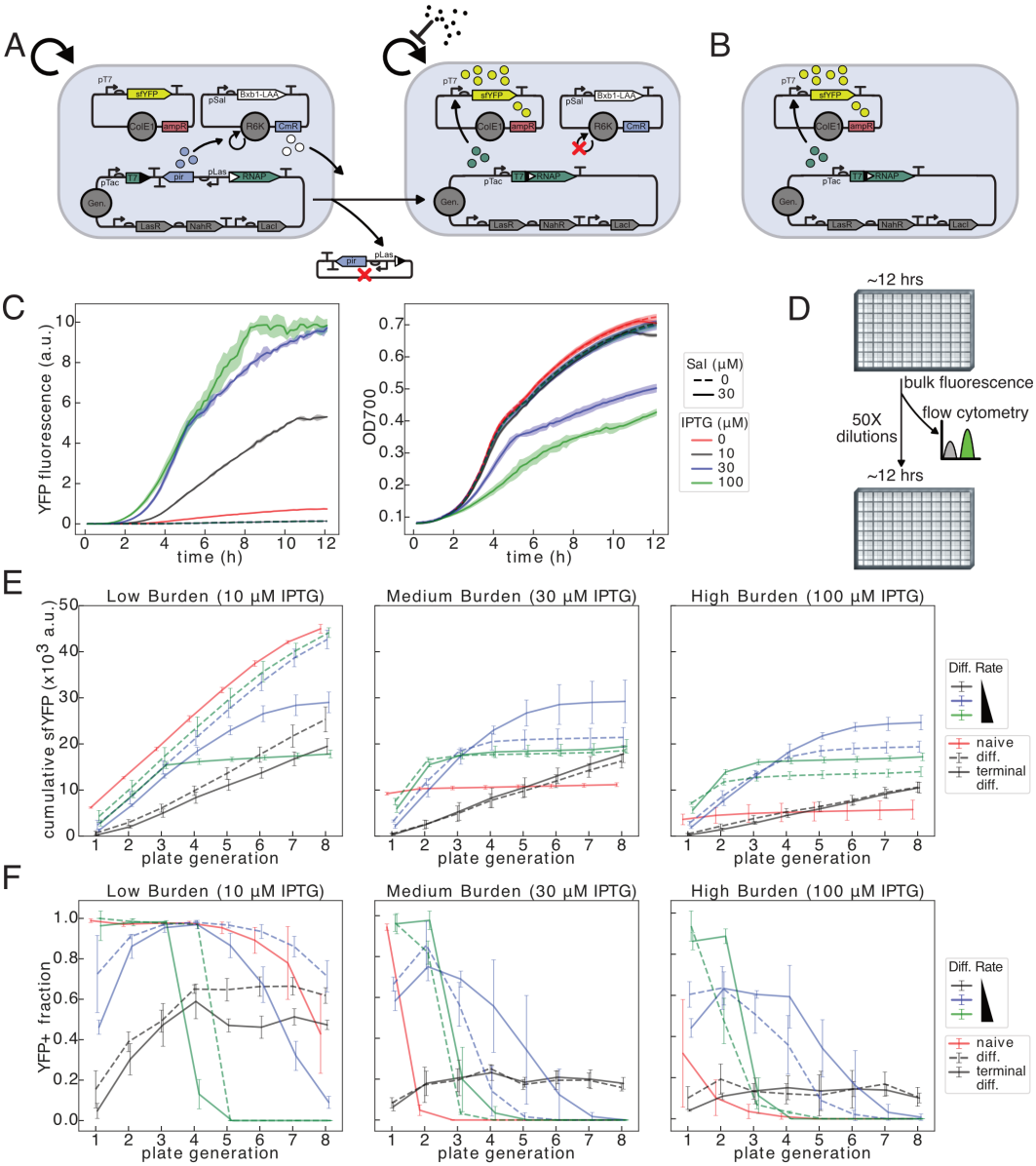

Figure 2.5: Differentiation improves duration and output from burdensome T7 driven expression circuits. (A) Schematic of differentiation-activated T7 RNAP-driven expression. (B) Schematic of naive inducible T7 RNAP-driven expression. Circuit is identical to a differentiated cell, but lacking the R6K plasmid. (C) Batch culture experiments of JS006 with circuit depicted in (A). Cells are grown in M9CA media + carb/chlor/1 μM Las-AHL, with or without 30 μM salicylate, with varying concentrations of IPTG. Curves are means +/- standard deviation of three replicates. (D-F) JS006 with circuit above grown in M9CA + carb/3 μM Las-AHL, +/- chlor, in varying concentrations of sal and IPTG. Cells diluted 50X into 300 μL total volume every ~12 h for 8 total growths. (E) Cumulative total production plotted is the sum of endpoint fluorescence values from 12 h plate reader growth. (F) Samples taken immediately after growth were analyzed by flow cytometry, and fraction sfYFP positive cells is plotted. (E-F) Means of four total replicates from two independent experiments +/- SD.

that achieved by naive production at 10 μM IPTG when integrase induction is sufficiently high (>5 μM salicylate). Here the delay in achieving ~100 percent differentiated producer cells is counteracted by a similar delay in accumulation of non-producers (Figure 2.5E-F).

With higher burden production, the benefit of differentiation for both circuit duration and total production becomes apparent. With a low differentiation rate (5 μM salicylate) and 30 μM IPTG, 15.9 +/- 0.02 percent are still producing sfYFP after 8 plate generations, greatly extending the duration of expression, with total production over the experiment being 16281 +/- 1508 a.u. (Figure 2.5E-F). A higher differentiation rate (7.5 μM salicylate), results in a greater total production (21414 +/- 2138 a.u.; ~1.9x naive) but decreased duration of expression. Increasing the differentiation rate further (10 μM salicylate) pushes this rate further from an apparent optimum, decreasing total production to 18552 +/- 381 a.u. (~1.7x naive). This benefit for total production is enhanced at the highest induction level of IPTG, with ~3.4x and ~2.4x the naive production with 7.5 and 10 μM salicylate, respectively.

The effect of selecting against differentiated cells, as in our modeling, is dependent on both the differentiation rate and the expression burden. At low burden (10 μM IPTG), this selection decreases total output in comparison to both naive and differentiation, with total production of 29010 +/- 2243 and 17822 +/- 832 a.u.(~0.6x and ~0.4x naive) with 7.5 and 10 μM salicylate, respectively. However, with higher burden production, differentiation with selection facilitates a total production of 29224 +/- 4629 a.u. with 7.5 μM salicylate, ~2.6x naive production and ~1.3x of differentiation without selection. As with differentiation alone, this benefit is exaggerated at the highest induction of IPTG, with production ~4.3x naive production, and ~1.3x that of differentiation alone.

In addition to evaluating the performance of these two differentiation architectures with respect to naive production, it is also useful to characterize the mutations which destroy circuit function for each case. For the naive and differentiation without chloramphenicol selection cases, colonies isolated after eight plate generations contained plasmid capable of expressing sfYFP when transformed into cells expressing T7 RNAP, but these cells failed to express sfYFP when transformed with the same expression construct encoded on a pSC101-chlor plasmid. It is apparent in both of these cases – although specific mutations were not identified – that a mutation on the genome has disrupted the expression of T7 RNAP, equivalent to the burden mutation described in our modelling. For the case of differentiation

with restricted cell division, it is apparent from the continued chloramphenicol resistance, loss of sfYFP expression, and insensitivity to induction with salicylate, that a mutation equivalent to the differentiation mutation (or alternatively both a burden mutation and a selection mutation) we posited in our modeling has occurred. Though mutations disrupting the expression of the integrase – either directly on integrase expression cassette encoded on the R6K plasmid or at the level of the transcription factor $NahR^{AM}$ – would accomplish this, we first examined the differentiation cassette itself. Apart from a ~1.3kb deletion which destroyed the attP site, we also observed two independent mutations in which the integrase-recombination event appears to have inverted rather than excised the intervening sequence, destroying both integrase attachment sites (supplementary sequences). Though we observed apparent differentiation mutations in the differentiation cassette in these three cases, we did not so in several others, and additional sequencing is required to determine all possible sources of mutation.

2.6 Discussion

With inspiration from bacterial reproductive and metabolic division of labor observed in nature, we developed a synthetic differentiation system that allows fractional tuning of progenitor and differentiated cells through inducible integrase-mediated differentiation and conditionally restricted cell division in differentiated cells. We applied this system to T7 RNAP-driven expression of fluorescent protein, and demonstrated that differentiation can improve total production output and duration of production for high burden production circuits, qualitatively matching deterministic modeling results. Further, we demonstrated that limiting the capacity of differentiated cells to undergo cell division was counterproductive for both total production and duration of production at low burden, but can increase both metrics at high burden relative to naive and differentiation architectures if the differentiation rate is appropriately tuned.

In modeling our differentiation architectures, we saw that the benefit of differentiation with and without restricted cell division relative to naive expression depended heavily on the expression burden. Specifically, the performance (both total production and duration) of differentiation with and without restricted cell division relative to naive production improves with increased burden. Though this trend agrees qualitatively with experimental results, it does not match exactly. While in our modeling differentiation and particularly differentiation with selection are ineffective or harmful with 10 percent burden, experimentally we see a strong benefit

for both total production and duration of production at 30 μM IPTG, corresponding to a ~12.9 percent growth penalty. This incongruity may be a result of cells being diluted from stationary phase (in all but the first plate generation) directly into media containing inducer. It has previously been shown that the cost/burden of unneeded protein production is elevated during the first few cell divisions following dilution from stationary phase [29]. Given that cells will undergo 5-6 cell divisions after a 50X dilution, an increase in effective burden for the first several cell divisions extends through much of the growth and may largely explain this difference between our modeling and experiments. If we instead perform experiments using higher fold dilutions or in continuous culture, this difference may be diminished.

Apart from effects due to burden, our modeling also revealed that the performance of our two differentiation architectures depended on the relative mutation rates. In particular, differentiation with restricted cell division is less sensitive to an increased rate of burden mutation, and the benefit of restricted cell division relative to differentiation alone is apparent with higher burden mutation rates and lower differentiation rates. Though direct quantification of mutation rates is complicated by potentially variable cost/burden of production across growth phase following cell dilution, we can infer that the differentiation mutation rate is likely not orders of magnitude larger than the burden mutation rate, and may indeed by on the same order or lower. We draw this conclusion because our modeling reveals that differentiation with selection tends to perform better than differentiation alone only when the burden mutation rate is of equal or greater order of magnitude than that of the differentiation mutation. While this may not be surprising, we recognize that a differentiation mutation may be facilitated by errors during integrase-mediated recombination, the rate of which has not been quantified to our knowledge.

While here we have demonstrated that limiting cell divisions in differentiated cells through selection with chloramphenicol can provide a modest improvement to total production and duration of production with respect to differentiation alone, it may be possible to increase this benefit with minor adjustments to our circuit architecture. First, because we are using chloramphenicol – which acts on the ribosome to inhibit protein synthesis – to select against differentiated cells, we may be unnecessarily inhibiting production. By using an alternative selectable marker such as *mFabI* and selecting with triclosan which inhibits lipid synthesis, we may be able to remove this inhibition of production in differentiated cells while maintaining inhibition of cell proliferation, thereby increasing total production [30]. Additionally, though

both differentiation architectures may benefit from reduction in the differentiation mutation rate, terminal differentiation is more sensitive to this mutation. Though this rate may be reduced by optimizing the circuit sequence following more careful analysis of mutations which occur in this case, we may alternatively reduce the effective mutation rate by requiring two or more mutations to occur to destroy differentiation potential. This may be accomplished by integrating two modified copies of the differentiation cassette, or by having two redundant differentiation mechanisms utilizing orthogonal integrases.

This first demonstration of utilizing synthetic differentiation to improve performance of burdensome functions, though reminiscent of examples of division of labor found in bacteria, differs in important ways which may limit the benefit gained from its implementation. Most obvious is that the function we are expressing in differentiated cells – T7 RNAP and fluorescent protein – is entirely non-essential. Replacing or supplementing this unneeded metabolic load with a function that is beneficial or essential for both progenitor and differentiated cells, thereby making the progenitor cells dependent on the differentiated cells, could improve the evolutionary stability of such differentiation architectures. Future more successful implementations of synthetic differentiation and division of labor may be closer to natural examples in this respect, however incorporating this feature into a differentiation architecture would likely prove insufficient. In an investigation of somatic differentiation in a synthetic system in yeast, Wahl and Murray demonstrate that multicellularity was critical for the evolutionary stability of the system, with unicellular systems being susceptible to cheaters [31]. They further speculate that this phenomenon explains the lack of the germ-soma distinction in unicellular species. However, because it is possible to engineer systems using biology that would never evolve naturally, this should not dissuade us from utilizing differentiation in addressing evolutionary stability.

While existing strategies can aid greatly in reducing the rate of mutations in engineered circuits, the only general strategies to reduce the rate at which mutations are selected are to reduce the burden of expression [1–3], or alternatively to integrate numerous copies of a genetic construct on the genome [10]. Reducing the burden of expression may not be a viable option for the production of toxic proteins or metabolites, or for certain industrial applications in which maximizing production is essential for economic viability. As well, though genomically integrating many copies of a construct may be effective at limiting the generation and selection

of non-productive mutant cells, cells expressing particularly burdensome or toxic functions will have impaired or destroyed ability to proliferate, potentially rendering long duration or continuous production nonviable. Further, applications in which engineered cells are not growing in mono-culture in a laboratory environment, but instead must compete with cells in a complex microbial community, will require consideration of this competition rather than simply the competition between functional and non-functional engineered cells.

The differentiation architectures we describe here represent a qualitatively new strategy for addressing the constraints imposed on synthetic biology by evolutionary forces, and may be applied generally to diverse circuits and functions. With the implementation of division of labor through differentiation, we can remove selective pressure for mutations relieving burden in a subset of the population, mitigate fitness defects in these progenitor cells, and sacrifice a tunable fraction to production with little regard for the burden or toxicity of the function. Circuit architectures utilizing differentiation can allow longer duration production – potentially in continuous culture – without the uniform growth defect common to all other strategies, and show promise in improving the evolutionary stability of engineered functions.

2.7 Model implementation

Naive expression circuit

In the naive implementation of a circuit, we have a population of cells (X_P) that are producing the protein/product of interest with some rate constant β, and as a result have some burden/growth penalty that is reflected in a growth rate $\mu_P < \mu_N$. Here μ_P is the growth rate of producers, and μ_N is the growth rate of non-producer cells. Producer cells can mutate with some probability λ_{m_B} (mutation that removes burden) to non-producer cells (X_N) that have a growth rate μ_N. Looking at this in the deterministic case in continuous culture with a carrying capacity K, and assuming the rate of production is limited in the same manner as the growth rate of the cells, we have that

$$\frac{dX_P}{dt} = \mu_P X_P \frac{K - X_P - X_N}{K} - \lambda_{MB} X_P - D X_P, \tag{2.1}$$

$$\frac{dX_N}{dt} = \mu_N X_N \frac{K - X_P - X_N}{K} + \lambda_{MB} X_P - D X_N, \tag{2.2}$$

$$\frac{dP}{dt} = \beta_P X_P \frac{K - X_P - X_N}{K}. \tag{2.3}$$

Differentiation circuit

With differentiation architecture, there are now two possible mutation types which could occur:

1. Production mutation (λ_{MB}): Mutation which inactivates the production capability of the cell. When a progenitor cell with this mutation differentiates, it will be a non-producer and growth at rate μ_N.

2. Differentation mutation (λ_{MD}): Mutation which inactivates the progenitor cell ability to differentiate. It will then grow at rate μ_N and differentiate at rate 0.

With the implementation of differentiation alone where differentiated cells continue to divide indefinitely, differentiated cells may incur this first type of mutation (λ_{M_B}) to yield cells that grow with rate μ_N, and whose differentiation does not have any effect. In writing out the system of ODEs to describe this, the subscript PP will be used to denote progenitor cells that are capable of differentiating into productive cells, N for non-producers, and DP for differentiated producers. As the differentiation architecture itself requires very minimal protein expression, we make the simplifying assumption that all non-producer cells grow at the same WT growth rate. This yields

$$\frac{dX_{PP}}{dt} = \mu_N X_{PP} \frac{K - X_{PP} - X_P - X_N}{K} - \lambda_{diff} X_{PP} - (\lambda_{M_B} + \lambda_{M_D}) X_{PP} - D X_{PP}, \tag{2.4}$$

$$\frac{dX_{DP}}{dt} = \mu_P X_{DP} \frac{K - X_{PP} - X_P - X_N}{K} + \lambda_{diff} X_{PP} - \lambda_{M_B} X_{DP} - D X_{DP}, \tag{2.5}$$

$$\frac{dX_N}{dt} = \mu_N X_N \frac{K - X_{PP} - X_P - X_N}{K} + (\lambda_{M_B} + \lambda_{M_D}) X_{PP} + \lambda_{M_B} X_{DP} - D X_N, \tag{2.6}$$

$$\frac{dP}{dt} = \beta_P X_P \frac{K - X_{PP} - X_P - X_N}{K}. \tag{2.7}$$

Terminal differentiation circuit

With the addition of selection against differentiated cells through selection for the R6K plasmid, the number of cell divisions is limited. Though differentiated cells retain limited proliferative potential, they will ultimately cease dividing and die, and we will therefore refer to this as *terminal* differentiation. With this circuit architecture, we now consider three types of mutations that may occur:

1. Production mutation (λ_{M_B}): Mutation which inactivates the production capability of the cell. When a stem cell with this mutation differentiates, it will be a non-producer and growth at rate μ_N.

2. Differentation mutation (λ_{M_D}): Mutation which inactivates the stem cells ability to differentiate. It will then grow at rate μ_N and be incapable of differentiation.

3. Selection mutation (λ_{M_S}): Mutation which occurs in differentiated cells (either producers or non-producers) which allows cells to evade selection and proliferate indefinitely.

In writing out the ODEs for this case, P will denote a progenitor cell capable of differentiating, D will denote a differentiated cell, and M will denote a mutated progenitor cell that can no longer undergo differentiation. The subscript N will indicate that a stem cell will differentiate into a non-productive differentiated cell, or is currently a non-productive differentiated cell, and similarly P for productive. Finally, the differentiated cells are not all equal, but are distinct in the number of divisions they have remaining before diluting the plasmid through cell division to the point where antibiotic resistance is insufficient to allow further proliferation. The cell produced from the differentiation event ($i = 0$) will have n divisions remaining, and when this cell divides, there will be one fewer cell with n divisions remaining, and 2 more cells with $n - 1$ divisions remaining ($i = 1$). When $i = n$, cells will no longer divide, and this is equivalent to cell death. Though the behavior of cells that are no longer able to divide (whether or not they continue to produce the protein/product of interest) would likely depend both on the type of selectable marker encoded on the R6K plasmid and the protein/product that is being produced, using chloramphenicol resistance as the selectable marker likely would prevent protein production in cells that are no longer able to divide.

As with the differentiation architecture, we make the assumption that all non-producer cells growth with the same WT growth rate. In the modeling of this architecture, we also must decide how selection will affect the growth rate of the dif-

ferentiated cells. Though we could imagine cell growth slowing due to insufficient chlorR expression as the plasmid copy number is reduced through cell division, in modeling this architecture we make the assumption that selection only affects cells with no remaining cell divisions. Therefore the growth rate of differentiated cells capable of further cell division is determined solely by the expression burden. With these assumptions, we can now write out our system of ODEs. In the equations below, the total population (X_{TOT}) is given by

$$X_{TOT} = X_{PP} + X_{PN} + X_{MP} + X_{MN} + \sum_{0}^{n} X_{DP}(i) + \sum_{0}^{n} X_{DN}(i). \tag{2.8}$$

Progenitor cells are described by

$$\frac{dX_{PP}}{dt} = \mu_{PP} X_{PP} \frac{K - X_{TOT}}{K} - \lambda_{\text{diff}} X_{PP} - \lambda_{MB} X_{PP} - \lambda_{MD} X_{PP} - D X_{PP}, \tag{2.9}$$

$$\frac{dX_{PN}}{dt} = \mu_{PN} X_{PN} \frac{K - X_{TOT}}{K} - \lambda_{\text{diff}} X_{PN} + \lambda_{MB} X_{PP} - \lambda_{MD} X_{PN} - D X_{PN}. \tag{2.10}$$

Mutant progenitor cells are described by

$$\frac{dX_{MP}}{dt} = \mu_{MP} X_{MP} \frac{K - X_{TOT}}{K} + \lambda_{MS} \sum_{i=0}^{i=n-1} X_{DP} - \lambda_{MB} X_{MP} - D X_{MP}, \tag{2.11}$$

$$\begin{aligned} \frac{dX_{MN}}{dt} = {} & \mu_{MN} X_{MN} \frac{K - X_{TOT}}{K} + \lambda_{MD} \left(X_{PN} + X_{PP}\right) + \lambda_{MB} X_{MP} \\ & + \lambda_{MS} \sum_{i=0}^{i=n-1} X_{DN} - D X_{MN}. \end{aligned} \tag{2.12}$$

Differentiated producer cells are described depending on the number of cell divisions.

For $i = 0$,

$$\begin{aligned} \frac{dX_{DP}(0)}{dt} = {} & \lambda_{\text{diff}} X_{PP} - \mu_{DP}(0) X_{DP}(0) \frac{K - X_{TOT}}{K} - \lambda_{MB} X_{DP}(0) \\ & - \lambda_{MS} X_{DP}(0) - D X_{DP}(0). \end{aligned} \tag{2.13}$$

For $0 < i < n$,

$$\frac{dX_{DP}(i)}{dt} = 2\mu_{DP}(i-1)X_{DP}(i-1)\frac{K - X_{TOT}}{K} - \mu_{DP}(i)X_{DP}(i)\frac{K - X_{TOT}}{K} - \lambda_{MB}X_{DP}(i) - \lambda_{MS}X_{DP}(i) - DX_{DP}(i). \tag{2.14}$$

For $i = n$,

$$\frac{dX_{DP}(n)}{dt} = 2\mu_{DP}(n-1)X_{DP}(n-1)\frac{K - X_{TOT}}{K} - DX_{DP}(n). \tag{2.15}$$

Differentiated non-producer cells are similarly described.

For $i = 0$,

$$\frac{dX_{DN}(0)}{dt} = \lambda_{\text{diff}}X_{PN} - \mu_{DN}(0)X_{DN}(0) + \lambda_{MB}X_{DP}(0) - \lambda_{MS}X_{DN}(0) - DX_{DN}(0). \tag{2.16}$$

For $0 < i < n$,

$$\frac{dX_{DN}(i)}{dt} = 2\mu_{DN}(i-1)X_{DN}(i-1) - \mu_{DN}(i)X_{DN}(i) + \lambda_{MB}X_{DP}(i) - \lambda_{MS}X_{DN}(i) - DX_{DP}(i). \tag{2.17}$$

For $i = n$,

$$\frac{dX_{DN}(n)}{dt} = 2\mu_{DN}(n-1)X_{DN}(n-1) - DX_{DP}(n). \tag{2.18}$$

Finally, we can model the production, assuming that all productive differentiated cells are producing at the same maximum rate. While we could model the concentration, we are mainly interested in the total output, and therefore will only keep track of the total production. Production is then described by

$$\frac{dP}{dt} = \beta_P \left(\sum_{i=0}^{i=n} X_{DP}(i) + X_{MP} \right) \frac{K - X_{TOT}}{K}. \tag{2.19}$$

There may be two adjustments to this production equation, one of which will affect the effective total population size. Particularly in the case of chloramphenicol which acts by inhibiting translation, the differentiated cells of generation n would not be metabolically active, would be considered dead, and therefore would not contribute

either to the total effective population size or to the production of P. We would then have

$$\frac{dP}{dt} = \beta_P \left(\sum_{i=0}^{i=n-1} X_{DP}(i) + X_{MP} \right) \frac{K - X_{TOT}}{K}. \tag{2.20}$$

The total effective population size (X_{TOT}) would then be given by

$$X_{TOT} = X_{PP} + X_{PN} + X_{MP} + X_{MN} + \sum_{0}^{n-1} X_{DP}(i) + \sum_{0}^{n-1} X_{DN}(i). \tag{2.21}$$

The second would come from relaxing the assumption that all producer cells would produced P at the same rate. We could imagine instead that, particularly in the case of chloramphenicol, the production rate of P would vary with the growth rate as this to is affected by the inhibition of translation by the antibiotic. We then would have

$$\frac{dP}{dt} = \beta_P \sum_{i=0}^{i=n-1} X_{DP}(i) \frac{\mu_{DP}(i)}{\mu_{DP_{max}}} \frac{K - X_{TOT}}{K} + \beta_P X_{MP} \frac{\mu_{DP}(i)}{\mu_{DP_{max}}} \frac{K - X_{TOT}}{K}. \tag{2.22}$$

In the modeling done for this paper, we assume that all differentiated cells that are capable of additional cell divisions ($i < n$) grow and produce with the same rate. We also assume differentiated cells that can no longer divide are equivalent to dead cells, and neither produce or contribute to the total cell count.

2.8 Materials and methods

Strains and constructs

JS006 strain *E. coli* were used for all experiments, and constructs were assembled using 3G assembly. Constructs were genomically integrated with clonetegration using pOSIP-KO and pOSIP-CH [32]. pOSIP plasmids were double digested with BamHI and SpeI, and PCR purified. For assembly, 3G was used [11], however P1 and PX adapters were used in place of UNS1 and UNSX to allow compatibility with the pOSIP backbone. Modified adapters were used to generate the bicistronic transcriptional unit for LasR^{AM}/NahR^{AM} (UNS3D/UNS3B), and inverted *pir* transcriptional unit (UNS3E*/UNS3A*). Generation of modified MoClo [33] compatible

T7 RNAP parts, as well as degradation tagged Bxb1 integrase and UNS1-UNSX R6K-chlor plasmid backbone, is summarized in the Table 2.1. Sequences for Bxb1 integrase attachment sites attB and attP were obtained from Ghosh, et. al [34]. Sequences of all parts, primers, and final constructs can be found in the supplementary information: (https://www.biorxiv.org/content/10.1101/614529v2.supplementary-material).

Deterministic modeling of differentiation circuits

Circuits depicted in Figure 2.1A-C were modeled deterministically using systems of coupled ordinary differential equations. In all simulations, rates for differentiation and mutations are first-order with respect to cell number, and do not depend on growth rate. Exponential growth without carrying capacity was assumed for simulations in Figure 2.1 E-H. For Figure 2.2, circuits were simulated with logistic growth with a carrying capacity of 10^9 cells, with cells being diluted 50x when cells reached 95 percent of the carrying capacity. Production was modeled as being proportional to the ratio of specific growth rate (actual growth rate after accounting for effect due to carrying capacity) to maximum growth rate for the specific cell type, and production rate was 1 a.u. per cell per hour for all simulations regardless of burden. For differentiation with selection, a cell with n remaining cell divisions divides to generate two cells with $n - 1$ remaining cell divisions. Cells with one remaining cell division therefore divide into two cells which do not divide. This was equated to cell death, and these cells have zero production and do not count towards the carrying capacity. For comparing total production (Figure 2.2A) and duration of circuit function with >10 percent of cells being producers (Figure 2.2B), the differentiation rate was selected using optimization (three independent starting values: 0.001, 0.1, 1 h^{-1}) to maximize total production or duration. Jupyter notebooks describing and running all simulations are available on the Github repository listed in supplementary information.

Differentiation experiments

For experiments with differentiation cells, cells were grown from glycerol stock in 3 mL culture of M9CA glucose (Teknova M8010) with 34 μg/mL chloramphenicol, 100 μg/mL carbenicillin, and 1 μM Las-AHL. Overnight cultures were diluted 1:100 into the same media and grown ~2-3 hours to OD 0.2-0.4. To avoid cross-over of antibiotics and inducers, cells were pelleted (3500g for 10 min) before resuspending in M9CA with appropriate antibiotics (carb for differentiation, carb

+ chlor for differentiation with selection) to OD ~0.1. Control cells with naive inducible expression of T7 RNAP from the genome were treated as above but grown in M9CA glucose + carbenicillin. Cells at OD ~0.1 were diluted 1:10 into a total volume of 300 μL containing appropriate antibiotics (carbenicillin +/- chloramphenicol) and various inducer concentrations (IPTG, salicylate, Las-AHL). Cells were grown in 96-well square-well plate (Brooks MGB096-1-2-LG-L) at 37°C with maximum-speed linear shaking in a BioTek Synergy H1m. OD700, sfYFP fluorescence (503/540 nm excitation/emission; gain 61 and 100), sfGFP fluorescence (485/515 nm excitation/emission; gain 61 and 100), and mScarletI fluorescence (565/595 nm excitation/emission; gain 100) were measured at 10 minute intervals as appropriate. For long-term experiments, cells were diluted 1:50 after ~12 h growth into the same media conditions into a replicate plate. All data and Jupyter notebooks are available on the Github repository.

Flow-cytometry

Immediately after the conclusion of a 12 h growth, cells were diluted 1:50 into 100 μL 1X PBS for analysis with flow cytometry. Samples were run on a Miltenyi Biotech MACSQuant VYB Flow Cytometer equipped with Violet 405 nm, Blue 488 nm, and Yellow 561 nm lasers. sfYFP was measured with the 488 nm laser with 525/50 nm filter, sfGFP with the 405 nm laser with 525/50 nm filter, and mScarletI with the 561 nm laser with 661/20 nm filter. 50,000 ungated events were recorded for each sample, and results were analyzed with custom Python code available in the Github repository listed supplementary information. Briefly, peak locations were determined from KDE fits of ungated flow data, gaussian mixture models used to assign cells to peaks, and cells within peaks were designated positive or negative for the respective fluorescent protein using a chosen threshold for peak mean. For T7 RNAP differentiation experiments, peaks with mean log10(sfYFP) >2.5 were designated *on*. For mScarletI/sfGFP differentiation experiments, peaks with mean log10(mScarletI) >3 were designated as *differentiated*.

Identification of mutations in differentiation-activated T7 RNAP expression

Cells from the eighth plate generation were struck for single colonies. For the case of differentiation with selection (+chloramphenicol), cells from two independent wells from 7.5 μM salicylate/0 μM IPTG and 7.5 μM salicylate/30 μM IPTG were plated on LB agar + chloramphenicol, carbenicillin, and Las-AHL. Colony PCR using p4.186.primary.FOR and pOSIP.insert.REV was performed on four colonies

from each, two of which were sequenced with RW.pir.int.R and RW.pir.int.F2, and sequences were mapped to 'differentiation cassette split T7 RNAP'. This was done similarly for naive (0 and 30 μM IPTG) and differentiation without selection (7.5 μM salicylate/0 μM IPTG and 7.5 μM salicylate/30 μM IPTG), however PCRs were unsuccessful. The source of mutation was determined by (1) isolating plasmid DNA from two isolated colonies from each plate and transforming into cells with genomically integrated inducible T7 RNAP (naive production cells lacking the pT7-sfYFP construct) and (2) transforming a pSC101-chlor-pT7-BCD2-sfYFP-T2m construct into competent cells prepared from these same cells.

Part	Template	FOR	REV
U1-R6K-chlor-UX	pOSIP-CH	RW.U1.R6kbb.F	RW.UX.R6kbb.R
T7(attL)RNAP*	C95m	RW.T7RNAP.split2.FOR	RW.T7RNAP.split2.REV
T7(L)attB	T7(attL)RNAP	pseqF	RW.T7split.attB.bsaID.R
attPT7(R)	T7(attL)RNAP	RW.T7split.attP.bsaIB.F	pseqR
Bxb1LAA	Bxb1	pseqF	RW.bxb1.LAA.rev1
			RW.bxb1.LAA.rev2

Table 2.1: Part construction. * indicates part was assembled with a 1 piece Gibson after PCR

2.9 Supplementary Information

All data, and code for data analysis and modeling are available on the Github repository (https://github.com/rlwillia/Integrase-differentiation). Sequences of all parts, primers, and constructs, and a table with descriptions of mutations observed in experiments with T7 RNAP differentiation with selection, can be found in the supplementary excel file (https://www.biorxiv.org/content/10.1101/614529v2.supplementary-material).

2.10 Acknowledgements

We would like to thank Andy Halleran, Anandh Swaminathan, and Andrey Shur for productive conversations; Andy Halleran for providing code for analysis of flow cytometry data; Samuel Clamons for providing code for tidying and analyzing Biotek data; and Andrey Shur and Andy Halleran for providing cloning resources. NahR^{AM}, LasR^{AM}, LacI^{AM}, and their corresponding evolved promoters P_{SalTTC}, P_{LasAM}, and P_{Tac} were provided by Adam Meyer [23]. The CIDAR MoClo Parts Kit which includes various promoter, RBS, CDS, and terminator parts used in the constructs described were provided by Douglas Densmore (Addgene kit 1000000059). This research was supported by the Institue for Collaborative Biotechnologies through

grant W911NF-09-0001 and cooperative agreement W911NF-19-2-0026 from the U.S. Army Research Office. The content of the information on this page does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

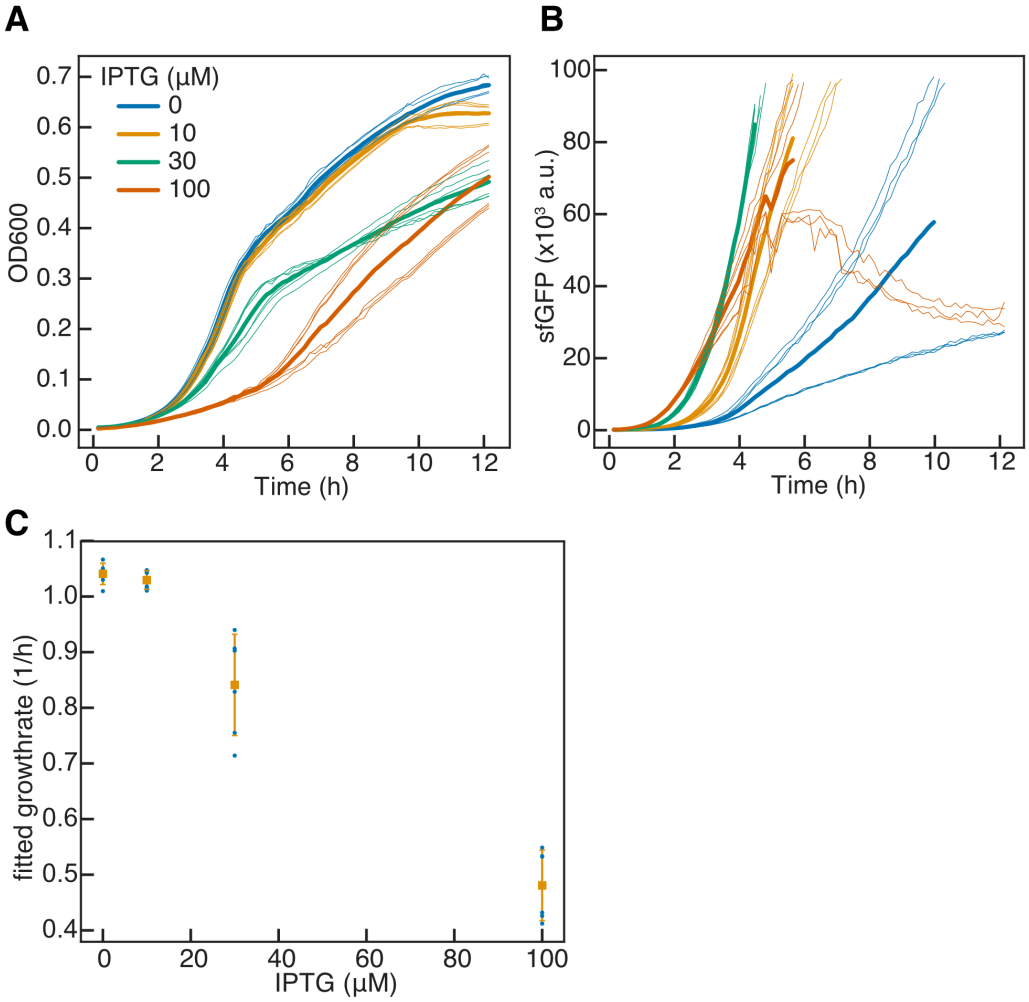

Figure 2.6: Characterization of burden level of naive T7 RNAP driven GFP. JS006 cells with naive IPTG inducible T7RNAP were transformed with a high copy ColE1 plasmid with AmpR. Cells were outgrown in LB with carbenicillin (100μg/mL) for 6 hours then diluted 1:50 into LB carb with varying concentrations of IPTG. (A) GFP production and (B) OD600 were measured every 10 minutes for six replicates grown at 37C in 0.3mL. (B) OD600 curves were trimmed to 30 percent of maximum OD600 achieved and used to fit an exponential growth model with noise floor, initial population, and growth rate parameters. Mean growth rate +/- SD of 6 replicates fitted separately plotted. Fitted growth rates are 1.04, 1.03, 0.84, and 0.48 h^{-1} for 0, 10, 30, and 100 μM IPTG, respectively, corresponding to ~1, ~19, and ~53 percent burden for induced relative to uninduced cases.

Chapter 3

IMPROVING THE EVOLUTIONARY STABILITY OF DIFFERENTIATION CIRCUIT ARCHITECTURES WITH REDUNDANCY

3.1 Introduction

In the previous chapter, we developed a tunable integrase-mediated differentiation system, applied it to T7 RNAP-driven expression, and demonstrated its utility in improving the evolutionary stability of burdensome functions in *E. coli*. This served as a useful initial demonstration of the viability of this strategy, but the ciruit design and implementation was lacking in several ways. Foremost among these shortcomings was the susceptibility of the differentiation circuit to be broken by a variety of single mutations. In the case of differentiation with unlimited growth of progenitor cells, a single mutation of the genomically integrated T7 RNAP expression construct would ablate the function, and in the case of terminal differentiation, a single mutation destroying an integrase attachment site or disrupting the expression of the transcriptional activator NahR would yield progenitor cells with minimal or no capacity to differentiate. Further, by placing the expression of the integrase responsible for catalyzing the differentiation event on a plasmid, it was inherently subjected to both more opportunities for mutation and to random plasmid partitioning effects which can hasten the expansion and fixation of beneficial mutations [11]. In the consideration of addressing these flaws and limitations, the exciting aspect to note is that by improving the evolutionary stability of this architecture, we are able to improve the evolutionary stability of any engineered function in general.

3.2 Reviewing intuition from deterministic modeling

Though we approach modeling this system in Chapter 2, we approach this again here to develop intuition behind various strategies of expressing burdensome functions. There are several differences in the approach to modeling these systems as will be discussed later. In the typical case where an entire population of cells inducibly or constitutively expresses a burdensome function, which we term naive expression (Fig 3.1A), we can model the system very simply. A population of cells produces an arbitrary gene product, and due to the burden of expression grows at a rate lower than wild-type. A mutation which disrupts the expression of this function

occurs at rate k_{MB}, relieving the burden, and restoring a wild-type growth rate. We model this system deterministically with carrying capacity limited growth in a chemostat with constant dilution, and see that both total production and duration of function decrease exponentially with the burden of expression, and similarly increasing dilution rate at a set burden decreases both metrics (Fig 3.1 D-G). The rate at which mutant cells growing at a wild-type growth rate will overtake slower growing producer cells increases exponentially with the difference in growth rate, and further if the growth rate of the producer cells is slower than the dilution rate, washout will occur if there is not mutational escape. Exponential increase in the burden mutation rate, however, results in only a linear decrease in duration and total production.

With a differentiation strategy, we instead begin with a population of progenitor cells, that because they do not express the burdensome function have a wild-type growth rate. These cells differentiate at some rate k_{diff}, becoming producer cells with a reduced growth rate, and as in the naive case incur a burden mutation at rate k_{MB}, relieving the burden and restoring a wild-type growth rate (Figure 3.1B). With the addition of this differentiation event, we also must consider differentiation mutations, the class of mutation which disrupts this process at rate k_{MD}. While the burden mutation may occur in both progenitor cells and differentiated producer cells, the differentiation mutation only occurs in progenitor cells. With the simplest differentiation strategy, both mutations lead to the same sink of non-producer cells which cannot differentiate. However, by adding an additional facet to this circuit that limits the number of divisions differentiated cells can undergo, both producer and non-producer differentiated cells will die. In this case of terminal differentiation, only the differentiation mutation leads to the sink of mutated progenitor cells which can no longer differentiate (Figure 3.1C).

Examining the behavior of these two differentiate circuit architectures with deterministic modeling, we see that the burden of the function and dilution rate effect performance much differently than in the naive case. While the total production and duration are relatively unaffected by burden in the case of terminal differentiation, differentiation alone approximates the naive case at low burden, and that of terminal differentiation at high burden. (Fig 3.1D-E). This is because the effective burden experienced by progenitor cells is not the burden of expression, but instead the rate of differentiation which functions as a death or dilution rate. Because of this, differentiation allows operation at a dilution rate higher than would be possible for naive

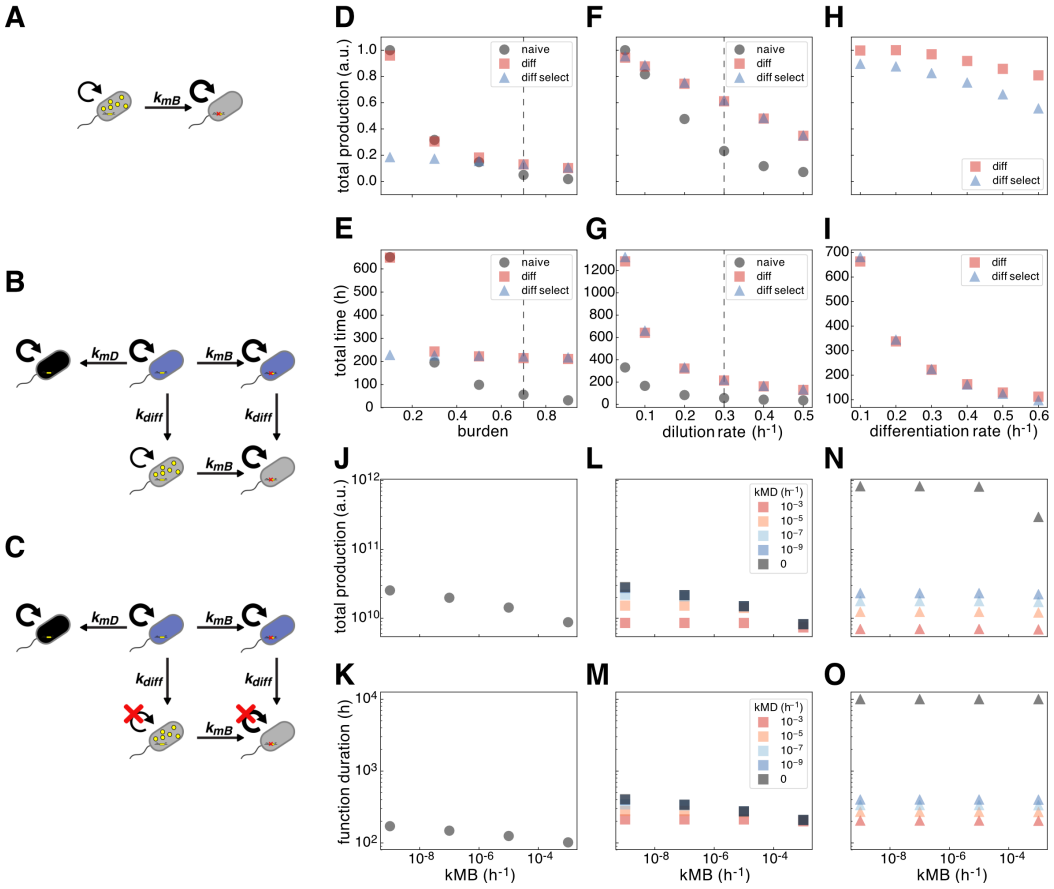

Figure 3.1: Deterministic modeling of strategies for expression of burdensome functions. (A-C) Grey cells containing product (yellow dots) are producers and grow at a slower rate than all other cells which are non-producers (A) Naive inducible or constitutive expression in which producer cells incur a burden mutation at rate k_{MB} (h^{-1}) which destroys expression and restores wild-type growth in non-producer cells. (B) Differentiation architecture in which progenitor cells (blue) differentiate at rate k_{diff} (h^{-1}) into differentiated cells (grey). Progenitor cells and differentiated producers incur burden mutation occurs at rate k_{MB} (h^{-1}), and progenitor cells incur differentiation mutations at rate k_{MD} (h^{-1}). (C) Terminal differentiation architecture. Same as (B), but differentiated cells can divide a limited number of times before dying (red X). (D-O) Deterministic simulations circuits in chemostat with constant dilution and carrying capacity limited growth. $K = 10^9$ cells, $\mu_P = 1\ h^{-1}$, $n_{\text{div}} = 4$. (D-I) $k_{MB} = k_{MD} = 10^{-7}\ h^{-1}$. (D-E) $D = 0.3\ h^{-1}$, $k_{\text{diff}} = 0.3\ h^{-1}$. (F-G) 30 percent burden, $k_{\text{diff}} = 0.3\ h^{-1}$. (H-I) 50 percent burden, $D = 0.3\ h^{-1}$. (J-O) 50 percent burden, $D = 0.2\ h^{-1}$, $k_{\text{diff}} = 0.3\ h^{-1}$,

expression given the burden level if the differentiation rate is appropriately chosen (Figure 3.1F-G). Increasing the rate of differentiation both increases the fraction of producer cells and therefore the effective burden experienced by the progenitor cells. This exponentially decreases the duration of function, but minimally effects the total production as it is compressed into a shorter period (Fig 3.1H-I). In examining the effects of varying the rates of the burden and differentiation mutations, we see in the case of differentiation alone that the mutation with the faster rate determines the duration of function and total production (Fig 1L-M). However with terminal differentiation, as we would expect from the circuit diagram, only the differentiation mutation can be the cause of circuit failure. As there is no selective pressure for burden mutations in the progenitor cell population, cells with this mutation accumulate linearly with time, and only affect total production if this rate is exceedingly high. Decreasing the rate of the differentiation mutation increases both metrics regardless of burden mutation rate, and critically, if this rate is zero, we achieve indefinite function (Fig 3.1N-O). Strikingly, terminal differentiation is a strategy which has the differentiation mutation as its singular Achilles' heel, and if this mutation is addressed, any function could be made evolutionarily stable.

3.3 Differentiation circuit development

In order to experimentally implement the differentiation strategy, we required (1) burdensome expression to be fully off in the progenitor cell population, (2) irreversible activation of an arbitrary function at a tunable rate, and in the case of terminal differentiation (3) means of limiting the growth of differentiated cells. Given the experience of the Murray lab with serine-integrases and the unidirectional nature of recombination characteristic of this class of proteins, we selected integrase mediated recombination as a means of activating our arbitrary function. As demonstrated in Chapter 2, we achieved means of both activating an arbitrary function of interest, and preventing leaky expression in the progenitor cells by separating the T7 RNAP coding sequence into two pieces which are brought together through a recombination event catalyzed by the integrase Bxb1.

In the case of terminal differentiation, it was important that this single recombination event would both activate the expression of the arbitrary function, as well as limit the capacity of the differentiated cell to grow and divide. To accomplish this, as discussed in Chapter 2 we co-opted the R6K plasmid system in which the π protein encoded by the *pir* gene is required for replication. Placing *pir* between the integrase attachment sites would then result in loss of π protein expression

upon differentiation, and subsequent dilution of the plasmid and plasmid encoded antibiotic resistance gene through cell growth and division, thereby allowing the proliferation of differentiated cells to be limited through antibiotic selection (Fig 3.2A). Accordingly, we see that induction of integrase expression results both in the expression of T7 RNAP-driven sfGFP, as well as loss of mScarletI fluorescence produced constitutively from the R6K plasmid (Fig 3.2C).

From our initial deterministic modeling, we observed that decreasing the rate or probability of the differentiation mutation improves the duration of function of the terminal differentiation circuit regardless of the rate of burden mutation. However, achieving the orders of magnitude improvement in mutation rate desired is not feasible simply by optimizing the circuit sequence to minimize mutation rate. We instead reasoned that instead of reducing the rate of mutation, increasing the number of independent mutations required to break the differentiation mechanism would yield more significant improvements. To this end, we envisioned a circuit design with two T7 RNAP differentiation cassettes in which the recombination of a single cassette would both activate the function and enable limiting the growth of differentiated cells. If a second identical cassette was integrated, recombination of both cassettes would be required to cease replication of the R6K plasmid and allow antibiotic selection mediated limitation of growth. Therefore a mutation preventing the recombination of one cassette would be sufficient to enable indefinite growth, giving opportunity for selection for burden mutations. However, if each differentiation cassette encoded a unique half of the π protein, a single recombination event would ablate the expression of functional pi protein and with it the replication of the R6K plasmid. We selected two candidate sites for splitting the protein, tagged the N- and C-terminal fragments with the N- and C terminal fragments of the Cfa intein [35], respectively, and functionally screened for R6K plasmid replication, yielding one functional split site (Figure 3.2B-C). Expression of the intein tagged fragments allows R6K plasmid replication, and inactivation of either the 5' fragment or 3' fragment through integrase-mediated recombination results in loss of the R6K plasmid (Figure 3.2C).

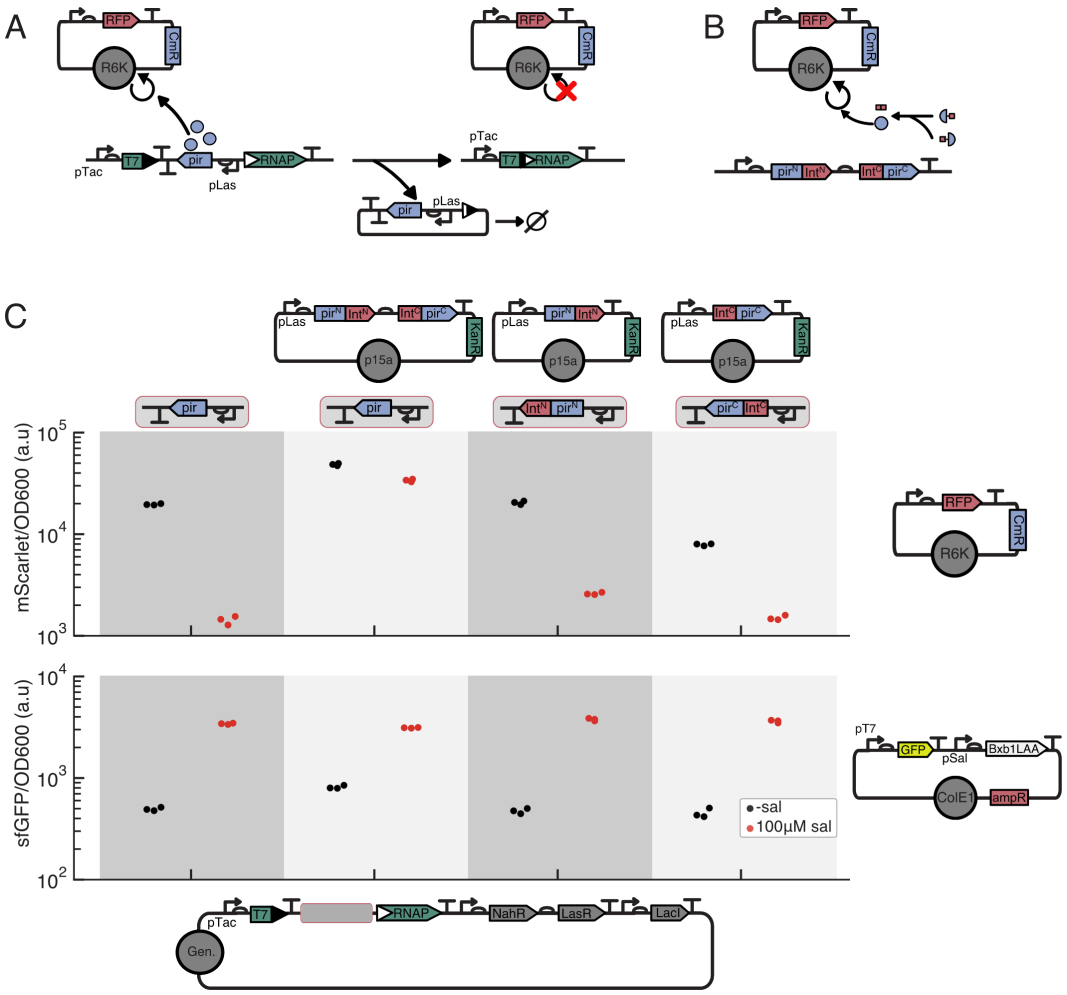

Figure 3.2: Integrase inactivation of π protein expression ablates R6K plasmid replication. (A) Schematic of integrase excision of *pir* resulting in loss of π protein and cessation of R6K plasmid replication. (B) Design of a split-π protein system in which each half is tagged with an intein fragment, and generate a full-length protein upon protein trans-splicing. (C) Strains harboring genomic integrations of circuits in which Bxb1 recombination simultaneously activates the expression of T7 RNAP and inactivates full-length or split π protein, and containing an R6K plasmid with chlorR and constitutive mScarletI w/ or w/o a p15a KanR plasmid encoding one or both halves of the π protein, were transformed with a ColE1 AmpR plasmid with pT7 GFP and P_{SalAM} Bxb1. OD normalized RFP (top) and RFP (bottom) were recorded after 18 h of growth in LB w/ 100 μg/mL carb, 34 μg/mL chlor, 30 nM Las-AHL, with (red) or without (black) 100 μM salicylate.

3.4 Experimental evaluation of differentiation and terminal differentiation strategies

To assess the capacity of differentiation strategies to improve the evolutionary stability of burdensome functions, we performed long-term experiments comparing burdensome T7 RNAP driven expression of a fluorescent protein in cells with one or two copies of inducible T7 RNAP to our single-cassette and two-cassette differentiation circuits. Both the single-cassette differentiation and two-cassette differentiation strains have two copies of inducible integrase, and critically all components in the naive and differentiation circuits were genomically integrated, ensuring precise copy number control and preventing effects due to plasmid partitioning (Figure 3.3A-C, 3.27). Experimental comparison of differentiation with terminal differentiation required only including chloramphenicol in the medium in the case of terminal differentiation, as without antibiotic present differentiated cells would grow unhindered after losing the R6K plasmid. Inducer and antibiotic conditions were uniform throughout 16 consecutive batch growths, with the degree of burden tuned with IPTG (P_{Tac} T7RNAP) and differentiation rate tuned with salicylate (P_{SalTTC} Bxb1-LAA). In determining the appropriate inducer concentrations for these experiments, we noted in a pilot experiment with 3 plate generations that while the behavior of 1x differentiation was minimally affected by the induction level of π protein expression with Las-AHL (10-300 nM; Figure 3.8, 3.10), 2x differentiation using the split-π protein was quite sensitive (Figure 3.9, 3.11). Specifically we observed R6K plasmid copy number as inferred through a constitutively expressed mScarletI to be more sensitive to Las-AHL induction in the range of concentrations tested, with higher copy number than 1x differentiation observed across all concentrations. As well, increased π protein induction negatively affected T7 RNAP-driven sfGFP in the first plate generation, and appears to result in a higher effective differentiation rate for a given level of integrase induction. When differentiation is induced at a high level in media containing chloramphenicol which selects for the R6K plasmid, a decrease in endpoint cell density is observed as cells are repeatedly diluted for batch growths (Figure 3.9). This affected was enhanced for 2x differentiation at lower induction levels of differentiation, particularly for higher burden expression. While we did not follow up on the potential mechanisms underpinning these effects, these experiments informed the choice of 10 nM Las-AHL for all subsequent experiments.

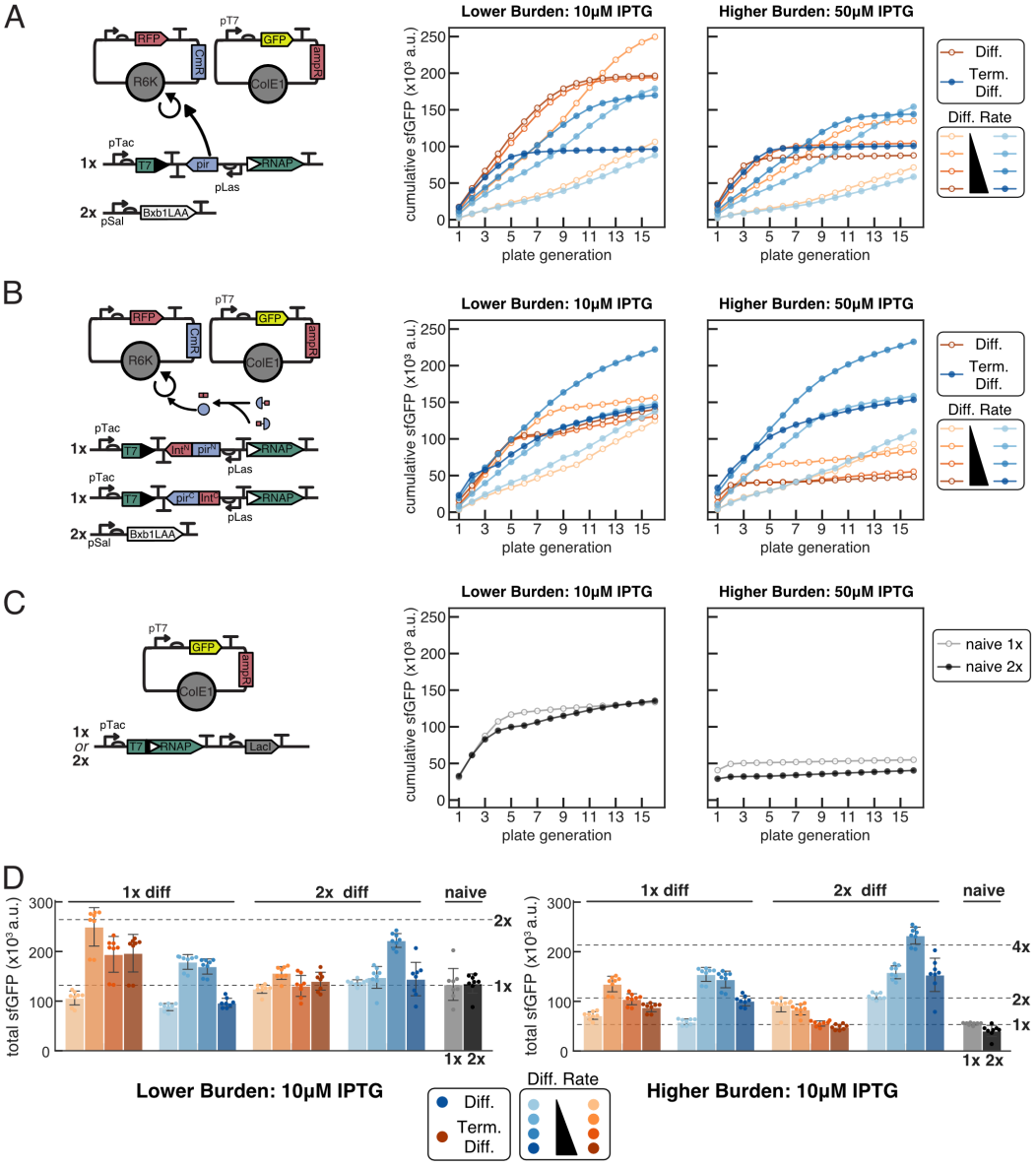
A
B
C
D
pT7
RFP
GFP
R6K
ColE1
CamR
AmpR
pTac
T7
pir
RNAP
pLas
pSal
Bxb1LAA
IntN
pirC
IntC
LacI
1x
2x
1x or 2x
Lower Burden: 10μM IPTG
Higher Burden: 50μM IPTG
cumulative sfGFP (x10³ a.u.)
plate generation
Diff.
Term. Diff.
Diff. Rate
naive 1x
naive 2x
total sfGFP (x10³ a.u.)
1x diff
2x diff
naive
Lower Burden: 10μM IPTG
Higher Burden: 10μM IPTG

Figure 3.3 *(previous page)*: Assessing the evolutionary stability of burdensome T7 RNAP driven expression from a high copy ColE1 AmpR plasmid. (A-C) 8 independent transformants were outgrown for 8 hours in LB media with appropriate antibiotics and inducers before 50x dilution into experimental conditions. Cells were grown in 96 well plates in 0.3 mL, diluted 50x every 8 h for 16 total growths, and 50 μL endpoint samples taken to measure OD700, sfGFP (485/515 nm), and mScarlet (565/595 nM) fluorescence in 384 well matriplates. Average cumulative sfGFP production plotted for each condition. (A-B) 1x differentiation and 2x differentiation. Each differentiation cassette additionally encodes NahR^{AM}, LasR^{AM}, and LacI^{AM} (Figure 3.27 for full circuit diagram). Cells were co-transformed with Cole1 AmpR pT7 GFP and R6KCmR-mScarlet, and plated on LB +carb/chlor/30 nM Las-AHL. Colonies were outgrown in LB +carb/chlor/10 nM Las-AHL before 50x dilution into experimental conditions in LB carb with chlor (blue, filled circles or without chlor (orange, open circles) with varying concentrations of salicylate (10, 15, 20, 30 μM) and IPTG (10, 50 μM). (C) Cells with one (naive 1x) or two (naive 2x) copies of genomically integrated T7 RNAP were transformed with Cole1 AmpR pT7 GFP, plated on LB + carb, and outgrown in LB + carb before dilution into experimental conditions. (D) Total cumulative production +/- SD after 16 growths for all strains in all conditions.

In both 1x and 2x differentiation circuits, regardless of the presence of chloramphenicol, lower differentiation rates allowed longer-lasting GFP production at a lower rate, high rates led to a more rapid cessation of production, and intermediate rates struck a balance (Fig 3.3A-B). With 1X differentiation, chloramphenicol selection (terminal differentiation) negatively affected total production in the lower burden case (+chlor: 178941 +/- 15858 a.u. vs. -chlor: 249601 +/- 41448 a.u., 10 μM IPTG/15 μM sal), but marginally improved it in the higher burden case (+chlor: 154290 +/- 15067 a.u. vs. -chlor: 134953 +/- 16694, 50 μM IPTG/15 μM sal). With 2x split-pir differentiation, while the benefit of chloramphenicol selection was somewhat apparent with lower burden (+chlor/10 μM IPTG/20 μM sal: 222140 +/- 14846 a.u. vs. -chlor/10 μM IPTG/15 μM sal: 156590 +/- 13646 a.u.), this was much greater in the higher burden case (+chlor/50 μM IPTG/20 μM sal: 232596 +/- 18047, vs. -chlor/50 μM IPTG/10 μM sal: 92770 +/- 14940 a.u.).

Comparing the performance of the differentiation circuits to that of 1x and 2x naive inducible T7 RNAP demonstrates that differentiation and terminal differentiation can outperform naive expression in both the lower and higher burden case, but that this benefit is exaggerated with higher burden. Specifically, the best performing circuit in the lower burden case (10 μM IPTG) in terms of total production achieved was 1x

differentiation without chlor selection, yielding ~1.9x (15 μM sal) that of 1x naive with the same IPTG induction. In the higher burden case (50 μM IPTG), the best performing circuit was 2x split-pir differentiation with chloramphenicol selection, yielding ~4.2x (20 μM sal) the total production of 1x naive with the same IPTG induction. Including chloramphenicol for the case of terminal differentiation has two opposing effects on the amount of production that will be achieved over the lifetime of circuit function. The first effect is negative, as limiting the growth of differentiated producer cells will limit the total amount of production achieved by any individual differentiated cell lineage. The second effect is positive, as this selection limits the growth of differentiated cells which have mutated or otherwise decreased or ceased expression of the function of interest, preventing the takeover of mutated cells and extending the duration of circuit function. Because these effects are opposing, one may dominate the other depending on the characteristic parameters. In the case of 1x differentiation with low burden, the former affect clearly dominates, and terminal differentiation under-performs. With higher burden, the later effect becomes more important and terminal differentiation marginally surpasses differentiation without selection.

The case of 2x differentiation is somewhat more complicated, both because two mutations would be required to yield mutated progenitor cells incapable of differentiating, and because an individual cell could have activated one or both copies of T7 RNAP. Without chloramphenicol selection, cells have more opportunity to activate both copies of T7 RNAP and therefore increase the amount of burden associated with production. The combination of these factors – requiring two mutations to yield mutated progenitors and having higher effective burden in a larger subset of differentiated cells when selection is absent – favor the performance of terminal differentiation, and particularly so when the burden of the function is high.

Intriguingly, when we compare the performance of 1x and 2x naive inducible expression, this redundancy did not provide any benefit. If we just consider genomic sources of mutational inactivation of function, we would expect a benefit as two sequential mutations would be required to inactivate the expression of T7 RNAP. Given that we did not observed this, and that the overwhelmingly majority of the burden associated with T7 RNAP expression will come not from its own expression, but from the transcription and translation of GFP, we considered plasmid associated mechanisms for inactivating function. While this could certainly be caused by mutation, we noted that plasmid loss would also accomplish this and can occur at

a much higher rate [36, 37]. As well, though we are selecting for the presence of the plasmid with carbenicillin, this choice of antibiotic resistance is problematic. As the mechanism of ampicillin/carbenicillin resistance is degradation of the antibiotic, resistance can be shared, and indeed sensitive cells can co-exist with resistant cells in the presence of antibiotic, with the resistant fraction scaling with antibiotic concentration [38].

In order to test whether this mechanism of plasmid loss and shared antibiotic resistance was preventing the benefit of a second cassette of T7 RNAP from being observed, we first investigated whether it could be occurring in our system. Both to test this mechanism and to select any alternative selectable marker to use for the ColE1 plasmid, we constructed an identical plasmid to ColE1 AmpR pT7 GFP which instead had kanamycin resistance (Cole1 KanR pT7 GFP), and performed co-culture experiments of naive 1x and the parental strain JS006 transformed with these plasmids (Figure 3.4). As only naive 1x cells produce appreciable levels of GFP from these plasmids, we could observe that cells with only AmpR could allow cells with only KanR to grow in LB with both kan and carb, while cells with only KanR did not allow cells with only AmpR to grow in the same condition. Additionally, spent media from AmpR cells could allow KanR cells to grow in LB with kan and carb, while the opposite was not true. We therefore concluded that the choice of AmpR on the ColE1 plasmid was likely allowing loss of expression through plasmid loss and shared antibiotic resistance, while the same mechanism would not hold with KanR (Figure 3.4). While in their study Yurtsev et.al. identified several experimental factors – antibiotic concentration, dilution factor between batch growths, and the initial ratio of sensitive and resistant cells – which affect the dynamics and steady state population distribution of antibiotic resistance plasmids, they also noted that increased burden associated with the plasmid would increase the fraction of antibiotic sensitive cheater cells. Because the plasmid of interest in our system encodes a potentially very burdensome function that is inducible, we would expect the growth of cheater cells to be favored increasingly so as the burden is increased.

In addition to experimentally investigating this, we wished to investigate this phenomenon computationally for all our differentiation and naive strains. We produced models for each circuit, and ran stochastic simulations which approximated our experimental design with a range of burden and differentiation rates. We then compared the results of models when non-genomic sources of mutation or loss of

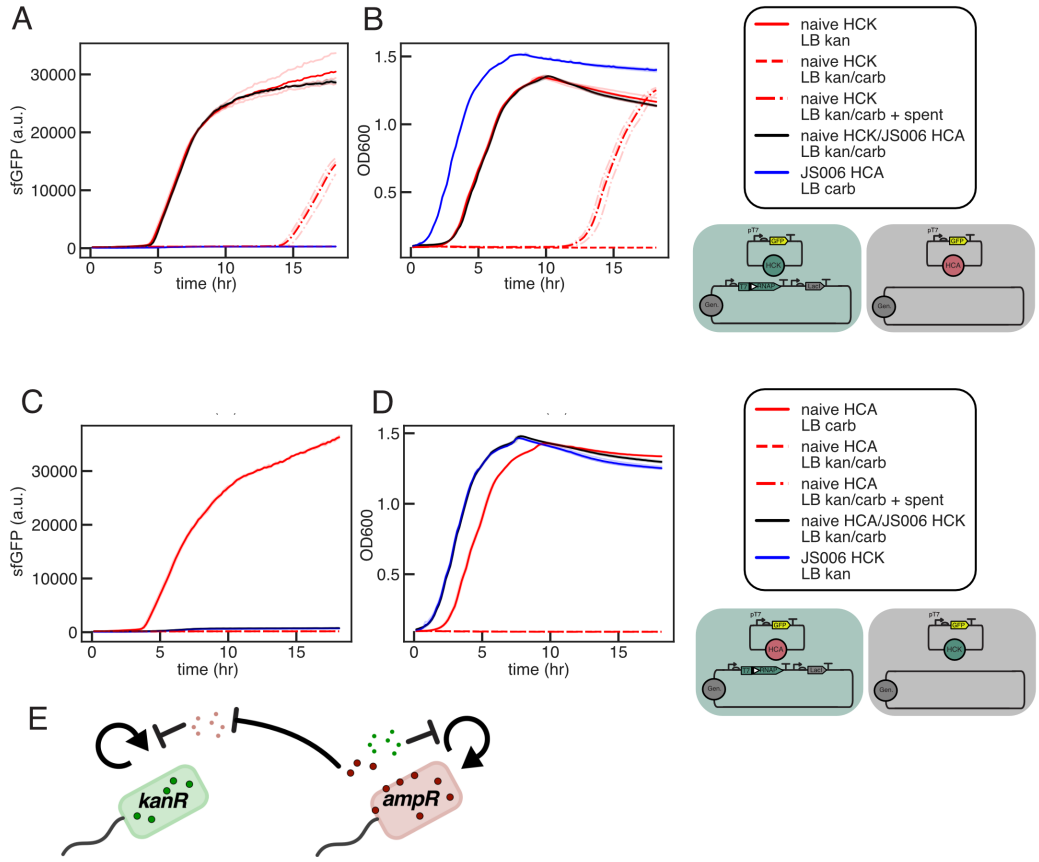

Figure 3.4: Assessing communal antibiotic resistance for carbenicillin and kanamycin. JS006 and naive 1x cells were transformed with HCA pT7 GFP or HCK pT7 GFP, and colonies outgrown for 8 h in LB with kan (50μg/mL) or carb (100μg/mL). Spent media was harvested by pelleting JS006/HCA pT7 GFP or JS006/HCK pT7 GFP and filtering supernatant (0.2μM). Cultures were diluted 1:50 into 0.2mL experimental monocultures or co-cultures in triplicate in 96 well plates, and OD600 and sfGFP (485/515nM) monitored over 18 h growth. (A-B) naive 1x + HCK pT7 GFP and JS006 + HCA pT7 GFP grown in mono- and co-culture. JS006 alone (blue) grows (B) but does not produce GFP (A). naive + HCK grows and produces GFP in LB Kan, in co-culture with JS006 HCA in LB carb/kan, and with delay in LB carb/kan/10 percent JS006 HCA spent media, but not in LB kan/carb. (C-D) naive 1x + HCA pT7 GFP and JS006 + HCK pT7 GFP grown in mono- and co-culture. (E) Schematic of shared antibiotic resistance. AmpR cells degrade carb inside the cell, in the periplasm, as well as in the medium, allowing amp sensitive cells to grow. Antibiotic resistance from KanR cells however is not shared significantly enough to allow growth of sensitive cells.

function were neglected, to models which incorporated stochastic plasmid loss and antibiotic resistance degradation; as well as to models which considered plasmid mutation (see Model Implementation for detailed description). In simulating cells with one or two copies of inducible T7 RNAP in both the naive and differentiation cases, assumptions were required about the relative burden levels and production rates. In the case of differentiation, the producer growth rate was the growth rate of a cell with one activated cassette of T7 RNAP, and for consistent rule implementation which would result in positive growth rates, the burden of the second copy produced a proportionate decrease in growth rate. For example, if a non-producer grows at rate 1 and a cell with one cassette active grows at rate 0.5, a cell with two cassettes active would grow at rate 0.25 ($\mu_N(\mu_P/\mu_N)^2$). Production then was assumed to increase linearly with the decrease in growth rate.

From these simulations, we recapitulate several observations from the initial deterministic modeling of the general strategies, and from our experiments. For both differentiation and terminal differentiation with one and two copies, we see that lower differentiation rates allows longer duration production at a lower rate, high differentiation rates lead to faster loss of production particularly for higher burden, and intermediate rates strike a balance and achieve the most total production (Figure 3.17, 3.21-3.26). At low burden (higher μ_P), terminal differentiation performs worse, but better as the burden increases (Figure 3.17). We also see that naive expression performs comparatively well at low burden relative to high burden. For 2x naive expression we model both the case where one cassette alone yields the growth rate μ_P and its corresponding production rate (2x*), and where the two cassettes together yield the growth rate μ_P (2x). As expected, at low burden and high differentiation rate, 1x differentiation without selection approximates the performance of both 1x and 2x naive, and 2x differentiation without selection approximates the performance of 2x* (Figure 3.17). While we do not address experimentally the affect of number of divisions post-differentiation, we do so computationally, and see that while increasing the number of divisions benefits total production achieved with low burden, this benefit subsides with increased burden (Figure 3.18).

Further, the redundancy and mutational robustness provided with 2x differentiation improves performance relative to the one cassette case for both differentiation and terminal differentiation. However, while computationally this benefit holds in these pairwise comparisons across all burdens for both differentiation and terminal differentiation, experimentally we observe 1x differentiation outperforming 2x

differentiation with both IPTG inductions when chloramphenicol is not included. Though this disparity could in part be due to our assumptions regarding growth rate and production rate in cells with 1 vs 2 activate T7 RNAP cassettes, we reasoned this could also be due to the plasmid-based mechanisms previously discussed. Indeed, we see that incorporating stochastic plasmid loss in conjunction with a model of antibiotic degradation and growth inhibition can negatively effect performance of the naive and differentiation architectures in a burden dependent manner, and that this effect is much greater in the two cassette case (Figure 3.17). Intuitively this makes sense, as even though the rate of plasmid loss is faster than of mutation, the difference in amount of time required to generate (1) a cell with no plasmid and (2) a cell with all copies of T7 RNAP mutated is much greater in the two cassette case. Therefor we would expect incorporating plasmid loss to more significantly impact the performance of redundant architectures.

Strikingly, this effect does not affect the performance of terminal differentiation architectures. We see this both in the total production achieved (Figure 3.17), and in tracking the population of cells which have lost the plasmid (Figure 3.21-3.26). When there is no antibiotic degradation, we see no accumulation of cells lacking the plasmid for any circuit, but with a sufficiently high rate of antibiotic degradation, we see a transitory rise in the fraction of cells that have lost the plasmid for naive and differentiation architectures, but not for terminal differentiation (Figure 3.19-3.26). As these cells require cells that have retained the plasmid to degrade antibiotic, they do not completely take over the population, but instead are eventually displaced by mutated cells which retain the plasmid. For differentiation, this effect is only observed if both the burden and differentiation rate is sufficiently high, and is never seen with terminal differentiation.

Similarly, we can instead neglect antibiotic degradation, and just consider plasmid mutations. Though this would be more accurately modeled by explicitly considering plasmid copy number, mutation of individual plasmids, stochastic partitioning, and the dependency of burden and production on the fraction of mutated plasmids, this would yield a prohibitively complex model, so we instead consider loss of function from a plasmid mutation to occur with a single event. The results of simulations which incorporate plasmid mutation largely align with the previously results considering antibiotic degradation and plasmid loss (Figure 3.28). Specifically, we see that incorporation of plasmid mutation negatively affects the performance of naive and differentiation circuits with two cassettes. As the mutation rates we consider

are equal to or slower than the burden mutation rate, the one cassette circuits are relatively unaffected. As well, instead of the transitory increase we observed with plasmid loss, plasmid mutations become fixed (Figure 3.19-3.26). Critically, the robustness of terminal differentiation circuits holds true when considering plasmid mutations, and we observe no accumulation of cells with mutated plasmids or any effect on production.

With knowledge that effects due to plasmid loss and shared antibiotic resistance should be much smaller of a factor with the KanR version of the ColE1 plasmid, we again compared the performance of our naive, differentiation, and terminal differentiation circuits (Figure 3.5). In contrast to the AmpR case, with KanR we now observed a benefit from including a second cassette in the naive case, with 2x naive yielding ~2.2x the production of 1x naive in the lower burden case, and ~1.6x in the higher burden case (Figure 3.5C). While 2x naive outperformed all differentiation circuits in the low burden case, this was not so with high burden, where 2x terminal differentiation achieved ~2.9x (20μM sal/50 μM IPTG/+chlor, Figure 3.5D). In agreement with our modeling where we observed a larger negative affect from incorporation of plasmid loss and antibiotic degradation with the two cassette naive and differentiation circuits, we see an improvement of 2x differentiation and 2x naive relative to their respective single-cassette counterparts when switching from AmpR to KanR. If the dominant factor were plasmid mutation instead of plasmid loss, we would not have expected this change in antibiotic selection to have the observed affect. Strikingly, terminal differentiation circuits were not benefited by reducing plasmid based effects by switching from AmpR to KanR, highlighting the robustness of this architecture to non-genomic factors affecting burdensome expression.

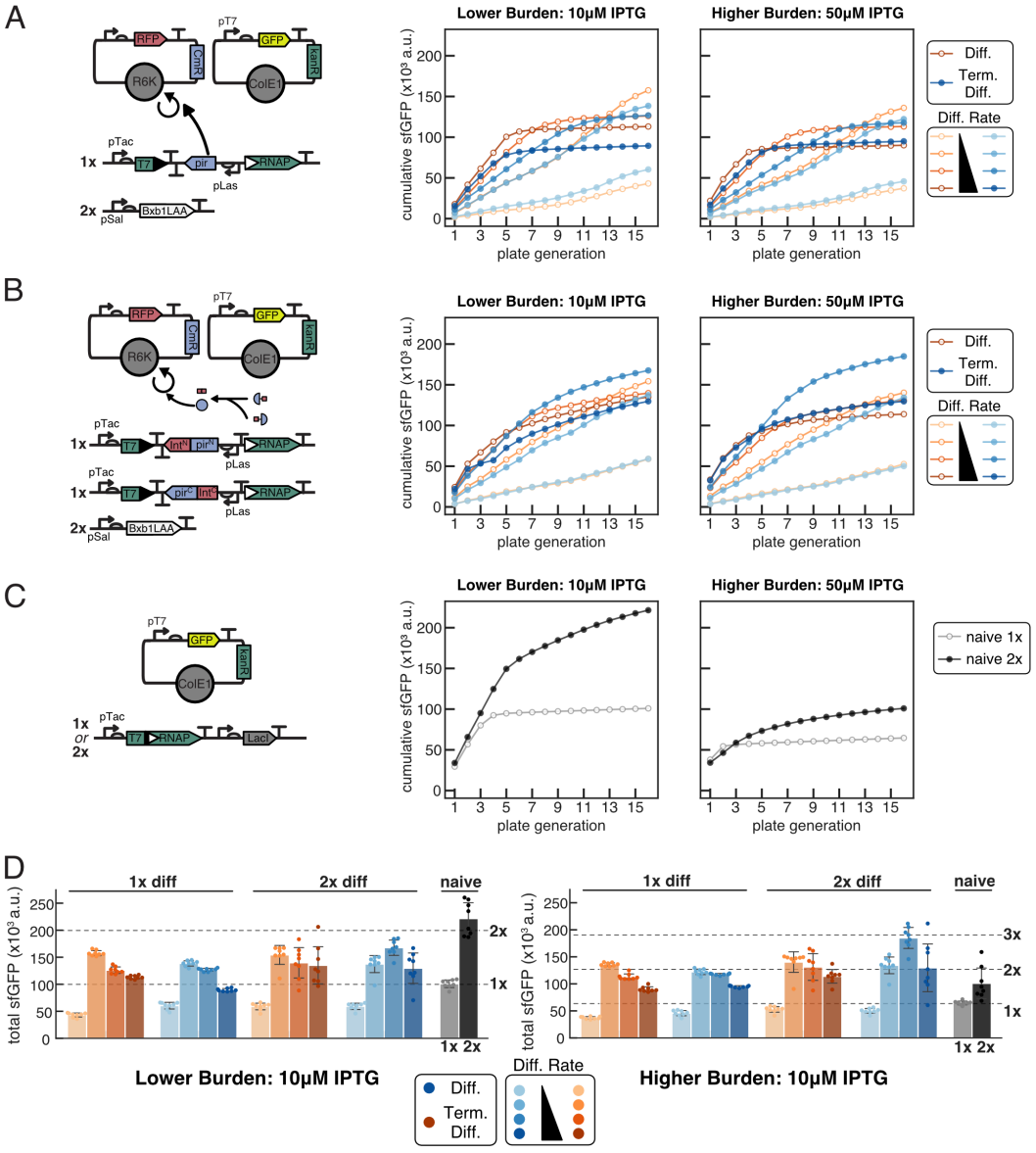
A
B
C
D
pT7
RFP
GFP
CmR
KanR
R6K
ColE1
pTac
1x
2x
T7
pir
RNAP
pLas
pSal
Bxb1LAA
Int^N
pir^N
pir^C
Int^C
LacI
or
Lower Burden: 10μM IPTG
Higher Burden: 50μM IPTG
cumulative sfGFP (x10³ a.u.)
plate generation
Diff.
Term. Diff.
Diff. Rate
naive 1x
naive 2x
1x diff
2x diff
naive
total sfGFP (x10³ a.u.)
3x

Figure 3.5 *(previous page)*: Assessing the evolutionary stability of burdensome T7 RNAP driven expression from a high copy ColE1 KanR plasmid. (A-C) 8 independent transformants were outgrown for 8 hours in LB media with appropriate antibiotics and inducers before 50x dilution into experimental conditions. Cells were grown in 96 well plates in 0.3 mL, diluted 50x every 8 h for 16 total growths, and 50 μL endpoint samples taken to measure OD700, sfGFP (485/515 nm), and mScarlet (565/595 nM) fluorescence in 384 well matriplates. Average cumulative sfGFP production plotted for each condition. (A-B) 1x differentiation and 2x differentiation. Each differentiation cassette additionally encodes NahR^{AM}, LasR^{AM}, and LacI^{AM} (Figure 3.27 for full circuit diagram). Cells were co-transformed with ColE1 KanR pT7 GFP and R6KCmR-mScarlet, and plated on LB +kan/chlor/30 nM Las-AHL. Colonies were outgrown in LB +kan/chlor/10 nM Las-AHL before 50x dilution into experimental conditions in LB kan with chlor (blue, filled circles or without chlor (orange, open circles) with varying concentrations of salicylate (10, 15, 20, 30 μM) and IPTG (10, 50 μM). (C) Cells with one (naive 1x) or two (naive 2x) copies of genomically integrated T7 RNAP were transformed with ColE1 KanR pT7 GFP, plated on LB + kan, and outgrown in LB + kan before dilution into experimental conditions. (D) Total cumulative production +/- SD after 16 growths for all strains in all conditions.

3.5 Differentiation enables expression of toxic functions

Intuitively and from modeling, the degree of burden or toxicity of the function of interest should not matter in allowing a differentiation architecture to improve its evolutionary stability if the function is maintained in a perfectly off state in the progenitor cells. Indeed, even a function with 99 percent burden should be maintained (Figure 3.17). To test this, we aimed to demonstrate that the differentiation circuit we developed could allow the production of a protein that will result in cell death: dnaseI. As progenitor cells do not produce any T7 RNAP, we reasoned that a T7 RNAP-driven dnaseI would not be expressed in the progenitor cells, allowing the encoded function to be replicated without toxicity or selective pressure for mutations. However, construction of a dnaseI expression plasmid identical to that of sfGFP yielded only mutated plasmids. Characterization of leaky expression from the pT7-GFP plasmid in the absence of T7 RNAP revealed fluorescence above background, explaining this inability to isolate functional plasmids (Figure 3.15). Incorporating two insulating terminators upstream of the T7 promoter mitigated leaky expression in the absence of T7 RNAP (Figure 3.15), and this insulation in conjunction with reducing the RBS strength allowed construction and isolation of a correctly sequenced dnaseI expression construct. While leak could have also been reduced by using the T7/lacO promoter and an additional source of LacI on the

expression plasmid, this would not eliminate leaky expression in progenitor cells upon induction of differentiation and T7 RNAP. Highlighting the importance of preventing leaky expression of toxic functions, transformation of 1x differentiation and 2x differentiation cells with insulated dnaseI plasmid yielded ~600 cfu and ~1000 cfu respectively, while naive 1x and naive 2x strains yielded 1 and 0 colonies respectively, compared to $> 10^4$ cfu for both when transformed with Cole1 AmpR pT7 GFP control (Figure 3.6A, Table 3.1).

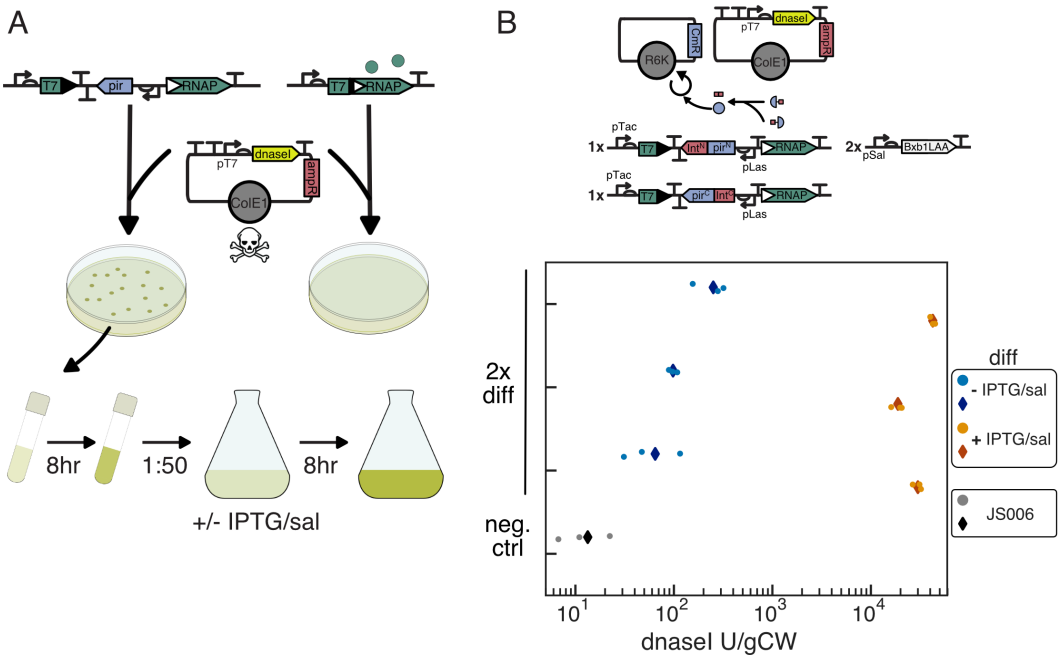

Figure 3.6: Production of functional dnaseI using integrase mediated differentiation. (A) Non-leaky T7 RNAP expression enables differentiation strains to replicate a plasmid encoding an insulated pT7 dnaseI cassette, while cells with leaky T7 RNAP cannot. (B) For assaying dnaseI production, 2x differentiation cells were co-transformed with R6KCmR-empty and Cole1 AmpR T13m T12m-pT7-B0032-dnaseI-T7T, and plated on LB + carb/chlor/30 nM Las-AHL. After 8 h outgrowth in LB + carb/chlor/10 nM Las-AHL, cultures were diluted 1:50 into 25 mL media +/- 10 μM IPTG/20 μM sal. Pellets for experimental cultures and JS006 negative control were harvested after 8 h, and lysate assayed for dnaseI activity (Figure 3.16). Lysates from three independent cultures for each condition, and one JS006 culture were assayed in triplicate.

To assess the capacity of differentiation to enable functional dnaseI expression, we co-transformed the 2x differentiation strain with the insulated dnaseI expression plasmid, and an empty R6KCmR plasmid. After outgrowth without induction, cultures were diluted into 25 mL cultures with or without induction with 20 μM salicylate and 10 μM IPTG. Afer 8h growth, un-induced cultures grew to cell

densities equivalent to JS006 control (7.2-7.6 g wet weight per liter vs. 7.6 ctrl), while induced cultures reached lower cell densities (2-4.4 gCW/L). Dilution 1:50 into fresh media yielded similar densities for the uninduced cultures after 8 h growth (7.2-7.6 gCW/L), while induced cultures had grown minimally after 8 h (OD600 < 0.05) and to a range of densities after 16 h total growth (2-6.8 gCW/L). The growth of the induced and uninduced cultures indicated that the dnaseI plasmid minimally affected growth when T7 RNAP is not expressed, and that 20 μM salicylate induction likely resulted in more complete differentiation in a large shaking culture in comparison to small volumes in 96 well microplates in previous experiments. Expression of functional dnaseI was quantified with an activity-based assay on lysate extracted from cell pellets (Fig 3.11), with activity measured equivalent to ~1.9-4.2$x10^4$ U/gCW (~3.7x10^4 – 1.9$x10^5$ U/L) in the three independent induced cultures and ~65-250 U/gCW (~500-1800 U/L) in the uninduced cultures, compared to ~13 U/gCW (~100 U/L) for the JS006 negative control (Fig 3.6B). This yield of dnaseI is on order with yields reported using T7 RNAP to drive the expression of recombinant dnaseI using the LacI repressible T7 promoter in Bl21(DE3)[pLysS] (1.5$x10^4$ U/L) and Bl21(DE3)[pLysE] (7.5$x10^4$ U/L) [39].

3.6 Discussion

Here we have developed architectures for implementing differentiation and terminal differentiation in *E. coli* for the expression of burdensome T7 RNAP-driven functions. Importantly, in our circuit design progenitor cells do not have an intact coding sequence for T7 RNAP, completely eliminating leaky T7 RNAP expression in progenitor cells. We computationally demonstrated that differentiation enables growth in a chemostat with constant dilution at a higher rate than would be possible than for naive expression of the same function with equivalent burden, which has important implications both for continuous bioproduction and deploying engineered bacteria in the gut microbiome or other environments with competition. Importantly, we demonstrated that limiting the growth of differentiated cells with terminal differentiation prevents mutations which inactivate the function of interest (burden mutations) from providing a selective advantage, removing this category of mutations as a cause of circuit failure. Finally we also demonstrated that reducing the rate of mutations which disrupt the process of differentiation improves the evolutionary stability of the terminal differentiation architecture regardless of the rate of burden mutations, and that completely eliminating this mutation would ensure indefinite circuit function.

Experimentally, we developed a differentiation-activated T7 RNAP architecture in which all circuit components that could mutate to disrupt the process of differentiation or expression of T7 RNAP were integrated on the genome, both ensuring exact copy number control and preventing plasmid partitioning effects from accelerating circuit failure [11]. With the goal of reducing the probability of disrupting the differentiation process through mutation and thereby increasing the longevity of the terminal differentiation architecture, we developed a novel split-π protein system through semi-rational splitting of the pir coding sequence and tagging the fragments with split-inteins. In long-term experiments with repeated dilutions of independent cultures, we compared the performance of 1x and 2x naive T7 RNAP-driven expression to 1x differentiation and 2x differentiation with or without selection against differentiated cells with chloramphenicol. We demonstrated that the rate, duration, and total amount of production could be tuned by varying the differentiation rate, with low differentiation rates enabling longer duration expression at a lower rate. Differentiation was particularly beneficial in comparison to naive expression with higher burden as expected from modeling, and the redundancy and robustness to differentiation mutations provide with the split-π protein terminal differentiation proved effective.

Further, we demonstrated both computationally and experimentally that effects due to instability of the ColE1 pT7 GFP expression plasmid and communal antibiotic resistance can negatively effect the performance of naive expression and differentiation without limited cell division, but that terminal differentiation circuits were immune to this effect. We further demonstrated computationally that the robustness of terminal differentiation circuits to burden mutations affecting T7 RNAP expression and to plasmid loss extend to the general case of plasmid mutations which disrupt the function of interest. Though genomic integration of functions is more time consuming and cumbersome than plasmid transformation, plasmid instability and considerations of cost for use of antibiotics and inducers in large cultures have often made genomic integration of constitutively expressed functions the preferred method for bioproduction in industry [10, 40]. However, because a system such as this with terminal differentiation where effects of plasmid instability can be substantially or entirely mitigated, we can potentially get the stability benefits of genomic integration with the ease of plasmid transformation.

Finally, because there should be no limit on the degree of burden or toxicity of a function expressed with our differentiation system (so long as the toxicity is limited

to the cells expressing the function), as a proof of concept we demonstrated that differentiation could enable the expression of functional dnaseI. In the course of this demonstration, we discovered that in the absence of leaky expression of T7 RNAP, non-T7 RNAP sources of leak could prevent isolation of correctly assembled dnaseI expression plasmids. While we mitigated this problem through the incorporation of insulating terminators to prevent transcriptional read-through from upstream of the T7 promoter, reducing the strength of the RBS was still required to isolate correctly sequenced plasmid. Improving this insulation and/or reducing any leaky expression that may be coming directly from the T7 promoter through directed evolution efforts may prove beneficial. As an additional benefit this would allow the use of higher strength RBS sequences without concern for leak, thereby enabling improved yields.

While the expression of toxic or highly burdensome products has long been of interest in bioproduction and synthetic biology, and effective strategies have been implemented to accomplish this [41], to our knowledge all existing strategies only work for single use batch culture inductions. The critical difference with our strategy of terminal differentiation is that progenitor cells continuously differentiate to replenish the population of cells expressing the toxic function, thereby allowing a toxic product to be produced continuously.

Though here we demonstrate the effectiveness of terminal differentiation and the stability benefit from requiring two instead of one mutation to break the differentiation activated delayed death architecture, this can be viewed as a proof of concept for the power that redundancy can provide in synthetic biology. In considering the scaling of this strategy to longer times and larger population sizes, we cannot help but imagine what can be done with a higher degree of redundancy. Scaling this specific architecture through further splitting of the π protein may be infeasible, but the tool-kit of synthetic biology certainly has means of allowing this strategy to scale further, through the inactivation of essential genes, activation of toxins, or otherwise.

3.7 Model implementation

Cell growth

For all simulations, we implemented carrying capacity limited growth of the form

$$\frac{dX_i}{dt} = \mu_i X_i \frac{K - X_{tot}}{K}, \tag{3.1}$$

where μ_i is the specific growth rate of cell type X_i, K is the carrying capacity (number of cells), and X_{tot} is the current total number of cells of all cell types. For simulations in Fig. 3.1, we modeled carrying capacity limited growth in a chemostat with constant dilution rate D (h^{-1}):

$$\frac{dX_i}{dt} = \mu_i X_i \frac{K - X_{tot}}{K} - D X_i. \tag{3.2}$$

Burden, differentiation, and integrase mutations

In order to generate the models for these simulations, we first generated all possible genotypes that could be present in the simulation. For the naive case, each genomically integrated cassette of inducible T7 RNAP can be in two states:

1. Producer (P): Cassette will enable production and have associated burden.

2. Non-producer (N): Cassette will not produce T7 RNAP and has no associated burden.

For 1x naive, there are only two genotypes (P, N), while for 2x naive, there are three genotypes (PP, PN, NN). Though we could explicitly model both PN and NP, this is unnecessary. For the naive case, there is only a single type of mutation, the burden mutation, which occurs at rate k_{M_B} (h^{-1}). Because mutations require DNA replication/cell division, this rate is further dependent on the current growth rate:

$$X_P \xrightarrow{k_{M_B}\mu_P \frac{K - X_{tot}}{K}} X_N \tag{3.3}$$

As the rate of the mutation is also proportional to the number of loci that could be mutated, for the general case of n cassettes we have that

$$X(P = i, N = n - i) \xrightarrow{i k_{M_B}\mu_P \frac{K - X_{tot}}{K}} X(P = i - 1, N = n - i + 1), \tag{3.4}$$

where i is the number of producer cassettes, and $n - i$ is the number of non-producer cassettes.

Differentiation brings two additional types of mutations, both of which act to disrupt the process of differentiation.

1. Differentiation mutation: Occurs at rate k_{M_D}, and disrupts the capacity for an individual cassette to undergo recombination.

2. Integrase mutation: Occurs at rate k_{M_I}, and disrupts integrase expression at one locus.

Before consideration of the integrase mutation, we can enumerate the possible genotypes for a single cassette:

1. PP : (P)rogenitor, (P)roducer. Cassette is in the un-recombined progenitor state, and would yield a producer cassette upon recombination.

2. PN : (P)rogenitor, (N)on-producer. Cassette is in the un-recombined progenitor state, and would yield a non-producer cassette upon recombination.

3. DP : (D)ifferentiated, (P)roducer. Cassette is in the recombined differentiated state, and is producing T7 RNAP.

4. DN : (D)ifferentiated, (N)on-producer. Cassette is in the recombined differentiated state, and is not producing T7 RNAP due to the burden mutation.

5. M- : (M)utated. Cassette is in the un-recombined progenitor state, but has incurred a differentiation mutation which prevents it from recombining. The Producer/Non-producer state is therefore not relevant and is neglected.

As each integrase cassette is either functional or not-functional in our model, we simply consider the number of functional integrase cassettes in our genotype. For example, the starting genotype for cells with two differentiation cassettes and two integrase cassettes would be (PPPP2). Mutation of one integrase cassette would then yield (PPPP1).

As with the burden mutation considered in the naive case, the rates of the burden and integrase mutations scale with the number of loci that could mutate, and depend on the current growth rate of the cells which could incur the mutation.

Terminal differentiation

In order to address the case of terminal differentiation, we explicitly modeled the number of cells divisions a differentiated cell could undergo. To illustrate this, we consider the case of a cell with two differentiation cassettes and two integrase cassettes (PPPP2). For differentiation without limited division, this would yield the genotype DPPP2. With terminal division, we include the subscript to indicate the

number of divisions the cell has undergone: ($DPPP2_0$). With the general case where there is a limit of n cell divisions, for $i < n$ we have

$$X_{DPPP2_i} \xrightarrow{\mu_P \frac{K - X_{tot}}{K}} 2X_{DPPP2_{i+1}}. \tag{3.5}$$

For $i = n$ we have

$$X_{DPPP2_i} \xrightarrow{\mu_P \frac{K - X_{tot}}{K}} \Phi. \tag{3.6}$$

Production and burden

In addition to tracking the population of producer cells, we also can track the production of an arbitrary product. To do so, we determine the production rate β ($\text{cell}^{-1}\text{h}^{-1}$) specific to each strain in the simulation, and assume further that production is also dependent on the current growth rate. For simplicity, we set $\beta = 1$. In the naive case, any genotype that has $P > 0$, and in the differentiation case, any genotype that has $DP > 0$, will be a producer. While in the single cassette case this is quite simple, for $n > 1$ cassettes we must address how to deal with the production rate for $P = 0, 1, ..., n$ for the naive case, and $DP = 0, 1, ..., n$ for the differentiation case. To do so, we will address production rate and burden simultaneously.

In the case of differentiation, we assign μ_N to be the specific growth rate of non-producer cells (DP=0), μ_P to be the growth rate of cells with a single cassette producing T7 RNAP (DP=1), and β to be the production rate for this genotype. Because growth rates must be non-negative, and cells have an inherent metabolic capacity, we assume each additional cassette will affect the growth rate proportionally according to

$$\mu_i = \mu_N \left(\frac{\mu_P}{\mu_N}\right)^{DP(i)} = \frac{\mu_P^{DP(i)}}{\mu_N^{DP(i)-1}}. \tag{3.7}$$

We similarly assume that the production rate does not linearly increase with the number of producer cassettes, but negatively correlates linearly with the growth rate according to

$$\beta_i = \beta \frac{\mu_N - \mu_i}{\mu_N - \mu_P}. \tag{3.8}$$

For $\mu_N = 1$ and $\mu_P = 0.5$ and two cassettes, production and growth rates would be

DP_i	μ_i	β_i
0	1	0
1	0.5	1
2	0.25	1.5

The naive case is treated identically with the number of P cassettes considered instead of DP with differentiation. However, because the initial genotype of cells with n cassettes has n producer cassettes, we must decide whether μ_P and β describe cells with $P = 1$ or $P = n$. We consider both cases in our stochastic simulations, where for the two cassette case "2x" indicates that μ_P and β describe the genotypes where $P = 2$, and "2x*" indicates these parameters describe the case of $P = 1$. For the later case, growth rates and production rates are determined as above. For the former case, the growth rate is calculated similarly by

$$\mu_i = \mu_N \left(\frac{\mu_P}{\mu_N} \right)^{DP(i)/n}, \tag{3.9}$$

where n is the total number of cassettes. The production rates are then determined as above.

Differentiation rates

We model the process of differentiation at a high level, and do not explicitly considering the underlying transfer functions describing the production of Bxb1 integrase or the rate of recombination given the number of intact and recombined cassettes. A mechanistic model would be difficult to implement in the context of these simulations, may prove prohibitively complex to implement with the relevant scales of population size, and is not necessary to capture the important features. Instead we set the tunable rate of differentiation $k_{\text{diff}}(\text{h}^{-1})$ that is constant throughout a given simulation. This rate is the maximum total rate of differentiation when all differentiation cassettes are in the progenitor state, and all integrase cassettes are functional. Reducing the amount of integrase expression by mutation of an integrase cassette will reduce the rate of differentiation, and we assume differentiation rate

varies linearly with the number of functional integrase cassettes. We also assume that the effective differentiation rate is not affected by growth rate. While this may not be completely accurate, decreased protein production rates when cells are growing more slowly will be counteracted by decreased dilution from cell growth and division. Therefore for the case of a single differentiation cassette we have

$$X_i \xrightarrow{k_{\text{diff}} \frac{I_i}{n_I}} X_j. \tag{3.10}$$

Where I_i is the number of functional integrase cassettes for genotype X_i, n_I is the total number of integrase cassettes, $\text{PP}_i = 1, \text{DP}_i = \text{PN}_i = \text{DN}_i = 0, \text{I}_i = 2$, and $\text{DP}_j = 1, \text{PP}_i = \text{PN}_i = \text{DN}_i = 0, \text{I}_j = 2$.

For numbers of cassettes greater than one, the total differentiation rate across all cassettes is k_{diff}, and therefore the differentiation rate for any individual progenitor cassette is k_{diff}/n. Here we implicitly assume that the state of one cassette does not affect the rate of differentiation of any other cassette. Generally, the rate of transitioning from one genotype to another through differentiation is given by

$$X_i \xrightarrow{k_{\text{diff}} \frac{PP_i}{n} \frac{I_i}{n_I}} X_j, \tag{3.11}$$

for the differentiation of a PP cassette, where $\text{PP}_j = \text{PP}_i - 1, \text{DP}_j = \text{DP}_i + 1, \text{PN}_j = \text{PN}_i, \text{DN}_j = \text{DN}_i, \text{I}_j = \text{I}_i$. Similarly, for the differentiation of a PN cassette, we have that

$$X_i \xrightarrow{k_{\text{diff}} \frac{PN_i}{n} \frac{I_i}{n_I}} X_j, \tag{3.12}$$

where $\text{PP}_j = \text{PP}_i, \text{DP}_j = \text{DP}_i, \text{PN}_j = \text{PN}_{i-1}, \text{DN}_j = \text{DN}_{i+1}, \text{I}_j = \text{I}_i$.

Plasmid loss, antibiotic degradation, and growth inhibition

With the above components of the model, we also incorporated features to describe loss of the ColE1 plasmid, degradation of the antibiotic in the medium, and growth inhibition of sensitive cells. We gathered initial parameters from a study investigating the role bacterial cheating in driving population dynamics of plasmids

with cooperative antibiotic resistance [38]. For a population of cells X_R which are resistant and degrade the antibiotic, the concentration of antibiotic is described by

$$\frac{dA}{dt} = -X_R \frac{V_{max}}{V} \frac{A}{A + K_m} + D(A_{in} - A), \tag{3.13}$$

where A (μg/mL) is the concentration of antibiotic in the medium; $V_{\max}$ (μg cell^{-1}h^{-1})is the maximum rate of antibiotic degradation; V (mL) is the volume of the culture; K_m (μg/mL) is the Michaelis constant, the concentration at which the rate of antibiotic degradation is half-maximal; A_{in} (μg/mL) is the concentration of antibiotic in the feed media, and D (h^{-1}) is the dilution rate. For all simulations, we modeled batch dilutions ($D = 0$). The rate of antibiotic degradation $V_{\max}$ as experimentally determined [38] was 10^6 molecules/cell/s (~2.5x10^{-6} μg/cell/h) for antibiotic resistance encoded on a low copy plasmid when ampicillin was used. This rate was used as 1x antibiotic degradation in the simulations. As experimentally the antibiotic resistance was encoded on a high copy ColE1 plasmid, we also examined the case of 5x this rate.

To incorporate plasmid loss in the model, all genotypes and reactions associated with the genotypes were duplicated, and designated as either R for resistant, or S for sensitive. Growth rates and production rates for sensitive cells are determined as previously described for resistant cells, however for sensitive cells, growth rates are set to that of non-producers (μ_N), and the production rate (β) to 0. While the growth rate of resistant cells is not affected by the concentration of antibiotic, the growth rate of sensitive cells was modeled using a Heaviside function:

$$\frac{dX_{i(S)}}{dt} = \mu_{i(S)} X_{i(S)} \frac{K - X_{tot}}{K} \mathcal{H}(MIC - A), \tag{3.14}$$

where MIC is the minimum inhibitory concentration of the antibiotic, and for $A \geq MIC, \mathcal{H}(MIC - A) = 0$; and for $A < MIC, \mathcal{H}(MIC - A) = 1$.

Finally, we modeled loss of the plasmid in the same manner as mutation, occuring at rate k_{PL} (h^{-1}), and again dependent on the current growth rate:

$$X_{i(R)} \xrightarrow{k_{PL}\mu_{i(R)}\frac{K - X_{tot}}{K}} X_{i(S)}, \tag{3.15}$$

where $X_{i(R)}$ and $X_{i(S)}$ have identical genotypes apart from the presence or absence of plasmid. Though copy number of the plasmid certainly varies continuously and is not binary as we model it here, this captures the general feature and is tractable to implement.

Plasmid mutation

After incorporation of plasmid loss and antibiotic degradation, modeling plasmid mutation could be modeled without additional modification to the models. We modeled this by setting the antibiotic concentration and antibiotic degradation rates to 0. While the plasmid of interest is present at a high copy number, and the fraction of mutated plasmid can vary and with it the associated production and growth rates, modeling this explicitly would not be tractable in the context of our models. Because random partitioning affects can result in a high fraction of mutated plasmids from a single initial mutation in a time scale much faster than acquiring additonal mutations [11], we make the simplifying assumption that a single mutation abolishes all expression with a single event.

Deterministic and stochastic implementations

In order to allow our models to be simulated either deterministically or stochastically, we implemented a manual ODE solver using Euler's method with time step 0.01 h that allowed selection of the simulation mode. For both deterministic and stochastic simulation, cell growth, production, and differentiation were modeled according to the previously described equations, and only mutation events were modeled stochastically. Because of the continuous nature of cell numbers produced using deterministic ODEs, cell numbers were rounded down for determining the number of cells which would mutate. To calculate the number of cells that would mutate in a given time step, we sampled randomly from a binomial distribution according to

$$n_{mut} = \mathcal{B}(N, kdt), \tag{3.16}$$

where n_{mut} is the number of cells that mutate in a given time step, N is the number of cells of the genotype considered for this reaction and the number of Bernoulli trials, k is the rate that was determined for this specific event, dt is the time step, and kdt is the probability of the event occuring in the time step. The number of cells is then subtracted from the source genotype, and added to the destination genotype.

In the simulations shown, we model dilution of batch cultures in a manner similar to our experiments. For deterministic modelling, at the end of each batch growth, the cell population for each genotype is divided by the dilution factor (d) for restarting the next batch growth. The antibiotic concentration for the start of the next batch growth is calculated as

$$A_* = A/d + (1 - 1/d)A_0, \tag{3.17}$$

where A_* is the concentration of antibiotic at the start of the next batch growth, and A_0 is the concentration of antibiotic in fresh media.

For dilution in stochastic simulations, the number of cells for each genotype in the next batch growth was determined by drawing from a binomial distribution according to

$$X_i^* = \mathcal{B}(X_i, 1/d). \tag{3.18}$$

Finally, because the metric of interest for production is the total amount of production and not the concentration of product, the amount of arbitrary product is not diluted but tracked continuously through subsequent batch growths. Production rate was assumed to vary with growth phase as in Chapter 2.

Deterministic modeling for Figure 3.1

Deterministic simulations in Fig 3.1 were run with models generated as described above considering the case of a single cassette for all circuits, neglecting the integrase mutation, and not considering plasmid mutation or antibiotic degradation. Models used carrying capacity limited growth with constant dilution, and were terminated after 10,000 h of simulated time.

3.8 Materials and methods

Strains and constructs

The wild-type *E. coli* strain JS006 was the base strain for the construction of all differentiation and naive circuit strains. Constructs were assembled with a combination of Golden Gate and Gibson assembly using 3G [42], and were integrated into the *E. coli* genome using clonetegration [32]. Because the R6K origin used for propagation of pOSIP plasmids from the clonetegration method of genomic integration is

the same origin in our differentiation architecture R6K plasmid, and the π protein driving the replication of this plasmid is expressed from the differentiation cassette being integrated, we needed to remove this origin from the pOSIP plasmids. We accomplished this by designing primers to PCR the pOSIP plasmids in two parts for use in Gibson assembly with circuits to be integrated, retaining the portions required for integration but removing the R6K origin. For Gibson assembly with linear pOSIP pieces, POS1 and POSX were used as terminal adapters instead of UNS1 and UNSX. The 1x naive strain and 1x differentiation strain was constructed by integration at the P21 (T) landing site, and the 2x naive and 2x split-pir differentiation strains by additional integration at the HK022 (H) landing site. 1x differentiation and 2x differentiation strains were integrated two additional times with the inducible Bxb1-LAA expression construct at the primary and secondary phage 186 (O) landing sites (Figure 3.27). Following transformation, integrations were checked via colony PCR with the pOSIP p4 primary corresponding to the landing site [32] and a reverse primer common to all pOSIP plasmids (5' ATTACTCAACAGGTAAGGCG 3'). Fidelity of integrations was checked with a combination of sequencing and functional screening prior to transformation with pE-FLP to excise the antibiotic resistance cassette and integration module, and integration of subsequent constructs [32]. Final strains for 1x and 2x naive, 1x differentiation 2x Bxb1, and 2x split-pir differentiaton 2x Bxb1 were whole-genome sequenced with minION using the Rapid Barcoding Kit (Nanopore SQK-RBK004) for verification. Reads were assembled with Flye (https://github.com/fenderglass/Flye/) and mapped to reference genomes containing intended genomic insertions in Geneious.

Modified MoClo [33] compatible parts for T7 RNAP, integrase attachment sites, and terminators were generated with standard molecular biology techniques (PCR, Gibson, oligo annealing and phosphorylation), and modified UNS adapters used for construction of polycistronic or inverted transcriptional units. The R6KCmR backbone was constructed with Golden Gate using an R6K origin amplified from the pOSIP plasmids. Sequences for Bxb1 integrase attachment sites attB and attP were obtained from Ghosh et.al.[34]. NahR^{AM}, LasR^{AM}, LacI^{AM}, and their corresponding evolved promoters P_{SalTTC}, P_{LasAM}, and P_{Tac} were provided by Adam Meyer [23]. The CIDAR MoClo Parts Kit which includes various promoter, RBS, CDS, and terminator parts used in the constructs described were provided by Douglas Densmore (Addgene kit 1000000059).

Differentiation experiments

Chemically competent cells were prepared from the naive and differentiation strains grown in LB without selection, with differentiation strains induced with 30 nM Las-AHL to allow pi protein expression for R6K plasmid replication. 1x and 2x naive strains were transformed with ColE1 AmpR pT7 GFP or ColE1 KanR pT7 GFP and plated on LB with 100 μg/mL carbenicillin or 50 μg/mL kanamycin. Differentiation strains were co-transformed with R6KCmR-mScarletI and ColE1 AmpR pT7 GFP or ColE1 KanR pT7 GFP, recovered in SOC with 30 nM Las-AHL, and plated on LB agar with 34 μg/mL chloramphenicol, 30 nM Las-AHL, and 100 μg/mL carbenicillin or 50 μg/mL kanamycin. Eight independent colonies were picked from each transformation, and grown at 37C in 300 μL LB in 96 square deep well plates for 8 hours. Naive strains were grown in LB with appropriate antibiotic, and differentiation strains grown in LB with chlor and carb or kan with 10 nM Las-AHL. Following outgrowth, cells were diluted 1:50 into experimental conditions. Cells were diluted every 8 hours for sixteen total growths in constant antibiotic and induction conditions, and sfGFP (485/515 nm), mScarlet (565/595 nm), and OD700 measured by taking 50 μL aliquots of endpoint culture and measuring in 384 well black wall clear bottom Matriplates. Average of two reads for each measurement in each well were used.

dnaseI expression and quantification

As an initial demonstration of the toxicity of dnaseI and of the importance for preventing leaky expression of T7 RNAP, chemically competent 1x and 2x naive strains, and 1x and 2x differentiation strains were transformed with 10 ng of HCA pT7 GFP or 10ng HCA with insulated pT7 dnaseI, and all or 10 percent plated on LB carb. Plates with more than 1000 colonies on the 10 percent plate were reported as $> 10^4$ cfu. For dnaseI expression experiments, 2x split-pir differentiation cells were co-transformed with an empty R6KCmR plasmid and the insulated ColE1 AmpR dnaseI expression plasmid, recovered in SOC with 30 nM Las-AHL, and plated on LB agar with carb/chlor/30 nM Las-AHL. Three independent colonies were inoculated into 3 mL LB cultures with carb/chlor/10 nM Las-AHL, outgrown for 8 hours at 37C, and diluted 1:50 into 25 mL media with or without 20 μM salicylate and 10 μM IPTG to induce differentiation and T7 RNAP expression. After 8 hours of growth, cultures were diluted 1:50 into the same conditions, and the remaining culture harvested. Wet weight of pellets after washing with PBS was recorded before storing at -20C. JS006 parental strain without the dnaseI expression

plasmid was grown similarly in LB without antibiotics and inducers for negative control. The second growth of uninduced cultures was harvested after 8 hours as before, and induced cultures allowed to grow for an additional 8 hours as minimal growth was observed after the initial growth.

Pellets were lysed via sonication of a 33 percent (w/v) cell suspension in 10 mM Tris pH 7.5/2 mM CaCl2 with protease inhibitor (Roche, 11836170001), cleared with centrifugation at 4C, and supernatant collected and kept on ice before assaying dnaseI activity. Buffers used for assay were as described in Kunitz [43], though to allow simultaneous measurement of many samples and to avoid problems we observed with background absorbance in crude cell lysate when performing the Kunitz assay, we developed a fluorescence-based assay similar to Vogel and Frantz [44]. Briefly, dnaseI assay buffer was prepared by diluting SYBR Safe (Invitrogen, S33102) 1:1000 into a solution of 100 mM sodium acetate/5 mM magnesium sulfate with 26.3 μg/mL calf thymus DNA (Sigma D1501). Assay plate was prepared by aliquoting 190 μL dnaseI assay buffer into 96 round-well clear bottom plates and equilibrating in the dark at 25C. Standards were prepared by adding various amounts of dnaseI (Invitrogen AM2222) to JS006 lysate diluted 1:10 in 0.85 percent NaCl. Samples to assay were diluted 1:10 or 1:50 in 0.85 percent NaCl, and 10μL of sample or standard pipetted with a multi-channel pipette into triplicate wells immediately before assay. The final amount of DNA per well was 5μg. After shaking briefly fluorescence (487/528 nM) was measured every minute for 2 hours at 25C. Fluorescence fold-change over the course of the two hour assay was used in fitting a standard curve (Fig 3.16), and dnaseI activity calculated from appropriate dilutions.

3.9 Acknowledgements

I would like to thank Andrey Shur for his help with minION sequencing, as well as the Sim Lab at the University of California Irvine for welcoming into their lab space to finish this work.

3.10 Supplemental Figures

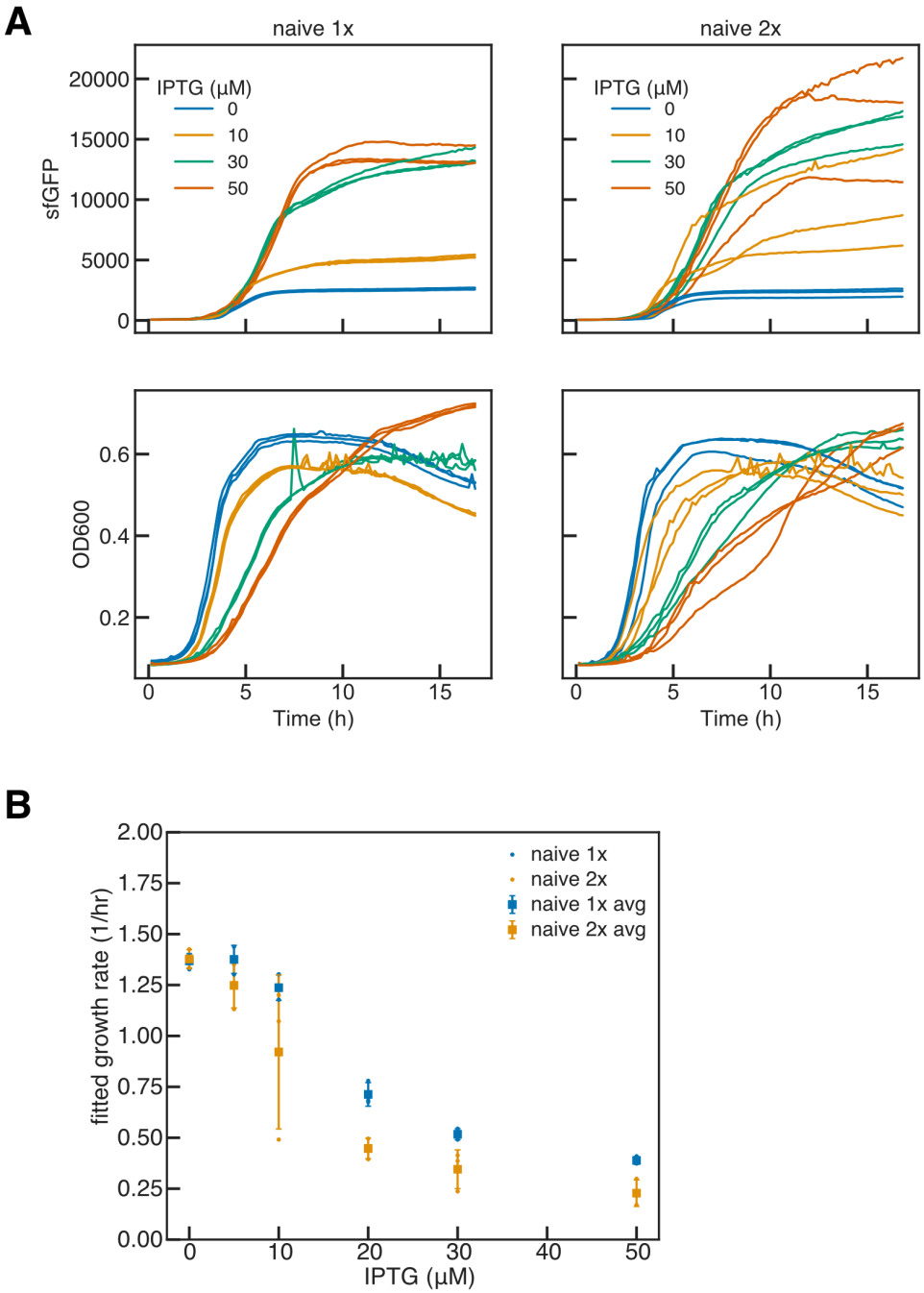

Figure 3.7: Characterization of burden level of naive T7 RNAP driven GFP. JS006 cells with 1 and 2 copies of naive IPTG inducible T7 RNAP were transformed with a high copy ColE1 plasmid with kanamycin resistance. Cells were outgrown in LB with kanamycin (50 μg/mL) for ~6 hours then diluted 1:50 into LB kan with varying concentrations of IPTG. (A) GFP production (top) and OD600 (bottom) were measured every 10 minutes for 1x naive (left) and 2x naive (right) grown in triplicate at 37C in 0.3 mL. (B) OD600 curves were trimmed to 60 percent of maximum OD600 achieved and used to fit an exponential growth model with noise floor, initial population, and growth rate parameters. Mean growth rate +/- SD of 3 replicates fitted separately plotted for naive 1x and 2x.

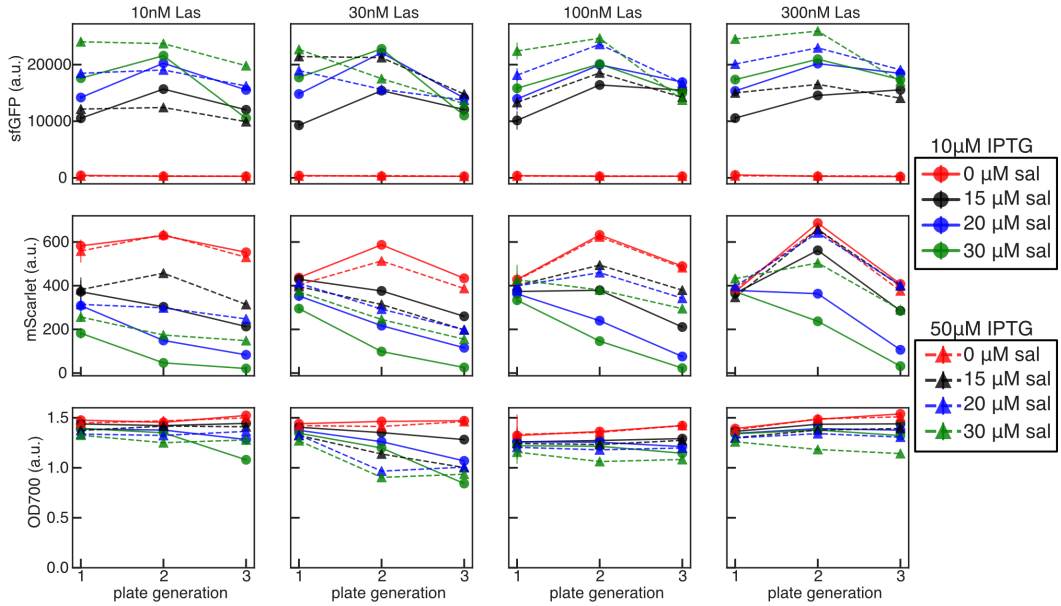

Figure 3.8: 1x differentiation with ColE1 AmpR pT7 GFP and R6KCmR-mScarlet; Las-AHL, salicylate, and IPTG gradient fluorescence and OD data by plate generation. Three replicate colonies outgrown for 8 hours in LB carb/chlor with varying concentrations of Las-AHL, then diluted 1:50 every 8 hours into 0.3 mL LB carb/chlor/Las with varying concentrations of sal and IPTG. Mean +/- SD of GFP (top), mScarlet (middle), and OD700 (bottom) plotted for three total growths. Color indicates sal concentration; 10 μM IPTG (circles, solid lines), 50 μM IPTG (triangles, dashed lines).

Strain	pT7 GFP	pT7 dnaseI
1x naive	$> 10^4$ cfu	1 cfu
2x naive	$> 10^4$ cfu	0 cfu
1x diff	$6.6x10^3$ cfu	590 cfu
2x diff	$> 10^4$ cfu	1090 cfu

Table 3.1: 50 μL chemically competent cells were transformed with 10 ng of ColE1 AmpR pT7 GFP or 10 ng ColE1 AmpR T13m-T12m-pT7-B0032-dnsaeI-T7T, and 5 μL or 50 μL plated on LB carb. $> 10^4$ indicates more >1000 colonies of plating 5 μL.

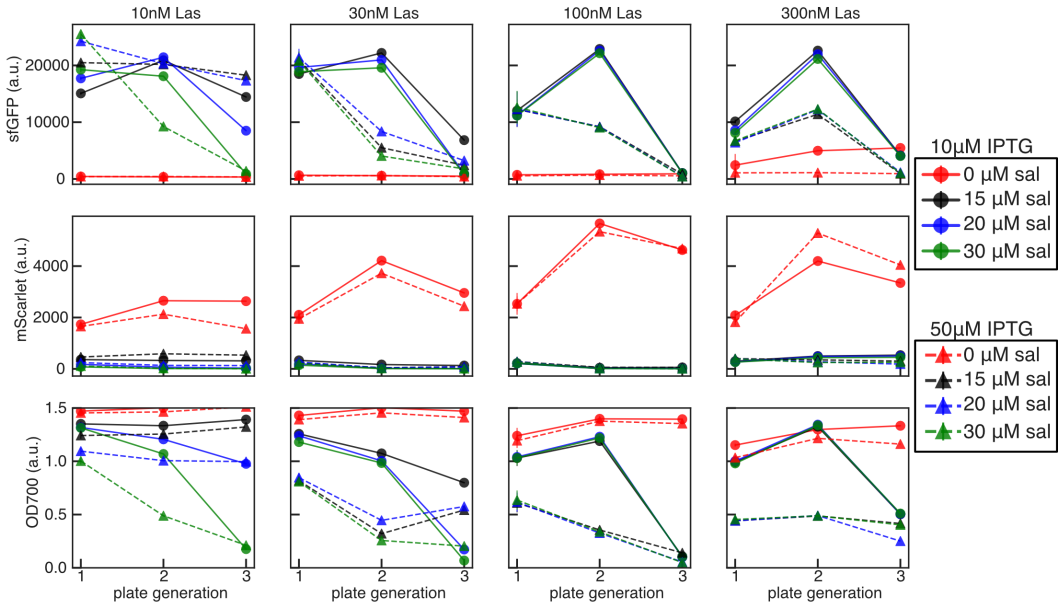

Figure 3.9: 2x differentiation with ColE1 AmpR pT7 GFP and R6KCmR-mScarlet; Las-AHL, salicylate, and IPTG gradient fluorescence and OD data by plate generation. Three replicate colonies outgrown for 8 hours in LB carb/chlor with varying concentrations of Las-AHL, then diluted 1:50 every 8 hours into 0.3 mL LB carb/chlor/Las-AHL with varying concentrations of sal and IPTG. Mean +/- SD of GFP (top), mScarlet (middle), and OD700 (bottom) plotted for three total growths. Color indicates sal concentration; 10 μM IPTG (circles, solid lines), 50 μM IPTG (triangles, dashed lines).

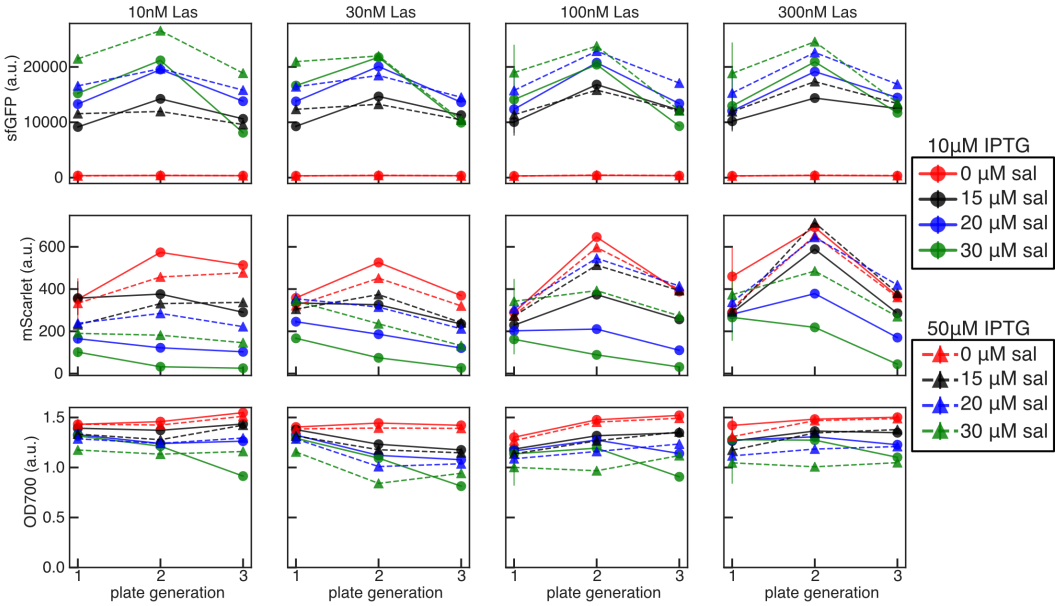

Figure 3.10: 1x differentiation with ColE1 KanR pT7 GFP and R6KCmR-mScarlet; Las-AHL, salicylate, and IPTG gradient fluorescence and OD data by plate generation. Three replicate colonies outgrown for 8 hours in LB kan/chlor with varying concentrations of Las-AHL, then diluted 1:50 every 8 hours into 0.3 mL LB kan/chlor/Las-AHL with varying concentrations of sal and IPTG. Mean +/- SD of GFP (top), mScarlet (middle), and OD700 (bottom) plotted for three total growths. Color indicates sal concentration; 10 μM IPTG (circles, solid lines), 50μM IPTG (triangles, dashed lines).

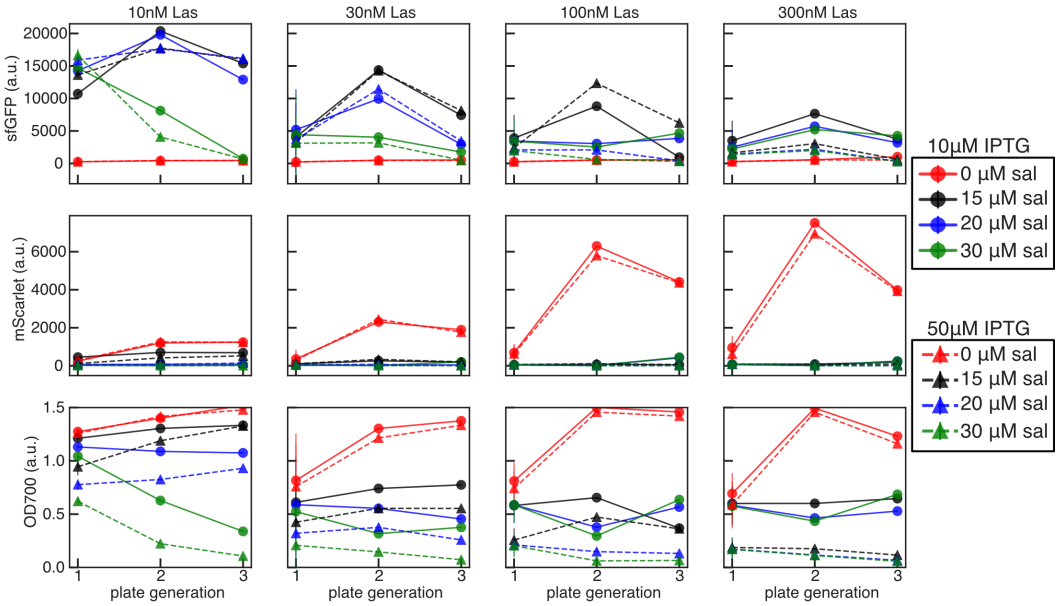

Figure 3.11: 2x differentiation with ColE1 KanR pT7 GFP and R6KCmR-mScarlet; Las-AHL, salicylate, and IPTG gradient fluorescence and OD data by plate generation. Three replicate colonies outgrown for 8 hours in LB kan/chlor with varying concentrations of Las-AHL, then diluted 1:50 every 8 hours into 0.3 mL LB kan/chlor/Las with varying concentrations of sal and IPTG. Mean +/- SD of GFP (top), mScarlet (middle), and OD700 (bottom) plotted for three total growths. Color indicates sal concentration; 10 μM IPTG (circles, solid line), 50 μM IPTG (triangles, dashed line).

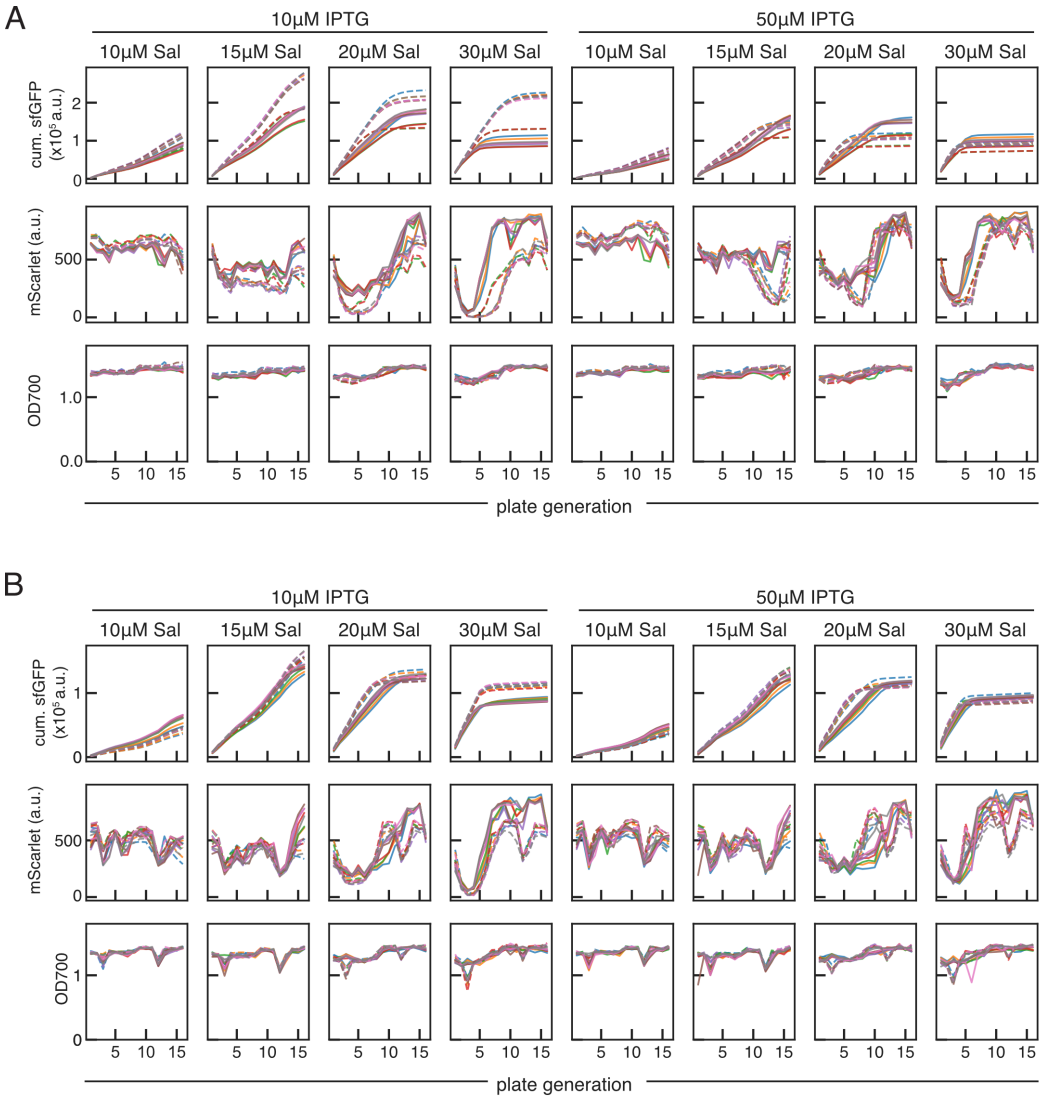

Figure 3.12: 1x differentiation evolution fluorescence and OD data by plate generation. (A-B) Cumulative endpoint sfGFP fluorescence (top row), endpoint mScarletI fluorescence (middle row) and OD700 (bottom row) for individual experimental replicates measured in 50 μL in 384 well matriplates. (A) Data from experiment with ColE1 AmpR pT7 GFP of cells grown in LB carb/chlor/10 nM Las (solid lines) and carb/10 nM Las (dashed lines) with varying concentrations of IPTG. (B) Data from experiment with ColE1 KanR pT7 GFP of cells grown in LB kan/chlor/10 nM Las (solid lines) and kan/10 nM Las (dashed lines) with varying concentrations of IPTG.

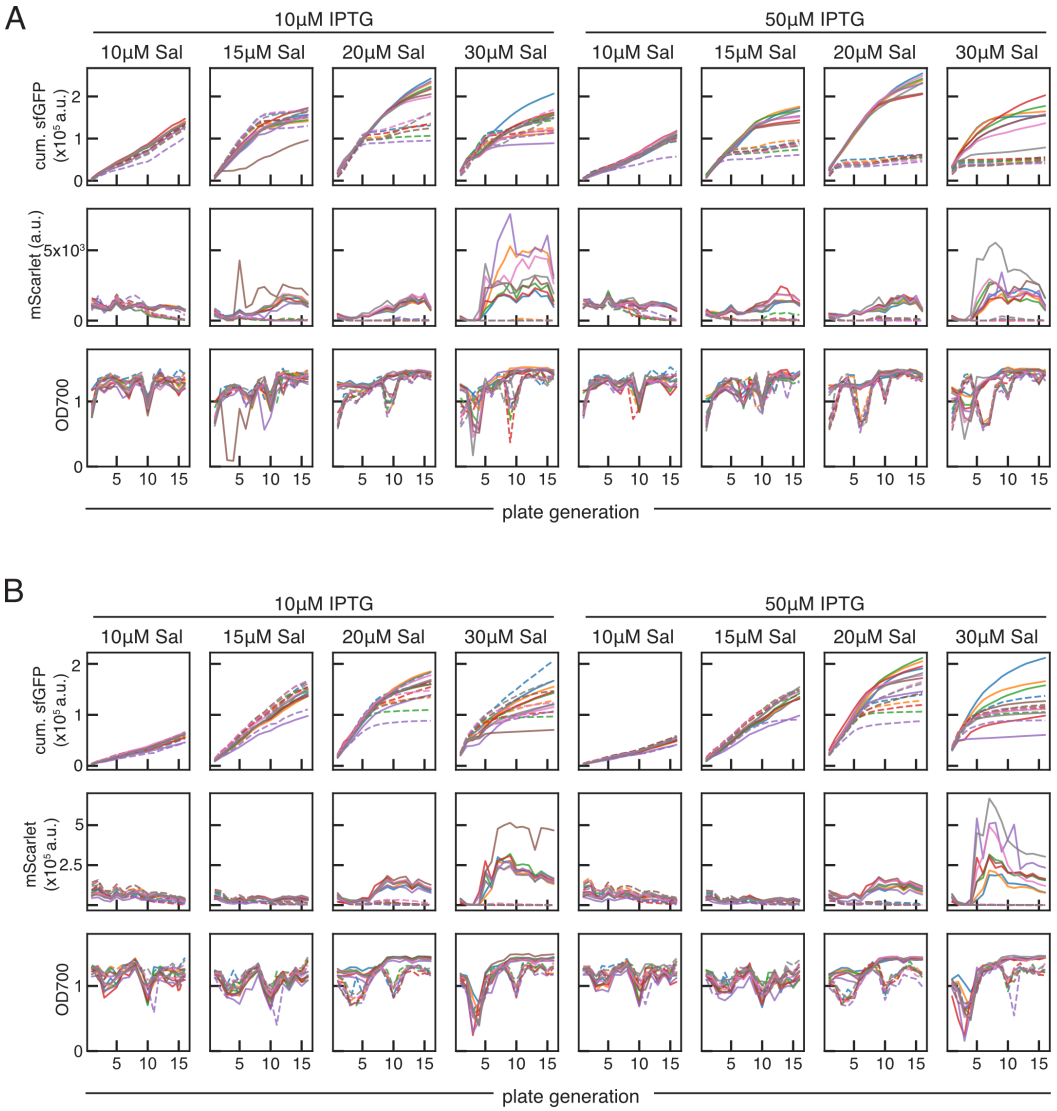

Figure 3.13: 2x differentiation evolution experiment fluorescence and OD data by plate generation. (A-B) Cumulative endpoint sfGFP fluorescence (top row), endpoint mScarletI fluorescence (middle row) and OD700 (bottom row) for individual experimental replicates measured in 50 μL in 384 well matriplates. (A) Data from experiment with ColE1 AmpR pT7 GFP of cells grown in LB carb/chlor/10 nM Las (solid lines) and carb/10 nM Las (dashed lines) with varying concentrations of IPTG. (B) Data from experiment with ColE1 KanR pT7 GFP of cells grown in LB kan/chlor/10 nM Las (solid lines) and kan/10 nM Las (dashed lines) with varying concentrations of IPTG.

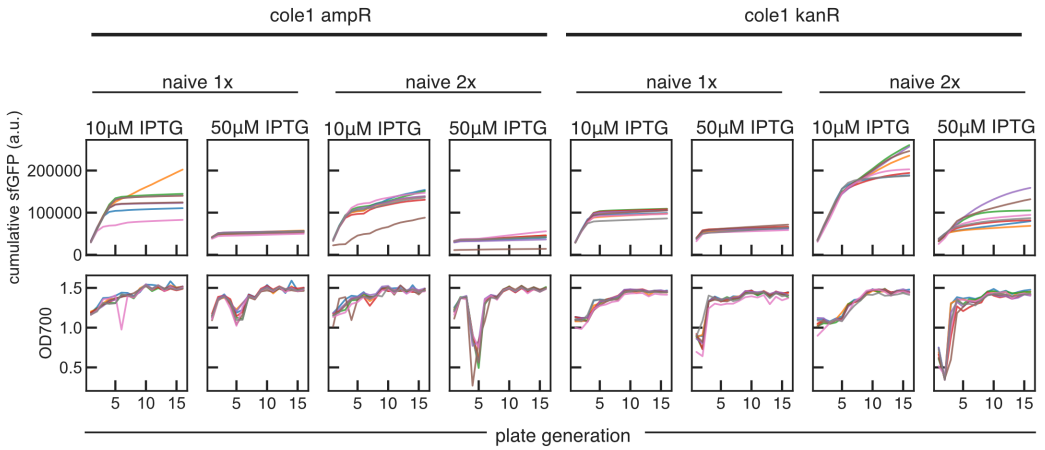

Figure 3.14: 1x and 2x naive evolution experiment fluorescence and OD data by plate generation. (A-B) Cumulative endpoint sfGFP fluorescence (top row) and OD700 (bottom row) for individual experimental replicates measured in 50 μL in 384 well matriplates. (Left) Data from experiment with ColE1 AmpR pT7 GFP of cells grown in LB carb with varying concentrations of IPTG. (B) (Right) Data from experiment with ColE1 KanR pT7 GFP of cells grown in LB kan with varying concentrations of IPTG.

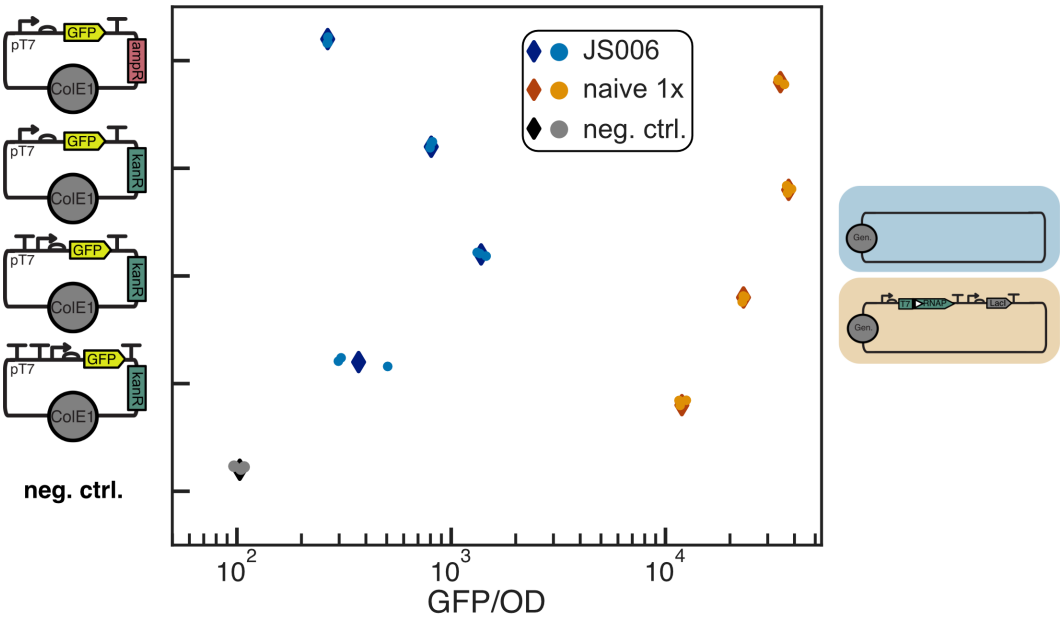

Figure 3.15: Assessment of leaky expression in the absence of T7 RNAP. JS006 and naive 1x cells were transformed with plasmids containing pT7-B0034-sfGFP-T7T, with 0, 1, or 2 insulating terminators. OD600 normalized sfGFP fluorescence reported, mean (diamonds) +/- SD of three replicates (circles) after 12 h growth compared to negative control JS006 lacking any GFP expression plasmid.

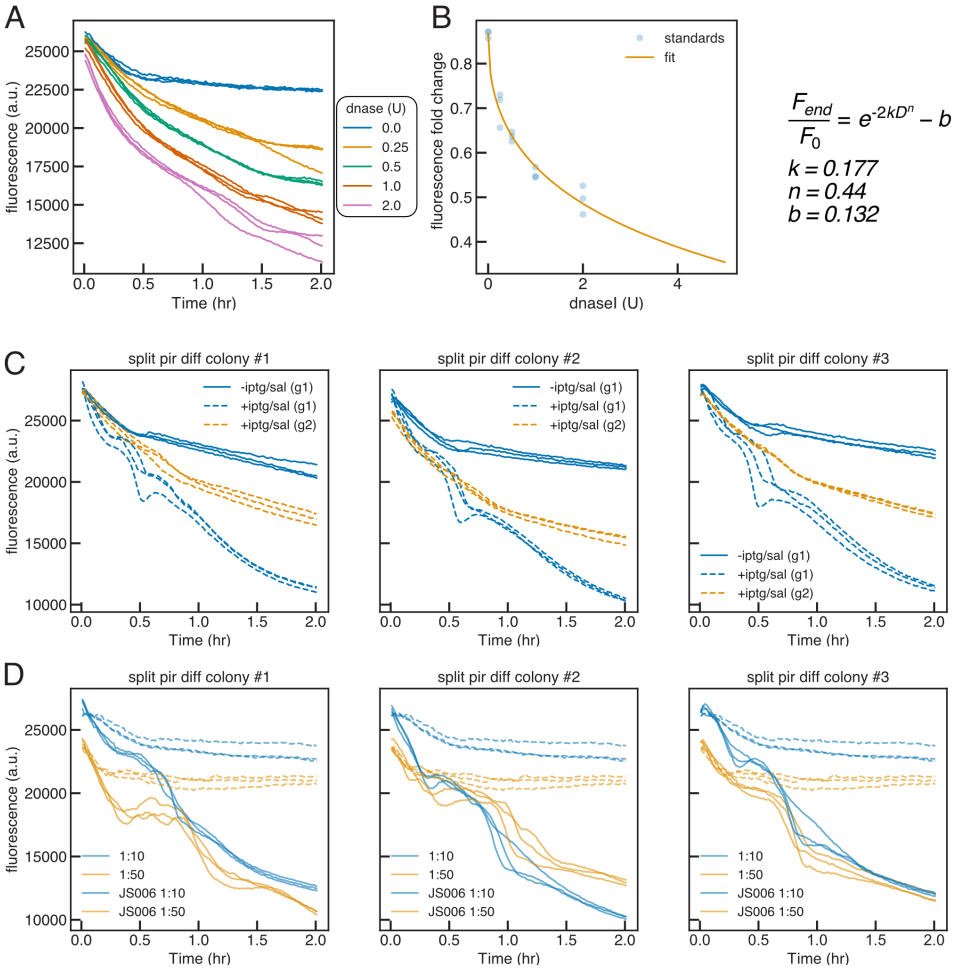

Figure 3.16: Fluorescence-based assay of dnaseI activity in cell lysate. (A) JS006 lysate diluted 1:10 in 0.85 percent NaCl, with 0, 0.25, 0.5, 1, or 2 U of dnaseI per 10 μL volume diluted sample. 10 μL sample added to 190 μL dnaseI assay buffer (100 mM sodium acetate/5 mM magnesium sulfate, 5 μg calf thymus DNA,1:1000 SYBR Safe). Fluorescence (487/528 nm) time-course of samples in triplicate measured every minute for 2 hours. (B) Fluorescence fold-change (endpoint/initial) used to a fit a model, where b describes background loss of fluorescence through photo-bleaching or other non-dnaseI related means of loss of fluorescence, k is the first-order rate constant describing the degradation of DNA by dnaseI, D is the concentration of dnaseI (U/rxn), n is a phenomenological constant which captures the non-linear relationship between dnaseI concentration and observed loss of fluorescence, and 2 is the time in hours for which the assay was ran. (C-D) Time-course traces of dnaseI assay of lysate from three independent experiments. (C) Assay performed on 1:10 dilutions of lysate from uninduced first growth (blue solid), and induced first (dashed blue) and second (dashed orange) growths. (D) Re-assayed lysate samples of induced first growth (solid) diluted 1:10 (blue) and 1:50 (orange), and JS006 lysate (dashed) diluted 1:10 (blue) and 1:50 (orange).

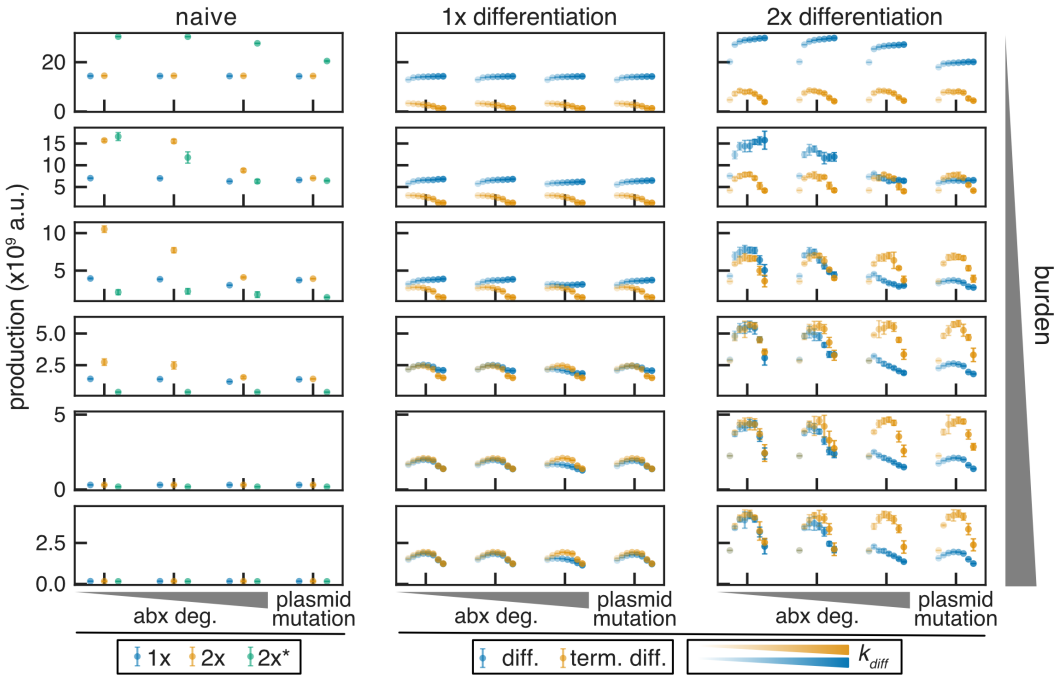

Figure 3.17: Stochastic simulations of burdensome production in 1x and 2x naive, differentiation, and terminal differentiation circuit architectures. Mean total production +/- SD of 8 stochastic simulations of 20 consecutive batch growths with 50x dilutions. $\mu_P = 2$ h^{-1}; 10, 30, 50, 70, 90, 99 percent burden (increasing top to bottom); $K = 10^9$ cells; $k_{MB} = k_{MD} = 10^{-6}$ h^{-1}; k_{diff} 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1 h^{-1}; $n_{\text{div}} = 4$. Production rate and burdens for 1x, 2x, 2x* as described in Model implementation. Simulations with antibiotic degradation are with plasmid loss rate $k_{PL} = 10^{-4}$ h^{-1}; 100 μg/mL antibiotic; MIC=1.1 μg/mL; and antibiotic degradation (V_{max}) rates of 0, ~2.52x10^{-6}, and ~1.26x10^{-5} μg/cell/h (left to right increasing abx deg). Simulations with plasmid mutation were modeled with $k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\text{max}} = 0$.

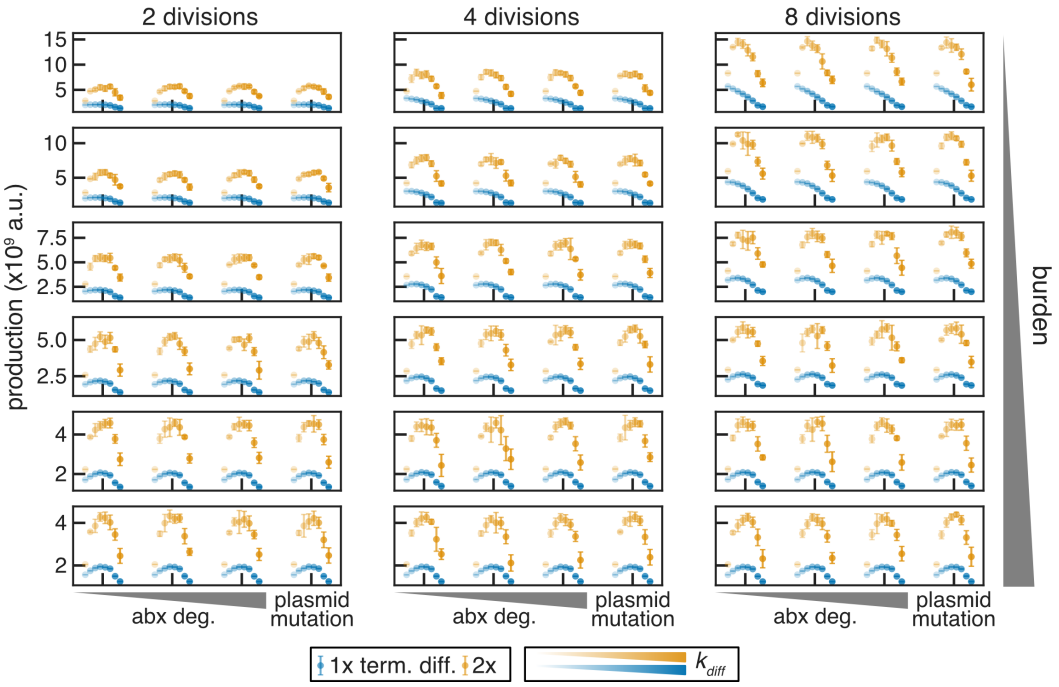

Figure 3.18: Stochastic simulations of burdensome production in 1x and 2x terminal differentiation architectures with varying n_{div}. Mean total production +/- SD of 8 stochastic simulations of 20 consecutive batch growths with 50x dilutions. μ_P = 2 h^{-1}; 10, 30, 50, 70, 90, 99 percent burden (increasing top to bottom); $K = 10^9$ cells; $k_{MB} = k_{MD} = 10^{-6}$ h^{-1}; k_{diff} 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1 h^{-1}; n_{div} of 2 (left), 4 (center) and 8 (right). Simulations with antibiotic degradation are with plasmid loss rate $k_{PL} = 10^{-4}$ h^{-1}; 100 μg/mL antibiotic; MIC=1.1 μg/mL; and antibiotic degradation (V_{max}) rates of 0, ~2.52x10^{-6}, and ~1.26x10^{-5} μg/cell/h (left to right increasing abx deg). Simulations with plasmid mutation were modeled with $k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\text{max}} = 0$.

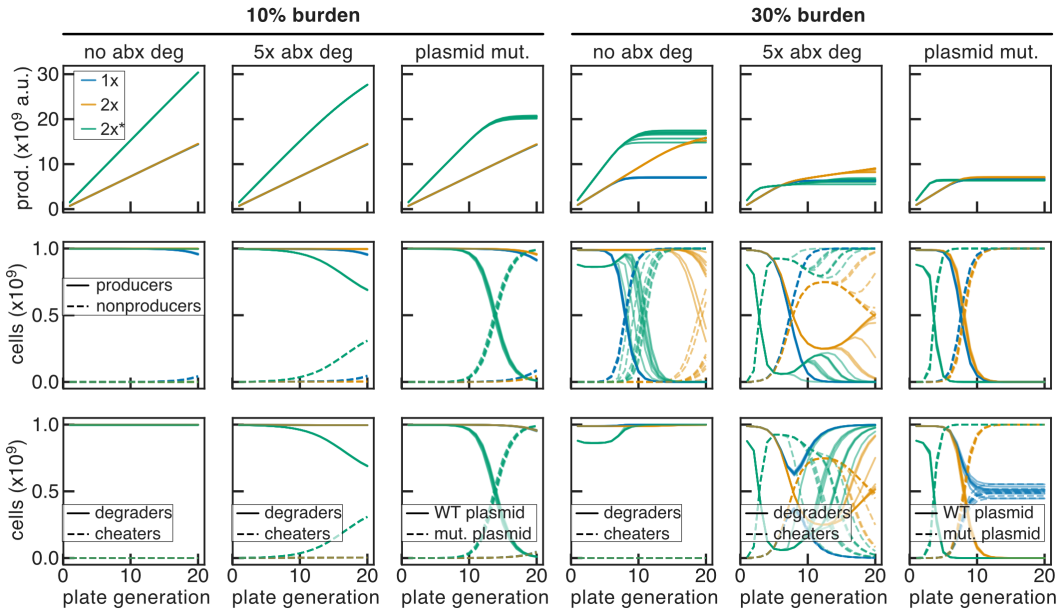

Figure 3.19: Stochastic simulations of 1x and 2x naive burdensome expression with 10 and 30 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

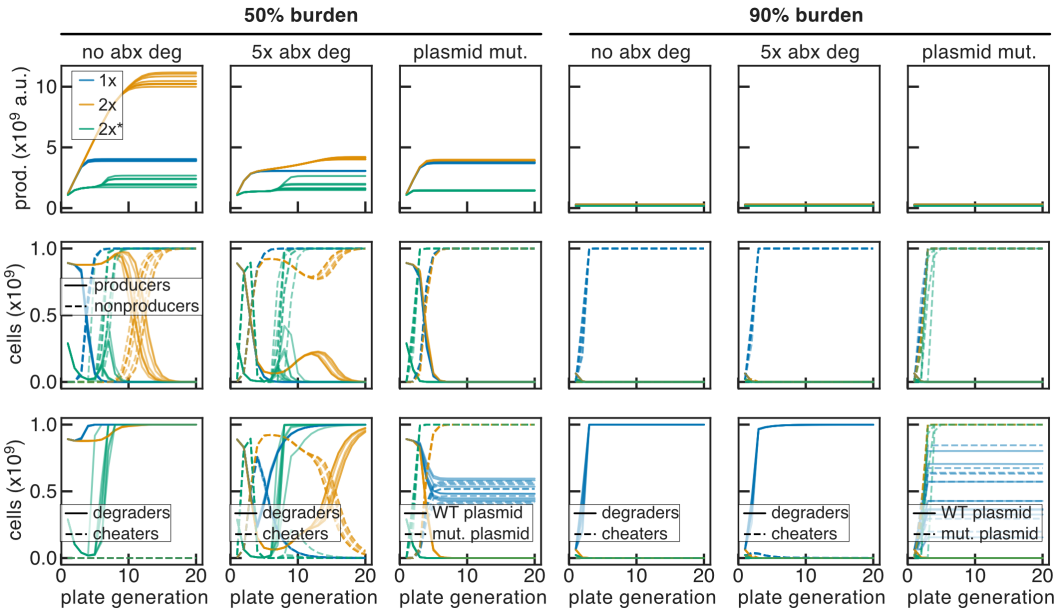

Figure 3.20: Stochastic simulations of 1x and 2x naive burdensome expression with 50 and 90 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

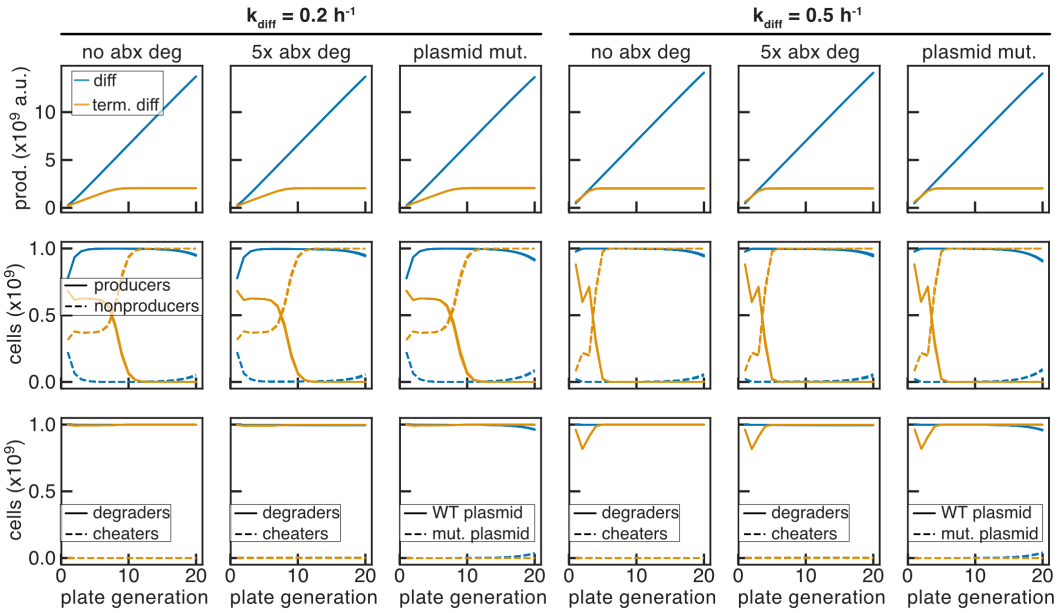

Figure 3.21: Stochastic simulations of 1x differentiation and terminal differentiation burdensome expression with 10 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

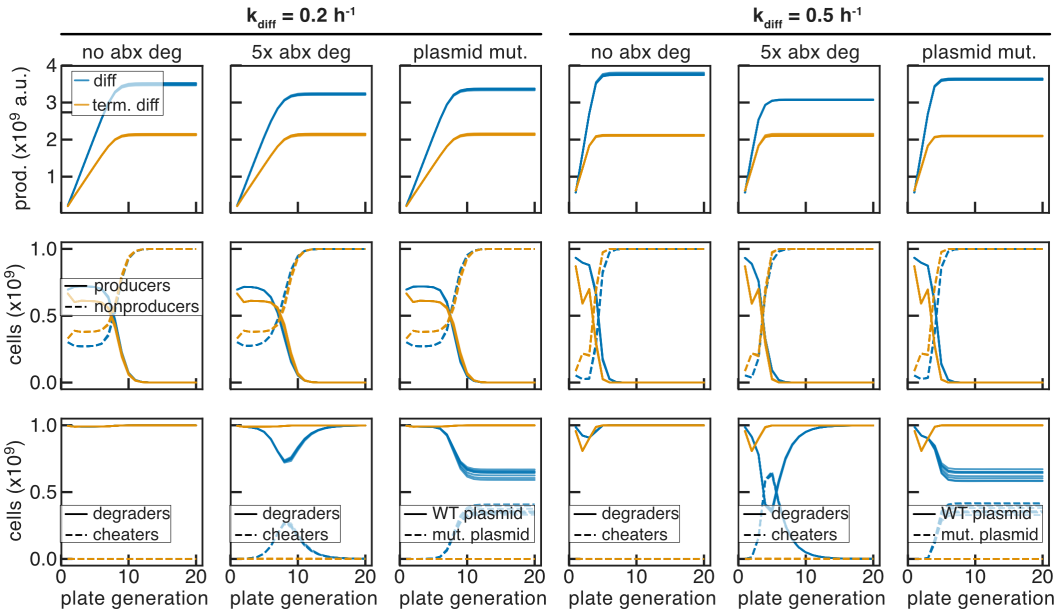

Figure 3.22: Stochastic simulations of 1x differentiation and terminal differentiation burdensome expression with 50 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

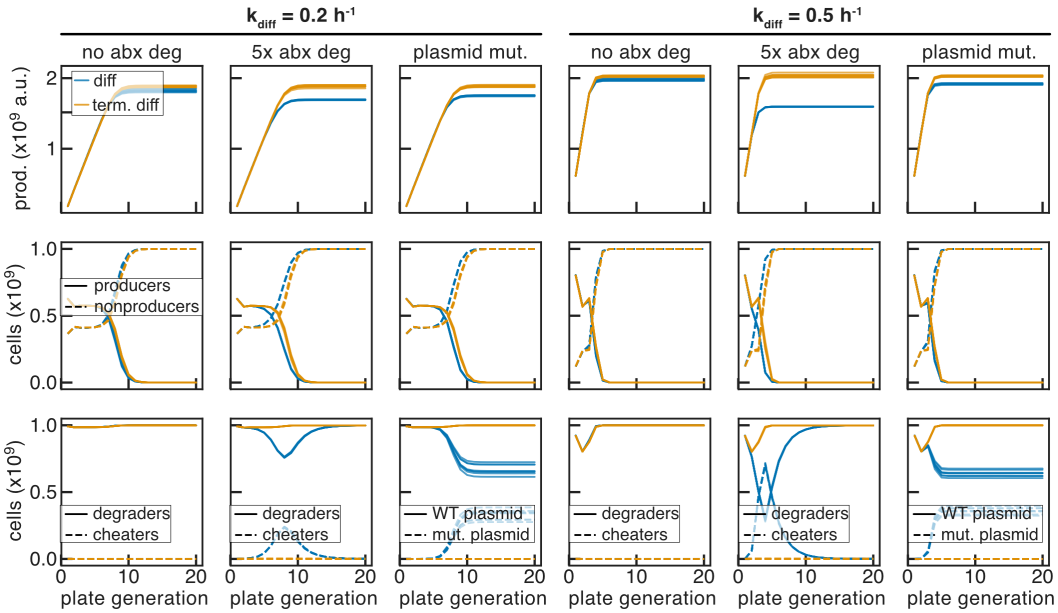

Figure 3.23: Stochastic simulations of 1x differentiation and terminal differentiation burdensome expression with 90 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

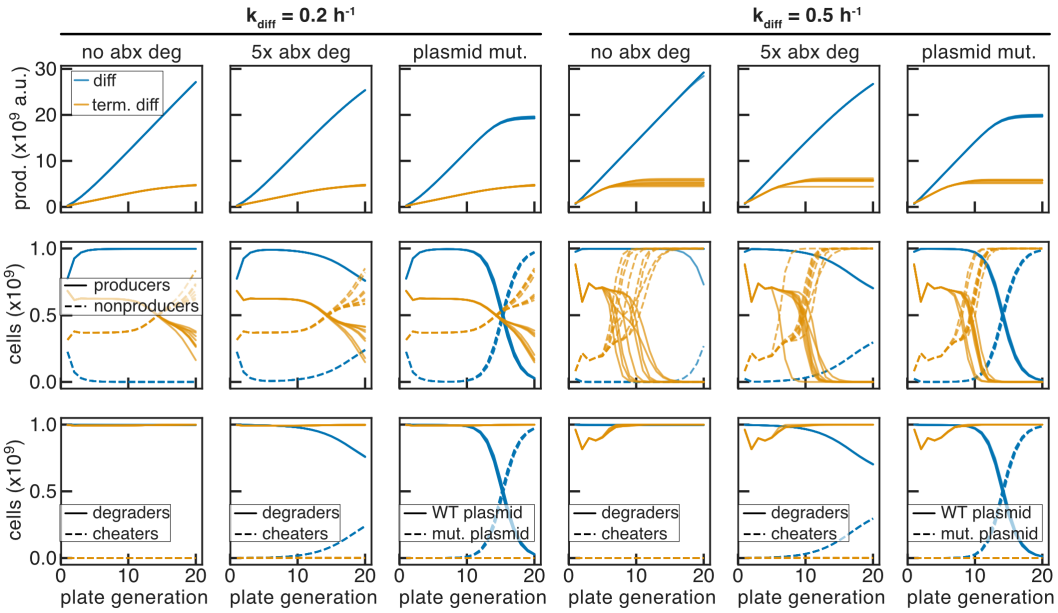

Figure 3.24: Stochastic simulations of 2x differentiation and terminal differentiation burdensome expression with 10 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

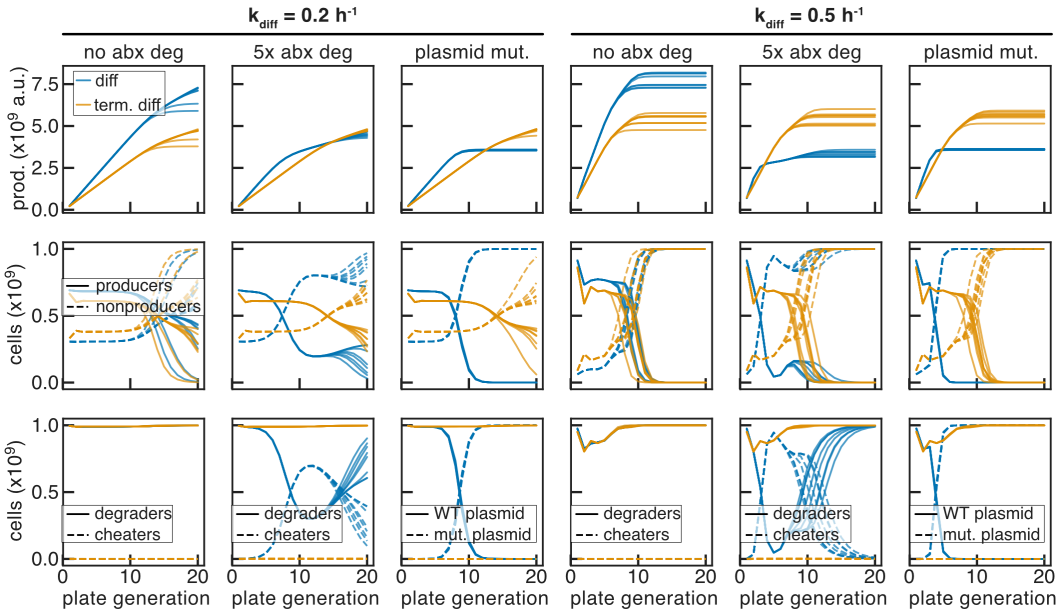

Figure 3.25: Stochastic simulations of 2x differentiation and terminal differentiation burdensome expression with 50 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

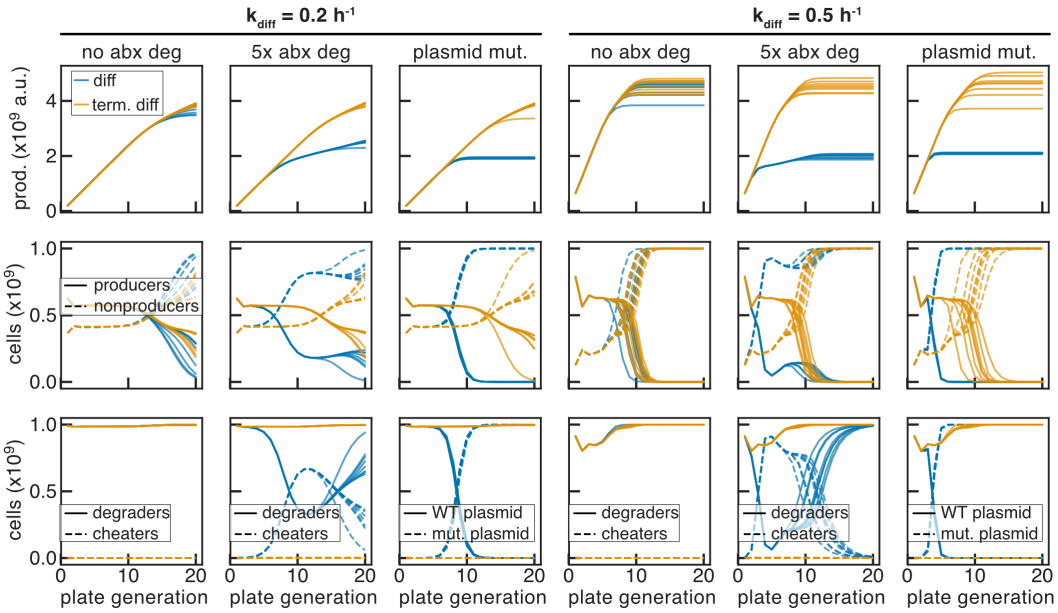

Figure 3.26: Stochastic simulations of 2x differentiation and terminal differentiation burdensome expression with 90 percent burden. Endpoint cumulative production (top); producer/non-producer fractions (middle); fraction retaining (degraders) or having lost (cheaters) the antibiotic resistance/expression plasmid (bottom) for simulations modeling plasmid loss ($k_{PL} = 10^{-4}$ h^{-1}) without antibiotic degradation (no abx deg; $V_{\max} = 0$) and with high level antibiotic degradation (5x abx deg; $V_{\max}$=~1.26x10^{-5} μg/cell/h); and fraction of cells with WT functional plasmid and mutatated non-functional plasmid (bottom) for simulations modeling plasmid mutation ($k_{PL} = 10^{-6}$ h^{-1}, 0 μg/mL antibiotic, and $V_{\max} = 0$) plotted for 20 consecutive batch growths. Simulations as described in Figure 3.17.

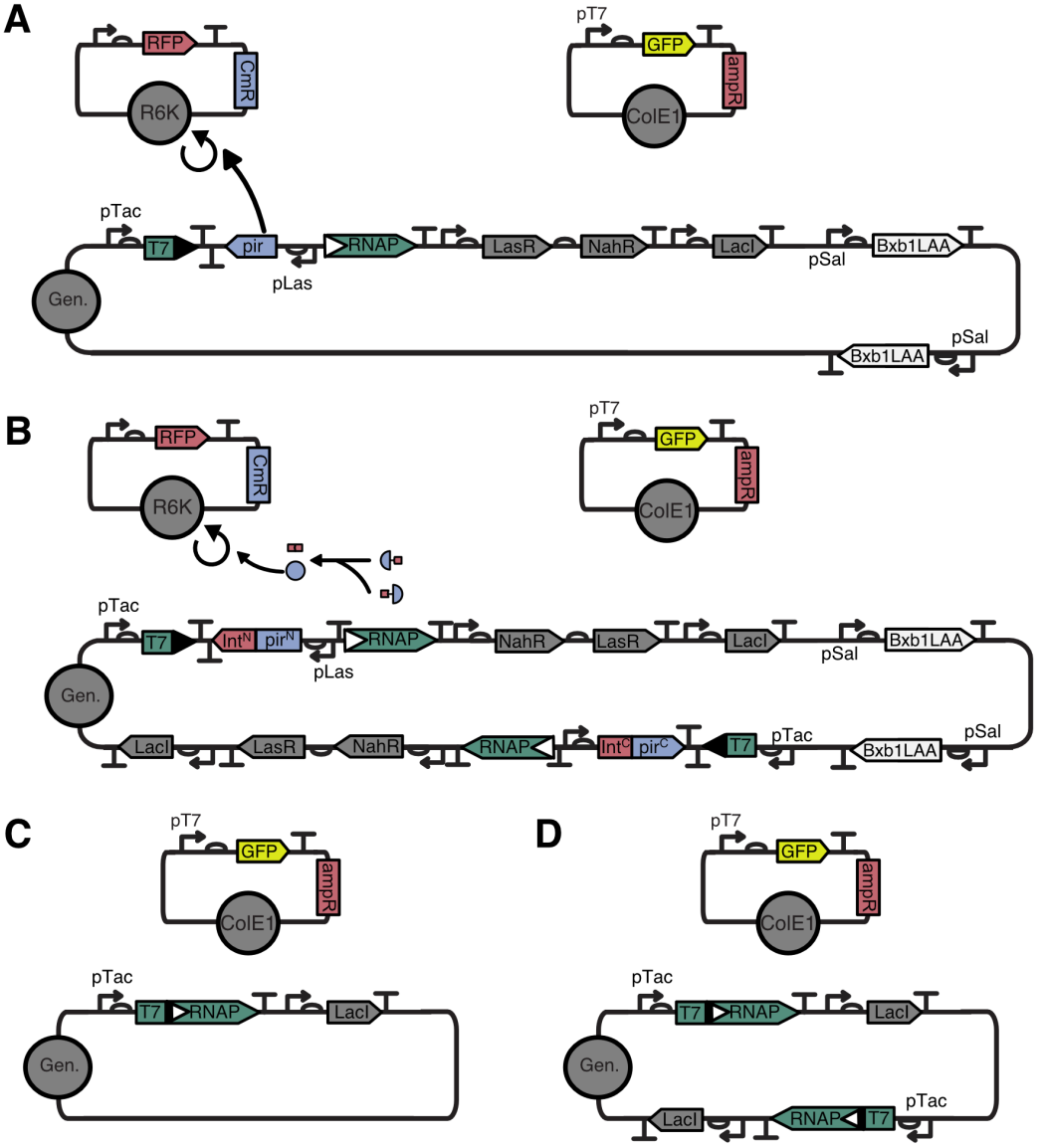

Figure 3.27: Full circuit diagrams for 1x and 2x naive and differentiation circuits as depicted in 3.3 and 3.5. (A) 1x differentiation circuit diagram. Differentiation cassette encoding excisable *pir*, recombination activated T7 RNAP, and NahR^{AM}, LasR^{AM}, and LacI^{AM}, is integrated with clonetegration [32] at the P21 (T) landing site. 2 copies of P_{SalTTC}-B0034-Bxb1LAA-T2m were integrated at the primary and secondary 186 (O) landing sites. (B) 2x differentiation circuit diagram. As in (A), with the the T site integration encoding the N-terminal portion of the split-π protein, and the second differentiation cassette integrated at the HK022 (H) site encoding the C-terminal portion of the split-π protein. (C) 1x naive expression cassette encoding inducible T7 RNAP and LacI^{AM} integrated at the T site. (D) As in (C) with second identical integration at the H site.

Chapter 4

DEVELOPMENT OF DIFFERENTIATION CIRCUIT ARCHITECTURES FOR SCALING IN TIME AND POPULATION SIZE

4.1 Introduction

In chapter 3, we improved the differentiation circuit architecture we developed in chapter 2 by incorporating a novel split-π protein which allowed a single recombination event in cells with two differentiation cassettes to both activate T7 RNAP expression and enable antibiotic selection to limit growth. We demonstrated both computationally and experimentally that this redundancy and robustness to mutation improves the evolutionary stability of the terminal differentiation architecture, and further demonstrated that this system could enable the expression of a toxic protein. However, despite the benefit provided by this degree of redundancy, we still observed circuit failure both experimentally and computationally with a small population size ($\sim 10^9$ cells). In this chapter, we investigate how this and additional circuit motifs can help in scaling the application of this system to longer times and larger population sizes. We further consider means of extending this redundancy in the terminal differentiation architecture beyond what we achieved with the split-π protein, and demonstrate a proof of concept experimental circuit design that should allow scaling to arbitrary n. In doing so, we are motivated by the important feature of the terminal differentiation architecture which protects burden mutations – those that disrupt the engineered functions of interest – from evolutionary forces. By addressing the evolutionary stability of the terminal differentiation architecture, we can address this for any biologically possible function.

4.2 Considerations for scaling terminal differentiation

In Chapter 3 we achieved a degree of robustness to differentiation mutations by generating a split π-protein, however in considering scaling this to a higher degree of redundancy, we considered several options. Importantly, the recombination of a single cassette must both activate the function of interest by joining the two halves of T7 RNAP as well as lead to delayed cell death or cessation of growth. While the π protein could potentially be split into more pieces, or an additional plasmid with a required replication factor used, this seemed both cumbersome and not scalable.

We additionally considered integrating the differentiation cassette flanking essential genes in the *E. coli* genome, with recombination activating cell death by excising the essential gene. Though feasible, this also seemed unnecessarily difficult and time consuming. In the ideal case, an identical construct could be integrated in the genome repeatedly, with each integration providing an additional layer of redundancy. To accomplish this, we reasoned that instead of inactivating an essential gene, a toxic or lethal gene could be activated.

While we initially considered considered using various toxins such as ccdb, ϕX174E, and gp2 which have been been harnessed in synthetic biology for population capping and composition control because of their ability to induce death and cell lysis [45–47], using such immediately lethal proteins is problematic in the context of our proposed circuit. In order for this strategy to be useful, differentiated cells must have sufficient time to express the function of interest before dying. While there are certainly means of inserting delay into the circuit to allow time for function before expression of the toxic protein, and we do propose one such strategy, the ideal mechanism of growth cessation would not immediately destroy cell integrity and metabolic activity. One such candidate which fulfills this criteria is provided by nature. In a demonstration of the utility of chromosome-free bacterial cells, the homing endonuclease I-ceuI was used to induce double stranded breaks in the genome, with subsequent genome loss mediated by endogenous nucleases. They demonstrated that these chromosome-free cells could be purified, and retain a degree of metabolic capacity for days to months when stored at 4C [48]. Importantly, the recognition sequence for I-CeuI is present in the *E. coli* genome seven times in the essential 23S rRNA gene *rrl*, making resistance to this endonuclease through mutation exceedingly unlikely or impossible. Before proceeding with developing a terminal differentiation circuit incorporating I-CeuI, we first tested an inducible I-CeuI integrated on the genome and verified its capacity to induce growth cessation (Figure 4.1). Additionally, cells with this construct fail to form colonies when induced.

4.3 ETERNAL: A terminal differentiation architecture with scalable redundancy

With I-CeuI expression as the mechanism for inducing growth cessation and the existing differentiation activated T7 RNAP architecture, we have all the necessary components for developing a terminal differentiation circuit with scalable redundancy. We therefore propose ETERNAL: *E*xcision-activated *T*7RNAP *E*xpression

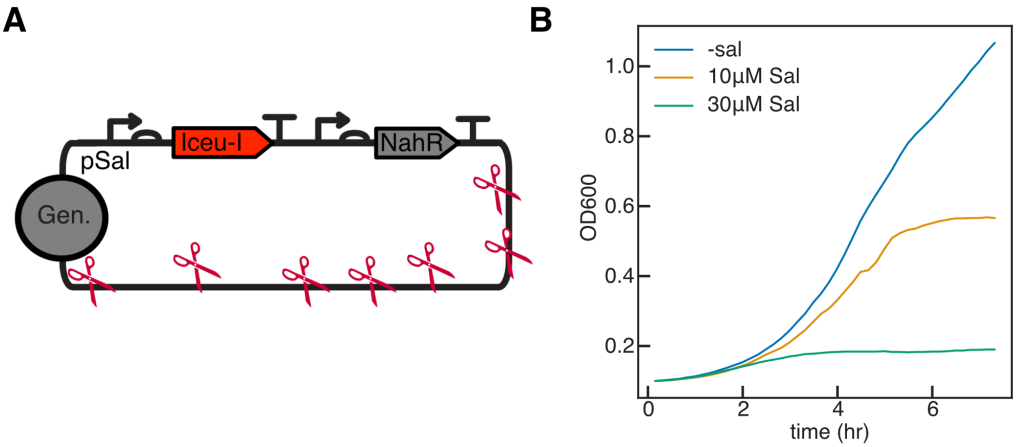

Figure 4.1: Inducible expression of I-ceuI homing endonuclease effectively stops growth. (A) Salicylate inducible I-CeuI integrated on the genome in a single copy with clonetegration [32]. (B) Cells outgrown in LB media for 8 h before dilution into fresh media in 300 μL cultures with varying concentrations of salicylate. OD600 monitored in Biotek plate reader.

with *R*edundant *N*uclease *A*ctuation of *L*imited-growth. The function of this proposed circuit is depicted in Figure 4.2. To scale the redundancy of this circuit, it is integrated in full in the genome n times. Each cassette would encode inducible Bxb1 integrase, recombination activated T7 RNAP, T7 RNAP-driven I-CeuI, and the transcription factors and accessory genes required for integrase and T7 RNAP induction (TFs araC and LacI^{AM}, and arabinose transporter araE). Because each component is present in each integration, n mutations would be required to fully break any step of the circuit. Induction with arabinose induces the expression of n copies of Bxb1 integrase, integrase can then recombine any of n cassettes to activate the expression of T7 RNAP, and T7 RNAP activates the expression of both the function of interest, and of all I-ceuI expression cassettes (Figure 4.2A). We further propose means of incorporating delay in the expression of I-CeuI through an intermediate step of T7 RNAP driven ϕC31 integrase expression, with ϕC31 then activating the expression of I-CeuI (Figure 4.2B). The amount of delay could be tuned with the expression strength of the second integrase, as well as by incorporating additional integrases in a cascade of expression as has recently been demonstrate [49].

As an initial demonstration of this circuit without incorporation of the integrase delay step, we genomically integrated the ETERNAL circuit (Figure 4.2A) lacking the I-CeuI expression cassette in a single copy (4.3A). We then integrated a T7 RNAP-driven I-CeuI expression cassette at a secondary location in the genome

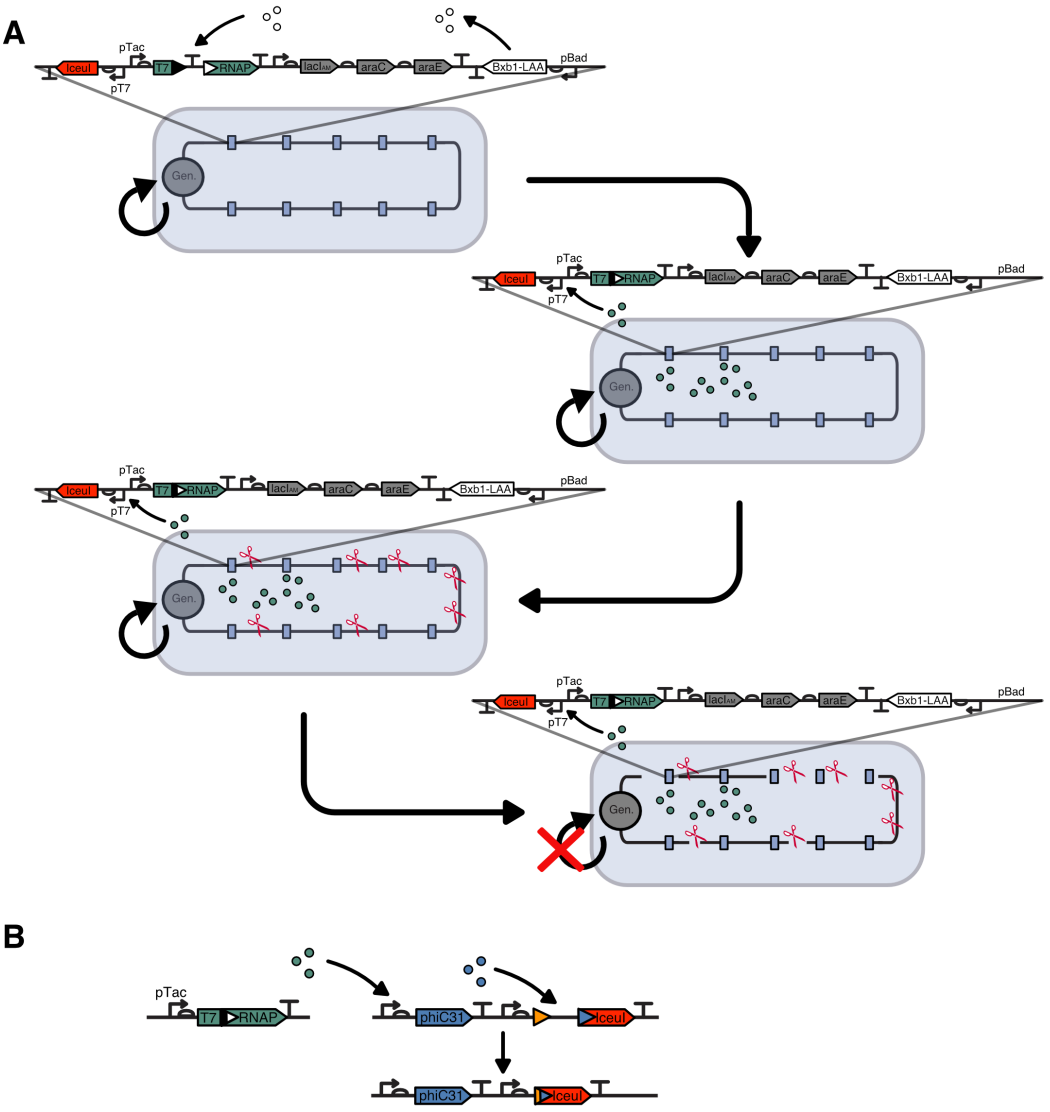

Figure 4.2: Excision-activated T7-RNAP Expression with Redundant Nuclease Actuation of Limited-growth (ETERNAL) circuit schematic. (A) Circuit schematic integrated on to the genome n times. (B) Addendum to ETERNAL circuit with T7 RNAP driven ϕC31 integrase which activates the expression of I-CeuI through recombination of its cognate attB and attP sites

(Figure 4.3B), and compared the growth and expression of T7 RNAP-driven GFP upon induction with arabinose (Figure 4.3C-D). Though the half maximal induction concentration for this promoter is reported as ~37μM [23], most of the dynamic range of arabinose concentration for inducing differentiation appears to be between 1 and 10 μM. To more fully use the dynamic range of the promoter for tuning differentiation rate, a lower strength RBS may desired. However, the differentiation rate is tunable, and as the presence of glucose appears to fully repress integrase expression and differentiation even with 100 μM arabinose, leaky integrase expression does not seem problematic with this expression strength. We clearly see GFP production dependent on integrase induction both with and without I-CeuI, and while there is minimal growth impact observed without I-CeuI, when I-CeuI is expressed we see evidence of complete or near complete differentiation in the plateau of cell growth with arabinose concentrations $\geq$10μM. As well, from the sharp increase in OD normalized GFP after the growth plateau, we see evidence that most GFP expression occurs after I-CeuI has caused growth arrest. Strikingly, OD normalized GFP expression is higher when I-CeuI is expressed at concentrations >10μM arabinose. Finally, when these cultures were diluted into the same conditions for a second batch growth, cells lacking I-CeuI again grow normally, while the growth of cells expressing T7 RNAP-driven I-CeuI is delayed increasingly with higher levels of integrase induction. This in conjunction with the diminished GFP production observed in the second growth indicates both that higher integrase induction results in more complete differentiation, and that with a single copy of the circuit integrated we are already likely seeing the expansion of cells that have incurred integrase or differentiation mutations.

In scaling this circuit to 2-4+ integrated copies, expression strength, and in particular expression leak must be considered. In an initial design of the circuit, we used the same promoter and transcription factor used in Chapters 2 and 3 (evolved P_{SalTTC} and $NahR^{AM}$ [23]). However, using a medium strength RBS (B0032) did not allow sufficiently high expression of Bxb1 integrase, and using a high strength RBS (B0034), though allowing sufficienty expression, resulted in leaky expression of the integrase in the absence of induction. Though some amount of leak could be tolerated, that which was observed would likely making constructing strains with additional integrations difficult. To rectify this problem, we instead used the P_{BAD} promoter as shown in Figures 4.2 and 4.3. The arabinose inducible P_{BAD} promoter is reported to have marginally lower leak than P_{SalTTC} [23], and additionally can be repressed by including glucose in the media. This repression in the presence of

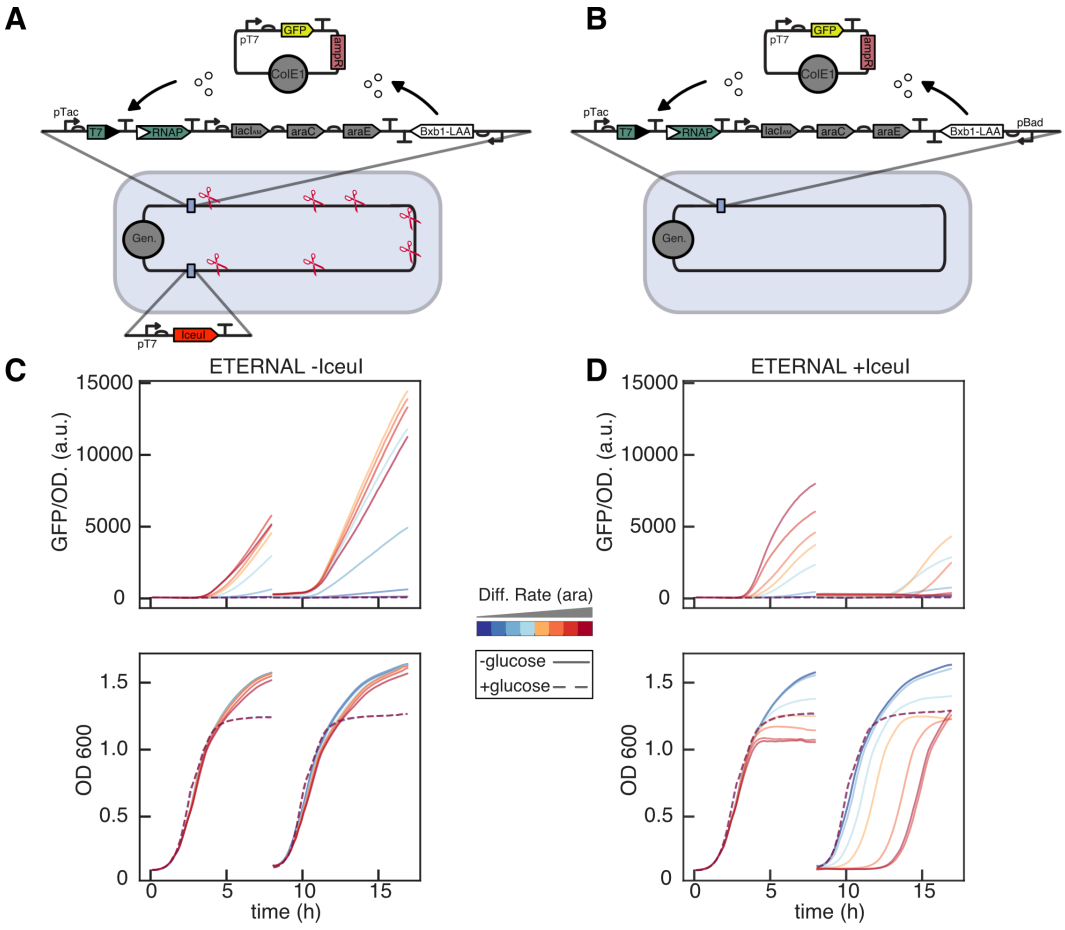

Figure 4.3: T7 RNAP-driven I-CeuI in ETERNAL allows terminal differentiation. (A) ETERNAL circuit as described in Figure 4.2 but lacking I-CeuI is integrated onto the genome in a single copy, and cells transformed with a ColE1 AmpR plasmid with T7 RANP-driven sfGFP. (B) As in (A) but with additional genomically integrated T7 RNAP driven I-CeuI cassette. (C) OD normalized GFP expression (top) and OD600 of cells (A) grown in 0.3mL LB carb with 10 μM IPTG and varying concentrations of arabinose (0, 1, 2.5, 5, 7.5, 10, 30, 100 μM). Cells diluted 1:50 after 8 hours into the same media conditions. (D) As in (C) but with circuit depicted in (B).

glucose will likely make construction of strains with numerous integrations more tractable than with P_{SalTTC}.

4.4 Model exploration of redundant architectures

Though experimentally the relationship of redundancy level to evolutionary stability has not yet been investigated for the terminal differenation ETERNAL circuit – or for differentiation or naive expression – beyond the two cassettes explored in Chapter 3, we can examine cases of three or more cassettes computationally. In Chapters 2 and 3 we conducted our experiments at relatively small scale ($\sim 10^9$ cells), and modeled the circuits in kind. However in considering extending the level of redundancy to 3+ cassettes, we must also consider the affect which population size will have on evolutionary stability. Though mutations are rare events, larger population sizes will sample these rare events more frequently, and therefore we might expect the level of redundancy to impact evolutionary stability differently at various scales of population size. For instance, we might expect the level of redundancy required to achieve a given level of evolutionary stability to be greater at larger population sizes. To investigate this explicitly, we extend the modeling framework developed in Chapter 3 to the cases of three and four cassettes, and examine the behaviour at scales equivalent to 100 μL (10^8 cells), 10 mL (10^{10} cells), and 1 L (10^{12} cells) cultures. As with modeling in Chapter 3, we also consider plasmid-based mechanisms driving loss of function, both plasmid mutations which disrupt the function, and plasmid loss with communal antibiotic degradation.

We first examine the effect of population size and redundancy on circuit performance in the absence of effects due to plasmid loss or plasmid mutation. Across all burden levels and all population sizes, increasing redundancy increases the evolutionary stability of the circuit as determined by total production over the course of 100 growths (800 hours, $\sim$1 month). We further see that increasing the population size tends to decrease the total production achieved, though not drastically, and not equally across redundancy levels. If we had modeled this system deterministally, there would be no difference in carrying capacity normalized production, and differences between different scales of population reveal effects due to the stochasticity of mutation. While for the single cassette case, a population size of 10^8 cells is sufficient to wash away stochastic effects, resulting in behaviour that appears deterministic, this is less true as the number of cassettes increases (Figure 4.4). The larger the population size is, the more deterministic the behavior of the system.

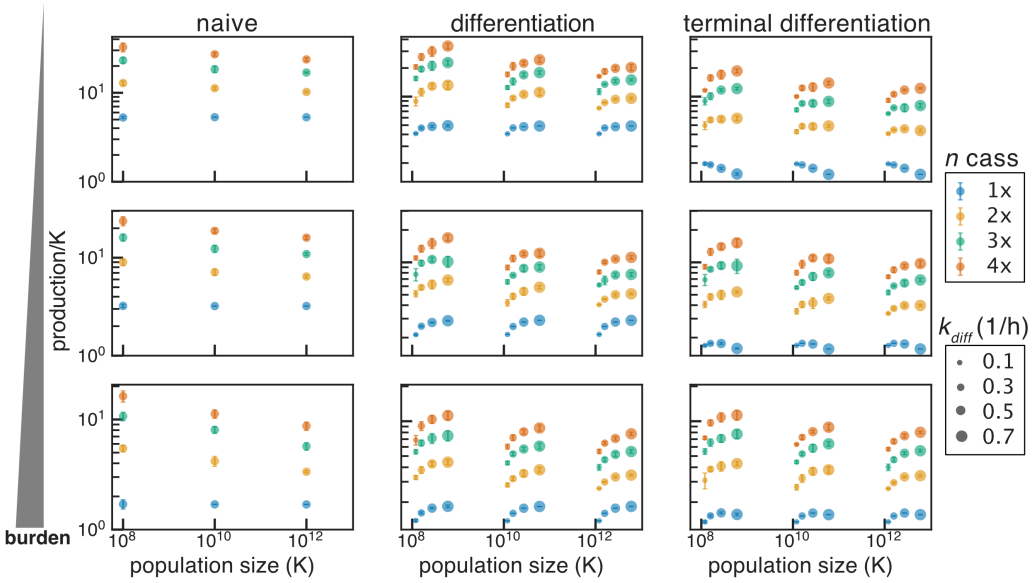

Figure 4.4: Stochastic simulations of naive, differentiation and terminal differentiation architectures without plasmid mutation or plasmid loss. Mean and standard deviation of 8 stochastic simulations normalized to population carrying capacity plotted for $K = 10^8$, 10^{10}, 10^{12}. Rates of burden mutation (k_{MB}), differentiation mutation (k_{MD}), and integrase mutation (k_{MI}) are 10^{-6} h^{-1}, antibiotic concentration is 100 μg/mL, antibiotic degradation rate ($V_{max} = 0$), MIC=1.1 μg/mL, plasmid loss rate ($k_{PL} = 10^{-4}$ h^{-1}), non-producer growth rate (μ_N=2 h^{-1}, ~20 minute doubling time). Simulations are of batch cultures diluted every 8 hours for 100 total growths. Columns are naive (left), differentiation (center) and terminal differentiation (right). Burden level increases down the row, with 30 percent (top), 50 percent (middle), and 70 percent (bottom). In all plots, color indicates copy number, and for differentiation circuits, differentiation rate is indicated by size, with rates 0.1, 0.3, 0.5, and 0.7 h^{-1} smallest to largest. Data were offset on the x-axis for clarity purposes.

As previously seen in Chapter 3, terminal differentiation is counterproductive with low burden functions, but becomes comparatively better with higher burden functions (Figure 4.4). Apart from this, it is also interesting to note the effect of differentiation rate in the case of terminal differentiation. In the case of a single cassette with lower burden (30 percent), increasing the differentiation rate beyond 0.1 h^{-1} decreases the total production. For higher burden levels there is an optimum above 0.1 h^{-1}, after which increasing it further decreases performance. However in the case of two or more cassettes, increasing the differentiation rate between 0.1 and 0.7 h^{-1} universally increases the total output. In our model we have made the assumption that differentiation rate is linearly dependent on the copy number of

functional integrase expression cassettes, meaning that a mutation of an integrase cassette will decrease the differentiation rate. We further assumed that for a given differentiation rate or integrase expression level, the differentiation rate of an individual cassette is not dependent on the state of the other cassettes, meaning that a differentiation mutation effecting one cassette will decrease the differentiation rate of the cell as a whole in a linear manner dependent on the copy number of the circuit. There therefore is additive selection for integrase and differentiation mutations, and the population of cells that have incurred such mutations will increase in abundance relative to cells that have not incurred such mutations due to a decrease in differentiation rate. This is similar to the naive case, where expression level and burden are also subject to additive selection. We will discuss this concept of additive selection further later in this chapter.

As in Chapter 3, we next consider the affect of plasmid loss in the context of shared antibiotic resistance, as well as plasmid mutation. Plasmid loss, a rare event that is common relative to the rate of mutation, is modeled with a rate of 10^{-4} h^{-1}, while plasmid mutation we model with a rate of 10^{-8} h^{-1}. Here we have made the simplifying assumption that a single mutational event makes all plasmids non-functional. In reality, any individual plasmid could mutate, making the rate of an initial mutation higher than would be for a single copy on the genome, and subsequent random plasmid partitioning and copy number fluctuations could then result in a cell with most or all plasmids mutated in relatively short order [11]. The lower rate we chose compared to the genomic mutation reflects a simplification where two processes are lumped together.

Though we observe similarities between plasmid loss and plasmid mutation, their different effects on both the naive and differentiation architectures are worth noting. With naive expression including plasmid loss and antibiotic degradation, the effect of population size and redundancy level is qualitatively similar to the case of neglecting this feature: Increasing population size decreases total production in a manner that depends on redundancy level (Figure 4.4-4.5). However, the benefit of increased redundancy is collapsed when considering plasmid loss, with this happening more severely at higher burden (Figure 4.6). As discussed by Yurtsev et.al. in their analysis of the dynamics of bacterial cheating in governing the population dynamics of β-lactam resistance plasmids, the distribution of cheaters and antibiotic degrading cells will tend towards that which maximizes population growth [38]. The consequence of this, as they note in the supplementary information, is that the more burdensome

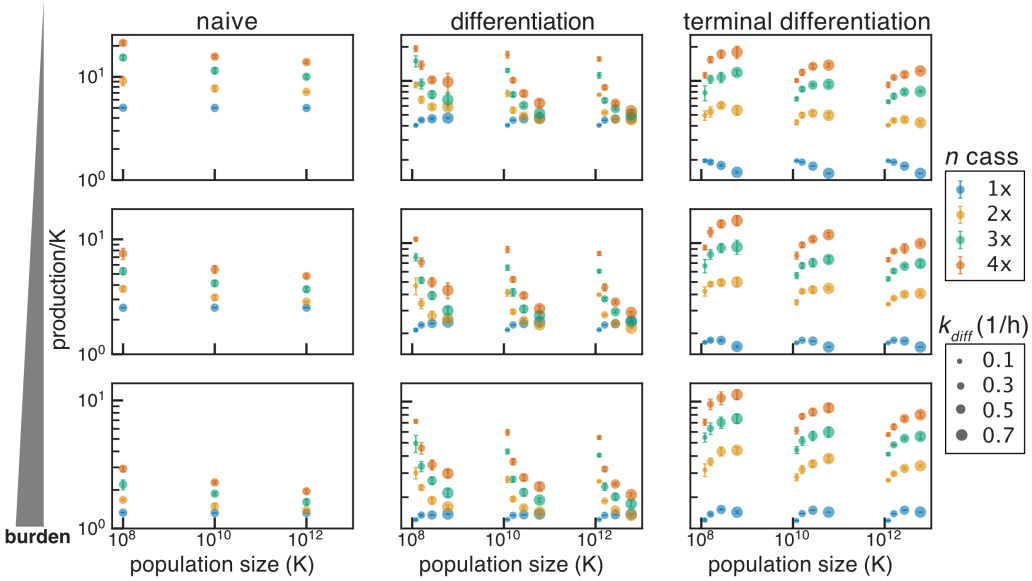

Figure 4.5: Stochastic simulations of naive, differentiation and terminal differentiation architectures with plasmid loss and antibiotic degradation. Mean and standard deviation of 8 stochastic simulations normalized to population carrying capacity plotted for $K = 10^8$, 10^{10}, 10^{12}. Rates of burden mutation (k_{MB}), differentiation mutation (k_{MD}), and integrase mutation (k_{MI}) are 10^{-6} h^{-1}; antibiotic concentration is 100 μg/mL; antibiotic degradation rate ($V_{max} \approx 1.26\text{x}10^{-5}$); MIC=1.1 μg/mL; plasmid loss rate ($k_{PL} = 10^{-4}$ h^{-1}); non-producer growth rate (μ_N=2 h^{-1}, ~20 minute doubling time). Simulations are of batch cultures diluted every 8 hours for 100 total growths. Columns are naive (left), differentiation (center) and terminal differentiation (right). Burden level increases down the row, with 30 percent (top), 50 percent (middle), and 70 percent (bottom). In all plots, color indicates copy number, and for differentiation circuits, differentiation rate is indicated by size, with rates 0.1, 0.3, 0.5, and 0.7 h^{-1} smallest to largest. Data were offset on the x-axis for clarity purposes.

having the plasmid is, the larger the fraction of the population will be cheaters. In the context of our model, this effect reveals itself in impacting the production of higher burden expression more severely than lower burden. The case of plasmid mutation, however, is strikingly different. Given the mutation rates we have chosen for plasmid mutation (10^{-8} h^{-1}) and genomic mutations (10^{-6} h^{-1}), we naturally see a benefit of redundancy at the small population size of 10^8 cells. However at larger population sizes, redundancy beyond two copies does nothing as the plasmid mutation will occur and take over before two or more of the slower mutations occur.

The differentiation architecture without limited division also fairs differently be-

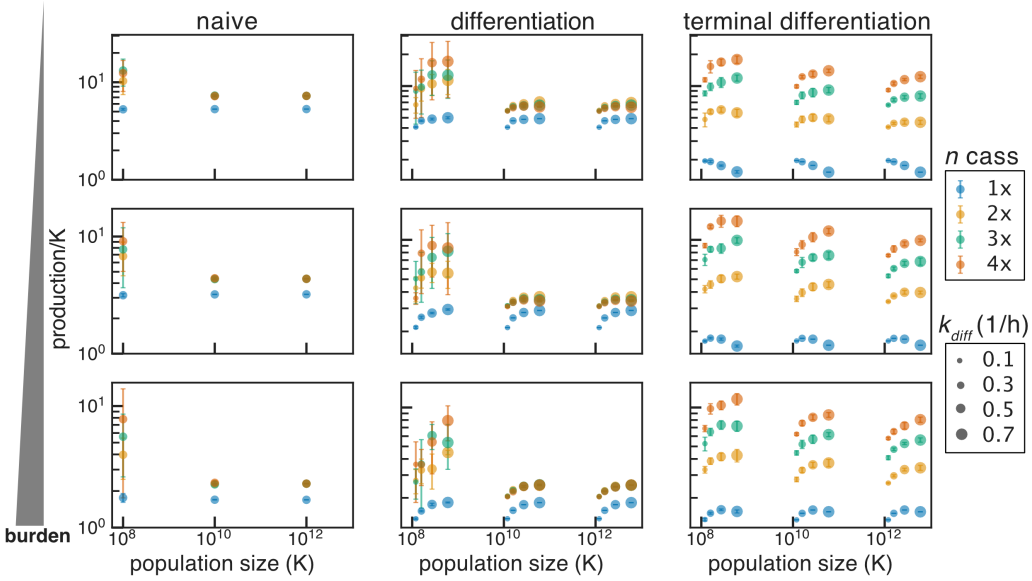

Figure 4.6: Stochastic simulations of naive, differentiation and terminal differentiation architectures with plasmid mutation. Mean and standard deviation of 8 stochastic simulations normalized to population carrying capacity plotted for $K = 10^8, 10^{10}, 10^{12}$. Rates of burden mutation (k_{MB}), differentiation mutation (k_{MD}), and integrase mutation (k_{MI}) are 10^{-6} h^{-1}; antibiotic concentration is 100 μg/mL; antibiotic degradation rate ($V_{max} = 0$), plasmid loss rate ($k_{PL} = 10^{-4}$ h^{-1}), non-producer growth rate (μ_N=2 h^{-1}, ~20 minute doubling time). Simulations are of batch cultures diluted every 8 hours for 100 total growths. Columns are naive (left), differentiation (center) and terminal differentiation (right). Burden level increases down the row, with 30 percent (top), 50 percent (middle), and 70 percent (bottom). In all plots, color indicates copy number, and for differentiation circuits, differentiation rate is indicated by size, with rates 0.1, 0.3, 0.5, and 0.7 h^{-1} smallest to largest. Data were offset on the x-axis for clarity purposes.

tween the cases of plasmid loss and plasmid mutation. In the limit of simultaneous differentiation of all cells, differentiation would be equivalent to naive expression. We would expect therefore that this architecture would behave more similarly to the naive case as the differentiation rate increases. This is indeed what we see when considering plasmid loss and antibiotic degradation (Figure 4.5-4.6). In contrast to what we observed without considering this feature, increasing the differentiation rate for all but the single cassette case acts to decrease the total production. As well, similar to the naive case, incorporating plasmid mutation collapses the total production achieved for number of cassettes greater than one (Figure 4.6).

In stark contrast to differentiation and naive designs, terminal differentiation is

completely unaffected by plasmid loss or plasmid mutation (Figure 4.5-4.6). This is true across all population scales and all redundancy levels. Though this naturally follows from the robustness of the terminal differentiation architecture to burden mutations we discussed in Chapters 2 and 3, it is important to note nonetheless. As a direct consequence of this, antibiotic selection for plasmid maintenance, particularly in the case of plasmids with dedicated partitioning systems where plasmid loss is less frequent, may not be necessary. As well, while genomic integration is now frequently the preferred method for bioproduction to improve copy number stability and obviate the need for antibiotic selection, terminal differentiation could provide both of these benefits with the convenience and ease of plasmid-based expression.

Additive selection drives successive mutations

In the simulations discussed thus far, there has been additive selection present in all circuits. In the naive case, the additive selection results from copy number dependent expression of the function. As each individual mutation decreases the burden of expression, cells with a single mutation will increase in abundance in the population due to the increased growthrate, providing fodder for accumulating additional mutations. In the case of differentiation, this additive selection can act both on burden mutations as in the naive case, as well as on differentiation or integrase mutations. In the latter case as discussed previously, both of these mutations result in a decreased differentiation rate which will result in an increased abundance of the genotype in the population. In the case of terminal differentiation however, the additive selection only exists for integrase and differentiation mutations. We see this intuition bear out in examining the winning genotypes that are most abundant at the end of the simulations. Without considering plasmid mutation and shared antibiotic resistance, in the naive case the winning genotype has universally mutated all T7 RNAP cassettes. With differentiation, we see a mixture of genotypes containing differentiated cassettes with burden mutations, cassettes with differentiation mutations, and integrase mutations. With terminal differentiation we do not see any burden mutations in cassettes in the progenitor or differentiated state, but only see integrase mutations and cassettes with differentiation mutations.

When we consider plasmid loss and shared antibiotic resistance, although this mechanism can negatively impact production, the expansion of cheaters is transitory and the end result in terms of winning genotypes is the same. However, with plasmid mutation for copy number two or greater, the plasmid mutation wins out universally at all but the smallest population size (10^8) for naive and differentiation (but not

terminal differentiation) architectures, where cells having mutated all cassettes can win out even with four cassettes. Because we have made the simplifying and inaccurate assumption combining the events of plasmid mutation and complete loss of function through plasmid partitioning into a single event with rate 10^{-8} h^{-1}, experimentally this population size affect with 10^8 would likely not hold true.

Because additive selection for mutations universally results in the successive accumulation of mutations in all of our circuits, it follows that regardless of the copy number in any of the circuit designs we describe here, the function will eventually be eliminated from the population. Certainly increasing the copy number will both increase the number of mutations needed to fully mutate the function and decrease the marginal fitness gain from each mutation, yielding a more evolutionarily stable function; but it will inevitably fail. Evolution will win out. However, practically speaking redundancy in all of these architectures can make meaningful improvements to the stability of engineered functions in regards to the application of synthetic biology to bioproduction and the deployment of engineered bacteria in the environment.

Though this is true, we imagine the case where selection is not additive, but recessive. Instead of individual mutations providing an increase in fitness which drives its expansion in the population, all copies of the function would be required to mutate in order to provide any difference in fitness or decrease in production. Halleran et.al. examined the impact of additive versus recessive selection in relation to the impact of plasmid partitioning on evolutionary stability, and demonstrated that only in the case of perfect partitioning as would be achieved by integrating multiple copies on the genome does recessive selection reveal its benefit [11]. We re-examine the impact of recessive versus additive selection here as a function of population size and copy number with and without the plasmid effects previously discussed (Figure 4.7). Here we see in the absence of plasmid mutation or communal antibiotic resistance, recessive selection universally outperforms additive selection (no distinction between additive and recessive with copy number 1). We further see the impact of copy number in the case of recessive selection as the size of the population varies. At low population size, 3 or more cassettes is sufficient, and 2 nearly so, to prevent the takeover of a fully mutated genotype in the course of 100 growths. Without the expansion of an initial mutant mediated through additive selection, cells with one mutation only gradually accumulate, making the population size which can incur additional mutations comparatively small. However, as the

population size increases, we see failure at higher copy numbers, with all but the case of 4 cassettes revealing fully mutated cells with a carrying capacity of 10^{12} cells. We also note that the degree of burden largely does not matter in this case because the growth rate of producer cells is sufficient to prevent washout due to excessive dilution. As well, the dominant factor here is the time required to generate the first fully mutated cell, at which point it will exponentially accumulate in the population. When we consider antibiotic degradation and plasmid loss, we see that recessive selection benefits the evolutionary stability similarly, though with burden decreasing the production achieved through its affect on the fraction of cheaters and degraders. Finally, when we consider plasmid mutation, as we might expect, recessive selection ceases to provide a benefit (Figure 4.7, right column).

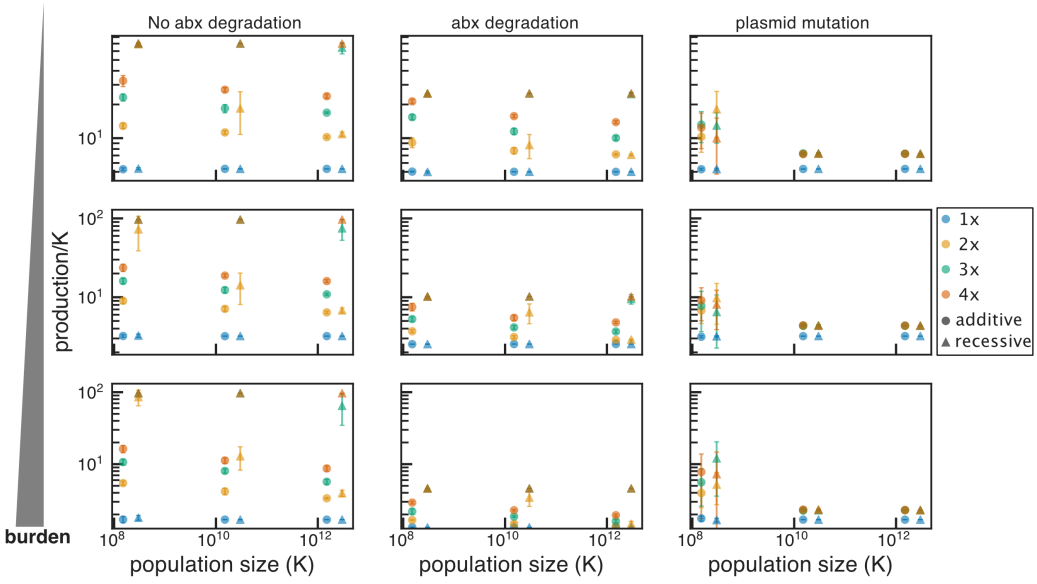

Figure 4.7: Comparison of naive expression with additive and recessive selection. Mean and standard deviation of 8 stochastic simulations normalized to population carrying capacity plotted for $K = 10^8, 10^{10}, 10^{12}$. Simulations are of batch cultures diluted every 8 hours for 100 total growths. Parameters for simulations without antibiotic degradation (left column), with antibiotic degradation (center column), and with plasmid mutation (right column) as described in Figures 4.4-4.7. Burden level increases down the row, with 30 percent (top), 50 percent (middle), and 70 percent (bottom). In all plots, color indicates copy number. Additive selection (circles), and recessive selection (triangles). Data were offset on the x-axis for clarity purposes.

4.5 Can recessive selection be achieved in a terminal differentiation circuit?

Though we clearly see the benefit of recessive selection with naive expression, this architecture is susceptible to plasmid-based effects that decrease production and/or lead to complete circuit failure. Though the differentiation architecture without limiting the division of differentiated cells is susceptible to these same effects, terminal differentiation is not. It follows that if we could remove the dependence of differentiation rate on the number of functional integrase cassettes and the number of cassettes in the progenitor state, we could achieve recessive selection and drastically improve the evolutionary stability of the architecture as demonstrated with naive expression. In this section, we consider if achieving or approaching recessive selection is feasible, and what would be required to do so.

The first requirement to achieving this is to remove the dependence of differentiation rate on the copy number of functional integrase expression cassettes. How could this be achieved? Rationally it seems that in order to accomplish this, we would need to maintain the same expression level of integrase regardless of the number of functional and mutated integrase copies. Segall-Shapiro et.al. demonstrated that copy number independent expression could be achieved by expressing a TALE repressor at the same copy number as a gene of interest, with increase in the copy number of the gene of interest being compensated by a proportional increase in the non-cooperative repressor [50]. When considering implementing this or something similar in the context of the terminal differentiation circuit, however, we realize that this copy number independent expression would not be robust to mutation. While perhaps copy number independent expression could be achieved for varying numbers of fully functional integrase cassettes, mutations disrupting integrase expression would not effect the expression of the repressor and would still be susceptible to additive selection. If a strategy enabling copy number independent selection were to be effective, it must be robust to mutations disrupting integrase expression. Though there may additional possible strategies we have not considered to approach this, the only route to achieve mutational robustness we can think of is to have the integrase directly regulate its own expression through negative autoregulation. As integrases are necessarily DNA binding proteins, this seemed a feasible function for these proteins to perform. Before investigating experimentally if this approach is fertile, we consider what we can achieve with negative autoregulation. We first write an equation describing the dynamics of an arbitrary gene X with negative autoregulation at copy number c, dimensionless production rate β, and hill coefficient n. Here the concentration of X has been non-dimensionalized by its binding constant:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \frac{\beta\, c}{1 + X^n} - X. \tag{4.1}$$

At steady state, we have that

$$X^{n+1} + X = \beta c. \tag{4.2}$$

In considering the meaning of this, we note that if n is high, this first term will dominate the left-hand side if X is large, and the steady state concentration of X will scale with $c^{\frac{1}{n+1}}$. In the limit of infinite cooperativity, there is therefore copy number independent expression, with higher n decreasing copy number dependence. Because X is in units of its binding constant, this means that the concentration of X must be in the concentration regime where it is effectively repressing itself. Though repressors are not infinitely cooperative and there is no guarantee of copy number independent expression, this strategy could reduce dependence on copy number in a manner that is mutationally robust if in the necessary parameter regime, and perhaps reduce the effect of additive selection.

The first and foremost criteria for this to work is that Bxb1 can function as a repressor. Having fulfilled that, it ideally has a hill coefficient greater than 1. In order to use Bxb1 as a repressor, we would require the site not be susceptible to recombination. Bxb1, in the absence of a directionality factor, catalyzes recombination unidirectionally, converting attB and attP sites to attL and attR sites, with the latter pair being catalytically dead. We therefore would use an attL or attR site to mediate repression. Fortuitously, Bxb1 binds as a dimer with stronger affinity to attL and attR sites (Kd 15 nM) than to the attB and attP sites (Kd 70 nM) [34]. This fact suggests that if the concentration of integrase is in the regime where recombination is occurring, it will likely also be in the regime where it is able to mediate repression. As well, that Bxb1 binds as a dimer further suggests that the hill coefficient will be greater than 1.

In an initial test of the capacity for Bxb1 to function as a repressor, we characterized the expression of a consitutive sfGFP containing the attL or attR site between the RBS and promoter driving the fluorescent protein with various induction levels of the integrase (Figure 4.8). For both the attL and attR sequence, we see GFP fluorescence decreasing as a function of integrase expression, with the attR sequence having both

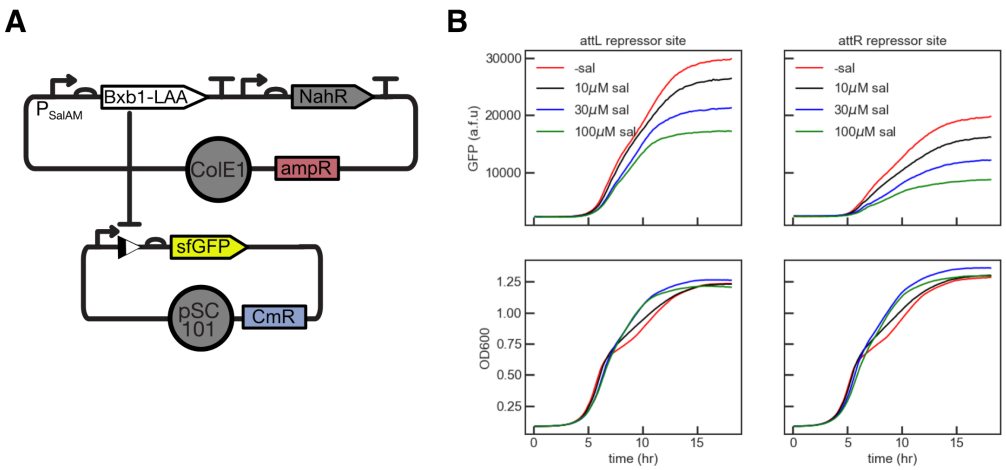

Figure 4.8: Bxb1 can mediate repression through attL and attR sites. (A) Bxb1 repression circuit schematic. Bxb1 integrase is inducibly expressed from a low copy ColE1 plasmid with AmpR. Constitutively expressed sfGFP cassette on on a pSC101 plasmid with chloramphenicol resistance contains either the attL or attR integrase attachment site between the promoter and RBS. (B) Time course GFP fluorescence (top) and OD600 (bottom) of cells containing the constitutive sfGFP construct with attL (left) or attR (right) attachment site with varying concetrations of salicylate.

lower fluorescence in the uninduced case, and a larger fold change. Though this does not guarantee this will function well with an inducible promoter or in the context of our differentiation circuit to reduce copy number dependence, it is give hope that this strategy could be successful.

If we assume that mutationally robust copy number independent expression of integrase could be achieved through negative autoregulation by Bxb1, would this be sufficient to achieve recessive selection in the context of terminal differentiation? As we discussed previously, to achieve recessive selection, we must eliminate the dependence of differentiation rate on both the copy number of functional integrase expression cassettes, as well as the number of cassettes in the progenitor state. With this strategy, we have just addressed the former. If we model the case where we have just achieved copy number independent expression of integrase, we see no benefit in the context of terminal differentiation. However, if we examine the genotypes which result from the evolution of 100 generations, we see the result of our game of whack-a-mutation. Instead of seeing a mixture of genotypes containing both differentiation and integrase mutations across all population sizes and redundancy levels, it instead follows the trend observed in the naive case of recessive selection. However, instead

of resulting in prolonged function at redundancy levels and population sizes which preclude or make unlikely the accumulation of successive mutations without additive selection, we see the selection for differentiation mutations which still have additive selection.

The additive selection for differentiation mutations we observe here results from the implicit assumption we made in our model that the state of each cassette does not affect the recombination rate of any other cassette. However, in actuality there may be some dependence, as the dynamics governing the free concentration of integrase and the rate of recombination are complex, depending on both the expression level of integrase and on the copy number of integrase attachment site substrates. Though when Artavanis et.al. through modeling and in cell-free TX-TL expression demonstrate a decrease in the delay of recombination as the concentration of substrate increases, they do so in a regime of high integrase expression where all recombination occurs within two hours [51]. In the differentiation system we have implemented here, the differentiation rate is substantially lower, and dependent on the steady state concentration of integrase protein. In this regime, sequestration of integrase monomers or dimers by integrase attachment sites may lower the effective concentration of integrase present in the cell, reducing the rate of differentiation of other cassettes. Because presumably differentiation mutations disrupt one or both integrase attachment sites, this may be a built in mechanism of compensating to some degree for differentiation mutations by increasing the effective concentration of integrase. Though this an intriguing potential mechanism, investigating its potential to affect the performance of our differentiation circuit would require a detailed mechanistic model and we will not address it in the context of this thesis. As well, without experimentally investigating a circuit such as ETERNAL at higher levels of redundancy, we will not know a priori if the assumption we made in our modeling which led to this additive selection for differentiation mutations holds true. It seems almost certain that there will be additive selection for integrase mutations, therefore if we perform deep sequencing on populations of cells with increasing numbers of ETERNAL cassettes after long-term evolution and see prevalence of explicitly differentiation mutations which disrupt integrase attachment sites, we may take this as evidence in favor of additive selection for such mutations. If on the other hand we see exclusively integrase mutations, this means only that on the spectrum of additive to recessive selection, integrase mutations fall closer to the former than differentiation mutations. Mutation and natural selection will find the lowest hanging fruit first.

In final thoughts here to addressing this, we consider the underlying reason for this additive selection: Each differentiation cassette relies on a common pool of integrase which we assume is induced at a constant level. The integrase concentration sets the total differentiation rate, and each differentiation cassette contributes to this rate. Perhaps one route for circumventing or delaying failure from additive selection is for each cassette to have its own pool of integrase. As many orthogonal integrases exist and have been demonstrated functional in *E. coli* [52], using distinct integrases for each cassette, or having n cassettes be split between some number of distinct integrases, could aid in addressing this foreseeable limitation. While additive selection would still remain a factor when viewing the differentiation rate of the cell as a whole, the distinct integrases could be induced sequentially after mutation and additive selection have destroyed the previous differentiation circuit. As well, experimental factors, such as varying the induction of differentiation over time [53], continuously introducing fresh ETERNAL cells, or maximally inducing differentiation periodically and re-inoculating at a timescale which would preclude the completion of additive selection, could further improve the performance of the ETERNAL system in a setting of continuous bioproduction.

Model implementation

Models for naive, differentiation, and terminal differentiation architectures were generated as described in Chapter 3, and extended to the cases of 3 and 4 cassettes. For models which included antibiotic degradation, the 5x rate described in Chapter 3 was used ($\sim$1.26x10^{-5} μg/cell/h [38]), and plasmid loss rate k_{PL} of 10^{-4} h^{-1}. For the case of plasmid mutation, simulations were modeled without antibiotic or antibiotic degradation, with $k_{PL} = 10^{-8}$ h^{-1}. For all simulations the non-producer growth rate was set to 2 h^{-1}, corresponding to a doubling time of approximately 20 minutes. For naive expression, the burden level modeled describes the growth rate of cells with all cassettes intact, with burden mutations proporionately decreasing production and burden as described in Chapter 3. For simulation of naive expression with recessive selection, the growth rate and production rate are unaffected except when all cassettes had mutated, at which point production rate is 0 and growth rate is that of non-producer cells. Code for Python ODE models was generated with custom Python script which allows user specification of number of cassettes, burden, circuit type, mutation rates etc. and will be made available. Simulations were ran on the Caltech HPC.

4.6 Materials and methods

Strains and plasmids

The wild-type JS006 strain was used as the base strain for all genome integrations. For construction of the ETERNAL circuit depected in Figure 4.3, the pIT5 system, an improved version of clonetegration, was used to genomically integrate the constructs [54]. I-CeuI was ordered as a gGlock from IDT, and transcriptional units for the circuit were generated with 3G assembly ([42]), and assembled with the a PCR linearized pIT5 plasmid backbone with Gibson assembly. All circuits constructed were assembled with 3G assembly.

Plate reader experiments

Plate reader experiments were all conducted in LB media with appropriate antibiotics. Cells were grown in 96-well square-well plate (Brooks MGB096-1- 2-LG-L) at 37°C with maximum-speed linear shaking in a BioTek Synergy H1m. Cell density (OD600) and GFP fluorescence (485/515 nm) were measured every 10 minutes.

4.7 Acknowledgements

I would like to thank the Sim Lab at the University of California Irvine for welcoming into their lab space to pursue this work.

Chapter 5

CONCLUSION

In this thesis, we have developed an integrase-mediated differentiation architecture which enables the evolutionary stability of burdensome or even toxic functions to be improved. When differentiation is combined with limiting the capacity of differentiated cells to grow indefinitely, the critical feature of robustness to mutations or plasmid-based effects which inactivate the function of interest emerges. In much of this thesis, we considered how to improve this terminal differentiation architecture by incorporating redundancy and developed circuit architectures which would allow this redundancy to be scaled. While we first and foremost considered the benefit of differentiation and terminal differentiation circuits for improving evolutionary stability, such an architecture has also been proposed for improving the efficiency of bioproduction on its own without consideration of evolutionary forces. Posed as a control optimization problem for maximizing bioproduction with differentiation being controlled by light induction, they demonstrated differentiation to be superior to induction [53]. We also naturally see the applicability of differentiation strategies to metabolic engineering. While tools of metabolic engineering like flux balance analysis can inform genetic modification of strains to improve the yield of valuable chemicals [55], these strategies naturally must be concerned with the growth of the organism to some degree. However, with a strategy of differentiation where producer cells are continuously replenished and sacrificed, genomic and metabolic knobs could be tuned to maximize yield without regard for the long-term viability of the cells. CRISPR/Cas systems have been demonstrated to allow activation and repression in $.coli$, and have been applied in metabolic engineering efforts, and could be co-opted in this context [56, 57]. Though these examples are not exhaustive, they do serve to highlight the potential of terminal differentiation architectures to enable new directions and applications of synthetic biology.

As this thesis concludes, I note a common theme that appears in many quotes, so I will not provide any direct attribution: Evolve or die. In the course of the work in this thesis, it seems that this axiom works the other way as well: Die or evolve. Any error-prone replicating system that does not die will inevitably evolve, and the only sure-fire way to stop your bacteria from evolving is surely fire (or bleach). However, while evolution is indeed an unstoppable force, we can do a pretty good

job at slowing it down.

BIBLIOGRAPHY

1. Glick, B. R. Metabolic load and heterlougos gene expression. *Biotechnology Advances* **13,** 247–261 (1995).

2. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering* **4,** 12. ISSN: 1754-1611. http://www.jbioleng.org/content/4/1/12 (2010).

3. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* **26,** 787–793. ISSN: 10870156 (2008).

4. Jack, B. R. *et al.* Predicting the Genetic Stability of Engineered DNA Sequences with the EFM Calculator. *ACS Synthetic Biology* **4,** 939–943. ISSN: 21615063 (2014).

5. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Molecular BioSystems* **10,** 1668–1678. ISSN: 17422051 (2014).

6. Pósfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312,** 1044–1046. ISSN: 00368075. arXiv: arXiv:1011.1669v3 (2006).

7. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R. & Pósfai, G. Low-mutation-rate, reduced-genome Escherichia coli: An improved host for faithful maintenance of engineered genetic constructs. *Microbial Cell Factories* **11,** 1–13. ISSN: 14752859 (2012).

8. Sleight, S. C. & Sauro, H. M. Visualization of evolutionary stability dynamics and competitive fitness of Escherichia coli engineered with randomized multigene circuits. *ACS Synthetic Biology* **2,** 519–528. ISSN: 21615063 (2013).

9. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nature Methods* **15.** ISSN: 1548-7091 (2018).

10. Tyo, K. E., Ajikumar, P. K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnology* **27,** 760–765. ISSN: 10870156 (2009).

11. Halleran, A. D., Flores-bautista, E. & Murray, R. M. Quantitative characterization of random partitioning in the evolution of plasmid-encoded traits, 1–14 (2019).

12. Yang, S., Sleight, S. C. & Sauro, H. M. Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits. *Nucleic Acids Research* **41,** 1–7. ISSN: 03051048 (2013).

13. Minty, J. J. *et al.* Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences* **110,** 14592–14597. ISSN: 0027-8424. https://www.pnas.org/content/110/36/14592 (Sept. 2013).

14. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnology* **33,** 377–383. ISSN: 1087-0156. http://dx.doi.org/10.1038/nbt.3095 (2015).

15. Roell, G. W. *et al.* Engineering microbial consortia by division of labor. *Microbial Cell Factories* **18,** 1–11. ISSN: 1475-2859. https://doi.org/10.1186/S12934-019-1083-3 (2019).

16. West, S. A. & Cooper, G. A. Division of labour in microorganisms: An evolutionary perspective. *Nature Reviews Microbiology* **14,** 716–723. ISSN: 17401534. http://dx.doi.org/10.1038/nrmicro.2016.111 (2016).

17. Van Gestel, J., Vlamakis, H. & Kolter, R. Division of Labor in Biofilms: the Ecology of Cell Differentiation. *Microbiology Spectrum* **3,** 1–24. ISSN: 2165-0497 (2015).

18. Kumar, K., Mella-Herrera, R. A. & Golden, J. W. Cyanobacterial heterocysts. *Cold Spring Harbor Perspectives in Biology* **2,** 1–19. ISSN: 19430264 (2010).

19. Landy, A. Dynamic, Structural, and Regulatory Aspects of lambda Site-Specific Recombination. *Annual Review of Biochemistry* **58,** 913–941. ISSN: 0066-4154. http://www.annualreviews.org/doi/10.1146/annurev.bi.58.070189.004405 (June 1989).

20. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 8884–8889. ISSN: 00278424 (2012).

21. Hsiao, V., Hori, Y., Rothemund, P. W. & Murray, R. M. A population-based temporal logic gate for timing and recording chemical events. *Molecular Systems Biology* **12,** 869 (2016).

22. Guiziou, S., Ulliana, F., Moreau, V., Leclere, M. & Bonnet, J. An Automated Design Framework for Multicellular Recombinase Logic. *ACS Synthetic Biology* **7,** 1406–1412. ISSN: 21615063 (2018).

23. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology* **15,** 196–204. ISSN: 15524469. http://dx.doi.org/10.1038/s41589-018-0168-3 (2019).

24. McGinness, K. E., Baker, T. A. & Sauer, R. T. Engineering Controllable Protein Degradation. *Molecular Cell* **22,** 701–707. ISSN: 10972765 (2006).

25. Rakowski, S. A. & Filutowicz, M. Plasmid R6K replication control. *Plasmid* **69,** 231–242. ISSN: 0147619X. https://linkinghub.elsevier.com/retrieve/pii/S0147619X13000267 (May 2013).

26. Borkotoky, S. & Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *International Journal of Biological Macromolecules* **118,** 49–56. ISSN: 18790003. https://doi.org/10.1016/j.ijbiomac.2018.05.198 (2018).

27. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the shine - dalgarno sequence and the translation initiation codon of escherichia coli m RNAs. *Nucleic Acids Research* **22,** 4953–4957. ISSN: 03051048 (1994).

28. Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. & Voigt, C. A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Molecular Systems Biology* **10,** 742–742. ISSN: 1744-4292 (2014).

29. Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth. *Molecular Cell* **38,** 758–767. ISSN: 10972765. http://dx.doi.org/10.1016/j.molcel.2010.04.015 (2010).

30. Jang, C. W. & Magnuson, T. A Novel Selection Marker for Efficient DNA Cloning and Recombineering in E. coli. *PLoS ONE* **8,** 1–7. ISSN: 19326203 (2013).

31. Wahl, M. E. & Murray, A. W. Multicellularity makes somatic differentiation evolutionarily stable. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 8362–8367. ISSN: 10916490 (2016).

32. St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. *ACS Synthetic Biology* **2,** 537–541. ISSN: 21615063 (2013).

33. Iverson, S. V., Haddock, T. L., Beal, J. & Densmore, D. M. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology* **5,** 99–103. ISSN: 21615063 (2016).

34. Ghosh, P., Pannunzio, N. R., Hatfull, G. F. & Gottesman, M. Synapsis in phage Bxb1 integration: Selection mechanism for the correct pair of recombination sites. *Journal of Molecular Biology* **349,** 331–348. ISSN: 00222836 (2005).

35. Stevens, A. J. *et al.* Design of a Split Intein with Exceptional Protein Splicing Activity. *Journal of the American Chemical Society* **138,** 2162–2165. ISSN: 15205126 (2016).

36. Wein, T., Hülter, N. F., Mizrahi, I. & Dagan, T. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature Communications* **10,** 1–13. ISSN: 20411723. http://dx.doi.org/10.1038/s41467-019-10600-7 (2019).

37. Chen, S., Larsson, M., Robinson, R. C. & Chen, S. L. Direct and convenient measurement of plasmid stability in lab and clinical isolates of E. coli. *Scientific Reports* **7,** 1–11. ISSN: 20452322 (2017).

38. Yurtsev, E. A., Chao, H. X., Datta, M. S., Artemova, T. & Gore, J. Bacterial cheating drives the population dynamics of cooperative antibiotic resistance plasmids. *Molecular Systems Biology* **9,** 1–7. ISSN: 17444292. http://dx.doi.org/10.1038/msb.2013.39 (2013).

39. Doherty, A. J., Connolly, B. A. & Worrall, A. F. Overproduction of the toxic protein, bovine pancreatic DNasel, in Escherichia coli using a tightly controlled T7-promoter-based vector. *Gene* **136,** 337–340. ISSN: 03781119 (1993).

40. Wang, Z. K. *et al.* Improving the Intensity of Integrated Expression for Microbial Production. *ACS Synthetic Biology* **10,** 2796–2807. ISSN: 21615063 (2021).

41. Kato, Y. Extremely low leakage expression systems using dual transcriptional-translational control for toxic protein production. *International Journal of Molecular Sciences* **21.** ISSN: 14220067 (2020).

42. Halleran, A. D., Swaminathan, A. & Murray, R. M. Single Day Construction of Multigene Circuits with 3G Assembly. *ACS Synthetic Biology* **7,** 1477–1480. ISSN: 21615063 (2018).

43. Kunitz, M. Crystallin desoxyribonuclease. Isolation and General Properties. Spectrophotometric method for the measurement of desoxyribonuclease activity. *The Journal of Gernal Physiology,* 349–362 (1949).

44. Vogel, B. & Frantz, S. Determination of DNase activity by degradation of ethidium bromide–DNA complexes using a fluorescence plate reader. *Analytical Biochemistry* **471,** 73–79. ISSN: 10960309. http://dx.doi.org/10.1016/j.ab.2014.11.013 (2015).

45. You, L., Cox, R. S., Weiss, R. & Arnold, F. H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428,** 868–871. ISSN: 00280836 (2004).

46. McCardell, R. D., Huang, S., Green, L. N. & Murray, R. M. Control of bacterial population density with population feedback and molecular sequestration. *bioRxiv.* ISSN: 2692-8205 (2017).

47. McCardell, R. D., Pandey, A. & Murray, R. M. Control of density and composition in an engineered two-member bacterial community. *bioRxiv.* ISSN: 2692-8205 (2019).

48. Fan, C. *et al.* Chromosome-free bacterial cells are safe and programmable platforms for synthetic biology. *Proceedings of the National Academy of Sciences,* 201918859. ISSN: 0027-8424. http://www.pnas.org/lookup/doi/10.1073/pnas.1918859117 (2020).

49. Kim, T., Weinberg, B., Wong, W. & Lu, T. K. Scalable recombinase-based gene expression cascades. *Nature Communications* **12,** 1–9. ISSN: 20411723. http://dx.doi.org/10.1038/s41467-021-22978-4 (2021).

50. Segall-Shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. *Nature Biotechnology* **36,** 352–358. ISSN: 15461696 (2018).

51. Artavanis, G., Hsiao, V., Hayes, C. & Murray, R. The role of single occupancy effects on integrase dynamics in a cell-free system. *bioRxiv,* 059675 (2016).

52. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nature Methods* **11,** 1261–1266. ISSN: 15487105 (2014).

53. Weill, E. *et al.* Optimal control of an artificial microbial differentiation system for protein bioproduction. *2019 18th European Control Conference, ECC 2019,* 2663–2668 (2019).

54. Hao, N., Chen, Q., Dodd, I. B. & Shearwin, K. E. The pIT5 Plasmid Series, an Improved Toolkit for Repeated Genome Integration in E. coli. *ACS Synthetic Biology* **10,** 1633–1639. ISSN: 21615063 (2021).

55. Antoniewicz, M. R. A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. *Metabolic Engineering* **63,** 2–12. ISSN: 10967184. https://doi.org/10.1016/j.ymben.2020.11.002 (2021).

56. Fontana, J. *et al.* Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. *Nature Communications* **11,** 1–11. ISSN: 20411723. http://dx.doi.org/10.1038/s41467-020-15454-y (2020).

57. Kim, B., Kim, H. J. & Lee, S. J. Regulation of Microbial Metabolic Rates Using CRISPR Interference With Expanded PAM Sequences. *Frontiers in Microbiology* **11,** 1–9. ISSN: 1664302X (2020).