

NORTHWESTERN UNIVERSITY

Elucidating the Causes of Heterogeneity in Bacterial Genome Replication and Conjugative

Transfer Leveraging Orthogonal Transcriptional Control

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree of

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences

By

Andrew Hilz Scarpelli

EVANSTON, ILLINOIS

December 2016

ABSTRACT

Elucidating the causes of heterogeneity in bacterial genome replication and conjugative transfer
leveraging orthogonal transcriptional control

Andrew Hilz Scarpelli

A disconnect exists between the behavior of transcriptional and genetic regulation in bacterial systems when comparing phenotypic patterns and dynamics of a population to those observed at the single cell level. In this thesis, we developed a number of tools and assays to better understand, overcome, and predict the challenges of heterogeneity within a population with the goal of enabling a better understanding of fundamental biology and the use of this understanding for engineering novel tools in *E. coli*. I describe our exploration of whether the strategy of conditional spatial sequestration of transcriptional regulators, a means by which bacteria naturally regulate native gene expression, could be engineered to create novel regulatory networks. We engineered an orthogonal conditional spatial sequestration system for transcriptional regulators which allows for the accumulation of said regulators within a cell that can be rendered functional in response treatment by a small molecule. I then describe our attempts to gain better understanding the regulation of the transfer machinery in the F plasmid, specifically why only a subset of cells within a donor population initiates conjugation. In particular, we examined the role of TraJ in regulating P_Y , the promoter associated with transcription regulation of most conjugative machinery. Through the development of two assays quantifying TraJ induction of P_Y and conjugative rates at a single cell level, we determined that while TraJ is required for induction from P_Y , expression of TraJ does not directly correlate to induction of P_Y , nor does the addition of orthogonal TraJ expression lead to an increase in conjugative rates. We then leveraged these observations to engineer a novel orthogonally regulated conjugative transfer system. Finally, I describe our work on the creation of a model to aid in the analysis and of design of genetic circuits for genomic integration by predicting how distributions of genomic DNA evolve within bacterial cultures in relation to changes in growth. To inform this model, we generated a library of genomic distribution snapshots throughout an entire growth curve of multiple batch cultures to calibrate and validate a predictive, agent-based model capable of capturing variation in *E. coli* genome copy number across multiple phases of growth.

Acknowledgements

First and foremost, I would like to thank my adviser, Joshua Leonard. It has been a humbling learning experience working with Josh. I am grateful for the opportunity he gave me to work in this lab.

I would like to thank my committee, whose patience, which I know I tested, is a testament to their abilities as mentors: Erik Sontheimer, Eric Weiss, Michael Jewett, and Keith Tyo. Their feedback and suggestions undoubtedly improved my science immeasurably.

I cannot speak highly enough of my colleagues in the Leonard lab with whom I have had the privilege to work with. My colleagues Taylor Dolberg, Michelle Hung, Joe Muldoon, Shirley Samuel-Abraham, Kelly Schwartz, Danny Wells, Andrew Younger, and Alia Zander are all incredible and inspiring scientists. The two doctoral students who started the lab fresh, Yishan Chuang and Nichole Daringer, have two very distinct and potent personalities, but set a unified supportive tone and atmosphere for the lab and created the community that the lab has become. I hope to have the personal charm of Dr. Chuang and the work ethic of Dr. Daringer in my future career. A special thanks needs to be added for Rachel Dudek, whom I have shared the entirety of my graduate experience with. We are in many ways polar opposites, but I don't think there was anyone better to have walked this path with.

To my friends outside of the lab, thank you for giving me a place to breathe when work became overwhelming. To the CHIditarod, thank you for giving me a way to help my community when I felt I had nothing left to give. To MORE and Science in Society, thank you for providing me an outlet to share my love of science outside of the lab. Lastly, I owe my family the world; I would not be anything if it weren't for the sacrifices of my grandparents, the thoughtfulness of my mother, the humor of my father, the humbling presence of my sister, and the love of them all.

Table of Contents

Abstract	2
Acknowledgements	3
List of figures	7
Chapter 1: Introduction	16
1.1 Spatial sequestration as gene regulation	17
1.1.1 Spatial sequestration as a natural bacterial mechanism for gene regulation	17
1.1.2 Artificial spatial sequestration of transcription factors	18
1.2 F-plasmid <i>tra</i> gene regulation and engineering novel regulation lateral gene transfer	20
1.2.1 Conjugation: mechanism and biological relevance	20
1.2.2 The F Plasmid as a model system and <i>tra</i> operon regulation	21
1.2.3 TraJ regulation and the heterogeneity in conjugative populations	22
1.3 Modeling genome copy number dynamics	22
1.3.1 Phenotypic heterogeneity resulting from variation in maintenance of genetic elements	23
1.3.2 Modeling <i>E. coli</i> cell cycle in exponential growth	25
1.3.3 Toward Modeling <i>E. coli</i> cell cycle across multiple phases of growth	27
Chapter 2: Evaluating whether conditional spatial sequestration is sufficient for mediating regulation of bacterial gene expression	29
2.1 Introduction	29

2.2 Materials and methods	31
2.2.1 Growth conditions and strains	31
2.2.2 DNA constructs	32
2.2.3 Microplate fluorescence assays	33
2.2.4 Microscopy and spheroplast synthesis	33
2.2.5 Flow cytometry	34
2.2.6 Western blotting	34
2.3 Results and Discussion	36
2.3.1 Protease-Alleviated Spatial Sequestration (PASS) Platform Design	36
2.3.2 Inducible PASS Construct Expression and Cleavage	41
2.3.3 Subcellular Localization of PASS Constructs	44
2.3.4 PASS Regulated Repression of Gene Expression	46
2.3.5 PASS-Regulated Induction of Gene Expression	51
Chapter 3: Quantitatively profiling the determinants of conjugative transfer initiation using orthogonal transcriptional control	59
3.1 Introduction	59
3.2 Materials and methods	63
3.2.1 Growth conditions and strains	63
3.2.2 DNA constructs	64
3.2.3 Microplate fluorescence assays	64
3.2.4 Flow cytometry	64
3.2.5. Transconjugant quantification	65
3.3 Results and discussion	65
3.3.1 Developing a minimum P _Y promoter	65

3.3.2 Regulation of P _Y -driven transcription by orthogonal TraJ	68
3.3.3 Initiation of conjugation via orthogonal transcriptional control	80
Chapter 4: Predicting bacterial genome copy number dynamics under variable growth regimes using a novel quantitative model	87
4.1 Introduction	87
4.2 Materials and methods	90
4.2.1 Growth conditions and strains	91
4.2.2 Bacterial genome quantitative PCR	91
4.2.3 Genomic labeling and flow cytometry	92
4.3 Results and discussion	93
4.4 Summary and future directions	107
Chapter 5: Conclusions and Recommendations	109
5.1 Chapter 2: Engineered spatial sequestration	109
5.2 Chapter 3: Regulation of F plasmid conjugation	111
5.3 Chapter 4: Genomic heterogeneity simulator	114
References	115

Index of Figures

Figure 2.1: PASS concept and mechanism of action. The proposed mechanism is as follows: tethering a transcriptional regulator to the inner face of the cytoplasmic membrane prevents its ability to regulate target gene expression; protease-mediated cleavage of the PASS construct liberates the transcriptional regulator to repress or activate its cognate promoter sequence. In this study, recombinant PASS constructs included a periplasmic mCherry ectodomain, a transmembrane α -helix derived from *E. coli* ATP synthase subunit B, the cleavage sequence for tobacco etch virus protease (TEV), and either the tetR or λ CI transcriptional regulator domains. Released tetR represses the constitutive pTet promoter, and released CI activates the conditional pRM⁺ promoter, to modulate output gene (GFP) expression. 38

Figure 2.2. Regulation of reporter constructs by soluble repressor and activator. (A) Cells were transformed with the tetR-regulated reporter plasmid and tetR expression plasmid, as indicated, and induced with 1 % (w/v) arabinose. (B) Cells were transformed with the CI-regulated reporter plasmid and CI expression plasmid, as indicated, and induced with 1 % (w/v) arabinose. Samples undergoing exponential growth were analyzed, blanked, and normalized as described in Chapter 2.2.3. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 10^{-5}$ and ** $p < 10^{-10}$, as calculated for a two-tailed paired Student's t-test. Abbreviations: TR, tetR. 40

Figure 2.3. Expression and proteolytic processing of PASS constructs. (A) Cells were transformed as indicated and induced with varying concentrations of IPTG. Fluorescence was quantified as in Figure 2.2. Experiments were conducted in biological triplicate, and error bars

indicate standard deviations. (B - D) Cells were transformed and induced, as indicated, with 1 mM IPTG and/or 1% (w/v) arabinose, and lysates were analyzed by N-terminal (mCherry) or C-terminal (6xHis) labeling. Protein standards are given in kilodaltons (kD). 43

Figure 2.4. PASS construct localization at the cytoplasmic membrane. (A) Micrographs of cells transformed with pTet-GFP, pLacIQ-mTR, and pBAD-TEV Protease treated with 1 mM IPTG. mCherry fluorescence is localized to the periplasmic area while GFP fluorescence is diffuse throughout the cytoplasm. (B) Cells transformed with pLacIQ-mTR or pLacIQ-mCI and induced with 1 mM IPTG; shown in mCherry channel. (C) Spheroplasts generated from cells in panel (B).

45

Figure 2.5. PASS-regulated gene repression by tetR. Cells were transformed and induced as indicated, and fold-change was quantified as in Chapter 2.2. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 10^{-10}$ and ** $p < 10^{-13}$, as calculated for a two-tailed paired Student's t-test. Abbreviations: mC, membrane-bound mCherry (mTR cleavage product); TR, tetR. 48

Figure 2.6. PASS-regulated gene repression in individual cells. (A) Cells transformed with pTet-GFP, pLacIQ-mTR, and pBAD-TEV were induced as indicated and analyzed by flow cytometry. (B) Mean fluorescence intensity (MFI) in GFP for mCherry-positive cells from panel (A). Experiments were conducted in biological triplicate, and error bars indicate standard deviations. Normalized MFI GFP values were determined by calculating MFI GFP for mCherry-

positive cells in the test sample, calculating MFI GFP for mCherry-positive cells in the uninduced sample, and then dividing the former by the latter. * $p < 0.01$ and ** $p < 0.005$, as calculated for a two-tailed paired Student's t-test. 50

Figure 2.7. PASS-regulated gene repression by λ CI. Cells were transformed and induced as indicated, and fold-change was quantified as in Methods. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 10^{-3}$ and ** $p < 10^{-5}$, as calculated for a two-tailed paired Student's t-test. Abbreviations: mC, membrane-bound mCherry (mCI cleavage product). 53

Figure 2.8. PASS-regulated gene induction in individual cells. (A) Cells transformed with pRM+-GFP, pLacIQ-mCI, and pBAD-TEV were induced as indicated and analyzed by flow cytometry. (B) Mean fluorescence intensity (MFI) in GFP for mCherry-positive cells from panel (A). Experiments were conducted in biological triplicate, and error bars indicate standard deviations. Normalized MFI GFP values were determined by calculating MFI GFP for mCherry-positive cells in the test sample, calculating MFI GFP for mCherry-positive cells in the uninduced sample, and then dividing the former by the latter. * $p < 10^{-2}$ and ** $p < 10^{-4}$, as calculated for a two-tailed paired Student's t-test. 55

Figure 3.1. Native F plasmid transfer machinery regulation and the construction of a P_Y reporter system. (A) Schematic representation of the regulatory region of the *tra* operon and its key regulators. Drawing is not to scale. Further information provided in the text. (B) Schematic

representation of our P_Y reporter system. Upon the addition of arabinose, TraJ is expressed and able to act upon P_Y resulting in expression of GFP. (C) Comparison of GFP induction when transcribed from different copy numbered plasmids. TraJ expression only leads to GFP induction on low copy number plasmids. Data shown was collected on a microplate reader. (D) Comparisons of GFP induction when under the control of P_Y variants of different lengths. Two potential variations on the full length sequence of P_Y were examined for their ability to drive expression of GFP in the presence of TraJ. Data shown were collected on a microplate reader. 67

Figure 3.2. Regulation of P_Y -driven transcription by TraJ. (A) GFP fluorescence as an output for induction by induced orthogonal systems in response to arabinose. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 0.05$ and ** $p < 0.005$ as calculated for a two-tailed paired Student's t-test. Data shown was collected on a microplate reader. (B) The effect of arabinose induction during stationary phase. Experiments were conducted in biological triplicate, and error bars indicate standard deviations for Mean GFP/OD600. Data shown was collected on a microplate reader. (C) Comparisons of an induced and uninduced pJ and reporter containing population via flow cytometry. (D) Mean Fluorescent Intensities as an output for induction in a population by induced orthogonal systems in response to arabinose via flow cytometry and percentage of cells expressing GFP. Cells examined via flow cytometry are comparable to trends as seen in panel (A) when considering mean fluorescences of populations. 70

Figure 3.3. Quantitative coupling between TraJ expression and P_Y induction in single cells.

(A) Schematic representation of pJaR reporter system. Upon addition of arabinose, TraJ is transcribed along with monomeric RFP (mRFP). TraJ is then capable of acting upon P_Y to induce GFP expression. (B) GFP fluorescence as an output for induction by induced orthogonal systems in response to arabinose detected via flow cytometry. pJaR leads to GFP induction in similar patterns to that of pJ and pJam. (C) Raw flow cytometry data of pJaR reporter system in the absence of arabinose and three replicates in the presence of 1% arabinose by volume. (D) Lack of correlation between GFP and mRFP (TraJ) expression. No significant correlation between GFP and mRFP expression could be found using various fittings. Shown here is a linear distribution of GFP-POSITIVE cells and a linear fit, with an R² value below 0.2.

74

Figure 3.4. Regulation of native conjugation machinery by exogenous TraJ.

(A) Schematic representation of pShuttle/pOX38 (pOX38;oriTf::CmR) conjugation reporter system. Conjugative machinery expressed by pOX38 is capable of transferring pShuttle, a plasmid that contains both a GFP expression cassette as well as the native F OriT. Upon forming a mating pair with a donor cell that contains an mRFP expression plasmid, conjugation can occur, resulting in a transconjugant expressing both mRFP and GFP. (B) Observed colonies from donor and transconjugant cells. (C) Transconjugant cells expressed both observable mRFP and GFP, while donor cells only expressed GFP. Observation of transconjugants by flow cytometry. Populations of recipient cells paired with populations of donors or mock “donor” with an immobile GFP plasmid instead of pShuttle were co-cultured for 5 h and then run on flow cytometry. The increased rate of doubly GFP and mRFP + cells appears in the system with functional donors, indicating the

fitness of this assay. (D) Comparison of quantification of mating assays by colony count or flow cytometry. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. (E) Observed mating efficiencies with the addition of pJ to donor population cells. TraJ expression was induced at various times relative to co-culture with recipient cells. Experiments were conducted in biological triplicate from three distinct colonies, the averages of those three colonies were then averaged, and error bars indicate standard deviations from initial averages.

78

Figure 3.5. Orthogonal regulation of conjugative transfer. (A) Schematic representation of novel, orthogonally regulated conjugation reporter system. Conjugative machinery is expressed from pHeadless after the induction of TraJ expression from an orthogonal expression plasmid (Shown in image as pJ). (B) Observed mating efficiencies by flow cytometry of orthogonal conjugative systems. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. No treatment condition was determined to be significantly different, with $p < 0.05$ as determined by a two tailed Student's t-test.

82

Figure 4.1. Injection-based strategy for connecting the HMG simulator to empirical growth data. This cartoon summarizes the process by which empirical growth data (e.g., a measured OD vs. time curve) is used to “drive” the HMG simulator via the volume injection method, where the open circles represent the sections of the growth curve where DNA distributions were measured. Thus in this illustration, the simulation would contain three independent steps: (1) The region of exponential growth is identified. This exponential growth rate is used to drive the HMG simulation

from a single cell inoculate to a diversified population of exponentially growing cells; (2) During post-exponential growth, the OD curve is used to calculate the rate at which the overall cell volume (of the population) is increasing; (3) At each time point, the calculated rate of volumetric change (per cell) is “injected” into each cell in the population, each of which advances its cell state via the HMG algorithm outlined in Figure 4.1. The dashed rectangles indicate that during each time step of the simulation, a random subset of 5000 cells is taken forward into the subsequent time step of the simulation in order to keep simulations computationally tractable.

95

Figure 4.2. Heterogeneous Multiphasic Growth (HMG) simulation algorithm. This figure summarizes the algorithms used to advance our agent-based simulation of bacterial growth. This algorithm marries our “injection” model for driving growth based upon experimentally measured growth curves with either the original CH model of bacterial replication (ignoring the dashed boxes) or an extended version of the CH model which incorporates the effects of *recA* mutation (including the dashed boxes). In each time step of the simulation, each cell is advanced through the 5 indicated processes: (1) Growth, (2) Opening of origin(s) of replication, (3) DNA replication and DNA degradation, (4) Segregation, and (5) Cell division. Gray boxes indicate the steps in the algorithm where noise is applied to the cell cycle.

97

Figure 4.3. Training of the HMG simulator framework. The HMG simulator was “fed” growth curves for TOP10 cells grown in LB, shaken at 230RPM or 23RPM, and simulated DNA distributions were compared with those which were measured empirically. The measured DNA distributions shown here each represent a single experiment, each of which is representative of

two or more independent experiments. The first column within each heat map represents the exponential growth phase (indicated by *), and all subsequent time points represent post-exponential growth. The simulator was run using two different models: the first model was based upon a prior description of exponential growth (1), which omits any consequences of *recA* mutation, and the second (updated and optimized) model incorporated our description of the consequences of *recA* mutation with parametric optimization. Similarity scores indicate the degree to which each prediction matches the observed DNA distribution. The solid lines on the two bottom panels represent the mean similarity score across the time course, and the shaded boxes represent the standard deviation of these scores across the time course.

100

Figure 4.4. Validation of the HMG simulator framework. The HMG simulator was “fed” growth curves for TOP10 cells grown in M9, shaken at 230RPM or 23RPM, and simulated DNA distributions were compared with those which were measured empirically. The measured DNA distributions shown here each represent a single experiment, each of which is representative of two or more independent experiments. The first column within each heat map represents the exponential growth phase (indicated by *), and all subsequent time points represent post-exponential growth. The simulator was run using two different models: the first model was based upon a prior description of exponential growth (1), which omits any consequences of *recA* mutation, and the second (updated and optimized) model incorporated our description of the consequences of *recA* mutation with parameters optimized based upon growth in LB (i.e., using the same updated and optimized model described in Figure 4.3). The solid lines on the two bottom

panels represent the mean similarity score across the time course, and the shaded boxes represent the standard deviation of these scores across the time course.

101

Figure 4.5. HMG simulator-based prediction of gene dosage effects. The updated and optimized HMG simulator was used as a testbed to predict gene dosage dynamics over a range of hypothetical growth curves (left column). Here we track three genomic loci (blue, red, and green rectangles), located at various positions relative to *oriC* (pink circle). Numbers accompanying the chromosome maps in the top row indicate the relative distance of each locus from *oriC*, in each scenario, on a scale where 1.0 is completely distal (e.g., the primary *Ter* site, teal rectangle). Each predicted trajectory represents the mean copy number of each locus per cell, averaged over 100 independent simulations, with error bars representing one standard deviation. For each hypothetical growth curve (left column), each shaded area is labeled with the doubling rate calculated for that window of growth. Each simulation was inoculated (initiated) under conditions of exponential growth, using the doubling rate calculated for the first indicated period of exponential growth (gray shading), and thereafter simulations proceeded using the injection method through the remainder of the growth curves.

105

CHAPTER 1

Introduction

Even within genetically identical populations of bacteria, dramatic phenotypic variation may exist between individual cells (2). This heterogeneity can result in very different behaviors of cells within a population and create a disconnect between what is seen on a population and a cellular level (3). In this thesis, I describe a number of bacterial systems that display heterogeneity and explore the relevant biology governing this behavior, with the goal of harnessing such microbial systems for applications in biotechnology. In chapter 2, I discuss a system of spatial sequestration as a mechanism of gene regulation. Bacteria commonly use spatial sequestration of transcriptional regulators to accumulate these regulators for rapid induction of genes required for acclimation to new and changing growth conditions. In order to better understand the limitations and mechanics of natural systems of this sequestration process, we engineered a novel spatial sequestration system and characterized the ability of said system to coordinate a response within a population of bacteria. We designed and implemented an orthogonal system in which transcription factors can be accumulated within a cell and only made functional upon induction with a small molecule. In chapter 3, I explain my work characterizing the heterogeneous expression of conjugative machinery from the F plasmid, a well characterized model system of conjugation. Specifically, I examined the role of TraJ, the main transcriptional regulator of conjugation, and how its expression contributes to this heterogeneity of conjugative machinery expression during bacterial growth. In chapter 4, I describe my contributions to the development of an agent based simulation tool for describing genomic distribution throughout various phases of growth for *E. coli* cultures. This agent based model will serve as a predictive tool for genomic copy number distribution within bacterial populations, utilizing a simplified representation of DNA replication

and cell division. Our model will accurately predict the distribution of genome copy number ensembles throughout multiple phases of growth, a capability lacking in current genomic distribution models.

1.1 Spatial sequestration as a mechanism of gene regulation

1.1.1 Spatial sequestration as a natural bacterial mechanism for gene regulation

Bacteria utilize mechanisms of conditional spatial sequestration to regulate a number of functions, including gene expression. A number of different gene regulation systems have been reported that utilize this technique including well studied examples such as MalT, σ_E , and Mlc (4-6). Spatial sequestration in bacteria usually consists of localizing transcription factors along the inner membrane of the cytoplasm, restricting their ability to influence gene regulation. When particular conditions unique to each system arise, the presence of maltose for MalT, an accumulation of misfolded porins for σ_E , or the absence of glucose for Mlc, these transcription factors are released from their localization fully functional from their localization and immediately can act as transcriptional regulators (6-8).

One of the best characterized natural spatial sequestration systems of transcriptional regulation can be found in the maltose uptake regulation system in *E. coli* (9, 10). In this system, MalE binds to maltose in the periplasm and taxis the sugar to a MalFGK₂, a transporter complex that spans the inner membrane between the periplasm and cytoplasm (7). Prior to interactions with MalE, the transporter complex sequesters the transcriptional regulator MalT to the cytoplasmic inner membrane (11). When MalE interacts with the complex, it causes a conformational change that results in the release of MalT, which is then free to act as a genetic regulator on a number of

genes with *mal* promoters (12). Thus the cell sequesters MalT in the absence of maltose, but it allows for MalT-directed transcription in the presence of maltose.

Another well-characterized but distinct natural spatial sequestration system is associated with the release of the stress factor σ_E (13). In sustained growth conditions, σ_E is bound to the anti-sigma factor RseA, which also spans the membrane between the periplasm and cytoplasm (14, 15). As misfolded porins accumulate, RseA binds in the periplasm, leaving the protein open to proteolysis by DegS (16). The degradation of RseA in the periplasm leads to alterations in the protein allowing for cleavage by YaeL in the cytoplasm (17). This second cleavage event results in the release of σ_E from its sequestration with RseA on the membrane, and σ_E then diffuses into the cytoplasm where it is able to recruit RNA polymerases and initiate the expression of a number of stress response genes (18).

A final natural system of conditional spatial sequestration is associated with Mlc and glucose transport (6). In the absence of glucose, Mlc acts as a soluble repressor in the cytoplasm, repressing genes associated with glucose uptake and metabolism (19). When glucose is taken up by the cell via the transporter PtsG, the transporter undergoes dephosphorylation (20). This dephosphorylation allows for the binding and sequestration of Mlc, resulting in alleviation of its repression of a number of genes associated with glucose metabolism (21).

Each system provides a relevant example of how conditional spatial sequestration allows for rapid changes in the transcriptional control in a bacterial cell in response to changing environmental or cellular conditions.

1.1.2 Artificial spatial sequestration of transcription factors

Conflicting evidence exists as to the feasibility of engineering artificial spatial sequestration of transcription factors in *E. coli*. For example, fusion of the binding domain of the

protein that sequesters the previously mentioned Mlc global repressor to LacY permease on the inner face of the cytoplasmic membrane in *E. coli* led to inactivation Mlc-mediated regulation of transcription (22). If the Mlc binding domain were instead fused to a cytoplasmic expressed protein, Mlc retained its ability to repress genes, strongly suggesting that the spatial sequestration is solely responsible for the inactivity of Mlc when bound to the inner membrane. Conversely, the localization of LacI to the same location using the M13 bacteriophage coat protein VIII failed to preclude LacI-mediated inhibition of LacI-regulated genes (23). This study specifically called into question whether spatial sequestration could be used to explain any of the natural cases in which transcription factors are temporarily housed at the inner membrane as the explanation for their lack of function, postulating instead that their sequestration is steric or conformational, not spatial. Despite this conflicting evidence as to the feasibility of engineering spatial sequestration-based transcriptional regulation in *E. coli*, engineering a synthetic conditional spatial sequestration-based system would provide a number of benefits. In particular, because transcriptional regulators could be pre-synthesized, such a system could confer a rapid response to a desired stimulus. However, the use of conditional spatial sequestration as a tool has not been explored.

In Chapter 2, I describe our previously published work describing the implementation and characterization of a framework of engineering conditional spatial sequestration of transcription factors in *E. coli* (24). We used a completely orthogonal spatial sequestration system to help characterize the limits of natural spatial sequestration system and to explore the potential of spatial sequestration as a tool for engineering more complex systems. Our system utilizes an exogenous protease for proteolytic release of an engineered transcription factor from a protein tether to the bacterial inner membrane. We then characterize our protease-alleviated spatial sequestration (PASS) in its capacity to function as either a transcriptional activator or repressor.

Chapter 2 is mainly adapted with permission from Pitner, Scarpelli, & Leonard, Copyright 2015, American Chemical Society (24).

1.2 F-plasmid *tra* gene regulation and engineering novel regulation lateral gene transfer

1.2.1 Conjugation: mechanism and biological relevance

Bacterial conjugation is an important process that is of relevance to a number of aspects of bacterial biology and is one of only three ways in which bacteria can laterally transfer genetic material (25). Conjugation consists of the transfer of genetic material usually encoding conjugative functionality from one bacterial cell directly into another recipient cell (26). Conjugation is initiated by the assembly of a secretion system in donor cells, which are host cells capable of conjugation, and the creation of a pilus, a long filament of protein able to attach to other (recipient) cells (27). The pilus is retracted, and the attached recipient cell is drawn towards the donor to form a mating pair (28). Mating pairs experience an outer membrane fusion, and conjugative machinery mediates transport of single-stranded copies of genetic information associated with the secretion system from the original donor cell into a recipient (29). The two cells then separate, leaving both cells genetically capable of transferring genetic information onto new recipients (30).

Conjugation has been shown to play a key role in a number of processes of biological and clinical relevance. Conjugation has been well documented as the means by which antibiotic resistance and virulence genes are spread within a population (31, 32). Conjugation is thought to be one of the main concerns in the spread of such genes within a clinical environment (29, 33). Conjugation also plays a key role in the evolution of bacterial genomes (34). Conjugation has long been associated with the transfer of conjugative and mobilizable plasmid between distinct different families of both gram negative and gram positive bacteria, and has been known to integrate into genomes of wide array of species, leading to transfer of genomic information between species (26, 35). Conjugation is seen as the main player in acquisition of large chromosomal regions that play

an important role in the survival of bacteria in newly colonized environments (36). The acquisition of these elements can be derived from diverse sources, as conjugation from many conjugative plasmids has been shown to span large numbers of species, with many conjugative plasmids able to transfer to and from distantly related organisms (25). Therefore, better understanding how conjugation is regulated could provide insights into fields ranging from microbial ecology and evolution to infectious disease management.

Conjugation is also used as a tool, enabling transfer of genetic information not only between members of the same species, but across species and even kingdoms (32, 37). Conjugation has been shown to be essential in the life cycle of the bacteria *Agrobacterium tumefaciens*, which transfers a virulence gene into plant roots as to create a response beneficial to the bacteria (38). This discovery has led to the observation of conjugation capabilities between bacteria and fungi (39), bacteria and human cells (40), and even bacteria and isolated mammalian mitochondria (41), which has led to interest in using conjugation as a means to engineer mammalian mitochondria (42). Conjugation has been used to introduce novel genetic circuits into diatoms, and is seen as a potential system to engineer algae for the production of fuels and designer molecules (37).

The existence and importance of conjugation has been known for almost seventy years (43), and the first discovered conjugative plasmid, the F plasmid, or Fertility factor, has been used consistently as a model system for the exploration of how conjugation is regulated, how it functions, and its implications in bacterial systems growth and dynamics (30). The F plasmid was fully sequenced in 2000 and most encoded genes have been at least partially characterized (27). The F plasmid is a relatively large plasmid, just shy of 100 kb, and it encodes the majority of genes required for conjugation in a single operon, called the *tra* operon (32).

1.2.2 The F Plasmid as a model system and *tra* operon regulation

Although the components of the F plasmid and its conjugative machinery have been well studied, a number of questions about the specific functions on the parts continue to exist. Specifically relevant to the work I describe in this thesis, much is known about the key factors regulating transcription of genes essential for transfer, but the exact mechanisms governing regulation have not been fully elucidated (44). Regulation of the *tra* operon from its upstream promoter P_Y has been extensively studied, and a number of key players have been elucidated, including the key player TraJ (32, 45). TraJ is encoded on a gene located outside of the *tra* operon on the F plasmid, and shares the role of activating transcription of most *tra* genes with native regulator ArcA (46). TraJ is only functionally expressed during exponential growth, and its expression is silenced outside of that phase by a number of different host proteins, many of which also compete against TraJ to suppress expression of the *tra* operon. With all that is known about TraJ, a number of question about its function still exist, including: Can TraJ modulate gene expression outside of exponential growth? What role does the concentration of TraJ play on in its ability to regulate gene expression and promote conjugative transfer? Under what conditions does TraJ most robustly promote gene expression from P_Y? As described below, each of these question is explored in this thesis.

1.2.3 TraJ regulation and the heterogeneity in conjugative populations

In the investigations described in Chapter 3, we sought to address these questions identified above by examining the function of TraJ in a quantitative, single-cell fashion. To this end, we developed a P_Y GFP reporter system as well as various constructs enabling inducible expression of TraJ via the application of a small molecule. Previous studies have used P_Y reporter systems to explore TraJ effects on transcription, but have used reporter systems that only provide information on a population level, which would be incapable of capturing heterogeneity within the population

(45). With this reporter system we were able to probe what role expression of TraJ has in mediating the heterogeneity of P_Y induction and what is the correlation between TraJ expression and P_Y induction.

We also wanted to explore the effect of orthogonal TraJ on the rate of conjugation. TraJ mutants are known to decrease the rate of conjugation in previously described systems (47). If having less functional protein decreased conjugative rates, we decided to investigate whether adding more would lead to an increase in conjugative rates. We also decided to explore when the timing of this additional expression of TraJ would result in the greatest effect on conjugative rates. Lastly, we wanted to quantify the ability of a completely orthogonal transcription control region would play in a novel conjugative system. We engineered a system to explore in what context we could induce or observe conjugative transfer with such as system.

1.3 Predicting heterogeneity and dynamics in genomic DNA content of bacterial populations across variable growth regimes

1.3.1 Phenotypic heterogeneity resulting from variation in maintenance of genetic elements

For coordination of expression of engineered protein system in bacteria, variation between cells in expression can make prediction and characterization of function difficult, which illustrates a need for increased stability of the copy number regulation of introduced genetic elements. This need for stability typically motivates the integration of engineered genetic elements into the genome as an alternative to relying upon expression from plasmids, as plasmid copy number may vary from cell to cell and subsets of a population experience plasmid loss over time (48). However, copy number of genes encoded within the *Escherichia coli* genome fluctuates and becomes elevated well above a single copy throughout rapid growth (regularly exceeding 8 copies per cell),

and in a manner dependent upon environmental factors such as nutrients and cell densities (49, 50). The replication of the genome in *E. coli* is also regulated by a number of factors and has a large role in how genes are expressed at different loci (51). Nonetheless, genomic integration of engineered gene circuits is still preferential for the exact reason of relative copy number stability (52).

While the genome is more stably regulated than many plasmid systems (50), it is not maintained at a single copy throughout different phases of growth (53). The replication of the genome takes about 40 minutes, with some variation depending on strain and growing conditions (54), and this slow rate of replication is one of the possible reasons for genomic copy number's heightened stability compared to smaller plasmids capable of completing multiple rounds of replication in a relatively short time. However, the replication of the genome and cell division are controlled via very distinct mechanisms, thus the ratio between copies of complete genomes per cell can highly vary in different conditions.

DNA replication begins at the origin of replication, *OriC* (54). This initiation is facilitated by DnaA (55). DnaA localizes to a series of specific binding sites and upon binding causes changes to the DNA structure within the *OriC*, and, after accumulating above a specific threshold on the DNA while complexed with ATP, begins the process of initiating replication (54, 56). Bound DnaA and ATP complexes allow the recruitment of the helicase DnaB, which is the first step in replicating the genome (57). Replication occurs in both directions from the *OriC*, and continues until termination, a process that is coordinated by interactions between the protein Tus and one of 10 *Ter* sequences along the genome (58). Tus complexes with *Ter* sites in an asymmetric fashion, which enables the complex to function as a direction-specific terminator of replication. When a replication fork reaches a Tus complex from one direction, the Tus complex will dissipate and replication will continue (59). If the replication fork reaches the Tus complex from the other

direction, however, the Tus complex will remain intact, inhibiting further replication until the opposite replication fork is able to complete replication from the other direction.

The accumulation of genomes within a cell during rapid growth means that the replication of DNA has to be decoupled from cell division. The regulation of cell division in *E. coli* is less well understood than the regulation of DNA replication, but a number of essential regulators of cell division have been identified (60). The first protein to localize at the point of division is FtsZ, a tubulin homologue that forms a ring around the nucleoid of the cell (61). This ring can form well before the two distinct genomic copies separate into distinct regions. The ring then complexes with the membrane-bound ZipA, tethering the ring to the membrane in preparation to pinch the membrane during division(62). This leads to the recruitment of a number of proteins to form the divisome, which contracts the membrane while actively transporting chromosomes to opposite daughter cells (63).

Interestingly, regulation of cell division and genome replication are actually largely decoupled. Initiation of cell division is largely associated with cell size in a mechanism that is currently not fully understood (64). Replication of the genome is largely initiated by the accumulation of DnaA and the energy state of the cell (55). While cell size has some effect on the internal concentrations of DnaA, these two mechanisms are mainly independent, leading to the accumulation of genomes during growth. While this decoupling enables *E. coli* to grow rapidly when nutrients are abundant, such the cell division time can be less than the time required to replicate a single genome (53), this decoupling also makes predicting variations in genomic DNA content more challenging.

1.3.2 Modeling *E. coli* cell cycle in exponential growth

While the understanding of these two processes – chromosome replication and cellular division – has grown since early models of genomic copy distributions in a population, modeling has relied on broader understandings of what governs these two mechanisms and largely avoided specific mechanistic details. Instead, since the earliest attempts to model the relationship between genome count and cell division have simply approximated these two processes into two terms for describing their role in the bacterial cell cycle (65, 66). C , the first of these terms, represents the time associated with average genomic replication. Details of the mechanism of chromosome replication are lumped into a general term giving an average rate of DNA synthesis. The other term, D , represents the average time that passes between the conclusion of replication with at least two distinct genomes and cell division. Each term is modified to account for with the rate at which growth occurs as well as other factors such as temperature and nutrient constraints (53).

A number of models for genomic distribution have been created at steady state growth, some based on predicted events within cells, some based on cell size distributions and some based on cell age distributions within a population (67-69). A Monte Carlo simulation described in 1995 by Keasling et al. used an elegant system where dividing cells were linked to a membrane such that cells only escaped the membrane upon division into new daughter cells to synchronize the age of cells in a collected population accurately predicted distributions of cell size, DNA content, and chromosome segregation patterns in exponentially growing cells. This model, however is unable to predict distributions outside of exponential growth, and is unable to predict how changes in their environment will affect these distributions. Newer models have combined similar approaches for predicting distributions with more detailed methods such as observation of genomic content by flow cytometry, but are also limited to exponential growth (53, 70). Even papers exploring more diverse growing conditions have limited their models to steady state growth, limiting the potential use to cells in the same growth state (71). In experimental cultures outside of simulations, growth rate are constantly changing throughout a run, and therefore, new design tools are needed to

evaluate and predict the consequence of genomic copy number heterogeneity in non-steady state cultures and synthetic biology experiments.

1.3.3 Toward modeling *E. coli* cell cycle across multiple phases of growth

In the investigation I described in Chapter 4, we set out to assess the feasibility and limitations of a comprehensive model of genome copy regulation and dynamics throughout multiple phases of growth in a bacterial culture. Working with Declan Bates and Melchior du Lac at the University of Warwick, we collaborated on designing, calibrating, and validating an agent-based computational model for predicting distributions of genomic content within ensembles of individual cells, throughout multiple phases of growth. While other models already exist for genome distribution at varying rates at steady state, but our model was designed with the intention of following cells out of exponential growth and all the way into stationary phase. This added complexity would allow for the design and prediction of function for novel genetic circuits integrated in the genome whose functionality was relevant or required in multiple phases of growth, such as those requiring genetic induction after exiting exponential growth or whose function are designed for batch culture usage. In our investigation, the computational modeling has been led by Melchior du Lac, and my contribution was largely in the collection of data for training and calibrating our models, as well as providing guidance throughout the process of model development to ensure that assumptions and mechanisms included in the model were well-grounded in biology.

Our collaborative effort led to the design of designed an agent-based simulation framework termed the Heterogeneous Multiphasic Growth (HMG) simulator. The data used to build and calibrate our simulator was largely collected via a methodology I developed. Experiments using this methodology conducted by Andrew Younger and myself generated a series of experimental

data sets quantifying genomic DNA distributions for cells grown in batch cultures, sampling multiple phases of growth, different growth media, and different shaking rates. These cells were stained and examined for genome content using single cell analysis to get a picture of population distributions of genome dynamics. We built our simulator based on a subset of this data, and validated it using other runs from the remaining sets. We have also investigated whether our model can predict genome replication dynamics in experiments quite divergent from those used to calibrate the simulator. For this work, I generated a series of data sets using experiments run in shake flasks during various phases of growth to probe the limits of the predictive ability of our model.

Chapter 4 is mainly adapted with permission from Du Lac, Scarpelli, Younger, Bates & Leonard, Copyright 2016, American Chemical Society (72).

CHAPTER 2

Evaluating whether conditional spatial sequestration is sufficient for mediating regulation of bacterial gene expression

2.1 Introduction

Conditional spatial sequestration of transcription factors is a strategy widely employed by prokaryotes to achieve gene regulation. Among the best characterized examples is the *E. coli* mechanism for regulated uptake of maltose (4, 7). In this system, the MalFGK₂ maltose transporter spans the cytoplasmic membrane, and in its resting state, this transporter sequesters the transcriptional activator MalT at the cytoplasmic membrane (11). When MalE binds maltose in the periplasm, this complex binds to the transporter and promotes a series of ATP hydrolysis-linked conformational changes (73), which eventually result in the release of MalT from the cytoplasmic face of the transporter. MalT then drives downstream mal genes in a manner dependent on endogenous maltotriose (a product of glycogen degradation (10)). Thus in this case, conditional localization of the transcriptional regulator is mediated by conformation-dependent protein-protein interactions.

A distinct mechanism for controlling localization of transcriptional regulators is regulated intramembrane proteolysis (RIP), which is conserved from prokaryotes to eukaryotes (13). In RIP, the sequential proteolysis of extra- and intra-membrane segments of a transmembrane protein leads to the cytoplasmic release of a transcription regulator (8). One such system is the σ_E stress response in *E. coli*, whereby accumulation of misfolded outer membrane porins is sensed to drive expression of stress response genes, such as chaperones(15). In this system, σ_E is initially bound to and

inhibited by the cytoplasmic membrane antisigma factor, RseA (30, 74). Binding of misfolded porins to RseA renders the periplasmic domain of RseA labile to cleavage by DegS, and this cleavage subsequently renders the cytoplasmic domain of RseA labile to cleavage by YaeL (17, 18, 75). Following this second cleavage event, sequestration of σ_E is alleviated and it diffuses from the membrane to recruit RNA polymerase to promoters of stress-response genes. Thus in both the σ_E RIP system and the MalT system, conditional control of gene expression is mediated by alleviation of spatial sequestration of a pre-synthesized transcription factor.

Importantly, these natural systems illustrate that conditional spatial sequestration of transcription factors enables rapid responses to changes in environmental or cellular state, since such responses do not require novel synthesis of transcriptional regulators to coordinate downstream gene expression. Thus, spatial conditional sequestration may also be useful for engineering synthetic biology technologies ranging from biosensing to coordination of engineered metabolic pathways. However, to date it remains unclear how or whether conditional spatial sequestration may be engineered. In some cases, artificial sequestration of native transcriptional regulators has proven sufficient to mediate spatial control of transcription. For example, targeting the transcriptional repressor Mlc to the inner face of the *E. coli* cytoplasmic membrane, by genetically fusing Mlc to the transmembrane protein LacY permease, was sufficient to de-repress Mlc-controlled genes (22). In contrast, targeting the LacI repressor to the inner face of the *E. coli* cytoplasmic membrane, by genetically fusing LacI to bacteriophage M13 coat protein, did not impede LacI-mediated repression of a tac operator-promoter located either on a plasmid or integrated into the chromosome (23). As proposed by Gorke et al., it is possible that these seemingly contradictory observations derive from distinct mechanism by which tethering regulates

the activity of transcriptional regulators or their interactions with DNA. However, it remains unknown whether other regulators commonly used in synthetic biology may be amenable to spatially-regulated control. Moreover, how or whether conditional spatial sequestration of transcriptional regulators may be engineered has not been explored.

Here we describe a platform for investigating and implementing conditional spatial sequestration of transcriptional regulators. We describe a mechanism for conditional tethering, in which expression of an exogenous (non-native) protease mediates proteolytic release of engineered transcriptional regulators from the inner face of the *E. coli* cytoplasmic membrane. We also demonstrate that this Protease-Alleviated Spatial Sequestration (PASS) mechanism can robustly mediate either conditional transcriptional activation or conditional transcriptional repression. Thus, this work both provides new insights into a biologically important facet of microbial gene regulation and establishes a new class of conditional regulation for the microbial synthetic biology toolbox.

2.2 Materials and Methods

2.2.1 Growth Conditions and strains

Cells were grown in Lysogeny broth (LB) Lennox formulation (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) for preparatory steps, and in supplemented M9 (M9 minimal medium with 0.4% glycerol, 0.2% casamino acids, and 1 mM thiamine hydrochloride) for characterization steps. All experiments were performed at 37°C. Singly-transformed cultures included 50 µg/ml

kanamycin or 34 $\mu\text{g}/\text{mL}$ chloramphenicol; doubly-transformed cultures included 37.5 $\mu\text{g}/\text{mL}$ kanamycin and 17 $\mu\text{g}/\text{mL}$ chloramphenicol. All experiments were performed in TOP10 competent cells (Life Technologies), which have genotype: F- *mcrA* Δ (*mrr*-*hsdRMS*-*mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *araD139* Δ (*araleu*)7697 *galU* *galK* *rpsL* (StrR) *endA1* *nupG*.

2.2.2 DNA constructs

Primers were purchased from Life Technologies (Carlsbad, CA) and IDT (Coralville, IA), and *E. coli*-optimized GeneArt Strings for mCherry_{TM}, TM_{PS}_{CI}, and TM_{TEV} Protease were ordered from Life Technologies. All other coding sequences were obtained from the Spring 2010 Registry of Standard Biological Parts Distribution.

2.2.3 Microplate fluorescence assays

Colonies were grown overnight in supplemented M9 medium with appropriate antibiotics and shaking, diluted to an OD₆₀₀ of 0.05 in 3 mL pre-warmed M9 medium, and after reaching an OD₆₀₀ between 0.4-0.6, cultures were again diluted to an OD₆₀₀ of 0.05 in 2 mL of pre-warmed M9 medium containing appropriate inducers and grown for 30-60 min. Three 180 μL replicates per culture were then transferred to black-walled clear-bottom 96-well plates (Corning) and placed in a Synergy H1MFD multi-mode microplate reader (BioTek). OD₆₀₀, mCherry fluorescence, and GFP fluorescence were measured every 15 min for ~ 10 h while shaking at 37°C. Monochromator excitation/emission settings were 585/615 nm for mCherry and 485/515 nm for GFP. To

distinguish GFP-mediated fluorescence from autofluorescence, the fluorescence/OD₆₀₀ value calculated for untransformed TOP10 cells was subtracted from each sample to generate “blanked” measures of GFP fluorescence/OD₆₀₀, which we termed, “GFP/OD₆₀₀”. To calculate the mean value of (GFP fluorescence intensity per OD₆₀₀) for each sample (here, Mean GFP/OD₆₀₀), four consecutive time points from late exponential phase were selected for each sample. Late exponential phase was analyzed to ensure that cells were at steady state, rather than exhibiting transient responses to the inducer; all time points analyzed were at least 3 hours after inducer addition. For each case analyzed, Mean GFP/OD₆₀₀ was averaged over four time points for each of three biological replicates. More specifically, analysis of growth for microplate analysis was examined using a semi-log plot of log₂(OD₆₀₀) vs. time. Exponential phase growth was determined by identifying sequential data points that appear co-linear on such a plot. Time points were generally selected towards the later end of exponential phase growth in order to maximize signal to noise in measured fluorescence. We observed no pronounced construct-specific impacts on growth rate (not shown). To correct for non-specific effects of inducers on reporter gene expression, for each genetic configuration analyzed in Figures 2.5 and 2.7, fold-change in Mean GFP/OD₆₀₀ upon inducer addition was normalized to the fold-change in Mean GFP/OD₆₀₀ elicited by adding the same inducer(s) to a strain harboring only the corresponding reporter plasmid (pTet-GFP or pRM+-GFP). Thus, by definition, fold-change for the inducer-free case for each genetic configuration was set to unity.

2.2.4 Microscopy and spheroplast synthesis

For fluorescent microscopy, agarose pads were prepared by adding 2% low-melt agarose (Lonza) to minimal M9 medium and heating until melted. 30 μ L of agarose solution were then added onto a depression slide and covered with another flat slide (Fisher Scientific). Following pad solidification (~10 min), 5 μ L of cell culture were added to the pad, covered with a coverslip, and sealed with clear fingernail polish. Slides were viewed on an inverted fluorescent Leica DM-IL LED microscope with a Leica HCX PL APO 100x/1.40 PH CS oil-immersion objective lens and high-resolution cooled Q-imaging CCD. Images in Figure 2.4 were sharpened using the DeconvolutionJ plugin for ImageJ to reduce optical distortion.

For spheroplast synthesis, whole cells were converted to spheroplasts based upon an established protocol (76). Briefly, cells were grown and induced as described for microplate assays, and samples were collected after 6 hours of growth in microplate format. 1 mL of induced cell culture was centrifuged at 10,000 x g for 4 minutes. The pellet was re-suspended in 100 μ L of 30 mM Tris hydrochloride buffer (Tris-HCl, pH 8.0) containing 0.1 mg/ml of chloramphenicol and 3 mM NaN₃. Cells were again pelleted, the supernatant was removed, and the pellet was suspended in 100 μ L of spheroplasting buffer (20% sucrose, 30 mM Tris-HCl (pH 8.0), 0.1 mg/mL chloramphenicol, 3 mM NaN₃). 10 μ L of a 1 mg/ml lysozyme solution (freshly dissolved in 0.1 M EDTA, pH 8.0) were added to each sample and incubated for 30 min at 0°C. Spheroplasts were separated from released periplasmic contents by centrifugation at 10,000 x g for 4 min.

2.2.5 Flow Cytometry

Cells were grown and induced as described for microplate assays, and samples were collected after 6 h of growth in microplate format. Samples were analyzed on an LSR II (BD), and mean fluorescent intensity (MFI) was calculated using FlowJo software (Treestar). A minimum of 4,000 individual cells (typically out of ~20,000 events) was analyzed per sample.

2.2.6 Western Blotting

Colonies were grown overnight in 5 mL culture tubes in supplemented M9 medium with appropriate antibiotics, diluted to an OD₆₀₀ of 0.05 in pre-warmed M9 medium, and after reaching an OD₆₀₀ between 0.4-0.6, cultures were again diluted to an OD₆₀₀ of 0.05 in pre-warmed M9 medium containing appropriate inducers. Samples were collected after reaching an OD₆₀₀ of at least 0.1, diluted to an OD₆₀₀ of 0.1 in M9 medium, and then combined with 2X Laemmli buffer. Samples were boiled at 95°C for 5 min. 30 µL per sample were loaded and run in pre-cast 12% (Figure 3.3D) or 4-15% gradient polyacrylamide gels (Bio-Rad). Gels were transferred to PVDF membranes (Bio-Rad) for 2 h at 100 V. Antibodies used for western blot analyses were anti-mCherry monoclonal (Abcam), anti-His6X polyclonal (Abcam), anti-rabbit HRP-conjugated secondary (Life Technologies), and anti-mouse HRP-conjugated secondary (Cell Signaling). Transferred membranes were blocked in blocking buffer (5% milk in TBST: 50 mM Tris, 150 mM NaCl, 0.1% Tween-20) for 1 h followed by primary antibody labeling in blocking buffer at 4°C overnight. Membranes were washed 3x with TBST and stained with secondary antibody in blocking buffer for 1 h at 4°C. Membranes were washed 3x with TBST, treated with ECL solution (Bio-Rad), and then exposed to films (GE).

2.3 Results and Discussion

2.3.1 Protease-Alleviated Spatial Sequestration (PASS) Platform Design

Figure 2.1 summarizes the platform developed for evaluating the feasibility of the Protease-Alleviated Spatial Sequestration (PASS) concept. The proposed mechanism is as follows: (1) the transcriptional regulator is tethered to the inner face of the cytoplasmic membrane by genetically fusing the regulator to an engineered single-pass integral membrane protein, separated by a sequence that is labile to cleavage by an exogenous protease (hereafter, the PASS construct)(7), expression of the exogenous protease enables it to cleave the PASS construct, liberating the transcriptional regulator to diffuse into the cytoplasm and bind DNA. Note that “exogenous” is used herein to indicate that the DNA encoding the protease is not native to the *E. coli* host, not to indicate that the protease is extracellular. In our system, PASS constructs included an N-terminal ectodomain based upon the monomeric red fluorescent protein mCherry (77), which was targeted for Sec-mediated transport to the periplasm via fusion to a MalE signal sequence (MalE_SS). The single-pass α -helical transmembrane domain was derived from the native *E. coli* ATP synthase subunit B, for which the structure is known (78). The protease derived from Tobacco etch virus (TEV) was selected based upon its high degree of sequence specificity (79, 80), and all PASS constructs included the canonical TEV substrate sequence, ENLYFQ/G, where the slash indicates the cleavage site. Notably, TEV has been harnessed for intracellular protein processing in *E. coli* in vivo, suggesting that off-target cleavage may not be problematic (81).

To evaluate the potential for PASS-mediated repression as well as induction of transcription, we selected a model transcriptional repressor and a model transcriptional activator. To evaluate repression, PASS constructs included the tetracycline-regulated repressor (tetR) and were evaluated with reporter plasmids driving the expression of GFP from the constitutive, tetR-regulated promoter pTet (Figure 2.1) (82). To evaluate activation, PASS constructs included a constitutively active transcriptional activator based on the CI transcription factor from λ phage (83). The λ CI protein activates the λ pRM promoter, and reporter constructs were developed based upon the pRM⁺ variant of this promoter in which OR3 was deleted to prevent repression of pRM⁺ at high concentrations of CI (84). Finally, 6xHis tags were appended to C-termini of transcription factors to facilitate biochemical analysis. PASS constructs were expressed in an IPTG-inducible fashion from the pLacIQ promoter, and TEV was expressed in an arabinose-inducible fashion from the pBAD promoter.

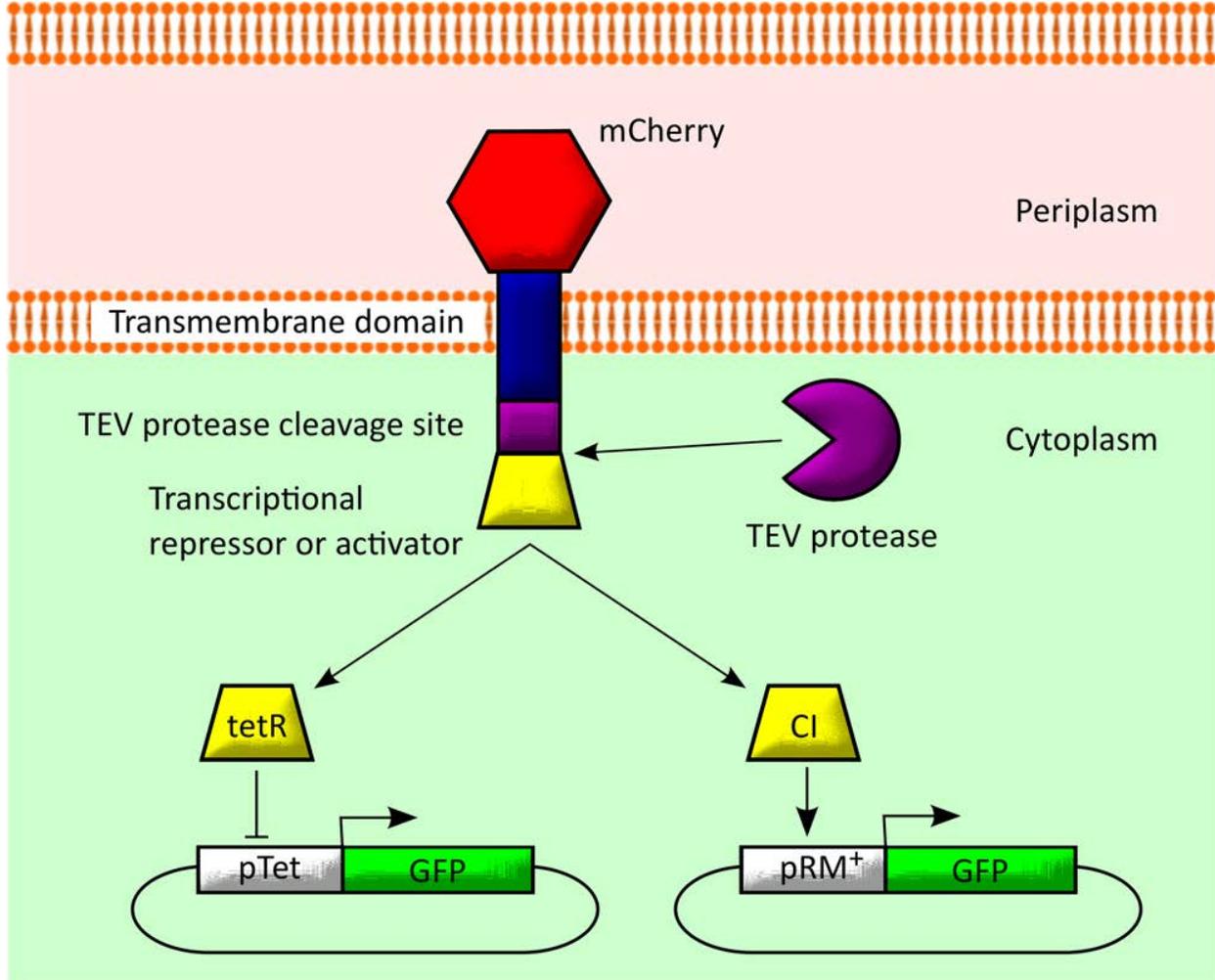
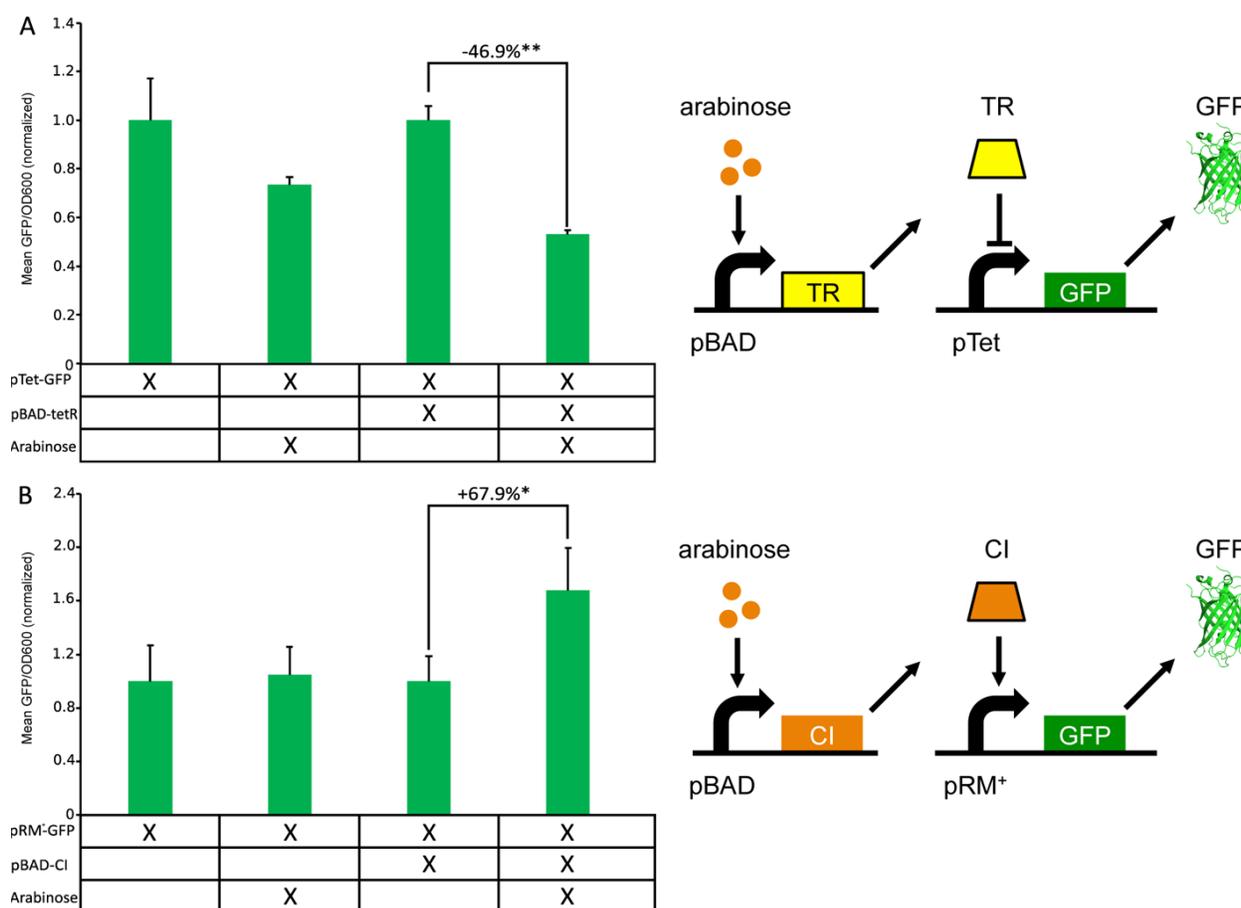


Figure 2.1: PASS concept and mechanism of action. The proposed mechanism is as follows: tethering a transcriptional regulator to the inner face of the cytoplasmic membrane prevents its ability to regulate target gene expression; protease-mediated cleavage of the PASS construct liberates the transcriptional regulator to repress or activate its cognate promoter sequence. In this study, recombinant PASS constructs included a periplasmic mCherry ectodomain, a transmembrane α -helix derived from *E. coli* ATP synthase subunit B, the cleavage sequence for tobacco etch virus protease (TEV), and either the tetR or λ CI transcriptional regulator domains.

Released tetR represses the constitutive pTet promoter, and released CI activates the conditional pRM⁺ promoter, to modulate output gene (GFP) expression.

To initially assess the quantitative range of our reporter systems, plasmids encoding arabinose-inducible tetR and CI (pBAD-tetR and pBAD-CI) were co-transformed with their cognate reporter plasmids (pTet-GFP and pRM⁺-GFP), induced, and evaluated by microplate-based fluorescence analysis (Figure 2.2). Both tetR and CI conferred significant repression or induction, respectively, of GFP expression, confirming operability of the reporter system. Arabinose did confer some tetR-independent reduction in GFP for pTet-GFP containing cells, but this reduction was substantially less than the tetR-mediated reduction in GFP expression. Nonetheless, subsequent microplate analyses were controlled and normalized to account for this effect. These soluble transcription factor controls also established benchmarks for interpreting the relative magnitude of subsequent PASS-mediated changes in reporter gene expression.



2.3.2 Inducible PASS Construct Expression and Cleavage

To evaluate expression of PASS constructs based upon membrane-bound tetR (mTR) or CI (mCI), fluorescence of the mCherry domain was assessed by microplate reader assays following induction with IPTG (Figure 2.3A). Most importantly, these data indicated that PASS constructs were expressed in a sufficiently stable fashion to enable mCherry folding and maturation. In addition, while IPTG treatment enhanced expression of both constructs, some expression was also evident in the absence of IPTG, presumably due to leaky expression from pLacIQ on high copy number plasmids (85). To evaluate whether PASS constructs were proteolytically processed in accordance with the proposed mechanism, we analyzed this process in greater detail focusing on the mTR constructs. To this end, mTR constructs were co-expressed with TEV (driven by pBAD-TEV) and evaluated by western blot (Figure 2.3B-D). N- and C-terminal fragments were detected via anti-mCherry and anti-6xHis antibodies, respectively. The full-length 55.6 kDa mTR construct was observed for all strains including the pLacIQ-mTR construct, in a manner that increased with IPTG but was not dependent on this inducer, which is consistent with the microplate assay analysis (Figure 2.3A). Arabinose-mediated expression of TEV dramatically increased the prevalence of the expected N- and C-terminal cleavage fragments of 31.4 and 24.2 kD, respectively, while decreasing prevalence of the full-length mTR band. The cleavage products were also present at low levels in the absence of arabinose, potentially due to slightly leaky expression of TEV protease.

Several unexpected bands were also observed. Expression of full-length mTR corresponded with appearance of a band at ~40 kD, which was observed in both anti-mCherry and anti-6xHis blots. Because overexpression of Male_SS-tagged mCherry leads to aggregation in the cytoplasm in a manner that blocks secretion (86), we hypothesized that under high levels of expression, some mTR aggregates in either the cytoplasm or after transport to the cytoplasmic membrane. If the aggregated mTR were partially degraded near the N-terminus, this would reduce construct size without ablating binding by either the anti-His6X antibody or the anti-mCherry antibody (which binds within amino acids 84-237 of mCherry; personal communication with Abcam, June 26, 2014). Notably, this proposed proteolysis did not result in liberation of substantial quantities of soluble tetR (Figure 2.3B), such that this effect is unlikely to impede evaluation of the PASS mechanism. The aggregate also appeared partially labile to TEV-mediated cleavage, yielding an mCherry-positive band of ~15-20 kD in the presence of TEV. As suggested by Figure 2.3A, substantial mTR was expressed in the absence of IPTG, and mTR expression increased with IPTG; in either case, induction of TEV with arabinose resulted in liberation of soluble tetR (Figure 2.3C,D). Altogether, these analyses indicated that mTR was generally expressed and processed as per the proposed PASS mechanism. Moreover, our conclusions pertaining to the proposed aggregation and proteolysis of mCherry are not specific to PASS constructs incorporating tetR, and thus this analysis likely applies generally to PASS constructs in which the periplasmic domain is based upon mCherry (including mCI).

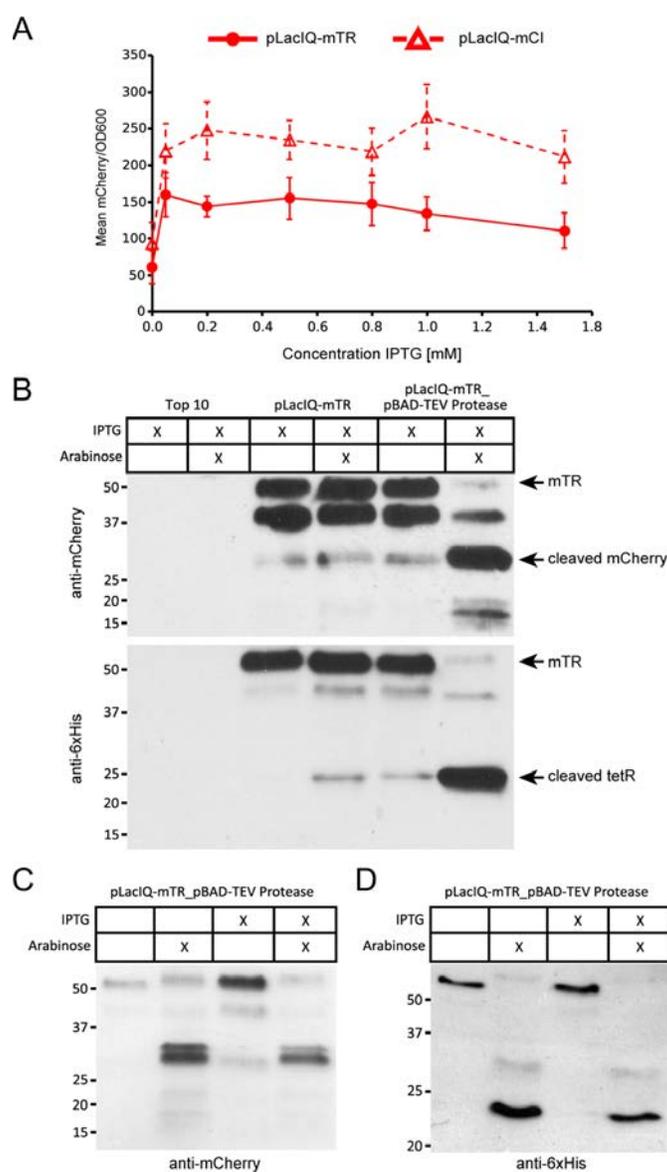


Figure 2.3: Expression and proteolytic processing of PASS constructs. (A) Cells were transformed as indicated and induced with varying concentrations of IPTG. Fluorescence was quantified as in Figure 2. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. (B - D) Cells were transformed and induced, as indicated, with 1 mM IPTG and/or 1% (w/v) arabinose, and lysates were analyzed by N-terminal (mCherry) or C-terminal (6xHis) labeling. Protein standards are given in kilodaltons (kD).

2.3.3 Subcellular Localization of PASS Constructs

To determine whether PASS constructs were successfully integrated into the cytoplasmic membrane, we next visualized cells by fluorescent microscopy. Cells co-expressing cytoplasmic GFP and mTR exhibited a mCherry-positive halo surrounding a GFP-POSITIVE core (Figure 2.4A). Thus, the mCherry domain was successfully secreted to the periplasm and refolded to reconstitute fluorescence. To determine whether this mCherry was associated with the cytoplasmic membrane (rather than simply secreted into the periplasm), we next generated spheroplasts by lysing the outer membrane to enable diffusion of soluble periplasmic species away from the cells (76). For cells expressing either mTR or mCI, membrane-associated mCherry was retained upon conversion of intact cells (Figure 2.4B) to spheroplasts, which exhibited characteristic rounded morphology (Figure 2.4C). Together with the western blot analysis, these data indicate that PASS constructs were efficiently integrated into the cytoplasmic membrane as designed, where they preferentially accumulated.

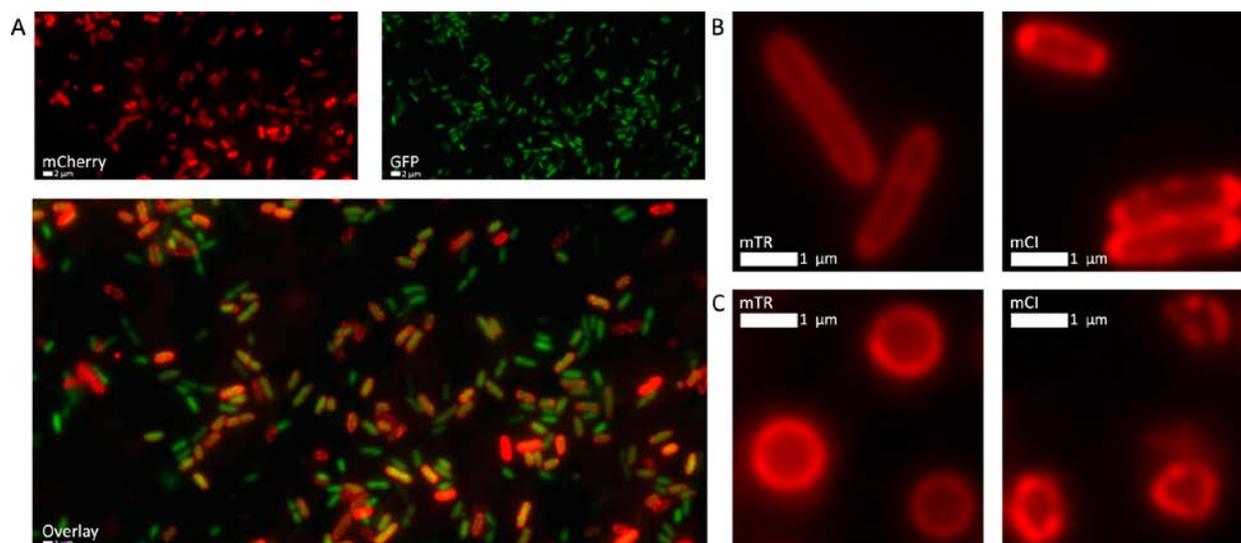


Figure 2.4: PASS construct localization at the cytoplasmic membrane. (A) Micrographs of cells transformed with pTet-GFP, pLacIQ-mTR, and pBAD-TEV Protease treated with 1 mM IPTG. mCherry fluorescence is localized to the periplasmic area while GFP fluorescence is diffuse throughout the cytoplasm. (B) Cells transformed with pLacIQ-mTR or pLacIQ-mCI and induced with 1 mM IPTG; shown in mCherry channel. (C) Spheroplasts generated from cells in panel (B).

2.3.4 PASS-Regulated Repression of Gene Expression

Having established that the PASS mechanism functions as designed, we next investigated whether mTR repressed reporter gene expression from pTet in a manner that depended on TEV expression (Figure 2.5). These experiments included a control series in which mCI was expressed in place of mTR, to control for potential non-specific impacts of PASS construct expression on GFP expression from the pTet-GFP reporter. Induction of TEV expression by arabinose led to a significant decrease in GFP expression, and the magnitude of this decrease was comparable to that mediated by soluble tetR (Figure 2.2). Repression was not enhanced by IPTG, although such treatment did increase expression of mTR (Figure 2.3). Thus, together, these data indicate that the low level of mTR expressed in the absence of IPTG provided sufficient tetR to maximally regulate the pTet promoter, and that TEV-mediated processing of mTR was efficient. Notably, no such IPTG- or arabinose- induced changes in GFP expression were observed for control cells expressing mCI in place of mTR. Moreover, these data indicate that subcellular sequestration of tetR to the inner surface of the cytoplasmic membrane via the PASS mechanism can limit this transcription factor's ability to suppress its cognate promoter. Notably, sequestration prevented repression of pTet encoded on high copy number plasmids; whether this holds true for chromosomally-integrated pTet remains to be determined. Our observations also differ from those in which LacI was tethered in a similar fashion (23). One potential explanation is that tethering tetR impacts its folding or DNA binding in a manner that is distinct from the mechanism by which tethering LacI impacts its regulatory capacity. Overall, tetR-based PASS constructs were functional and robust. One unexpected observation was that IPTG-induced expression of mTR (without arabinose) also

conferred a reduction in reporter GFP expression, at least at the population-averaged level. Since IPTG-mediated induction of mTR did not alter overall bacterial growth compared to control cells in which IPTG induced expression of mCI (not shown), a simple growth effect is unlikely to entirely explain these results. Thus, we next investigated these phenomena at the single cell level in order to elucidate the mechanism by which mTR regulated reporter gene expression.

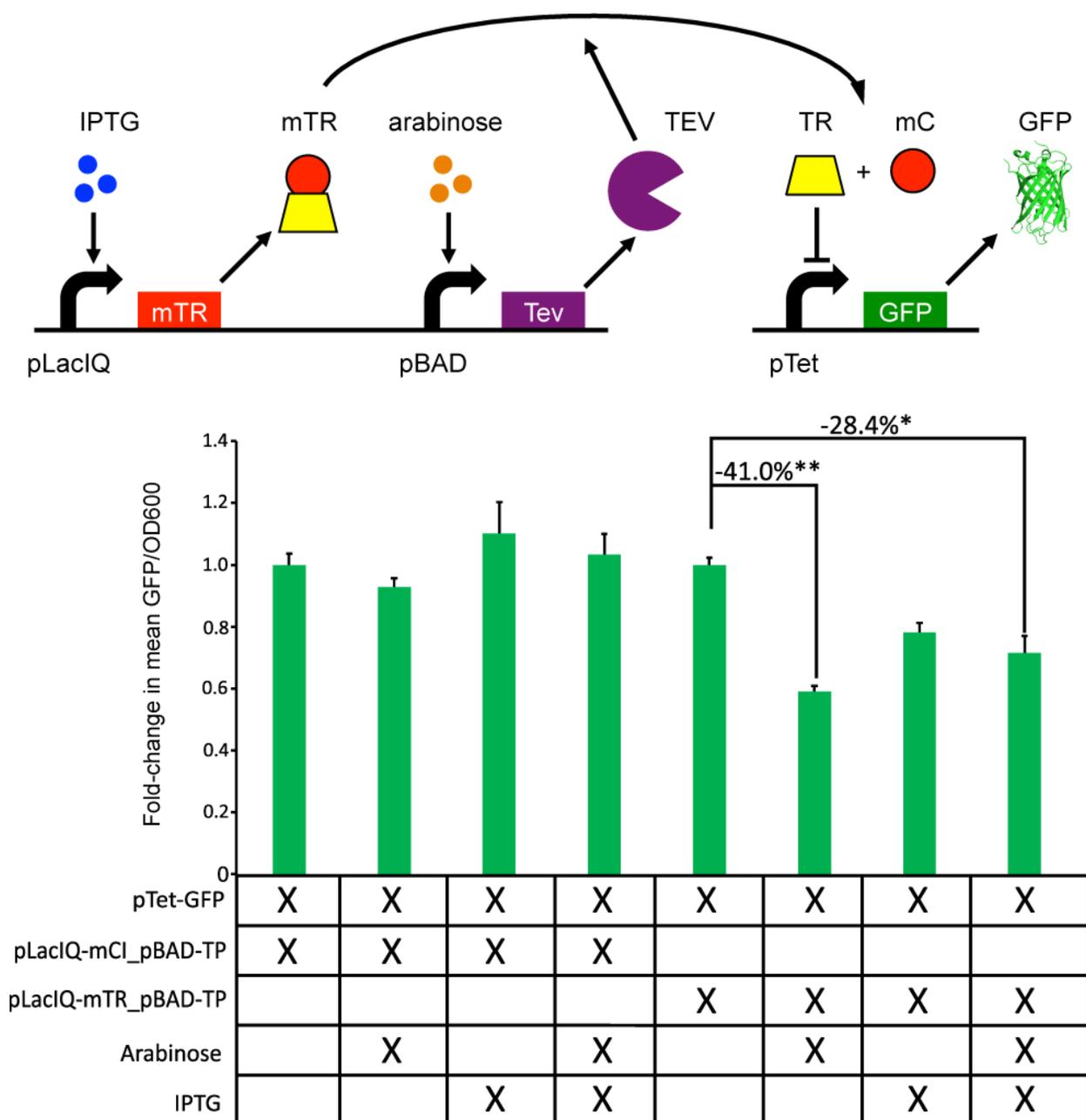


Figure 2.5: PASS-regulated gene repression by tetR. Cells were transformed and induced as indicated, and fold-change was quantified as in Methods. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 10^{-10}$ and ** $p < 10^{-13}$, as

calculated for a two-tailed paired Student's t-test. Abbreviations: mC, membrane-bound mCherry (mTR cleavage product); TR, tetR.

To evaluate whether individual cells within the population exhibited differential PASS-mediated conditional gene expression, we leveraged the fact that mTR constructs are fluorescent to analyze PASS function by flow cytometry. Consistent with western blot and microplate analyses, TEV-mediated cleavage of mTR drove repression of pTet-GFP, even at uninduced levels of mTR expression for which mCherry expression was indistinguishable from background (Figure 2.6A). Within the population of mCherry-positive cells, arabinose-induced expression of TEV drove a substantial suppression of pTet-GFP, for both basal and induced levels of mTR co-expression (Figure 2.6B), and the magnitude of this suppression surpassed that observed in population-averaged quantification by microplate assays (Figure 2.5). Thus, while IPTG-mediated induction of mTR may moderately suppress GFP expression in cells expressing pTet-GFP, our single cell analysis revealed that this effect was far less important than was the TEV-mediated release of tetR to suppress pTet via the proposed PASS mechanism (even in mCherry-positive cells). Although it is not clear why only a subset of cells expressed detectable levels of mCherry (as part of mTR), the PASS mechanism appeared to function robustly in this population. In addition, the arabinose (TEV)-responsive population was generally mCherry- in the absence of IPTG and mCherry-positive in the presence of IPTG. Thus, there may exist a subset of cells or cell states in which the PASS mechanism functions most robustly.

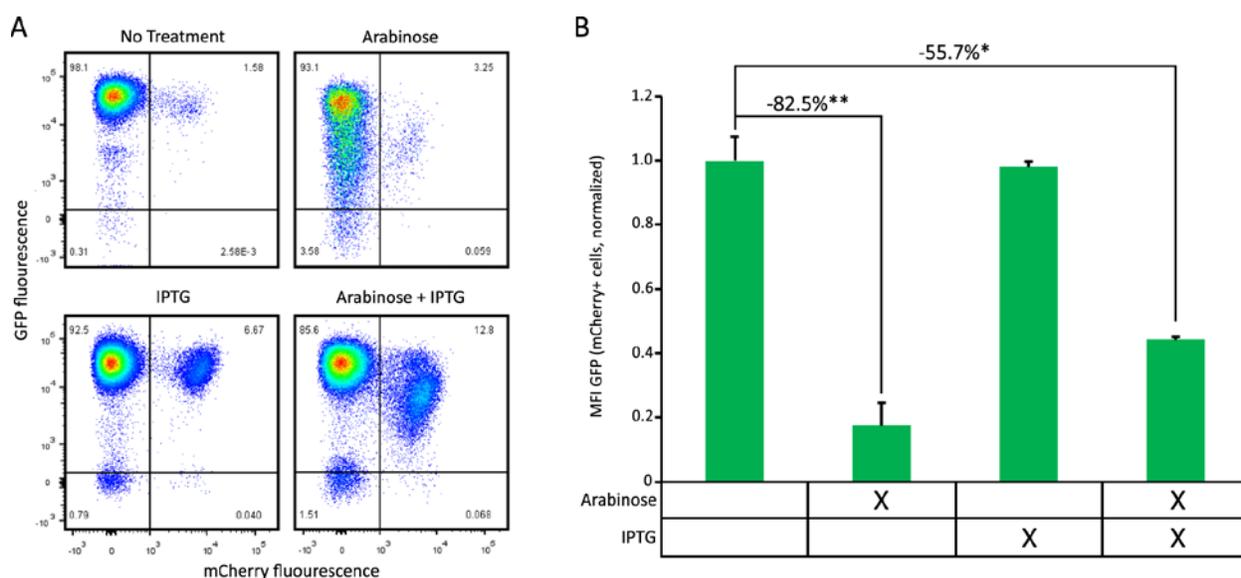


Figure 2.6: PASS-regulated gene repression in individual cells. (A) Cells transformed with pTet-GFP, pLacIQ-mTR, and pBAD-TEV were induced as indicated and analyzed by flow cytometry. (B) Mean fluorescence intensity (MFI) in GFP for mCherry-positive cells from panel (A). Experiments were conducted in biological triplicate, and error bars indicate standard deviations. Normalized MFI GFP values were determined by calculating MFI GFP for mCherry-positive cells in the test sample, calculating MFI GFP for mCherry-positive cells in the uninduced sample, and then dividing the former by the latter. * $p < 0.01$ and ** $p < 0.005$, as calculated for a two-tailed paired Student's t-test.

2.3.5 PASS-Regulated Induction of Gene Expression

Having established that the mTR PASS system mediated conditional repression of gene expression, we next investigated whether the mCI PASS system mediates conditional activation of gene expression. In general, the results mirrored those observed with mTR, and these experiments similarly included a control series in which mTR was expressed in place of mCI. Induction of TEV expression by arabinose led to a significant increase in GFP expression from pRM⁺ (Figure 2.7), and the magnitude of this increase was comparable to that mediated by soluble CI (Figure 2.2). TEV-mediated induction of gene expression did not require induction of mCI expression by IPTG, suggesting that leaky expression of mCI was sufficient to maximally activate the reporter in an arabinose (and thus TEV)-dependent manner. Moreover, these data indicate that subcellular sequestration of CI to the inner surface of the cytoplasmic membrane via the PASS mechanism can limit this transcription factor's ability to activate its cognate promoter. Single cell analysis by flow cytometry again indicated that only a subset of cells expressed mCI at levels detectable as mCherry-positive (Figure 2.8A), and significant TEV-mediated induction of GFP was observed in mCherry-positive cells (Figure 2.8B). To facilitate comparison, data in Figures 6 and 8 were collected with the same flow cytometer settings. The moderate induction of pRM⁺-GFP by CI (compared to basal expression of GFP from pTet-GFP) is consistent with prior reports (84) and Figure 2.2, such that although the magnitude of mCI-mediated conditional gene regulation was less than that observed with mTR, this limitation likely reflects upon the dynamic range accessible to the CI system rather than the robustness of the PASS mechanism. Overall, these data

collectively suggest that the PASS mechanism for conditional gene regulation may be applicable to a range of transcription factor platforms and mechanisms for achieving gene regulation.

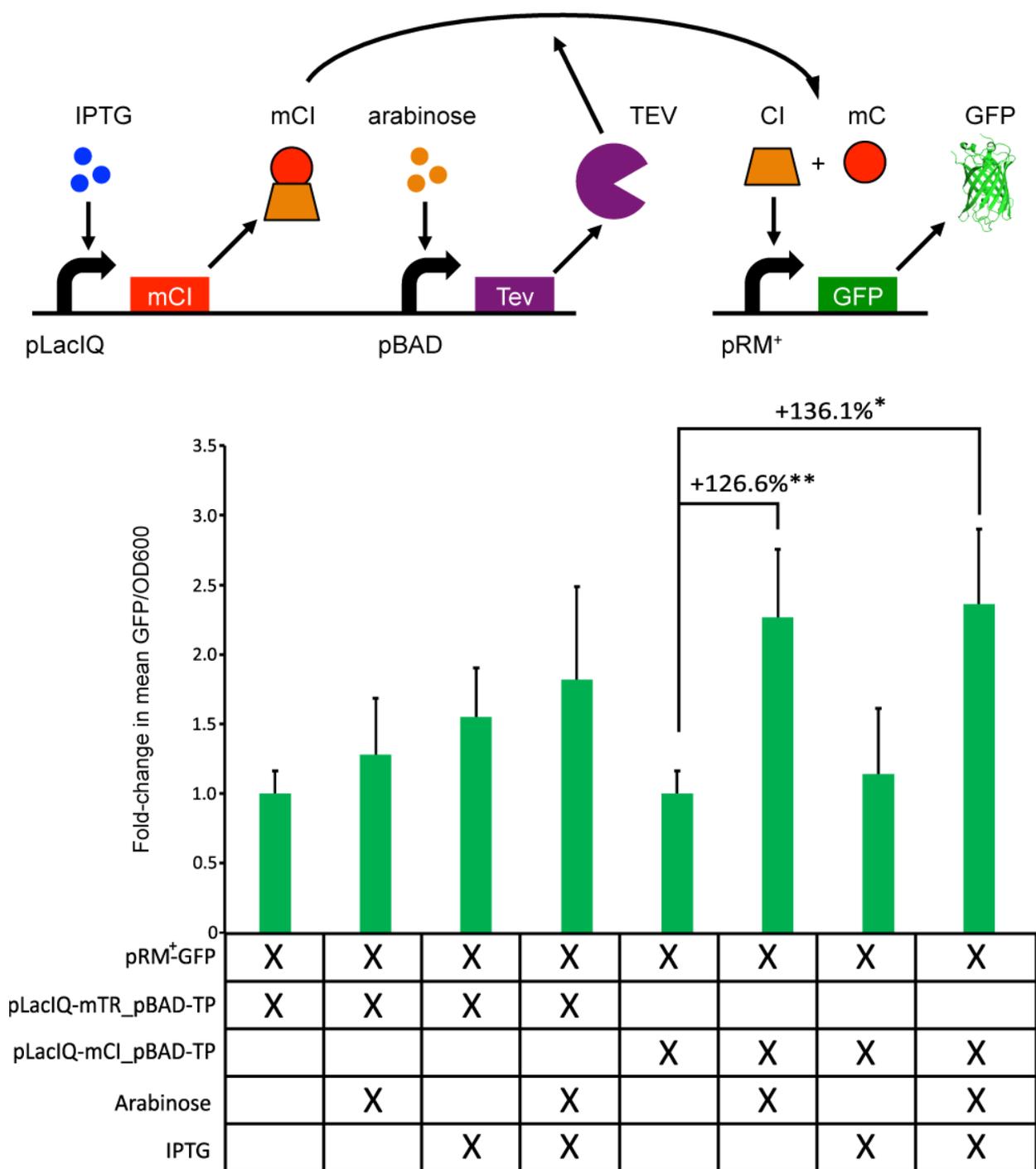


Figure 2.7: PASS-regulated gene repression by λ CI. Cells were transformed and induced as indicated, and fold-change was quantified as in Chapter 2.2. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 10^{-3}$ and ** $p < 10^{-5}$, as

calculated for a two-tailed paired Student's t-test. Abbreviations: mC, membrane-bound mCherry (mCI cleavage product).

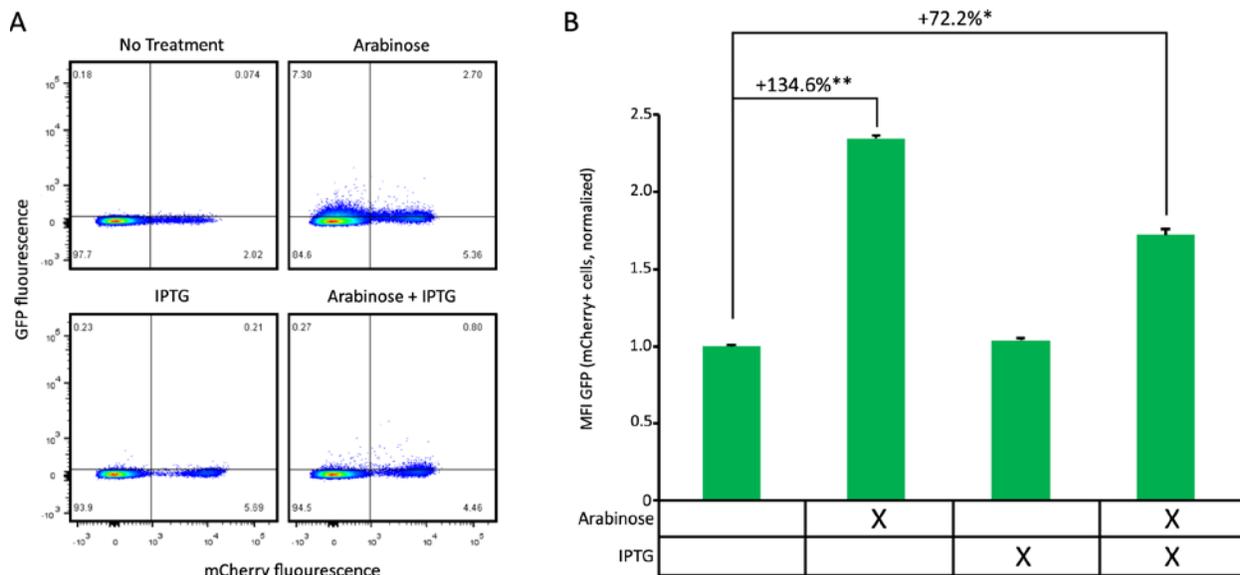


Figure 2.8: PASS-regulated gene induction in individual cells. (A) Cells transformed with pRM+-GFP, pLacIQ-mCI, and pBAD-TEV were induced as indicated and analyzed by flow cytometry. (B) Mean fluorescence intensity (MFI) in GFP for mCherry-positive cells from panel (A). Experiments were conducted in biological triplicate, and error bars indicate standard deviations. Normalized MFI GFP values were determined by calculating MFI GFP for mCherry-positive cells in the test sample, calculating MFI GFP for mCherry-positive cells in the uninduced sample, and then dividing the former by the latter. * $p < 10^{-2}$ and ** $p < 10^{-4}$, as calculated for a two-tailed paired Student's t-test.

2.3 Summary and future directions

The central objective of this study was to investigate a fundamental question in microbial gene regulation – is the engineered conditional sequestration of a transcriptional factor to the inner face of the cytoplasmic membrane a viable strategy for regulating gene expression in *E. coli*? We answered this query in the affirmative by establishing the feasibility of the PASS mechanism, and our results provide novel insights into microbial gene regulation and suggest strategies for harnessing this mechanism for biotechnology.

Given the feasibility of the PASS mechanism in our system, it is worth considering the conditions under which such regulation is possible. In this study, sequestration of either a repressor (tetR) or an activator (λ CI) precluded regulation of cognate promoters encoded on high copy number plasmids. Thus, this mechanism might be extensible to a range of regulator-promoter platforms, although the LacI system might represent an exception to this trend, (23) as discussed earlier. Multicopy plasmids with pUC19-derived origins, such as those on which reporters were encoded in this study, have been observed to cluster in subcellular foci (87), which could help explain why sequestration was a robust regulator of transcription factor activity in our investigation. However, reporter plasmids used to evaluate sequestration of LacI used the same pUC19-derived origin (pMB1) and demonstrated no inhibition of repression upon tethering LacI to the cytoplasmic membrane (23), so such plasmid clustering is seemingly not sufficient to mediate the effect of spatial sequestration of transcription factors. Subcellular spatial localization of chromosomal loci is highly regulated and dynamic (88), and whether the PASS strategy can be extended to regulation of chromosomal promoters, and potentially promoters at different chromosomal locations, requires further investigation. Thus, PASS also provides a novel

experimental tool for investigating gene regulation in the context of subcellular spatial organization and dynamics.

While this study focused on establishing the feasibility of the PASS mechanism, there now exist multiple opportunities for optimizing PASS performance and leveraging this mechanism for synthetic biology applications. For example, use of lower copy number reporter plasmids would likely improve fold-change in output gene expression upon protease expression, particularly for reporter plasmids regulated by sequestered repressors (e.g., pTet-GFP). Only a subset of cells expressed sufficient PASS constructs to appear mCherry-positive by flow cytometry, and it is possible that PASS expression may be more uniform if these constructs were expressed at a lower level (e.g., by using a less efficient ribosome binding site). Although our characterization experiments used inducible expression of TEV to “activate” the PASS construct, regulation of PASS could be made translation-independent to achieve the same rapid and robust responses exhibited by the natural systems discussed in the Introduction (e.g., the σ_E RIP and MalT systems), which inspired this investigation. To illustrate, we consider an application in which PASS is harnessed for biosensing. The TEV protease may be genetically split into fragments that individually lack catalytic activity, and catalytic activity may be reconstituted by bringing these fragments into proximity (89). By fusing split-TEV fragments to protein domains that associate only in the presence of a small molecule (90), protease activity is reconstituted in response to changes in cell state or environment. If a cell were engineered to constitutively express both these split-TEV biosensors and PASS constructs, then introduction of the small molecule analyte would trigger protease reconstitution, and these proteases could immediately act upon the large pool of pre-synthesized PASS constructs. Although such examples require direct experimental

investigation, such a PASS-based mechanism could enable biosensors with higher sensitivity and rapid responses due to the intrinsic catalytic signal amplification within this proposed scheme. Altogether, PASS presents a new strategy for engineering microbial gene regulation to achieve the desirable performance characteristics exhibited by natural mechanisms while also being amenable to modular protein and gene circuit engineering.

CHAPTER 3

Quantitatively profiling the determinants of conjugative transfer initiation using orthogonal transcriptional control

3.1 Introduction

Conjugative transfer of DNA between bacteria plays an important role in microbial ecology, evolution, and disease. Through this process of direct cell-to-cell transfer, the exchange of large portions of genomic DNA can drive evolution and even speciation (34), including the spread of antibiotic resistance and virulence genes (44), and conjugation can even enable transfer of genetic material between biological kingdoms (37). Conjugative plasmids have also been shown to aid in the creation of biofilms (91), and conjugative plasmids encode genes that serve as models and tools for the study of functions including genetic partitioning elements, anti-microbial agents, and non-coding RNA systems (92-95). Given the role of conjugation in these diverse processes, understanding how and when conjugation occurs remains an important goal.

Much of our understanding of conjugation is based upon investigation of the F plasmid, which was the first conjugative plasmid discovered (in 1947) and serves as the prototypical conjugative model systems (43, 96). Within this nearly 100 kb plasmid, the conjugative machinery is largely encoded in a single 33 kb region termed the *tra* operon, which includes at least 34 known genes (97). Transcription of this entire operon is driven from a single promoter upstream of the first operon gene, *traY*, which is termed P_Y. A few genes within the *tra* operon may also be expressed from promoters located within the operon, such as *traT*, which encodes a protein that

precludes conjugative transfer to F⁺ cells via a process termed surface exclusion (98). Nonetheless, conjugative transfer is understood to be regulated primarily by regulation of transcription from P_Y. Only two transfer genes are located on the F plasmid upstream of the *tra* operon – *traJ* and *traM* (97). *TraJ* functions as the main transcriptional regulator of P_Y, such that expression of *TraJ* is required to promote transcription from P_Y (45, 96). *TraM* mediates interactions between the OriT sequence on the F plasmid DNA and the secretory machinery (96), and *TraM* promotes interactions between host and *tra* proteins required to form the relaxosome and initiate transfer (99). Thus *TraM* is also required for conjugative transfer. Although the roles of *TraJ* and *TraM* are relatively well-characterized, the determinants of how and when these proteins are expressed are substantially more complicated and poorly understood.

Competing interests within the cell result in a complex regulatory system determining when conjugation occurs (97). Most of the machinery encoded by the *tra* operon contributes to the components or construction of a Type IV Secretion System, a protein complex capable of extending a large filament of protein capable of attaching to recipient bacterial cells. The creation of each system requires the synthesis of at least fourteen essential proteins and thereafter uses energy to constantly extend and retract once fully formed (100, 101). This secretion system is required for the transfer of conjugative plasmids, but the synthesis and energy use comes at a large cost to the host cell, putting the cell in a state of increased stress (102). Thus, the activation of expression of the *tra* operon is essential for the function of the F plasmid but detrimental to the health of the cell, explaining the need for a complex regulatory mechanism. On the F plasmid, this complex system involves P_Y, *TraJ*, *TraM*, other *tra* elements and a variety of host-associated factors (Figure 3.1). Dimers of *TraJ* bind *tra* DNA and mediate interactions with the host

transcription factor ArcA to promote expression of conjugative machinery; loss of either TraJ or ArcA leads to a complete loss of genetic transfer in most common genetic backgrounds (46, 103). TraJ also promotes transcription from P_Y by displacing Histone-like nucleoid structuring protein (H-NS), which is a host protein that otherwise suppresses transcription of the *tra* operon (45). H-NS binds to regions of DNA, silencing gene expression, and is associated with DNA segments with increased curvature and AT content (104). H-NS has long been associated with silencing of promoter regions that evolved outside of the *E. coli* genome, such as those of viral genomes and other selfish genetic elements (104). TraJ protein is degraded by stress-induced chaperonin GroEL and the protease HslVU, which leads to reduced levels of TraJ when these proteins are up-regulated upon exit from rapid growth (105, 106). Host proteins H-NS, cyclic AMP receptor protein (CRP), and leucine-responsive protein (lrp) each regulate transcription of *traJ*. Host factor for Q β replicase (Hfq) promotes active degradation of mRNA encoding TraJ (and TraM) either directly or by recruiting host RNases (107-110). While TraJ is translated primarily from an mRNA transcribed via its cognate promoter, P_J , TraJ may also be translated from polycistronic mRNA, including P_M -driven transcripts encoding upstream TraM, and P_M/P_J -driven transcripts encoding downstream *tra* genes (TraY, TraA, TraL, etc.) (109). Thus, TraJ is regulated at the transcriptional, post-transcriptional, and post-translation levels by a range of host mechanisms.

TraJ expression may also be regulated by plasmid-derived regulatory mechanism. In some conjugative plasmids, TraJ is regulated by an antisense system termed FinOP, which is encoded within the conjugative plasmid (95). In this system, FinP is an antisense RNA that is stabilized by binding to its partner protein FinO, which protects FinP from RNase-mediated degradation. The FinO-FinP complex inhibits translation of TraJ by binding to the 5' UTR of TraJ mRNA. However,

in most IncF plasmids, including the F plasmid, the *finO* gene is interrupted by an insertion element (IS3), which greatly diminishes the regulatory role of FinOP and leads to enhanced expression of conjugative machinery (95, 111, 112). Given this complex network of potential regulatory interactions, elucidating how and why conjugation is initiated, and what mechanism(s) limit conjugation under various conditions has proven challenging.

A question of particular importance for understanding the role of conjugation in bacterial evolution and ecology is this – why do some cells within a population initiate conjugation while others do not, and how is such heterogeneity in states influenced by both internal and external factors (113, 114)? In general, the efficiency of conjugation is substantially influenced by growth dynamics; both growth rate and growth phase contribute to the efficiency of transfer (115). Thus, conjugation may be suppressed due to silencing of transcription from the *tra* operon by negative regulatory factors, such as H-NS, which accumulate as cell growth slows (104). Other cellular regulators may respond to changes in growth conditions or stress to regulate conjugative machinery, such as host chaperonins and proteases such as GroEL and HslVU, which mediate degradation of TraJ (105, 106). Overall, it remains unclear how negative and positive regulation of conjugation interact to generate heterogeneity in the expression of conjugative machinery, and to what extent such heterogeneity is ascribable to limitations upon TraJ-mediated transcription from P_Y.

To elucidate the role of transcriptional control in the initiation and efficiency of conjugation, in this study we developed a novel system in which the expression of upstream *tra* genes *traJ* and *traM* was decoupled from P_Y-driven expression of the *tra* operon. This partially refactored system was harnessed to investigate fundamental, quantitative questions pertaining to

the regulation of conjugation, including: does TraJ regulate P_Y-driven expression in a digital or analog fashion? Does limited expression of TraJ limit the frequency of conjugative transfer? Can the initiation of conjugation be functionally decoupled from the processes that coordinate downstream aspects of conjugation?

3.2 Materials and methods

3.2.1 Growth conditions and strains

Cells were grown in Lysogeny Broth (LB) Lennox formulation (10 g/L of tryptone, 5 g/L of yeast extract, 5 g/L of NaCl) or supplemented M9 (M9 minimal medium with 0.4% glycerol, 0.2% casamino acids, and 1 mM thiamine hydrochloride) unless otherwise specified. Antibiotics were used at the following final concentrations: ampicillin (Amp), 100 µg/ml; kanamycin (Km), 50 µg/ml; and chloramphenicol (Cm), 34 µg/ml. All experiments were conducted using the commercially available Top 10 strain (F- *mcrA* Δ(*mrr*-*hsdRMS*-*mcrBC*) φ80*lacZ*Δ*M15* Δ*lacX74* *nupG* *recA1* *araD139* Δ(*ara-leu*)7697 *galE15* *galK16* *rpsL*(StrR) *endA1* λ-) (Invitrogen).

3.2.2 DNA constructs

Primers were purchased from Life Technologies (Carlsbad, CA) and IDT (Coralville, IA). pOX38/MC1041 *oriTf* knockout, (pOX38;*oriTf*::CmR), was obtained upon request from the Registry of Standard Biological Parts. pBelo BAC 11 and the F plasmid from strain ER2738 were

obtained from New England Biolabs (Ipswich, MA). All other coding sequences were obtained from the Spring 2010 Registry of Standard Biological Parts Distribution.

3.2.3 Microplate fluorescence assays

Colonies were grown overnight in supplemented M9 medium with appropriate antibiotics and shaking, diluted to an OD₆₀₀ of 0.05 in 3 mL of pre-warmed M9 medium, and after reaching an OD₆₀₀ between 0.3 and 0.6, cultures were again diluted to an OD₆₀₀ of 0.05 in 2 mL of pre-warmed M9 medium containing appropriate inducers. Three 200 μ L replicates per culture were then transferred to black-walled clear-bottom 96-well plates (Corning) and placed in a Synergy H1MFD multimode microplate reader (BioTek). OD₆₀₀ and GFP fluorescence were measured every 30 min for ~16 h while shaking at 37° C. Monochromator excitation/emission settings was 485/515 nm for GFP. For each sample, background fluorescence from M9 medium controls paired with appropriate concentrations of arabinose was subtracted to quantify GFP-specific fluorescence.

3.2.4 Flow cytometry

Cells were grown and induced as described for mating efficiency assays, and samples were collected after 5 h of growth. Cells were diluted 1:1000 in phosphate buffered saline with 25 mM Ethylenediaminetetraacetic acid (EDTA). Samples were analyzed on an LSR II (BD), and mean fluorescent intensity (MFI) was calculated using FlowJo software (Treestar). A minimum of 10,000 individual cells (typically out of ~50 000 events) was analyzed per sample.

3.2.5 Transconjugant quantification

A freshly plated single colony was inoculated into LB media containing appropriate antibiotics and grown at 37°C with shaking for 16 h. Cells were then diluted 50-fold and grown for 3 hours to an OD₆₀₀ of 0.2-1.0 with shaking. Cells were then diluted with pre-warmed LB to an OD₆₀₀ of 0.1. Donor and recipient cells were then co-cultured at 37° C without agitation for 5 hours to allow for conjugation and expression from transferred genetic components. At 5 hours post co-culture, cells were removed from incubation and analyzed by either flow cytometry (described below) or colony mating assays. For colony mating assays, cells were diluted 1:100,000 in supplemented M9 minimal media and plated onto LB plates with appropriate antibiotics. Plates were then grown for at least 16 hours and colonies were counted and scored as either donor or transconjugant based on fluorescent protein expression evaluated by fluorescent microscopy.

3.3 Results and discussion

3.3.1 Developing a minimal P_Y reporter construct

While P_Y reporter systems have previously been describe, the development of the GFP reporter used in this study was used to determine aspects of the required length and copy number to create a functional TraJ reporter. A series of two-plasmid systems for quantifying TraJ function as an activator of P_Y were created testing multiple features of the design space. TraJ was placed under transcriptional control of P_{BAD} in a medium copy number replicon plasmid (p15A) in cells

containing a compatible plasmid encoding GFP downstream of P_Y (Figure. 3.1B). In our initial construct, we defined “ P_Y ” as the sequence used in the P_Y LacZ reporter used by Will and Frost in 2006 (45). We also investigated whether copy number of the reporter plasmid impacted TraJ-mediated induction of transcription from P_Y . Thus, both a high copy number reporter plasmid (ColE1 origin) and a low copy number reporter plasmid (F1/P1 origin) were generated. Upon induction of TraJ with arabinose, significant induction of GFP expression was only observed in the low copy number plasmids (Figure 3.1C). The inability of TraJ to function as an activator of P_Y on a high copy number plasmid suggests that TraJ needs to be in excess of the promoter upon which it acts to promote transcription. In order to investigate whether our P_Y sequence comprised the minimal promoter that is inducible by TraJ, we also investigated a previously described shorter definition of P_Y . This “short” P_Y construct included only the region downstream from the first region of high curvature previously associated with H-NS-mediated repression (104). In contrast to our original P_Y construct, induction of TraJ expression did not induce transcription from the truncated P_Y (Figure 3.1D). This result suggests that sequence coding for TraJ, which is present only in our longer P_Y construct, mediates TraJ-dependent interactions that promote transcription from P_Y and cannot be attributed solely to TraJ-mediated alleviation of H-NS binding.

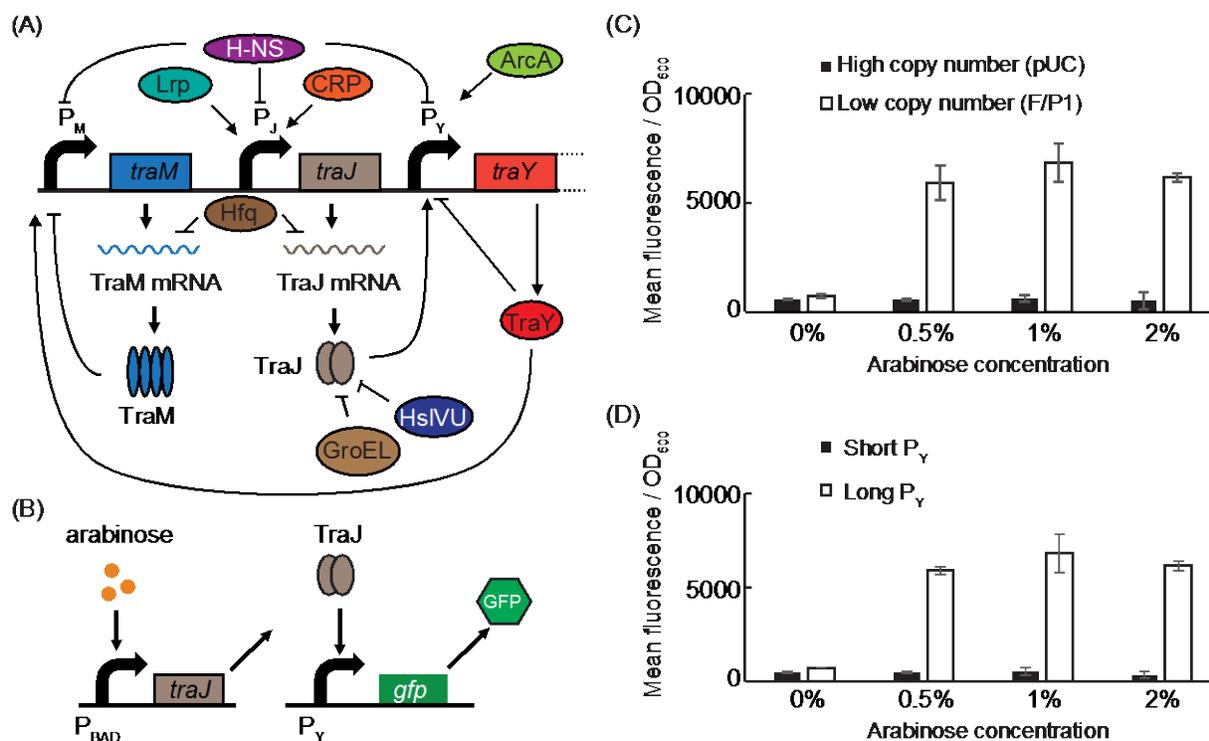


Figure 3.1: Native F plasmid transfer machinery regulation and the construction of a P_Y reporter system. (A) Schematic representation of the regulatory region of the *tra* operon and its key regulators. Drawing is not to scale. Further information provided in the text. (B) Schematic representation of our P_Y reporter system. Upon the addition of arabinose, TraJ is expressed and able to act upon P_Y resulting in expression of GFP. (C) Comparison of GFP induction when transcribed from different copy numbered plasmids. TraJ expression only leads to GFP induction on low copy number plasmids. Data shown was collected on a microplate reader. (D) Comparisons of GFP induction when under the control of P_Y variants of different lengths. Two potential variations on the full length sequence of P_Y were examined for their ability to drive expression of GFP in the presence of TraJ. Data shown was collected on a microplate reader.

3.3.2 Regulation of P_Y-driven transcription by orthogonal TraJ

TraJ is often transcribed co-cistronically with the upstream TraM or downstream genes of the *tra* operon in native systems, each of which can impact the efficiency with which TraJ is translated relative to these other components (109, 116). To investigate how such variations may impact TraJ-induced conjugation, we next explored how TraJ induced expression from P_Y was affected by inclusion of TraJ in various orientations within a multicistronic transcript. Along with our initial pJ construct, we included two constructs in which TraJ was expressed co-cistronically with another gene in these experiments. Upon induction with arabinose, bacteria with the plasmid encoding traJ-traM cistrons (labeled pJaM here forth) behaved similarly to those with traJ alone (pJ), both showing strong induction of GFP in the presence of arabinose (Figure 3.2A). Surprisingly, plasmids with traM-traJ cistrons (pMaJ) appeared to lose inducibility and instead constitutively expressed GFP.

We next explored whether orthogonal induction of TraJ could overcome the inactivation of P_Y in stationary phase. traJ transcripts are degraded by Hfq in stationary phase, so we were surprised to see that trends seen in exponential phase held true for stationary phase (Figure 3.2B). pJ and pJaM were both able to induce expression of GFP in the presence of arabinose. pMaJ drove constitutive expression of GFP and actually seemed to depress GFP expression in the presence of arabinose. The induction of P_Y by exogenous expression of TraJ has been shown in the past with a LacZ reporter, but the P_Y GFP reporter allowed for single cell analysis of this induction, giving us a tool to examine whether the implications of TraJ expression and action on P_Y on the single

cell level could contribute to explaining the stochastic nature of expression of the *tra* operon. To examine this phenomenon, we analyzed the cells via flow cytometry. The overall population mean fluorescence via this assay followed similar trends to what we see via a plate reader (Figure 3.2D, first panel). This single cell analysis allowed us to further explore the probability that a cell expressing TraJ would express consistently from P_Y . Looking at the overall patterns of GFP expression throughout an experiment, we were able to see two distinct populations of cells, a portion showing activation of P_Y and a population with no GFP expression (Figure 3.2C). After gating these two populations, we noted that only a subset of cells actually turned on P_Y GFP in even the highest expression levels. Less than 20% of cells in all reporter construct showed GFP expression. This shows that even when TraJ expression is induced, only a portion of cells are responsive to this activation.

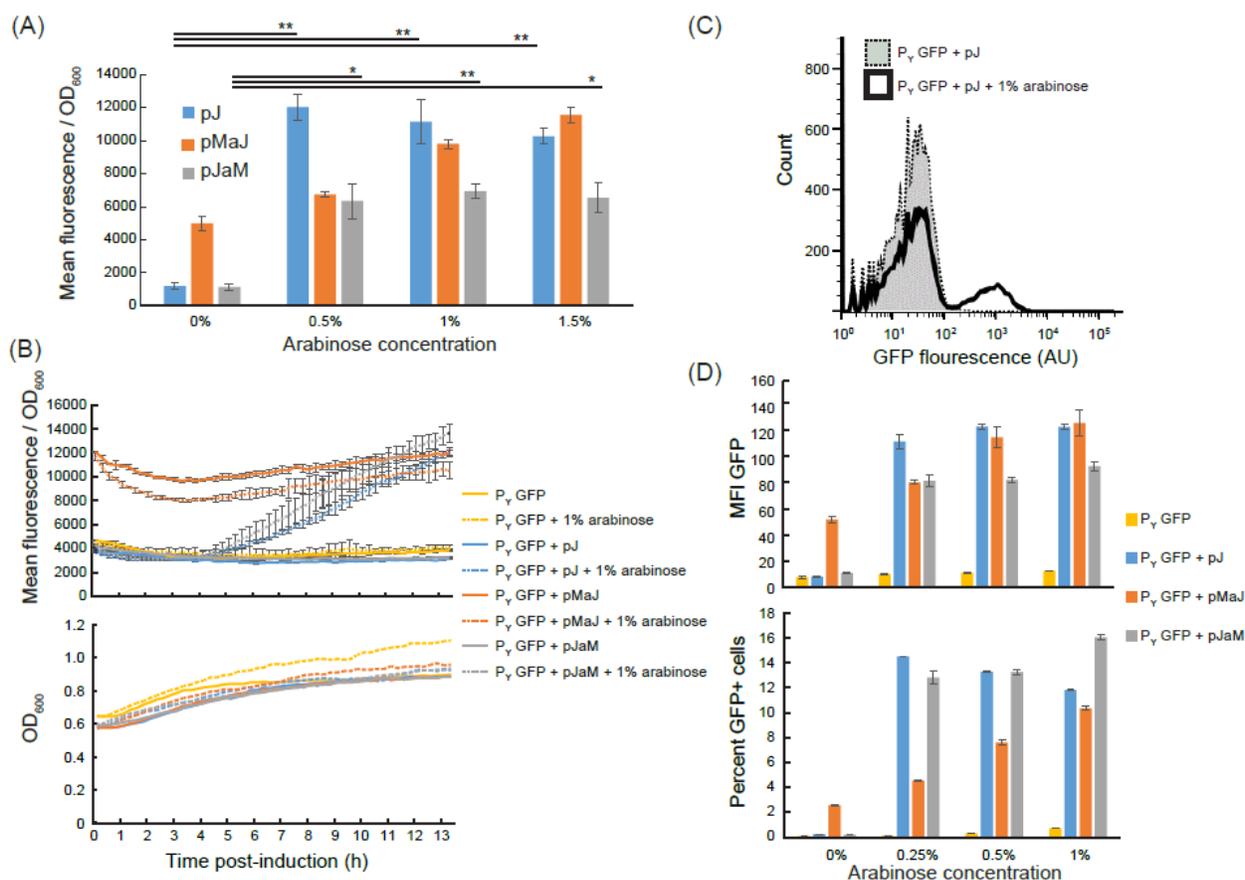


Figure 3.2. Regulation of PY-driven transcription by TraJ. (A) GFP fluorescence as an output for induction by induced orthogonal systems in response to arabinose. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. * $p < 0.05$ and ** $p < 0.005$ as calculated for a two-tailed paired Student's t-test. Data shown were collected on a microplate reader. (B) The effect of arabinose induction during stationary phase. Experiments were conducted in biological triplicate, and error bars indicate standard deviations for Mean GFP/OD₆₀₀. Data shown were collected on a microplate reader. (C) Comparisons of an induced and uninduced pJ and reporter containing population via flow cytometry. (D) Mean Fluorescent Intensities as an output for induction in a population by induced orthogonal systems in response to arabinose via

flow cytometry and percentage of cells expressing GFP. Cells examined via flow cytometry are comparable to trends as seen in panel (A) when considering mean fluorescences of populations.

Given the observed multimodal induction of P_Y-driven expression by TraJ, we next investigated whether the degree of TraJ expression impacted either the probability of inducing P_Y-driven transcription or the level of P_Y-driven transcription. We hypothesized if there exists a minimal expression level at which TraJ expression induces P_Y-driven reporter expression, then cells with low amounts of TraJ would be “off,” while all cells that are “on” would possess a level of TraJ that exceeds this threshold value. Similarly, if the TraJ regulates the extent of P_Y-driven expression (in addition to or instead of simply determining whether P_Y-driven expression is induced), then the level of TraJ expression would correlate with the level of P_Y-driven expression. To investigate these questions, a P_{BAD} traJ-mRFP construct (pJaR) was generated and evaluated for its potential to regulate P_Y *gfp* (Figure 3.3a). The pJaR construct consisted of a P_{BAD} regulated traJ cocistronically expressed with mRFP, such that mRFP expression serves as a proxy for TraJ expression at a single cell level. At a population level, pJaR induced P_Y in a manner similar to both pJ and pJaM, further confirming that the downstream gene on a traJ-led polycistronic transcript does not modulate TraJ-induced expression from P_Y (Figure 3.3B). Comparing mRFP fluorescence, our reporter for TraJ expression, and GFP fluorescence, our reporter for P_Y activation, illustrated that while GFP expression is not seen in the absence of mRFP, only a subset of mRFP expressing cells express GFP (Figure 3.3C). Therefore, we can conclude that when TraJ is being expressed, only a subset of cells with P_Y express the genes it promotes. Beyond this, the level of GFP expression does not correlate with mRFP expression; looking at only GFP expressing cells, a linear, log-linear, or exponential fit to the data could not be found with R² values in all three cases failing to rise above 0.2 (Figure 3.3D). The conclusion to be drawn from this is that the level of TraJ expression does not correlate with the level of induction from P_Y. Instead, our data

suggests that while TraJ is required to initiate expression from P_Y , it is only able to initiate this expression in a subset of cells when reaching a specific threshold in a subset of cells, and that exceeding that threshold does not enhance P_Y induction.

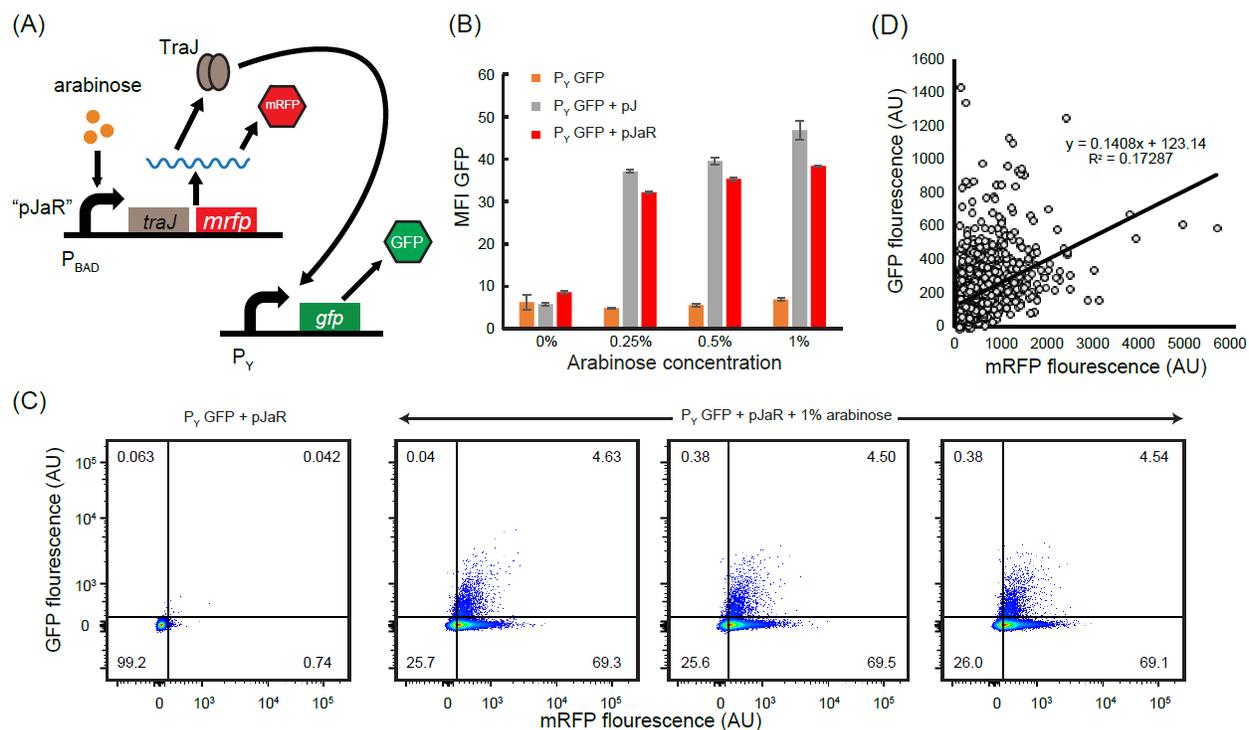


Figure 3.3: Quantitative coupling between TraJ expression and P_Y induction in single cells.

(A) Schematic representation of pJaR reporter system. Upon addition of arabinose, TraJ is transcribed along with monomeric RFP (mRFP). TraJ is then capable of acting upon P_Y to induce GFP expression. (B) GFP fluorescence as an output for induction by induced orthogonal systems in response to arabinose detected via flow cytometry. pJaR leads to GFP induction in similar patterns to that of pJ and pJam. (C) Raw flow cytometry data of pJaR reporter system in the absence of arabinose and three replicates in the presence of 1% arabinose by volume. (D) Lack of correlation between GFP and mRFP (TraJ) expression. No significant correlation between GFP and mRFP expression could be found using various fittings. Shown here is a linear distribution of GFP-positive cells and a linear fit, with an R^2 value below 0.2.

By examining the expression of TraJ (via mRFP) in cells that express GFP from P_Y , we observe three things: 1) TraJ expression is required for P_Y induction (Figure 3.3C), as GFP is not significantly expressed without TraJ induction. 2) TraJ expression does not always result in P_Y induction, as many cells appear mRFP-positive, but GFP-negative. 3) Within the population of cells inducing GFP expression from P_Y , the level of TraJ does not correlate with P_Y induction (Figure 3.3D). These three observations allow us to speculate on the behavior of the system. In an ideally behaved hyperbolic response, adding a given amount of a stimulant results in a correlated output response (117). However, our observed response is very different, with even a little expression of TraJ resulting in a strong response (GFP expression) in a subset of cells, and the fraction of cells that exhibit this strong response (High GFP expression) do not continue to grow as more stimulant (TraJ (represented by mRFP)) is added. Therefore, our observations are more indicative of binary response than of a graded response (118). This observed ultrasensitivity cannot be explained by TraJ alone; as more TraJ was added, we did not observe more cells converted to the “ P_Y on” state. Therefore, while TraJ works as a binary controller in this system, it is not the only component regulating P_Y expression and is not sufficient to induce P_Y . Indeed, this claim is consistent with prior observations, since TraJ is known to act in conjunction with a number of regulatory partners and competitors such as H-NS, ArcA, Irp, and CRP, all of which likely contribute to the fate decision of cells in terms of P_Y expression (45, 46, 107, 108). Additionally, in some redox states, TraJ is known to act differently due to structural changes resulting from environmental sensing of a PAS domain, meaning that TraJ itself may be functional in a subset of cells but ineffective or even act as a competitive inhibitor in others (47).

Since TraJ expression lead to a strong induction of P_Y in a subset of cells, we hypothesized that the addition of orthogonally regulated TraJ could lead to larger pool of active conjugation donor cells, and thus increase mating efficiencies within a population. To enable rapid quantification of a large number of cells in a population, we developed an assay for analysis of conjugation via flow cytometry (Figure 3.4A). Our method consists of the transfer of a high copy number shuttle plasmid, called pShuttle, harboring a constitutive GFP expression cassette. This plasmid is transferred by means of an immobile F derivative, which we obtained from the Registry of Standard Parts, pOX38; oriTf::CmR (for the sake of brevity is hereafter referred to as pOX38). This immobile host of conjugative machinery is able to transport pShuttle into other cells. To quantify this transfer, we used recipient cells expressing monomeric RFP from a medium copy number plasmid. To test if this method was feasible, we conducted a mating assay similar to one previously described in the literature (45) in which conjugation was allowed to occur in incubated, fresh media with donor and recipient strains, and then cells were plated onto LB agar plates. Because donor and recipient cells were distinguishable via fluorescence, we plated onto Ampicillin containing plates. pShuttle in this experiment encoded an ampicillin resistance gene. Transconjugant colonies were identified as colonies expressing both GFP and mRFP (Figure 3.2B), and mating efficiency was determined to be the number of mRFP and GFP co-expressing cells (transconjugants) to the number of GFP expressing cells (donors).

Samples were analyzed via flow cytometry. After gating out cells by forward and side scatter, we looked at the expression of GFP and mRFP within the population. For a negative control, we co-cultured “donor cells” with a pOX38 strain containing an immobile GFP encoding plasmid. pOX38 is capable of expressing the full conjugative machinery, and thus should still form

mating pairs, but does not contain the sequence for the *OriT*, the sequence of DNA associated with a functional relaxosome, and thus transfer. As can be seen from raw data (Figure 3.4C), while a number of dually expressing cells appear in the control experiment, this is a small fraction of those seen in cells with a functional donor. Samples from cultures were examined by colony based mating assays and flow based mating assays, showing similar rates between different runs (Figure 3.4D).

Using this flow based assay, the effect of the addition of orthogonally regulated TraJ to the rates of conjugation was explored. Mating cultures of donors and recipients were induced with arabinose at various time points relative to co-culture to determine the effect on conjugative rates. We tried multiple time points prior to co-culture to determine if pre-exposure would lead to an increase in TraJ promoted conjugative machinery would predispose cells to be more likely to conjugate. We also attempted induction at the time of co-culture and an hour post co-culture to determine if the addition of newly synthesized TraJ would boost conjugation. To overcome any clonal variation between runs, a series of colonies from both pJ inclusive and exclusive pOX38-Shuttle ensemble plates were averaged and compared (Figure 3.4E). Surprisingly, the addition of TraJ did not have a significant effect on conjugative rates determined by flow cytometry. Additional runs were conducted of this system and trends were consistent in multiple repeats of the experiments.

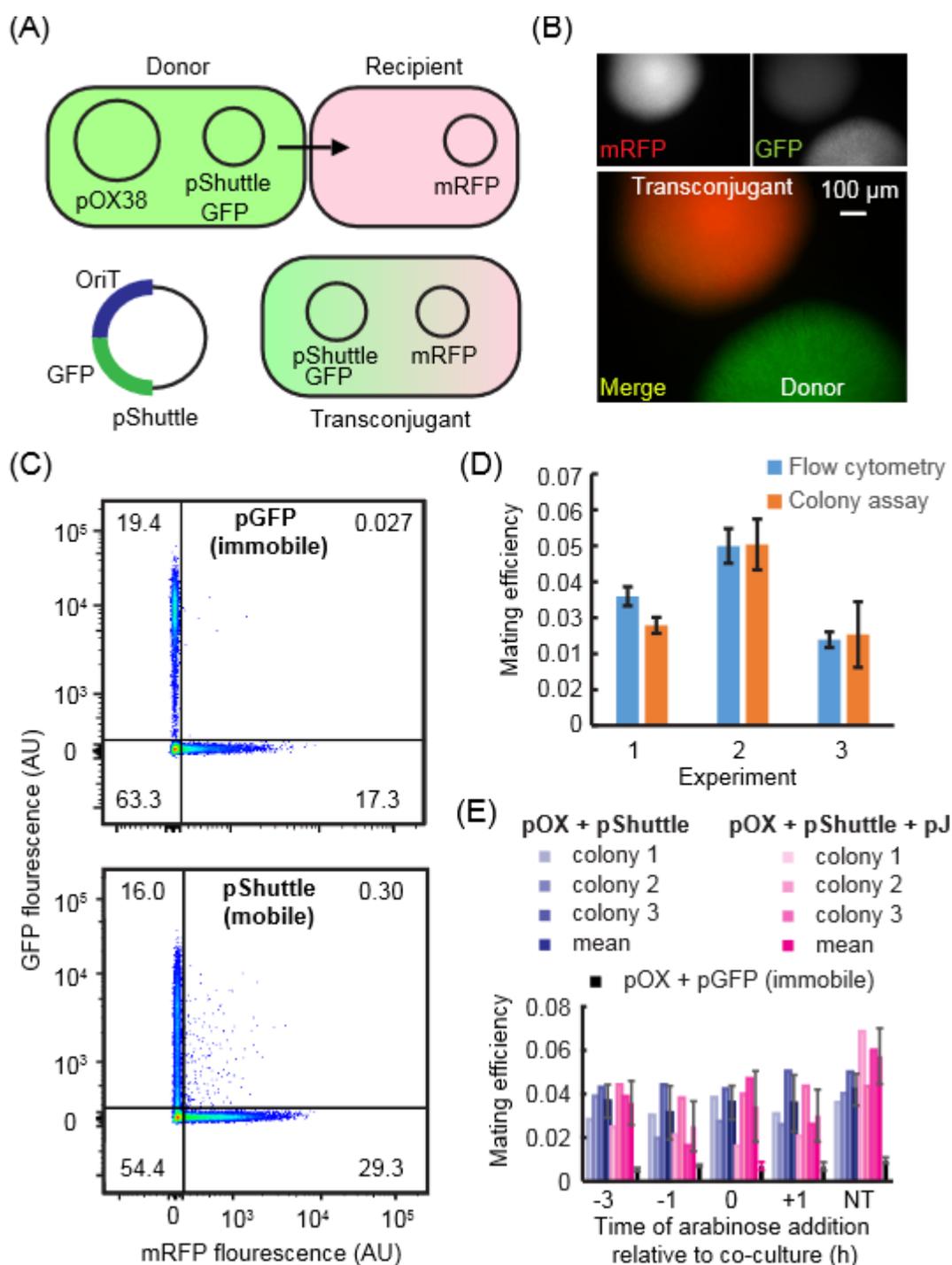


Figure 3.4: Regulation of native conjugation machinery by exogenous TraJ. (A) Schematic representation of pShuttle/pOX38 (pOX38;oriTf::CmR) conjugation reporter system. Conjugative machinery expressed by pOX38 is capable of transferring pShuttle, a plasmid that contains both a

GFP expression cassette as well as the native F OriT. Upon forming a mating pair with a donor cell that contains an mRFP expression plasmid, conjugation can occur, resulting in a transconjugant expressing both mRFP and GFP. (B) Observed colonies from donor and transconjugant cells. (C) Transconjugant cells expressed both observable mRFP and GFP, while donor cells only expressed GFP. Observation of transconjugants by flow cytometry. Populations of recipient cells paired with populations of donors or mock “donor” with an immobile GFP plasmid instead of pShuttle were co-cultured for 5 hours and then run on flow cytometry. The increased rate of doubly GFP and mRFP positive cells appears in the system with functional donors, indicating the fitness of this assay. (D) Comparison of quantification of mating assays by colony count or flow cytometry. Observed mating efficiencies by flow cytometry matched with rates seen by colony assays. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. (E) Observed mating efficiencies with the addition of pJ to donor population cells. TraJ expression was induced at various times relative to co-culture with recipient cells. Experiments were conducted in biological triplicate from three distinct colonies, the averages of those three colonies were then averaged, and error bars indicate standard deviations from initial averages.

3.3.3 Initiation of conjugation via orthogonal transcriptional control

The orthogonal regulator plasmids (pJ, pJaM, and pMaJ) were then used as a tool to engineer a novel conjugative ensemble of plasmids capable of lateral gene transfer without native regulatory elements. A large plasmid called pHeadless, consisting of the full *tra* operon under P_Y was placed in a bacterial artificial chromosome derived plasmid so that it could be maintained with antibiotic resistance alongside pShuttle as well as an orthogonal regulator plasmid. Cells were induced at various time points relative to co-culture and then examined via flow cytometry after 5 hours (Figure 3.5B). Because the pJ ensemble lacks TraM, a protein required for conjugation (119), it cannot create transconjugant cells and should be considered a negative control. Thus it serves as a negative control simply because it cannot transfer genetic information, and it should still be able to induce P_Y via TraJ. TraJ activation of the *tra* operons means that conjugative machinery could still enable the creation of a pilus, attach to other cells, and create mating pairs, a donor and recipient cell temporarily fused awaiting the completion of genetic transfer. Since this outcome of cells sticking together without a resulting transconjugant could occur if the *tra* operon was activated, the pJ system can be considered a stringent negative control. Notably, when pHeadless was paired with pJaM, significant conjugative transfer was observed, although such transfer did not require arabinose to induce high levels of TraJ and TraM (indeed conjugative transfer was somewhat higher in the absence of arabinose). In contrast, pairing pHeadless with either pMaJ or pJ alone did not result in significant conjugative transfer. Altogether, these results suggest that the low levels of TraJ (and TraM) expression driven by P_{BAD} , even in the absence of arabinose, were sufficient to confer functional induction of conjugative transfer. It is also possible

that if pMaJ drives high constitutive expression from P_Y in pHeadless as was the case for P_Y in our reporter construct, than this result also suggests that coordination of functional transfer may be impaired by high constitutive expression from P_Y. Finally, even though pJaM mediated functional induction of conjugation, the observed rates were lower than those observed for pOX38-mediated transfer. In sum, these results demonstrate that regulation of the TraJ and TraM may be decoupled from regulation of the *tra* operon, although maximal rates of conjugative transfer may be sensitive to the level (and perhaps timing) at which TraJ (and potentially TraM) are expressed.

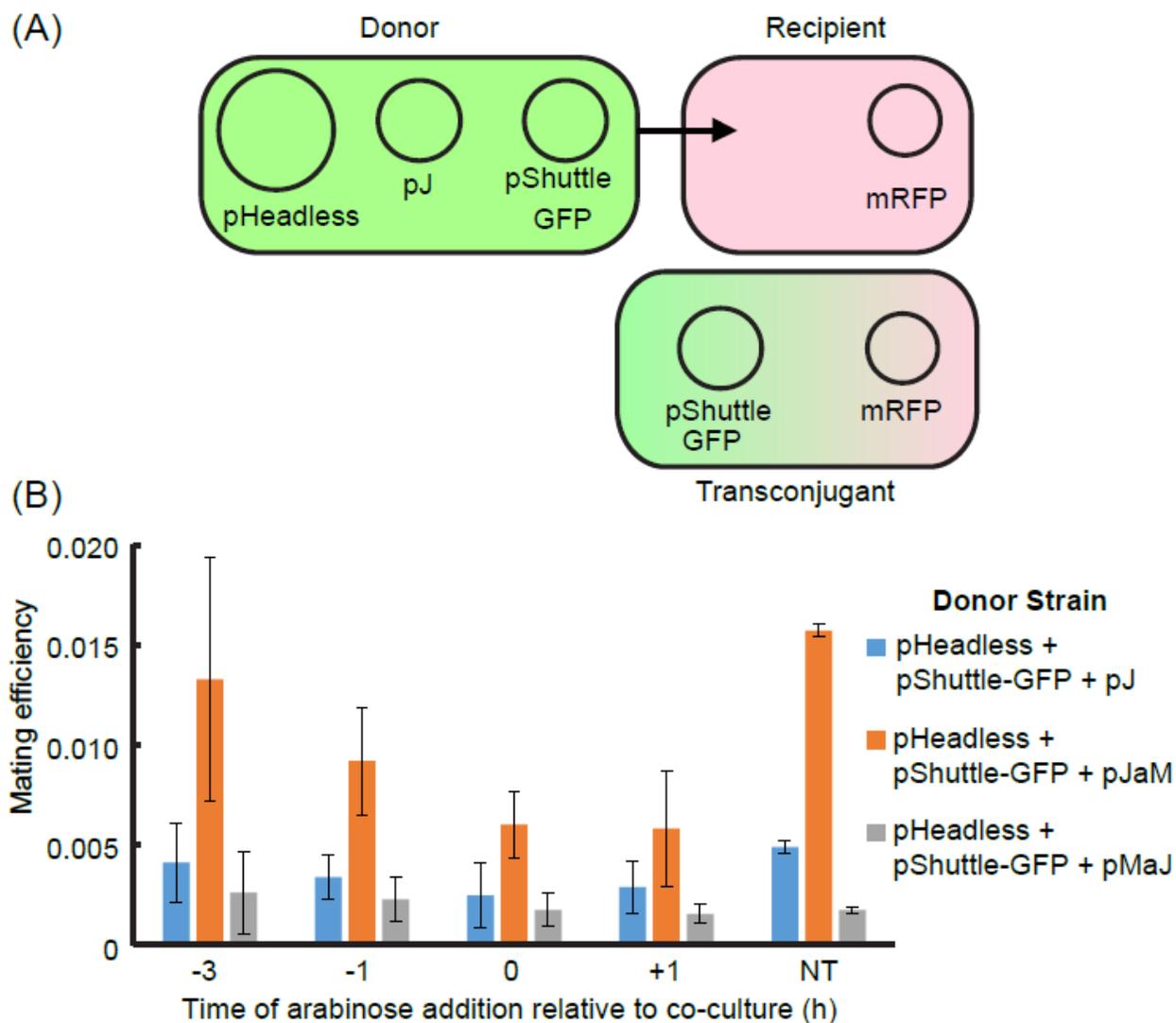


Figure 3.5. Orthogonal regulation of conjugative transfer. (A) Schematic representation of novel, orthogonally regulated conjugation reporter system. Conjugative machinery is expressed from pHeadless after the induction of TraJ expression from an orthogonal expression plasmid (Shown in image as pJ). (B) Observed mating efficiencies by flow cytometry of orthogonal conjugative systems. Experiments were conducted in biological triplicate, and error bars indicate standard deviations. No treatment condition was determined to be significantly different, with $p < .05$ as determined by a two tailed Student's t-test.

3.4 Summary and future directions

The first P_Y reporter was described in 1991 (120), and since then, the list of P_Y regulators identified has grown at a steady pace, yet we still understand little about how P_Y -driven expression is induced under different growth conditions. However, previously described systems have utilized enzymatic LacZ reporter systems, which reveal population average behavior but do not enable monitoring of cell-to-cell variability within the population. To meet this need and investigate how conjugative transfer is initiated at the single cell level, we developed a novel fluorescent reporter system. By determining that both a truncated P_Y reporter as well as a reporter on a high copy number plasmid failed to exhibit TraJ-inducible expression from P_Y (Figure 3.1C and 1D), we identified several novel features of TraJ- P_Y interactions. When paired with our observations that TraJ expression levels did not correlate with the degree to which transcription is induced from P_Y (Figure 3.3D), the fact that a high copy number plasmid fails yield a functional reporter suggest that a threshold concentration of TraJ is required to induce transcription from P_Y , but after that point, more TraJ does not confer more transcription from P_Y . Thus, TraJ appears to regulate P_Y -driven expression in a “digital”, rather than “analog”, fashion. We hypothesize that on the high copy reporter plasmid, more copies of P_Y per cell may result in a higher threshold that was not reached even under induced expression of TraJ. An alternative explanation is that induction of transcription from P_Y requires host factors that become limiting in the context of a high copy number plasmid, such that TraJ is not sufficient to induce transcription from P_Y . The fact that a truncated P_Y failed to work as a functioning reporter indicates that the coding sequence of TraJ, which is all that differs between the two reporter constructs, is important for conferring TraJ-

inducible regulation of P_Y . This hypothesis is also consistent with our observations comparing pJaM- and pMaJ- mediated regulation of either P_Y -GFP or pHeadless — while the coding sequence of TraM has never been shown to have an effect on expression of TraJ, multiple loci within the F plasmid's genetic elements have been shown to contribute to altered expression of transcripts found elsewhere on the plasmid (32). The removal of the intergenic region between *traM* and *traJ* in our construct may have either introduced or altered one such additional regulatory element.

Our key finding that a subset of cells expresses GFP from a P_Y reporter in the presence of TraJ expression strongly indicates that TraJ has a key role in the stochastic induction of expression of conjugative machinery, but heterogeneity in TraJ expression is not sufficient to explain heterogeneity in P_Y -driven expression. Other components of the regulatory mechanism must create conditions within individual cells that allow for a decision as to whether to activate conjugative machinery once a threshold level of TraJ has been reached. For instance, TraJ may fail to properly coordinate with ArcA, another protein required for induction of P_Y , in a subset of cells, or may fail to overcome repression by H-NS or targeted degradation.

It is important to note that the mechanism herein hypothesized for TraJ's regulation of P_Y did not closely resemble that of a bistable, hyperbolic, or ultrasensitive circuit. In a bistable circuit, adding inducer above a certain concentration leads to switch-like behavior (121). While there was indeed a threshold concentration of TraJ necessary to induce expression from P_Y , this did not lead to long term, stable expression, which would otherwise be expected from a bistable system. TraJ expression was required but not sufficient in order to drive expression. Our system is also not ultrasensitive. In a hyperbolic system, an increase in the stimulus should lead to an initial direct increase in the response (116). However, simply increasing the stimulus (TraJ), did not uniformly

lead to an increase in the output genes, and the magnitudes of expression of stimulus and response did not correlate. In an ultrasensitive system, expression of a key element above a specific threshold leads to induction of the system. Again, while TraJ was required for induction, TraJ was not sufficient in all cells to drive induction of P_Y . We did not observe a particular concentration of TraJ (or in this case, our fluorescent proxy for TraJ) that was sufficient to lead to consistent induction of P_Y , and therefore our observations are not explained by a simple model of ultrasensitivity (i.e., to TraJ).

Our final experiments exploring the possibility of orthogonal control of conjugation provide promising initial clues as to how one might engineer an experimentally-inducible lateral gene transfer platform. Conjugation is already being used as a tool for transfer between bacteria and across kingdoms (37, 39, 41), and creating an engineered and tunable transfer system would be a useful tool for introducing new genetic elements upon small molecule induction. This would prove useful when engineering a system where the introduction of genetic elements or gene expression is required at specific time intervals or when specific conditions have been met. Transfer only upon induction would allow for control over when genetic elements are introduced to a system, possibly allowing for the accumulation of precursors in a metabolic pathway or regulatory elements in a genetic circuit. This platform also comprises a useful new experimental tool for elucidating the role of key regulatory elements in isolation from the complex native regulatory network. Our results suggest that increasing rates of conjugative transfer under orthogonal control may be achieved by tuning the relative levels of both TraJ and TraM and placing the expression of TraJ (and potentially TraM) under the control of an inducible promoter with very low background expression. Since cells bearing pJaM and pHeadless exhibited the highest

conjugative transfer rates in the absence of arabinose, it is also possible that the optimal induction of conjugation may occur when TraJ (and perhaps TraM) are only induced to low, but finite, levels of expression. Improvement on this system may also come from adding other possible regulators or interaction partners to the set of induced genes, allowing for a stronger regulatory push. Altogether, this investigation provides new tools and insights for understanding and controlling how conjugation is regulated within individual cells.

CHAPTER 4

Predicting bacterial genome copy number dynamics under variable growth regimes using a novel quantitative model

4.1 Introduction

For many applications in microbial synthetic biology, optimizing a desired function, such as biosynthesis via an engineered metabolic pathway, requires careful and labor-intensive optimization of the degree to which various genes are expressed (122). Transgene copy number, genomic integration site, promoter strength, translational efficiency, and culture conditions all impact “function” in a manner that is difficult to predict and typically requires high throughput screening (123), (124), (125) or evolutionary selection for such properties (126). One challenge for predicting such effects or even interpreting typical characterization experiments is that in bacteria such as *E. coli*, genome copy number varies widely across different phases of growth, often exceeding 8 copies per cell during portions of a typical fermentation (49), (50), (71), (127). Genomic replication also impacts how and when endogenous genes are expressed from different loci (51), and in addition, growth rate affects gene expression at a global level (128). While such phenomena are relatively well-understood at a mechanistic level and benefit from decades of research in this area, there are important gaps in our quantitative understanding of such processes that limit our ability to predict or explain the impact of genomic copy number variation on engineered biological functions.

In *E. coli*, upon which this discussion will focus, DNA replication begins at the origin of replication, *oriC* (54), and is facilitated by DnaA (55). The initiation of replication is largely

regulated by DnaA, which accumulates at oriC via an ATP-dependent process to alter the conformation of oriC and recruits the helicase, DnaB, which initiates the process of replication (56), (129),(57). Replication proceeds in both directions from oriC and continues until termination, a process that is coordinated by interactions between the protein Tus and one of 10 Ter sequences along the genome (58). Tus complexes with Ter sites in an asymmetric fashion, which enables the complex to function as a direction-specific terminator of replication. When a replication fork reaches a Tus complex from one direction, the Tus complex will dissipate and replication will continue (59). If the replication fork reaches the Tus complex from the other direction, however, the Tus complex will remain intact, inhibiting further replication until the opposite replication fork is able to complete replication from the other direction. Thus, the concentration of active (ATP-bound) DnaA is generally understood to integrate the effects of cell size and other aspects of cell state in order to regulate genome replication, although these processes are only indirectly connected to the regulation of cell division. Indeed, decoupling the regulation of cell division and genome replication enables *E. coli* to grow rapidly when nutrients are abundant. The minimal time required to replicate the genome is about 40 minutes, but by initiating multiple rounds of replication prior to cell division, the doubling time can be substantially shorter than 40 minutes (54).

The regulation of cell division in *E. coli* is less well understood than is the regulation of DNA replication, but regulation is strongly coupled to cell size (24), and a number of essential regulators of cell division have been identified (130). The first protein to localize at the point of division is FtsZ, a tubulin homologue that forms a ring around the nucleoid of the cell (61). This ring can form well before the two distinct genomic copies separate into distinct regions. The ring then complexes with the membrane-bound ZipA, tethering the ring to the membrane in preparation to pinch the membrane during division (62). This leads to the recruitment of a number of proteins to form the divisome, which contracts the membrane while actively transporting chromosomes to

opposite daughter cells to complete the process of division (63). While the adaptive benefits of such decoupling between genome replication and cell division are well understood, this decoupling also presents challenges for predicting both dynamics and intercellular variations in genomic DNA content.

To address these issues, a number of theoretical models have been developed to predict variations in genomic DNA content. The most widely used model was first created in 1968 by Cooper and Helmstetter (hereafter, CH model) (65). The CH model was the first to formally describe the relationship between mass accumulation and chromosome dynamics of a bacterium, and it explained, among other things, the aforementioned dynamics associated with overlapping rounds of genome replication (65). The central observation motivating the CH model is that every DNA replication initiation event occurs at a fixed ratio between the number of origins of replication (copies of *oriC*) and the mass of the cell (131). To explain this observation, the CH model separates the cell cycle into three distinct phases. The “C” phase represents the time required to complete one round of genomic DNA replication. Once the cell completes at least one round of replication, the “D” phase represents the period during which the cell then undergoes segregation of the chromosomes into two daughter cells to complete cell division. If the doubling time is greater than the sum of the C and D periods, then another phase arises called the “B” phase, which is simply the time required for the cell to accumulate enough mass to initiate a new round of replication (132). By combining the CH model with a probability density function (PDF) describing the theoretical age distribution of a population growing exponentially (133), the DNA distributions of such a population can be calculated using the growth rate, C and D times (129). This strategy has been widely used to determine the C and D parameters, for example by fitting simulated DNA distributions to experimentally measured DNA distributions sampled from exponentially growing cultures (71),(53).

However, outside the scope of exponential growth the previously described PDF describing the age distribution of a population is no longer valid, since the previous assumption of unrestricted growth cannot be made. During a typical batch fermentation, cells exiting exponential growth experience a transition phase, a stationary phase, and finally a death phase, during which growth limiting factors becoming increasingly important. Furthermore cells grown upon complex media may exhibit multiple instances of such phases, as well as other dynamics (129),(53). To help explain growth restriction, various mechanistic growth models have been proposed (134), each of which is an extension of the Malthusian growth model with context-specific parameters that restrict growth in ways that may be loosely attributed to biological mechanisms (1). However, Malthusian models generally provide poor fits to experimentally observed growth curves for a wide range of growth conditions (135). Monod pioneered the early development of empirical models that better fit experimental growth curves, by mathematically linking the growth rate of a population of cells to particular growth-limiting substrates (136),(137). Except in the case of idealized three phase growth curves, however, all such models are generally highly inaccurate. Since the vast majority of synthetic biology characterization experiments utilize batch fermentations, such as shake flasks, and microtiter plate cultures, which exhibit complex growth dynamics particularly as cell densities become appreciable, new tools are required to predict and evaluate genomic DNA content under such conditions(134),(138).

To meet this need, we report a novel modeling strategy that leverages agent-based simulation and high performance computing to robustly predict the dynamics and heterogeneity of genomic DNA content within bacterial populations across variable growth regimes. We show that by directly feeding routinely collected experimental data, such as optical density (OD) time series, into our mechanistic simulations, our model predicts genomic DNA distributions that accurately reproduce those observed experimentally over a range of non-exponential growth conditions.

4.2 Materials and Methods

4.2.1 Growth Conditions and Strains

Cells were grown in Lysogeny Broth (LB) Lennox formulation (10 g/L of tryptone, 5 g/L of yeast extract, 5 g/L of NaCl) or supplemented M9 (M9 minimal medium with 0.4% glycerol, 0.2% casamino acids, and 1 mM thiamine hydrochloride) as specified, and all cultures were run at 37°C. Streptomycin was used at a final concentration of 50 µg/ml. All experiments were conducted using the commercially available TOP10 strain (F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ r-hsdRMS- Δ r-hsd nupG recA1 araD139 Δ (ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ -) (Invitrogen).

4.2.2 Bacterial genome quantitative PCR

Cells were grown overnight in 3 mL of M9 supplemented media and appropriate levels of streptomycin. Cells were prepared for growth using a BIOSTAT® B (Sartorius Stedim Biotech) benchtop bioreactor by adding 250 µL to 4 mL of M9 supplemented media and grown for 2 hours into exponential phase. Components of the bioreactor were autoclaved, assembled, media was added, and media was allowed to acclimate to proper temperature and aeration for one hour. 1 mL of cells grow in into exponential phase were added to each bioreactor. Samples were collected upon addition and after every half hour thereafter and oxygen levels and OD₆₀₀ were recorded. 5 µL of each culture was added to 1 mL of PBS and placed on a hot plate for 5 minutes, effectively killing and lysing all cells. Lysed cell solutions were then immediately placed in a -20 freezer until samples could be analyzed by quantitative PCR.

A protocol for analysis of TolC counts in a population was outlined in a paper by Wan et. Al. in 2011 (115). qPCR primers from that publication were used to quantify TolC copy numbers

in solution (Forward: CGACAAACCACAGCCGGTTA, Reverse: CAGCGAGAAGCTCAGGCCA). A plasmid with the coding sequence for TolC was constructed. A standard curve with that plasmid was created using a fivefold dilution series from 1 ng/ μ L solution. Samples were quantified

4.2.3 Batch culture growth, genomic labeling, and flow cytometry

For batch cultures, cells were grown overnight in 3 mL of LB and appropriate levels of streptomycin. Cells were then diluted fifty fold in 100 mL of growth media and placed in a shaking incubator. Cells were sampled prior to addition to growth media and upon addition, then every 30 minutes thereafter. Sampled cells were used to determine OD_{600} and treated for quantification of genomic distribution by flow cytometry.

For examination by flow cytometry, sampled cultures were diluted 5 μ L culture into 200 μ L ice-chilled PBS. For OD_{600} 's under 0.5, 1 μ L culture into 200 μ L ice-chilled PBS. for OD_{600} 's between 0.5 and 2.0, and 0.5 μ L culture into 200 μ L ice-chilled PBS. for OD_{600} 's over 2.0. 800 μ L ice chilled ethanol was then added to this solution. The solution was gently shaken, and the immediately spun with a microcentrifuge for 5 minutes at 1000xg. Ethanol solution was discarded and pelleted cells were then resuspended in 500 μ L chilled PBS, and spun a second time in the same conditions. Cells were then resuspended in 500 μ L chilled PBS with 1 μ g/mL DAPI, and immediately placed in a 4 C refrigerator until samples could be run on the flow cytometer. Cells were run on the flow within 24 hours of collection. Samples were analyzed on an LSR II (BD), and analyzed using FlowJo software (Treestar). A minimum of 2,000 individual cells (typically out of \sim 25, 000 events) was analyzed per sample

4.3 Results and Discussion

Most simulations of bacterial population dynamics start with a simple model of growth, typically framed at the single cell level, which is then expanded to predict the growth dynamics of populations (71), (53), (69). Our objectives are not well served by this approach, in that from a bioengineering standpoint, it would be useful to have a model that works in the reverse direction – starting from simple, experimentally measured growth curves (OD vs. time), such a model would enable one to infer the growth dynamics of the individual cells within such a population. To achieve this goal of describing chromosomal dynamics across a heterogeneous population, we designed an agent-based simulation framework termed the Heterogeneous Multiphasic Growth (HMG) simulator. This framework comprises two distinct innovations – an “injection growth” mechanism and a novel agent-based description of the bacterial cell cycle.

The “injection growth” mechanism was designed to enable us to relate experimentally measured growth curves to the growth of individual cells within the simulation. To implement this mechanism, at every time step, we calculate the population’s instantaneous volumetric changes (where, in this case “volumetric” refers to the collective cellular volume of the population), and distribute the changes in volume equally among all members of a population (assuming a well-mixed system). This mechanism thus enables us to match the growth of simulated cells to the growth of the measured population without restricting our analysis to the limited window or assumption of balanced growth, such that the injection growth model may be used across various phases of growth (134), (139), (138). Because population dynamics under exponential growth have been extensively studied and robustly mathematically defined, we leverage this knowledge to initiate our simulated population of cells such that the simulated DNA distributions and cell states match those expected for exponentially growing cells (49), (140), (69). In practice, this requires us to first identify the section of the experimental growth curve in which cells are growing

exponentially, and we developed a partially automated strategy for doing so. We then initiate our simulations using this measured exponential growth rate, and the model is advanced under conditions of Malthusian growth to generate a population of simulated cells that represent a distribution of states observed during exponential growth. We hypothesized that if such an injection model were coupled to a suitably mechanistic description of DNA replication and cell division, it may be possible to predict genomic DNA distributions beyond exponential phase growth.

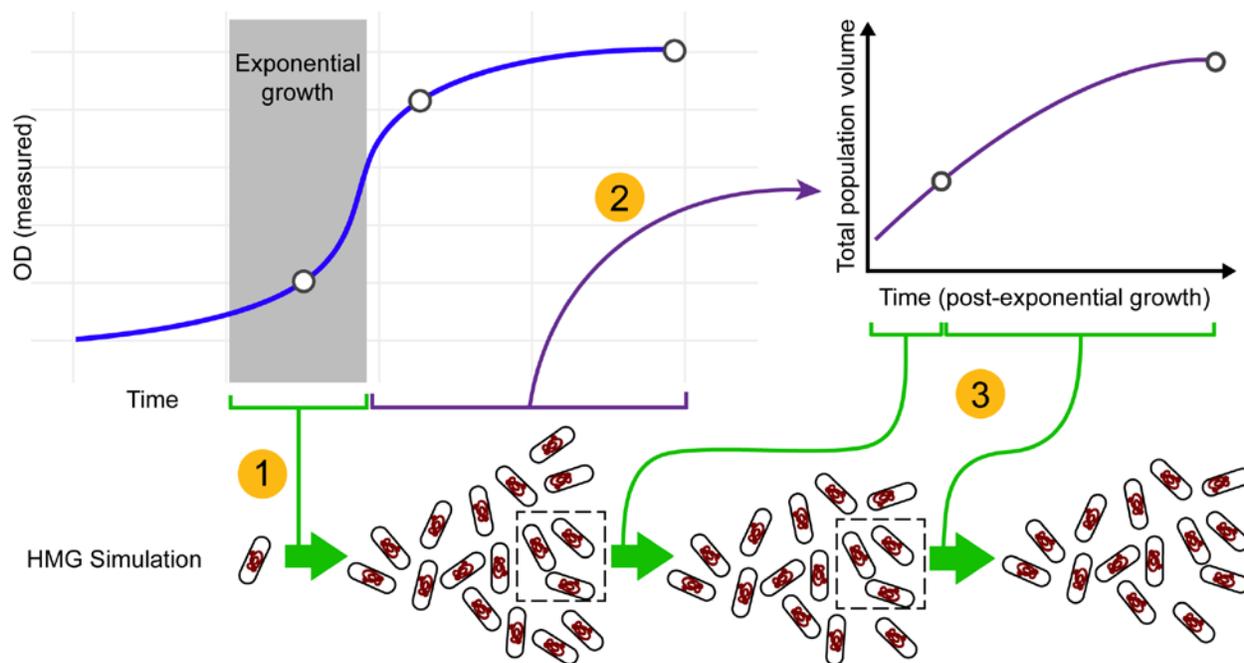


Figure 4. 1. Injection-based strategy for connecting the HMG simulator to empirical growth data.

This cartoon summarizes the process by which empirical growth data (e.g., a measured OD vs. time curve) is used to “drive” the HMG simulator via the volume injection method, where the open circles represent the sections of the growth curve where DNA distributions were measured. Thus in this illustration, the simulation would contain three independent steps: (1) The region of exponential growth is identified. This exponential growth rate is used to drive the HMG simulation from a single cell inoculate to a diversified population of exponentially growing cells; (2) During post-exponential growth, the OD curve is used to calculate the rate at which the overall cell volume (of the population) is increasing; (3) At each time point, the calculated rate of volumetric change (per cell) is “injected” into each cell in the population, each of which advances its cell state via the HMG algorithm. The dashed rectangles indicate that during each time step of the simulation, a random subset of 5000 cells is taken forward into the subsequent time step of the simulation in order to keep simulations computationally tractable.

The second innovation of the HMG simulator comprises a novel agent based model of the bacterial cell cycle, wherein the growth and each individual cell in a population is simulated in parallel. The central algorithm describing this model is summarized in Figure 4. 2. The advantage of performing population simulations using this method is twofold. First, biological noise can be accurately captured in a predictive and mechanistically meaningful manner. Secondly, we can examine the dynamics of individual cells in a population to elucidate their role in the overall dynamics of a population, which facilitates both model development and utilization. As described above, the overall goal of our modeling framework is to take an experimentally measured growth curve (OD vs. time) as an input and predict the dynamics and distributions in genomic DNA content over time; therefore, we explicitly do not attempt to predict growth as a function of any experimental parameter.

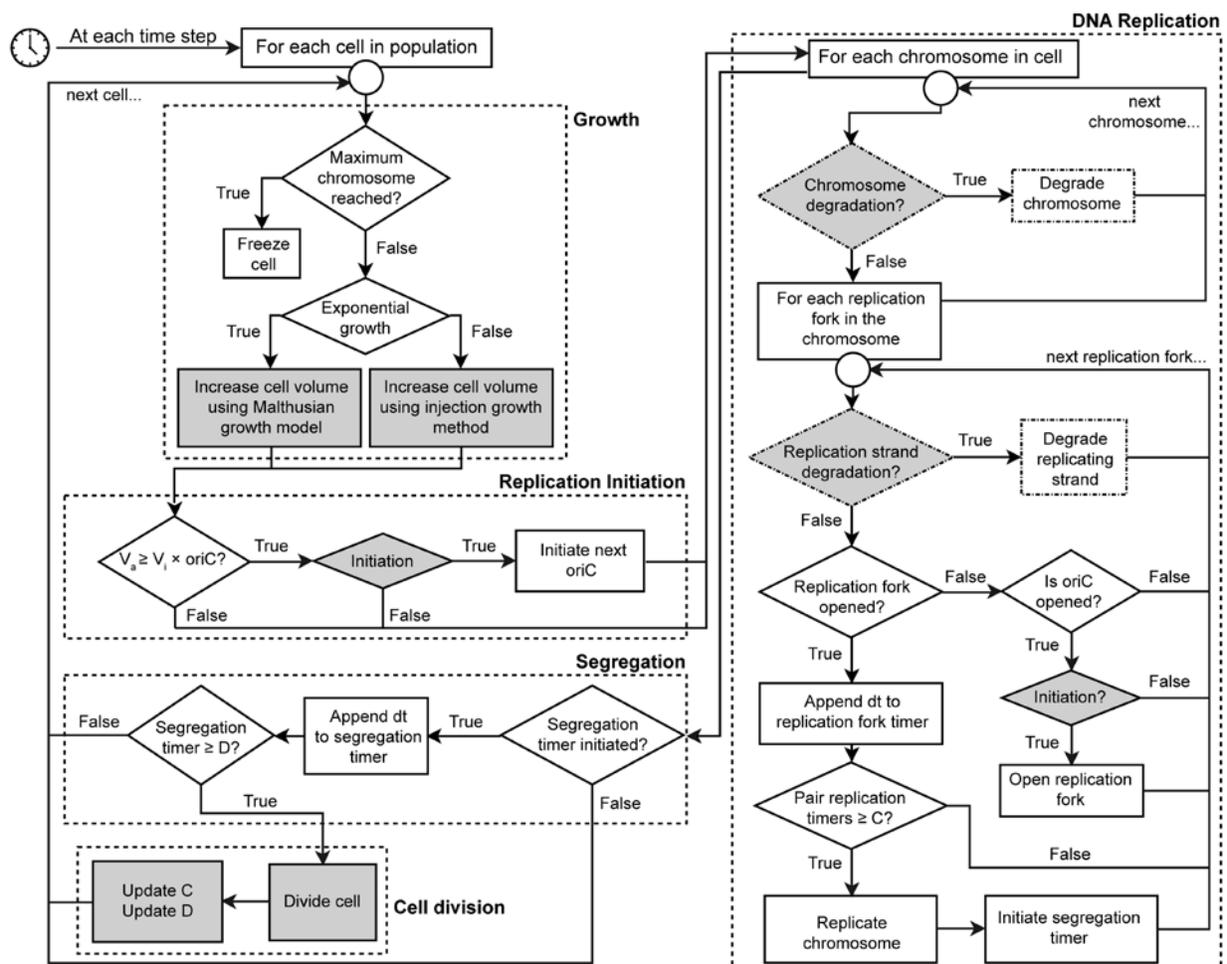


Figure 4. 2: Heterogeneous Multiphasic Growth (HMG) simulation algorithm. This figure summarizes the algorithms used to advance our agent-based simulation of bacterial growth. This algorithm marries our “injection” model for driving growth based upon experimentally measured growth curves with either the original CH model of bacterial replication (ignoring the dashed boxes) or an extended version of the CH model which incorporates the effects of *recA* mutation (including the dashed boxes). In each time step of the simulation, each cell is advanced through the 5 indicated processes: (1) Growth, (2) Opening of origin(s) of replication, (3) DNA replication and DNA degradation, (4) Segregation, and (5) Cell division. Gray boxes indicate the steps in the algorithm where noise is applied to the cell cycle.

After calculating the volume injection rate, each individual cell of our agent-based simulation otherwise follows the standard CH model, including the following key steps: (1) When the volume of a cell reaches the critical initiation volume (V_i), all *oriC* in that cell are deemed competent for replication initiation (or “open”); (2) Replication at each *oriC* is stochastically initiated to represent asynchronicity between independent chromosomes as well as overlapping rounds of replication events on a single chromosome; (3) When the C period “timer” elapses for any replication event in a cell, the D period segregation “timer” starts; (4) When the D period timer elapses, the cell splits into two daughter cells, with each partially or completely replicated chromosome being randomly assigned to one daughter or the other in a symmetrical fashion. At the time of cell division, new cell cycle parameters (C & D timers) are assigned to each daughter cell and Gaussian noise is applied (as illustrated by the grey boxes in Figure 4.1). To validate our injection model framework, we fed previously reported C & D parameters associated with a particular growth rate into our HMG simulation and confirmed that the DNA distribution predicted under conditions of exponential growth matched those generated by a validated Monte Carlo simulation of the canonical CH model (69).

Our HMG simulator also captures the fact that many laboratory bacterial strains contain a mutant version of the *recA* gene (*recAI*) which has greatly reduced DNA-dependent ATPase activity (141), which is the case for the strain used in this study (TOP10). As critical mass is reached, all *oriC* loci in the cell open simultaneously, such that in a WT population of cells, the vast majority of individual cells contain 1, 2, 4, 8.... (i.e. 2^n) chromosome copies per cell (132). Populations of *recAI* cells, on the other hand, contain individuals with whole numbers (1, 2, 3, 4 etc..) of chromosome copies per cell (142). Our current understanding is that this phenomenon likely stems from a dysfunctional DNA repair mechanism (142, 143) (144), although the exact mechanisms by which the lack of functional RecA impacts the replication and repair mechanism is still the subject of intensive research (145). Therefore, we decided to capture the impacts of

RecA deficiency by including only high-level consequences that are generally believed to lead to the observed aberrant chromosome copy number phenotype: in our simulation, when DNA damage occurs in RecA-deficient cells, the inability to repair this damage by homologous recombination results, stochastically, in degradation of either the replicating strand or the entire replicating chromosome. Although there is no direct evidence that *recA* mutation-associated DNA degradation varies with growth rate (142), (146), there is evidence that mutant *recA* phenotypes are exacerbated by faster growth rates, so a reasonable explanation is that faster growth leads to more replications forks, and therefore more stalled replication forks, and therefore more instances in which lack of RecA leads to DNA degradation (147).

We next developed a strategy for calibrating our HMG simulator to experimentally generated growth curves that include both exponential and post-exponential growth phases. When considering exponential growth under various conditions, the times for replication (C period) and segregation (D period) are highly correlated with growth rate (53), (69). Overall, as the growth rate decreases, both C and D periods increase from some minimal value, and these relationships are relatively well-described by simple empirical functions.

To evaluate our HMG simulator against experimentally measured DNA distributions, we first measured DNA content across multiple phases and conditions of growth using the type of shake flask cultures typically used for routine characterization of engineered strains (Figure 3 and 4). As a base case implementation of our HMG simulator, we calculated C & D parameters using functions based upon analyzing exponential growth¹, and in this base case, we omitted any description of *recA*-related defects (Figure 4.2) (53), (69). As expected, this base case simulation produced relatively close matches to the experimental data during exponential growth and early stages of the transition to stationary phase (Figures 4.3-4.4). However, as the population departs further from exponential growth and approaches stationary phase, these predictions become increasingly inaccurate.

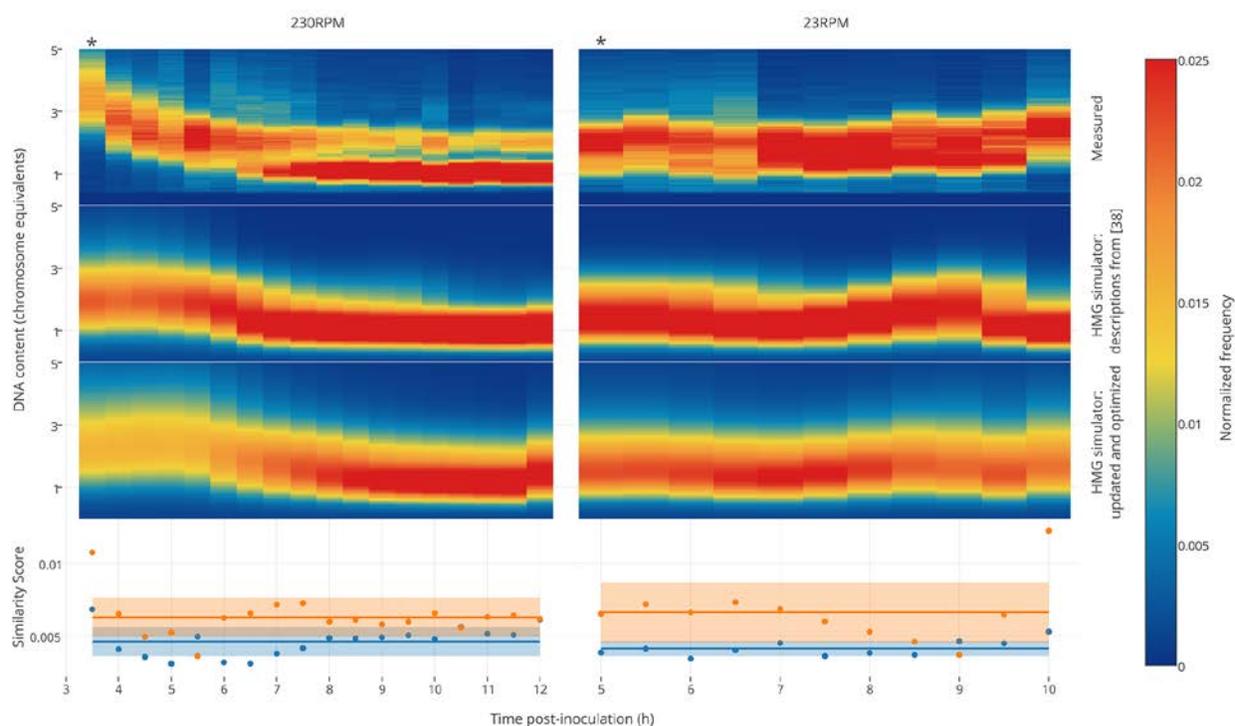


Figure 4.3: Training of the HMG simulator framework. The HMG simulator was “fed” growth curves for TOP10 cells grown in LB, shaken at 230RPM or 23RPM, and simulated DNA distributions were compared with those which were measured empirically. The measured DNA distributions shown here each represent a single experiment, each of which is representative of two or more independent experiments. The first column within each heat map represents the exponential growth phase (indicated by *), and all subsequent time points represent post-exponential growth. The simulator was run using two different models: the first model was based upon a prior description of exponential growth (69), which omits any consequences of *recA* mutation, and the second (updated and optimized) model incorporated our description of the consequences of *recA* mutation with parametric optimization. The solid lines on the two bottom panels represent the mean similarity score across the time course, and the shaded boxes represent the standard deviation of these scores across the time course.

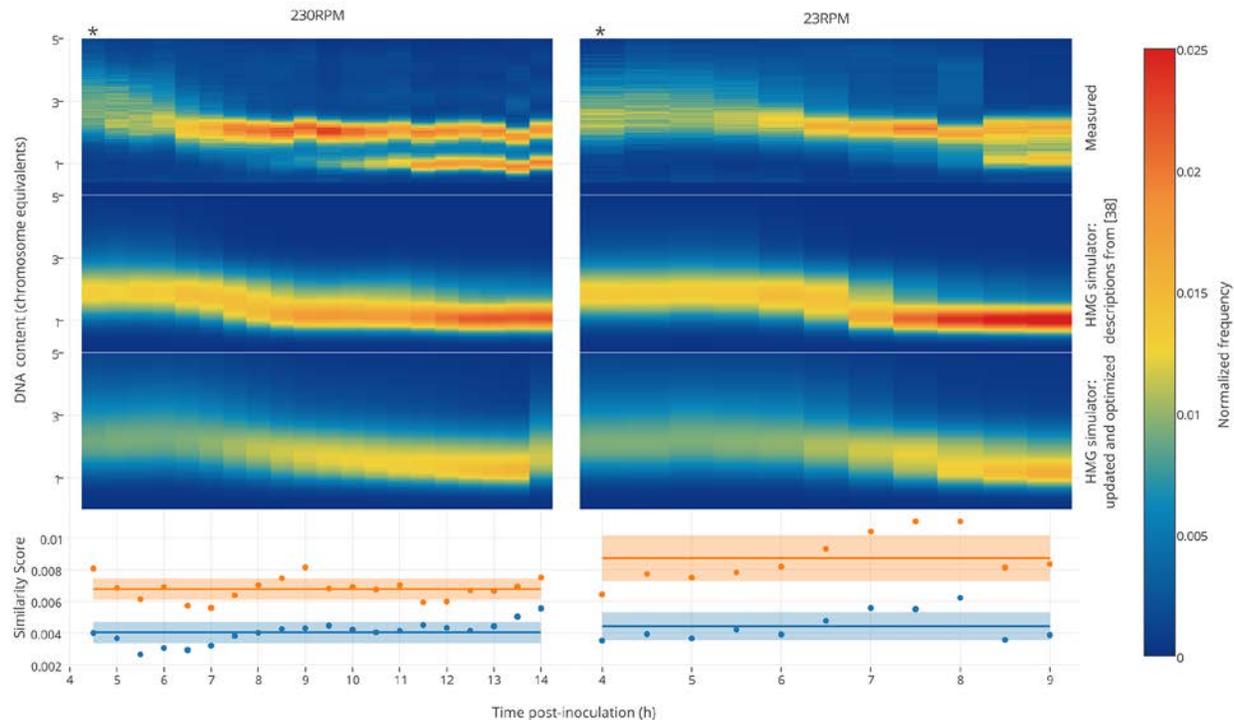


Figure 4.4: Validation of the HMG simulator framework. The HMG simulator was “fed” growth curves for TOP10 cells grown in M9, shaken at 230RPM or 23RPM, and simulated DNA distributions were compared with those which were measured empirically. The measured DNA distributions shown here each represent a single experiment, each of which is representative of two or more independent experiments. The first column within each heat map represents the exponential growth phase (indicated by *), and all subsequent time points represent post-exponential growth. The simulator was run using two different models: the first model was based upon a prior description of exponential growth (69), which omits any consequences of *recA* mutation, and the second (updated and optimized) model incorporated our description of the consequences of *recA* mutation with parameters optimized based upon growth in LB (i.e., using the same updated and optimized model described in Figure 4.3). Similarity scores indicate the degree to which each prediction matches the observed DNA distribution. The solid lines on the

two bottom panels represent the mean similarity score across the time course, and the shaded boxes represent the standard deviation of these scores across the time course.

To improve upon these predictions for post-exponential growth, we next attempted to improve the HMG simulator in several ways. First, we incorporated the simplified description of *recA*-associated DNA degradation. Next, we attempted to optimize both parameters associated with this DNA degradation as well as parameters for the functions relating C and D periods to growth rate. To this end, we utilized a Genetic Algorithm (GA) to fit simulation parameters to one subset of our data – experimentally measured DNA distributions for TOP10 cells grown in LB at two different shake rates (230RPM and 23RPM) (Figure 4.3). This updated and optimized HMG simulation exhibited a 26.72% and 37.7% improvement in accuracy for predicting measured DNA distributions for the 230RPM and 23RPM cases, respectively, as quantified by a similarity score.

We next investigated whether the updated and optimized HMG simulator could also predict genomic DNA dynamics for cells grown under conditions not included in the training data (i.e., as model validation), and to this end we examined cells grown in M9 medium. As shown in Figure 4.4, the updated and optimized HMG simulator generates excellent fits during both exponential and post-exponential growth, and the optimized parameters provide a significantly better prediction of the experimental data than was achieved using standard parameters from the literature (increases in accuracy were 40.34% and 49.15% for cells shaken at 230RPM and 23RPM, respectively, as measured by similarity score). This improvement in performance is also readily visualized, qualitatively, by comparing histograms corresponding to the data presented as heat maps in Figures 4.3 and 4.4, which clearly indicate the points at which the pure CH model, run under the HMG simulator, breaks down.

To illustrate the potential usefulness of our model as a design tool, we considered a scenario in which the designer wishes to insert three novel genes into the *E. coli* genome, and the HMG simulator is used to predict how insertion position impacts gene dosage dynamics across various growth conditions. For example, if these genes encoded three enzymes, then such dynamics could substantially impact the relative concentration of these enzymes relative to both one another and

to endogenous enzymes and other gene products. To this end, we generated theoretical growth curves representing different biologically realistic growth dynamics, and we used the optimized HMG simulator to predict gene dosage dynamics (Figure 4.5). The scenarios considered could represent fast and slow growth, growth on heterogeneous substrates (e.g. multiphase growth), and discontinuous growth conditions such as might be observed during metabolic adaptation (e.g., start-stop). As expected, genes located closer to *oriC* exhibit greater variation in mean copy number than do those distal from *oriC*, since genes located closer to *oriC* are present at higher copy numbers overall. While the simple fast and slow growth conditions both exhibit straightforward dynamics – gene dosage decreases with decreasing growth rate and approaches 1 at stationary phase – the ratio between gene copy numbers (both during and after exponential growth) varies substantially as a function of genomic integration locations. Moreover, during multiphase growth, which is observed during most typical characterization experiments, gene dosage (and relative gene dosage) varies substantially between the two time frames corresponding to distinct rates of exponential growth (0-200 min and 200-400 min). Finally, the start-stop growth conditions result in dramatic spikes in gene dosage as simulated cultures exited intermediate “pauses” to resume rapid growth.

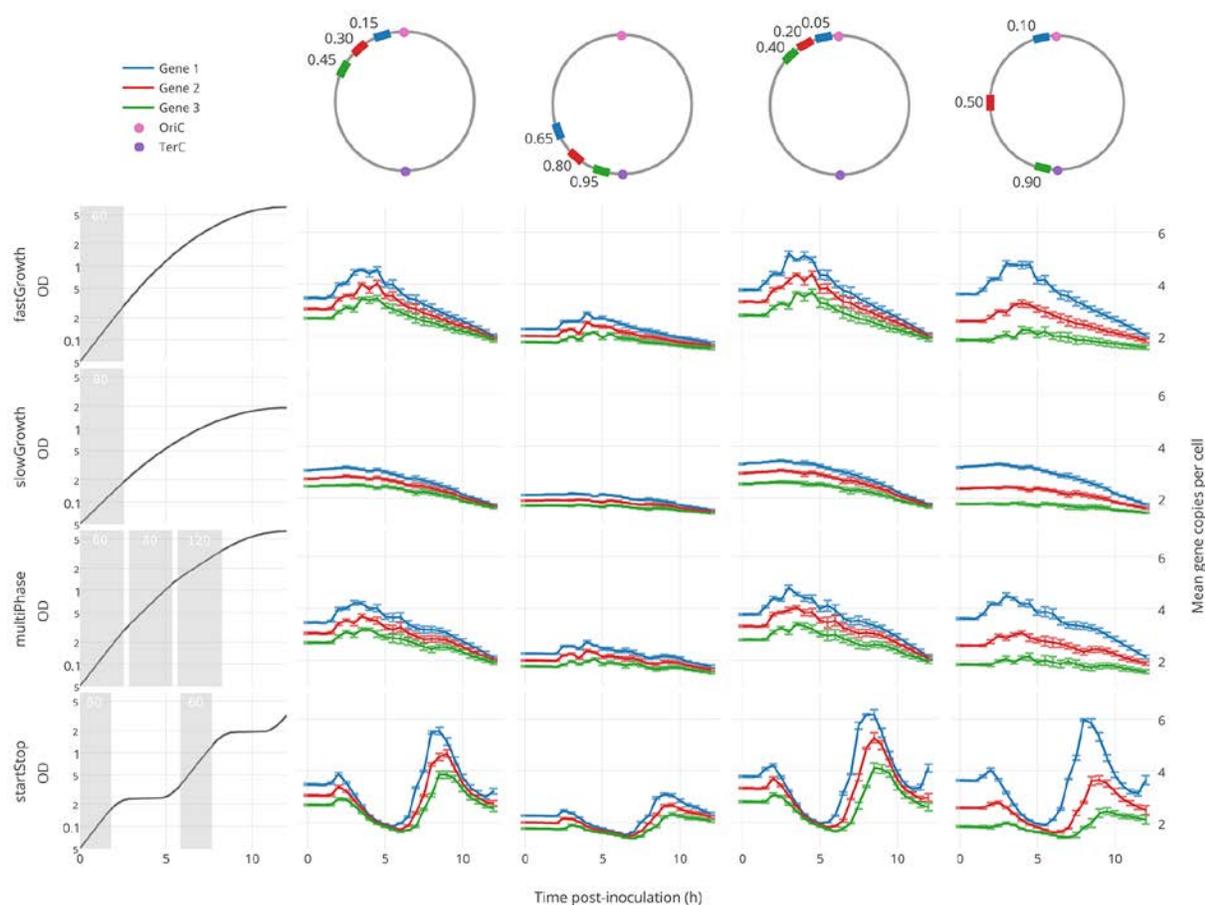


Figure 4.5: HMG simulator-based prediction of gene dosage effects. The updated and optimized HMG simulator was used as a testbed to predict gene dosage dynamics over a range of hypothetical growth curves (left column). Here we track three genomic loci (blue, red, and green rectangles), located at various positions relative to *oriC* (pink circle). Numbers accompanying the chromosome maps in the top row indicate the relative distance of each locus from *oriC*, in each scenario, on a scale where 1.0 is completely distal (e.g., the primary *Ter* site, teal rectangle). Each predicted trajectory represents the mean copy number of each locus per cell, averaged over 100 independent simulations, with error bars representing one standard deviation. For each hypothetical growth curve (left column), each shaded area is labeled with the doubling rate calculated for that window

of growth. Each simulation was inoculated (initiated) under conditions of exponential growth, using the doubling rate calculated for the first indicated period of exponential growth (gray shading), and thereafter simulations proceeded using the injection method through the remainder of the growth curves.

Each of these examples suggests strategies by which a synthetic biology practitioner might make use of the HMG simulator. As a hypothesis generation tool, the HMG simulator may be used to evaluate whether gene dosage may plausibly explain why an engineered function or pathway behaves differently under different growth regimes. Conversely, the HMG simulator may be used to design experiments to probe how gene dosage impacts the performance of a particular function. Ultimately, the HMG simulator may be paired with other design tools to facilitate the design of novel functions that operate in a desirable fashion over a range of growth conditions.

4.4 Summary and future directions

This investigation developed and validated a novel modeling approach to meet the need for synthetic biologists to be able to predict and evaluate the impact of genomic (and gene locus) copy number variation across a population of bacterial cells and across variable growth regimes. The results reported here have validated the overall HMG simulation strategy, and the injection growth model upon which it is based. Although we observed that a single calibrated version of the HMG simulator could make accurate predictions across a range of conditions, some conditions resulted in lower prediction accuracy for reasons that are not yet clear. Thus, further investigations are required to determine the extent to which any set of model calibrations enables predictions across genotypes, growth conditions, and other perturbations such as plasmid and transgene load. Ultimately, we expect that the HMG simulator will provide a powerful tool for the evaluation and design of synthetic microbial functions that perform robustly across a range of growth conditions (all software is freely available from the authors upon request). Another fundamental biological question that remains open and would be of interest is this – do there exist interactions between genomic distributions and the distributions of plasmids within a population? Variation exists

between cells in the maintenance of plasmids, and loss can occur in a subset of cells within a population (48). To add extra complexity, some plasmids and regulons thereon have been shown to share regulation with that of the genome and may be expected to track copy numbers accordingly, while other plasmids have shown to be regulated independent of the genomic replication process (148, 149). Expanding our experimental investigations and modeling efforts to integrate interactions between genomic replication, cell growth, and plasmid maintenance would increase the utility of our HMG framework, which may ultimately be extended to new challenges such as modeling the spread of conjugative elements within a population, expanding on the work presented in Chapter 3.

CHAPTER 5

Conclusions and Recommendations

5.1 Chapter 2: Engineered spatial sequestration

The goal of our study was to determine the feasibility of engineering a spatially sequestered transcription factor to the inner membrane of bacteria, only to be released from sequestration by protease activity, where transcription of said protease is induced by a small molecule. We proved that, in fact, we could use spatial sequestration in an engineered system, and that this system could be used for both activation and repression of genes within the cell.

The limits of PASS as a tool are only quantified at this point by the parts that were used in this study. PASS architecture could effectively sequester tetR and λ CI. From our observations, λ CI in particular does not induce strong activation of its associated promoter, and whether PASS architecture could accommodate sequestration with a stronger regulator/promoter pairing needs to be further explored. LacI proved to be still functional when sequestered to the inner membrane with M13 coat protein (23), and attaching LacI to PASS may shed light onto why that pairing failed to sufficiently inhibit the regulator while sequestered; is the regulator or the means of sequestration the reason for a lack of sequestration of activity?

There are a number of means that could be used to improve the function of PASS as a tool. As discussed earlier, noise is associated with expression from plasmids, particularly high copy number plasmids (3, 150). In our system, we used high copy number plasmids for encoding our GFP reporter. This may account for the leakiness of expression. Moving the genetic circuits to low copy number plasmids or onto the genome might improve signal to noise and make the system a more robust tool. However, changing the plasmid on which the reporter gene is expressed may

have an effect on the function of sequestration. Plasmids are known to cluster localize in specific locations within the cell (87), and it may be that PASS functioned as hypothesized based as much on the localization of the reporter as of the regulator.

In terms of understanding of how native sequestration systems work, of the natural systems described in chapter 2, our engineered system most closely approximates that of the σ_E . Both systems rely on an irreversible proteolytic degradation of the sequestration mechanism. In our system, protease activity is supplied by a cytosolic protease that is immediately active upon expression. With the degradation of RseA, protease activity is activated conditionally, occurring upon the accumulation of misfolded porins, which usually occurs when cells are under stress (5, 15). If our system were to be used to understand the dynamics of this native system or as a tool to replicate behavior of this system, the protease activity would have to be altered to function upon some cue, but fully transcribed when activated. One means of doing this would be to use a split protease that could only be assembled upon binding of both parts to a small molecule. Indeed, the protease used in this study from tobacco etch virus has been shown to be split and reconstituted when bound to a membrane with the antibiotic rapamycin (89).

As a tool, PASS has the potential to be useful in a number of conditions. In metabolic engineering, it is often useful to rapidly change regulation the metabolome at different points in growth to maximize growth at some points in an experiment and production of a desired product in another (151, 152). Spatial sequestration has been shown to function in natural systems for conditional activity of regulators of multiple genes such as Mlc and σ_E , changing the metabolome of the cell. Combining PASS with global regulators may allow for a new tool for shifting cellular metabolomes upon induction with a small molecule. Our characterizations also focused on using a repressor and an activator individually, but there is no reason that these two functions could not be carried out simultaneously. Combining these two systems or similarly engineered systems could allow for the activation of genes and repression of others simultaneously.

5.2 Chapter 3: Regulation of F plasmid conjugation

The finding of key importance to the study described is that TraJ, thought to be the key activator of expression from P_Y , is only capable of inducing expression from P_Y in a subset of cells. The reason for this is unclear, but the expression of TraJ does not correlate with expression from P_Y , meaning that the heterogeneity seen between cells in the expression of conjugative machinery is not due to expression level of TraJ. There are multiple other known activators of conjugation, including ArcA and feedback loops within the F plasmid (32, 46). There are also known to be a large number of competitors of TraJ enforcing suppression of P_Y (45). Interactions between all of these components may create cell states that are amenable to induction from P_Y in a subset of cells, but render other cells incapable of producing the proper transfer machinery.

There are a number of reasons that a cell would actively work to repress conjugation. The conjugative machinery requires massive energy commitments from the cell for expression and for proper continued function (26, 95). Because of this interplay between the needs of the cell and the programming to execute conjugation by the F plasmid, it becomes clear why a simple circuit of regulation where expression of TraJ results in conjugation would be unwise for not only the cells but also the F plasmid itself. If all conjugative donor cells were activated, the energy commitment would put these cells at a disadvantage that may result in extinction of the population of donors. Balancing conjugation with simple F plasmid maintenance is beneficial to both parties, and more complex regulation could come from either the host or other components of the F encoded machinery.

Our final experiment exploring the feasibility of an orthogonal control system for conjugation gives many clues as to the biology of conjugation. The fact that our system with pJaM

showed a transfer rate that was low comparative to F driven conjugation but reliably above background allows us to start at a functioning framework in trying to engineer a useful orthogonal conjugative tool. Induction was strongest with pJaM uninduced. This suggests that low expression of TraM or TraJ (or TraM and TraJ) may yield more meaningful upregulation of the conjugative machinery and conjugative rates themselves than strong expression from fully induced systems. As a means to improve this system in the future, TraM and TraJ may have to be decoupled to understand how expression of each individually affects the rate of conjugation. Based on this study, it may be reasonable to guess that the level of TraJ induction is not a strong determinant on whether cells express the machinery controlling conjugation. TraM may be having more of an impact on the rate of conjugation in this circumstance TraM is required in conjugation to complex with both the *OriT* DNA sequence and with the secretion system (41). Low levels of TraM might be required for proper coordination between its various partners and roles within the conjugative machinery.

The use of conjugation not only to understand native phenomena like the spread of antibiotic resistance or virulence factors but as a tool for lateral gene transfer has created a strong interest in using conjugation to transfer genetic elements from easily modified bacterial systems to more complex organisms (37, 40, 42). In order to improve efficiency of any engineered system to different organisms, more than regulation of conjugation needs to be modified. In particular, TraA, the protein that composes the pilus, could be modified to attach to membrane proteins of other organism (153). However before our system could be applied as a tool for inter-kingdom genetic transfer, understanding of the regulatory principals at play would have to be improved. Therefore, if the system described in this study were used as a tool, multiple parts would have to be engineered concurrently to design a feasible system for species to species transfer.

Much of the understanding of the interaction partners of TraJ in regulating P_{γ} have been discovered in strains with mutations of those partners. TraJ is not required for transfer in cells with

mutant H-NS (104). Conjugation cannot occur in cells without ArcA independent of TraJ expression (46). Screening the function of our engineered system in cells with mutations of other regulators may give a clearer picture of what aspects of regulation are limiting conjugation in our engineered system.

As a tool, our pOX38 and pShuttle pairing has strong potential to help elucidate the function of different components of the conjugative machinery. Traditional assays for mating efficiency relied on colony counting assays that had low sampling outputs and large time constraints. Using our conjugative reporter system would allow for rapid quantification of changes in conjugation rates in different mutational backgrounds or with different orthogonal regulators added. It may be a useful screen for testing more variants rapidly in determining how changes in relevant components of the conjugative machinery or regulators affect the rate of conjugation.

5.3 Chapter 4: Genomic heterogeneity simulator

For the work described in this paper, we constructed multiple distinct sets of data for the use in training, calibrating, and validating an agent based model as part of a collaboration with Andrew Younger, Declan Bates and Melchior Du Lac. With this agent based modeling based simulator, researchers can predict of genomic distributions in a population within the full runtime of an experiment. By simply supplying the OD₆₀₀ data collected from a run over time, information about the distributions of genomes throughout their experiment can be acquired and analyzed in order to better understand gene count dynamics throughout different phases of growth. As a tool the simulation should better help with design of gene circuits requiring the coordination between multiple specific gene loci. Ultimately, such understanding could allow for better coordination and control of engineered functions across variable growth conditions.

In future work, the HMG simulator can hopefully be further advanced in order to handle interactions between genetic elements outside of the bacterial genome. Incorporating these elements in future iterations of the HMG simulator may be relatively straightforward to implement within the mechanistic, agent-based modelling framework we developed. The HMG simulator will hopefully be able to incorporate different plasmids and their maintenance within a diverse population. This would be relevant to further research since plasmid maintenance has a lot of interesting biological features. While genomes cannot be lost (Cells need their genome), many cells within a population lose their plasmids, especially since plasmids often have an energy cost associated with their maintenance (150). Plasmids can cost competing cells valuable resources and slow growth, meaning the population dynamics and resource scarcity may lead to differing levels of plasmid loss. Depending on the replicon dictating plasmid maintenance, some plasmids sync their replication with the replication of the genome while others are replicating independently (148, 154). Simply adding a single, low copy plasmid to cells in a population would add large amounts of complexity. However, as seen routinely in this document, many engineered systems require multiple plasmids, and coordination between parts on each play a role in determining the function of the larger system. Adding the capability of predicting the dynamics between plasmids would be a useful tool and is a future goal of the collaboration between our labs.

Another goal for future iterations of our model is adding interactions between cells. As discussed at length in Chapter 3, horizontal gene transfer including conjugation has many important biological implications. Modeling how genetic information travels within a population would aide in studying relevant processes such as how antibiotic resistance spreads in clinical setting. With the potential to use conjugation machinery as a tool for transfer, having a model able to predict the distribution of conjugative plasmids within individual cells as well as their likelihood of transfer and distribution within a population. This tool would be useful in exploring what aspects of conjugating would need to be engineered for the optimization of transfer events.

References

1. Pruitt KM, DeMuth RE, & Turner ME, Jr. (1979) Practical application of generic growth theory and the significance of the growth curve parameters. *Growth* 43(1):19-35.
2. Casadesus J & Low DA (2013) Programmed heterogeneity: epigenetic mechanisms in bacteria. *J Biol Chem* 288(20):13929-13935.
3. Elowitz MB, Levine AJ, Siggia ED, & Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297(5584):1183-1186.
4. Bohm A & Boos W (2004) Gene regulation in prokaryotes by subcellular relocalization of transcription factors. *Curr Opin Microbiol* 7(2):151-156.
5. Rowley G, Spector M, Kormanec J, & Roberts M (2006) Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. *Nat Rev Microbiol* 4(5):383-394.
6. Nam TW, *et al.* (2001) The Escherichia coli glucose transporter enzyme IICB(Glc) recruits the global repressor Mlc. *EMBO J* 20(3):491-498.
7. Gorke B (2012) Killing two birds with one stone: an ABC transporter regulates gene expression through sequestration of a transcriptional regulator at the membrane. *Mol Microbiol* 85(4):597-601.
8. Ehrmann M & Clausen T (2004) Proteolysis as a regulatory mechanism. *Annu Rev Genet* 38:709-724.
9. Chen J (2013) Molecular mechanism of the Escherichia coli maltose transporter. *Curr Opin Struct Biol* 23(4):492-498.
10. Dippel R, Bergmiller T, Bohm A, & Boos W (2005) The maltodextrin system of Escherichia coli: glycogen-derived endogenous induction and osmoregulation. *J Bacteriol* 187(24):8332-8339.

11. Richet E, Davidson AL, & Joly N (2012) The ABC transporter MalFGK(2) sequesters the MalT transcription factor at the membrane in the absence of cognate substrate. *Mol Microbiol* 85(4):632-647.
12. Boos W & Shuman H (1998) Maltose/maltodextrin system of Escherichia coli: transport, metabolism, and regulation. *Microbiol Mol Biol Rev* 62(1):204-229.
13. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* 100(4):391-398.
14. Grigorova IL, *et al.* (2004) Fine-tuning of the Escherichia coli sigmaE envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the transmembrane anti-sigma factor, RseA. *Genes Dev* 18(21):2686-2697.
15. Dartigalongue C, Missiakas D, & Raina S (2001) Characterization of the Escherichia coli sigma E regulon. *J Biol Chem* 276(24):20866-20875.
16. Chaba R, *et al.* (2011) Signal integration by DegS and RseB governs the E-mediated envelope stress response in Escherichia coli. *Proc Natl Acad Sci U S A* 108(5):2106-2111.
17. Kanehara K, Ito K, & Akiyama Y (2003) YaeL proteolysis of RseA is controlled by the PDZ domain of YaeL and a Gln-rich region of RseA. *EMBO J* 22(23):6389-6398.
18. Walsh NP, Alba BM, Bose B, Gross CA, & Sauer RT (2003) OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* 113(1):61-71.
19. Plumbridge J (1998) Expression of ptsG, the gene for the major glucose PTS transporter in Escherichia coli, is repressed by Mlc and induced by growth on glucose. *Mol Microbiol* 29(4):1053-1063.

20. Lee SJ, Boos W, Bouche JP, & Plumbridge J (2000) Signal transduction between a membrane-bound transporter, PtsG, and a soluble transcription factor, Mlc, of *Escherichia coli*. *EMBO J* 19(20):5353-5361.
21. Kimata K, Inada T, Tagami H, & Aiba H (1998) A global repressor (Mlc) is involved in glucose induction of the ptsG gene encoding major glucose transporter in *Escherichia coli*. *Mol Microbiol* 29(6):1509-1519.
22. Tanaka Y, Itoh F, Kimata K, & Aiba H (2004) Membrane localization itself but not binding to IICB is directly responsible for the inactivation of the global repressor Mlc in *Escherichia coli*. *Mol Microbiol* 53(3):941-951.
23. Gorke B, Reinhardt J, & Rak B (2005) Activity of Lac repressor anchored to the *Escherichia coli* inner membrane. *Nucleic Acids Res* 33(8):2504-2511.
24. Pitner RA, Scarpelli AH, & Leonard JN (2015) Regulation of Bacterial Gene Expression by Protease-Alleviated Spatial Sequestration (PASS). *ACS Synth Biol*.
25. Ochman H, Lawrence JG, & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299-304.
26. Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EP, & de la Cruz F (2010) Mobility of plasmids. *Microbiol Mol Biol Rev* 74(3):434-452.
27. Arutyunov D & Frost LS (2013) F conjugation: back to the beginning. *Plasmid* 70(1):18-32.
28. Schroder G & Lanka E (2005) The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* 54(1):1-25.
29. Juhas M, Crook DW, & Hood DW (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol* 10(12):2377-2386.

30. Campbell EA, *et al.* (2003) Crystal structure of Escherichia coli sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol Cell* 11(4):1067-1078.
31. Alekshun MN & Levy SB (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell* 128(6):1037-1050.
32. Frost LS & Koraimann G (2010) Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol* 5(7):1057-1071.
33. Christie PJ, Whitaker N, & Gonzalez-Rivera C (2014) Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta* 1843(8):1578-1591.
34. Narra HP & Ochman H (2006) Of what use is sex to bacteria? *Curr Biol* 16(17):R705-710.
35. Sanderson KE (1976) Genetic relatedness in the family Enterobacteriaceae. *Annu Rev Microbiol* 30:327-349.
36. Bellanger X, Payot S, Leblond-Bourget N, & Guedon G (2014) Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev* 38(4):720-760.
37. Karas BJ, *et al.* (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* 6:6925.
38. Stachel SE, Timmerman B, & Zambryski P (1987) Activation of Agrobacterium tumefaciens vir gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region: requirement for 5' virD gene products. *EMBO J* 6(4):857-863.
39. Heinemann JA & Sprague GF, Jr. (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* 340(6230):205-209.
40. Kunik T, *et al.* (2001) Genetic transformation of HeLa cells by Agrobacterium. *Proc Natl Acad Sci U S A* 98(4):1871-1876.

41. Yoon YG & Koob MD (2005) Transformation of isolated mammalian mitochondria by bacterial conjugation. *Nucleic Acids Res* 33(16):e139.
42. Yoon YG, Koob MD, & Yoo YH (2010) Re-engineering the mitochondrial genomes in mammalian cells. *Anat Cell Biol* 43(2):97-109.
43. Tatum EL & Lederberg J (1947) Gene Recombination in the Bacterium *Escherichia coli*. *J Bacteriol* 53(6):673-684.
44. Boto L (2010) Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* 277(1683):819-827.
45. Will WR & Frost LS (2006) Characterization of the opposing roles of H-NS and TraJ in transcriptional regulation of the F-plasmid *tra* operon. *J Bacteriol* 188(2):507-514.
46. Strohmaier H, *et al.* (1998) Signal transduction and bacterial conjugation: characterization of the role of ArcA in regulating conjugative transfer of the resistance plasmid R1. *J Mol Biol* 277(2):309-316.
47. Arutyunov D, Rodriguez-Maillard JM, & Frost LS (2011) A PAS domain within F plasmid TraJ is critical for its function as a transcriptional activator. *Biochem Cell Biol* 89(4):396-404.
48. Boyd D, Weiss DS, Chen JC, & Beckwith J (2000) Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. *J Bacteriol* 182(3):842-847.
49. Akerlund T, Nordstrom K, & Bernander R (1995) Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of *Escherichia coli*. *J Bacteriol* 177(23):6791-6797.
50. Hill NS, Kadoya R, Chatteraj DK, & Levin PA (2012) Cell size and the initiation of DNA replication in bacteria. *PLoS Genet* 8(3):e1002549.

51. Jin DJ, Cagliero C, & Zhou YN (2012) Growth rate regulation in Escherichia coli. *FEMS Microbiol Rev* 36(2):269-287.
52. Santos CN & Yoshikuni Y (2014) Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. *Nat Protoc* 9(6):1320-1336.
53. Michelsen O, Teixeira de Mattos MJ, Jensen PR, & Hansen FG (2003) Precise determinations of C and D periods by flow cytometry in Escherichia coli K-12 and B/r. *Microbiology* 149(Pt 4):1001-1010.
54. Skarstad K & Katayama T (2013) Regulating DNA replication in bacteria. *Cold Spring Harb Perspect Biol* 5(4):a012922.
55. Messer W (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* 26(4):355-374.
56. Atlung T & Hansen FG (1999) Low-temperature-induced DnaA protein synthesis does not change initiation mass in Escherichia coli K-12. *J Bacteriol* 181(18):5557-5562.
57. Kurokawa K, Nishida S, Emoto A, Sekimizu K, & Katayama T (1999) Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in Escherichia coli. *EMBO J* 18(23):6642-6652.
58. Kamada K, Horiuchi T, Ohsumi K, Shimamoto N, & Morikawa K (1996) Structure of a replication-terminator protein complexed with DNA. *Nature* 383(6601):598-603.
59. Mulcair MD, *et al.* (2006) A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell* 125(7):1309-1319.
60. Osella M, Nugent E, & Cosentino Lagomarsino M (2014) Concerted control of Escherichia coli cell division. *Proc Natl Acad Sci U S A* 111(9):3431-3435.
61. Buddelmeijer N & Beckwith J (2002) Assembly of cell division proteins at the E. coli cell center. *Curr Opin Microbiol* 5(6):553-557.

62. Natale P, Pazos M, & Vicente M (2013) The Escherichia coli divisome: born to divide. *Environ Microbiol* 15(12):3169-3182.
63. Stouf M, Meile JC, & Cornet F (2013) FtsK actively segregates sister chromosomes in Escherichia coli. *Proc Natl Acad Sci U S A* 110(27):11157-11162.
64. Robert L (2015) Size sensors in bacteria, cell cycle control, and size control. *Front Microbiol* 6:515.
65. Cooper S & Helmstetter CE (1968) Chromosome replication and the division cycle of Escherichia coli B/r. *J Mol Biol* 31(3):519-540.
66. Helmstetter C, Cooper S, Pierucci O, & Revelas E (1968) On the bacterial life sequence. *Cold Spring Harb Symp Quant Biol* 33:809-822.
67. Tyson JJ (1989) Effects of asymmetric division on a stochastic model of the cell division cycle. *Math Biosci* 96(2):165-184.
68. Skarstad K, Steen HB, & Boye E (1985) Escherichia-Coli DNA Distributions Measured by Flow-Cytometry and Compared with Theoretical Computer-Simulations. *Journal of Bacteriology* 163(2):661-668.
69. Keasling JD, Kuo H, & Vahanian G (1995) A Monte Carlo simulation of the Escherichia coli cell cycle. *J Theor Biol* 176(3):411-430.
70. Grant MA, *et al.* (2011) DnaA and the timing of chromosome replication in Escherichia coli as a function of growth rate. *BMC Syst Biol* 5:201.
71. Stokke C, Flatten I, & Skarstad K (2012) An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS One* 7(2):e30981.
72. du Lac M, Scarpelli AH, Younger AK, Bates DG, & Leonard JN (2016) Predicting the Dynamics and Heterogeneity of Genomic DNA Content within Bacterial Populations across Variable Growth Regimes. *ACS Synth Biol*.

73. Bordignon E, Grote M, & Schneider E (2010) The maltose ATP-binding cassette transporter in the 21st century--towards a structural dynamic perspective on its mode of action. *Mol Microbiol* 77(6):1354-1366.
74. De Las Penas A, Connolly L, & Gross CA (1997) The sigmaE-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol Microbiol* 24(2):373-385.
75. Kanehara K, Ito K, & Akiyama Y (2002) YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. *Genes Dev* 16(16):2147-2155.
76. Ueguchi C & Ito K (1990) Escherichia coli sec mutants accumulate a processed immature form of maltose-binding protein (MBP), a late-phase intermediate in MBP export. *J Bacteriol* 172(10):5643-5649.
77. Shaner NC, *et al.* (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* 22(12):1567-1572.
78. Dmitriev O, Jones PC, Jiang W, & Fillingame RH (1999) Structure of the membrane domain of subunit b of the Escherichia coli F0F1 ATP synthase. *J Biol Chem* 274(22):15598-15604.
79. Dougherty WG, Parks TD, Cary SM, Bazan JF, & Fletterick RJ (1989) Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* 172(1):302-310.
80. Kapust RB, Tozser J, Copeland TD, & Waugh DS (2002) The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* 294(5):949-955.
81. Kapust RB & Waugh DS (2000) Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr Purif* 19(2):312-318.

82. Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25(6):1203-1210.
83. Michalowski CB & Little JW (2013) Role of cis-acting sites in stimulation of the phage lambda P(RM) promoter by CI-mediated looping. *J Bacteriol* 195(15):3401-3411.
84. Huang D, Holtz WJ, & Maharbiz MM (2012) A genetic bistable switch utilizing nonlinear protein degradation. *J Biol Eng* 6(1):9.
85. Glascock CB & Weickert MJ (1998) Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene* 223(1-2):221-231.
86. Dinh T & Bernhardt TG (2011) Using superfolder green fluorescent protein for periplasmic protein localization studies. *J Bacteriol* 193(18):4984-4987.
87. Pogliano J, Ho TQ, Zhong Z, & Helinski DR (2001) Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc Natl Acad Sci U S A* 98(8):4486-4491.
88. Javer A, *et al.* (2013) Short-time movement of *E. coli* chromosomal loci depends on coordinate and subcellular localization. *Nat Commun* 4:3003.
89. Wehr MC, *et al.* (2006) Monitoring regulated protein-protein interactions using split TEV. *Nat Methods* 3(12):985-993.
90. Williams DJ, Puhl HL, 3rd, & Ikeda SR (2009) Rapid modification of proteins using a rapamycin-inducible tobacco etch virus protease system. *PLoS One* 4(10):e7474.
91. Ghigo JM (2001) Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412(6845):442-445.
92. Ogura T & Hiraga S (1983) Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc Natl Acad Sci U S A* 80(15):4784-4788.
93. Bahl MI, Hansen LH, & Sorensen SJ (2009) Persistence mechanisms of conjugative plasmids. *Methods Mol Biol* 532:73-102.

94. Filutowicz M, *et al.* (2008) Bacterial conjugation-based antimicrobial agents. *Plasmid* 60(1):38-44.
95. Mark Glover JN, *et al.* (2015) The FinO family of bacterial RNA chaperones. *Plasmid* 78:79-87.
96. Gubbins MJ, Will WR, & Frost LS (2005) *The F-plasmid, a paradigm for bacterial conjugation* (Cambridge University Press, Cambridge, UK).
97. Frost LS, Ippen-Ihler K, & Skurray RA (1994) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* 58(2):162-210.
98. Rashtchian A, Crooks JH, & Levy SB (1983) traJ independence in expression of traT on F. *J Bacteriol* 154(2):1009-1012.
99. Ragonese H, Haisch D, Villareal E, Choi JH, & Matson SW (2007) The F plasmid-encoded TraM protein stimulates relaxosome-mediated cleavage at oriT through an interaction with TraI. *Mol Microbiol* 63(4):1173-1184.
100. Lawley TD, Klimke WA, Gubbins MJ, & Frost LS (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* 224(1):1-15.
101. Clarke M, Maddera L, Harris RL, & Silverman PM (2008) F-pili dynamics by live-cell imaging. *Proc Natl Acad Sci U S A* 105(46):17978-17981.
102. Grace ED, *et al.* (2015) Activation of the sigmaE-dependent stress pathway by conjugative TraR may anticipate conjugational stress. *J Bacteriol* 197(5):924-931.
103. Lu J, *et al.* (2008) Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol Microbiol* 70(1):89-99.
104. Will WR, Lu J, & Frost LS (2004) The role of H-NS in silencing F transfer gene expression during entry into stationary phase. *Mol Microbiol* 54(3):769-782.

105. Zahrl D, Wagner A, Tscherner M, & Koraimann G (2007) GroEL plays a central role in stress-induced negative regulation of bacterial conjugation by promoting proteolytic degradation of the activator protein TraJ. *J Bacteriol* 189(16):5885-5894.
106. Lau-Wong IC, Locke T, Ellison MJ, Raivio TL, & Frost LS (2008) Activation of the Cpx regulon destabilizes the F plasmid transfer activator, TraJ, via the HslVU protease in *Escherichia coli*. *Mol Microbiol* 67(3):516-527.
107. Starcic M, *et al.* (2003) The cyclic AMP-cyclic AMP receptor protein complex regulates activity of the traJ promoter of the *Escherichia coli* conjugative plasmid pRK100. *J Bacteriol* 185(5):1616-1623.
108. Camacho EM, Serna A, & Casadesus J (2005) Regulation of conjugal transfer by Lrp and Dam methylation in plasmid R100. *Int Microbiol* 8(4):279-285.
109. Will WR & Frost LS (2006) Hfq is a regulator of F-plasmid TraJ and TraM synthesis in *Escherichia coli*. *J Bacteriol* 188(1):124-131.
110. Wong JJ, Lu J, & Glover JN (2012) Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Mol Microbiol* 85(4):602-617.
111. Willetts N & Maule J (1986) Specificities of IncF plasmid conjugation genes. *Genet Res* 47(1):1-11.
112. Yoshioka Y, Ohtsubo H, & Ohtsubo E (1987) Repressor gene finO in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS3 into F finO. *J Bacteriol* 169(2):619-623.
113. Novotny CP & Fives-Taylor P (1974) Retraction of F pili. *J Bacteriol* 117(3):1306-1311.
114. Jacobson A (1972) Role of F pili in the penetration of bacteriophage fl. *J Virol* 10(4):835-843.

115. Wan Z, Varshavsky J, Teegala S, McLawrence J, & Goddard NL (2011) Measuring the rate of conjugal plasmid transfer in a bacterial population using quantitative PCR. *Biophys J* 101(1):237-244.
116. Wong JJW, Lu J, & Glover JNM (2012) Relaxosome function and conjugation regulation in F-like plasmids – a structural biology perspective. *Molecular Microbiology* 85(4):602-617.
117. Veitia RA (2003) A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biol Rev Camb Philos Soc* 78(1):149-170.
118. Louis M & Becskei A (2002) Binary and graded responses in gene networks. *Sci STKE* 2002(143):pe33.
119. Lu J, Fekete RA, & Frost LS (2003) A rapid screen for functional mutants of TraM, an autoregulatory protein required for F conjugation. *Mol Genet Genomics* 269(2):227-233.
120. Silverman PM, Wickersham E, & Harris R (1991) Regulation of the F plasmid traY promoter in Escherichia coli by host and plasmid factors. *J Mol Biol* 218(1):119-128.
121. Ferrell JE, Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14(2):140-148.
122. Cho C, Choi SY, Luo ZW, & Lee SY (2015) Recent advances in microbial production of fuels and chemicals using tools and strategies of systems metabolic engineering. *Biotechnol Adv* 33(7):1455-1466.
123. Santos CN, Regitsky DD, & Yoshikuni Y (2013) Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat Commun* 4:2503.
124. Moser F, *et al.* (2012) Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synth Biol* 1(11):555-564.

125. Wang HH, *et al.* (2012) Genome-scale promoter engineering by coselection MAGE. *Nat Methods* 9(6):591-593.
126. Tyo KE, Ajikumar PK, & Stephanopoulos G (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat Biotechnol* 27(8):760-765.
127. Gonze D (2013) Modeling the effect of cell division on genetic oscillators. *J Theor Biol* 325:22-33.
128. Klumpp S, Zhang Z, & Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139(7):1366-1375.
129. Skarstad K, Boye E, & Steen HB (1986) Timing of initiation of chromosome replication in individual *Escherichia coli* cells. *EMBO J* 5(7):1711-1717.
130. Osella M & Lagomarsino MC (2013) Growth-rate-dependent dynamics of a bacterial genetic oscillator. *Phys Rev E Stat Nonlin Soft Matter Phys* 87(1):012726.
131. Bremer H & Dennis PP (2008) Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* 3(1).
132. Abner K, Aaviksaar T, Adamberg K, & Vilu R (2014) Single-cell model of prokaryotic cell cycle. *J Theor Biol* 341:78-87.
133. Powell EO (1958) An outline of the pattern of bacterial generation times. *J Gen Microbiol* 18(2):382-417.
134. Lopez S, Prieto M, Dijkstra J, Dhanoa MS, & France J (2004) Statistical evaluation of mathematical models for microbial growth. *Int J Food Microbiol* 96(3):289-300.
135. Ferenci T (1999) 'Growth of bacterial cultures' 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. *Res Microbiol* 150(7):431-438.
136. Bren A, Hart Y, Dekel E, Koster D, & Alon U (2013) The last generation of bacterial growth in limiting nutrient. *BMC Syst Biol* 7:27.

137. Monod J (1949) Adaptation, mutation and segregation in relation to the synthesis of enzymes by bacteria. *Biochem J* 44(3):xix.
138. Zwietering MH, Jongenburger I, Rombouts FM, & van 't Riet K (1990) Modeling of the bacterial growth curve. *Appl Environ Microbiol* 56(6):1875-1881.
139. Buchanan RL WRDW (1999) When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves,. *Food Microbiol* (14):313-326.
140. Helmstetter CE (1968) DNA synthesis during the division cycle of rapidly growing *Escherichia coli* B/r. *J Mol Biol* 31(3):507-518.
141. Bryant FR (1988) Construction of a recombinase-deficient mutant recA protein that retains single-stranded DNA-dependent ATPase activity. *J Biol Chem* 263(18):8716-8723.
142. Skarstad K & Boye E (1993) Degradation of individual chromosomes in recA mutants of *Escherichia coli*. *J Bacteriol* 175(17):5505-5509.
143. Khan SR & Kuzminov A (2012) Replication forks stalled at ultraviolet lesions are rescued via RecA and RuvABC protein-catalyzed disintegration in *Escherichia coli*. *J Biol Chem* 287(9):6250-6265.
144. Kuzminov A & Stahl FW (1997) Stability of linear DNA in recA mutant *Escherichia coli* cells reflects ongoing chromosomal DNA degradation. *J Bacteriol* 179(3):880-888.
145. Kouzminova EA, Rotman E, Macomber L, Zhang J, & Kuzminov A (2004) RecA-dependent mutants in *Escherichia coli* reveal strategies to avoid chromosomal fragmentation. *Proc Natl Acad Sci U S A* 101(46):16262-16267.
146. Courcelle J & Hanawalt PC (2003) RecA-dependent recovery of arrested DNA replication forks. *Annu Rev Genet* 37:611-646.

147. Horii ZI, and Suzuki, K. (1968) Degradation of the DNA of Escherichia Coli K12 Rec- (JC1569b) after irradiation with ultraviolet light. *Photochemistry and photobiology* (8): 93-105.
148. Shields MS, Kline BC, & Tam JE (1987) Similarities in control of mini-F plasmid and chromosomal replication in Escherichia coli. *J Bacteriol* 169(7):3375-3378.
149. Keasling JD, Palsson BO, & Cooper S (1991) Cell-cycle-specific F plasmid replication: regulation by cell size control of initiation. *J Bacteriol* 173(8):2673-2680.
150. Paulsson J & Ehrenberg M (2001) Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys* 34(1):1-59.
151. Soma Y, Tsuruno K, Wada M, Yokota A, & Hanai T (2014) Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. *Metab Eng* 23:175-184.
152. Leonard E, Nielsen D, Solomon K, & Prather KJ (2008) Engineering microbes with synthetic biology frameworks. *Trends Biotechnol* 26(12):674-681.
153. Manchak J, Anthony KG, & Frost LS (2002) Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer. *Mol Microbiol* 43(1):195-205.
154. Keasling JD, Palsson BO, & Cooper S (1992) Replication of Mini-F Plasmids during the Bacterial Division Cycle. *Res Microbiol* 143(6):541-548.