Flagellar Biosynthesis in E.coli Is Regulated by a Cascade of Stochastic Pulses

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:37944999

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
The undersigned, appointed by the
Department of Molecular and Cellular Biology
have examined a dissertation entitled
"Flagellar biosynthesis in E.coli is regulated by a cascade of stochastic pulses"
presented by Jeongho Mark Kim,
candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature
Typed name: Prof. Nancy Kleckner

Signature
Typed name: Prof. Karine Gibbs

Signature
Typed name: Prof. Johan Paulsson

Date: January 9, 2017
Flagellar biosynthesis in *E.coli* is regulated by a cascade of stochastic pulses

A dissertation presented
by
Jeongho Mark Kim
to
The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biochemistry

Harvard University
Cambridge, Massachusetts
January 2017
Flagellar biosynthesis in *E. coli* is regulated by a cascade of stochastic pulses

**ABSTRACT**

Genetically identical populations of bacteria can exhibit significant phenotypic heterogeneity even in the absence of environmental variations. In bacteria such as *Bacillus* and *Salmonella*, flagellar motility has been found to be one such phenotype. By contrast, in *Escherichia coli*, flagella are traditionally thought to be synthesized continuously and homogeneously throughout exponential growth. In this system, flagella synthesis is regulated by a three-tiered transcriptional cascade: First, a master regulator FlhDC (Class 1) regulates the expression of genes that encode the flagellar basal body (Class 2). One Class 2 gene encodes the alternative sigma factor FliA which then regulates the transcription of Class 3 genes encoding the filament and chemotaxis machinery.

Here, I examined the transcriptional dynamics of flagellar genes in individual cells by a combination of fluorescent protein readouts and time-lapse microscopy. Unexpectedly, flagellar gene expression within a single cell fluctuated dramatically over time in stochastic pulses. I discovered that this behavior was obscured in many laboratory strains due to a previously described mutation. Genes within the same class pulsed synchronously; however, pulses ranged from more to less frequent in Classes 1, 2 and 3 respectively.

In turn, I investigated how pulses in gene expression were generated and propagated through the flagellar cascade. At both steps of the cascade, only a subset of upstream
fluctuations (or pulses) activated downstream genes. Analyzing the “dose-response”
relationships revealed a tri-phasic response: a “repressed”, an “ultrasensitive”, and a
“proportional” phase. I propose how this behavior might arise from a “molecular titration
model” and propose YdiV and FlgM, post-translational regulators of FlhD and FliA,
respectively, to be critical regulators.

My work also excludes some proposed molecules as regulators of pulsing and suggests
other molecules as potential secondary modulators of basic regulatory cascade. These results
contrast stochastic pulses in *E. coli* with previous examples of flagellar heterogeneity in
*Salmonella*.

In summary, this dissertation describes the discovery of stochastic pulsing in *E. coli*
flagellar synthesis. My work offers mechanistic insight into how these pulses propagate across
the transcriptional cascade and establishes flagellar synthesis in *E. coli* as a new model system
for the study of phenotypic heterogeneity.
# TABLE OF CONTENTS

**ABSTRACT** ................................................................................................................................. iii  
**ACKNOWLEDGEMENTS** ................................................................................................................ vii  
**CHAPTER 1: Introduction** ............................................................................................................. 1  
REFERENCES ........................................................................................................................................ 7  
**CHAPTER 2: Flagellar genes of *E. coli* are regulated in stochastic pulses during exponential growth** ........................................................................................................................................ 13  
INTRODUCTION ............................................................................................................................... 13  
RESULTS ............................................................................................................................................ 14  
DISCUSSION ....................................................................................................................................... 27  
REFERENCES ....................................................................................................................................... 32  
**CHAPTER 3: Generation and propagation of stochastic pulses across the flagellar transcriptional cascade** .................................................................................................................................................... 37  
INTRODUCTION ............................................................................................................................... 37  
RESULTS ............................................................................................................................................ 39  
DISCUSSION ....................................................................................................................................... 51  
REFERENCES ....................................................................................................................................... 54  
**CHAPTER 4: Mutations that affect flagellar gene expression during exponential growth of *E. coli*** .................................................................................................................................................. 57  
INTRODUCTION ............................................................................................................................... 57  
RESULTS ............................................................................................................................................ 59  
DISCUSSION ....................................................................................................................................... 68  
REFERENCES ....................................................................................................................................... 70  
**CHAPTER 5: Conclusions and Prospects** ..................................................................................... 74  
REFERENCES ....................................................................................................................................... 80  
**APPENDIX** .................................................................................................................................... 83  
MATERIALS AND METHODS ............................................................................................................ 83  
REFERENCES ....................................................................................................................................... 101
Alice laughed. “There’s no use trying,” she said: “one can’t believe impossible things.”
"I daresay you haven't had much practice,” said the Queen. "When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast."

-Lewis Carroll
ACKNOWLEDGEMENTS

This work would not have been possible without the extraordinary support and guidance I received from my mentors, friends and colleagues. First and foremost, I would like to thank my advisor, Professor Philippe Cluzel. He always encouraged me to pursue ambitious projects and patiently supported and reassured me when I faced difficult challenges. I am deeply grateful for the extraordinary opportunities he gave me that helped me grow as an independent thinker.

I was incredibly fortunate to be surrounded by wonderful lab mates. Arvind “Rasi” Subramaniam, Jeff Moffitt and Lisa Marshall gave me exceptional crash courses in programming, optics, microscopy and microfluidics that helped me become a more complete scientist. I am especially grateful to Enrique Balleza. My work benefitted not only from the biological tools he developed but also from the numerous hours we spent discussing our projects over shots of espresso. Thank you for being my brother-in-arms for so many years.

I owe a deep gratitude to Professor Johan Paulsson and Professor Richard Losick who graciously shared their microfluidic mother machine and loaned me precious time on their microscope. Their generosity dramatically accelerated my project and allowed me to see the proverbial light at the end of the tunnel. I am also immensely grateful to members of their lab, Matt Cabeen, Tom Norman, Nate Lord, Jon Russell and Laurent Potvin-Trottier. I want to thank Matt in particular for patiently training me and never tiring of my endless list of questions—I wish him the best in his future endeavors.

I had an extraordinary dissertation advisory committee. Professor Johan Paulsson not only generously offered me techniques developed in his lab but also gave me valuable insights into my dataset. Professor Karine Gibbs encouraged me look beyond the often insular world of *E.*
coli which led me to discover fascinating parallels and contrasts. Their constant support, feedback and advice over the years were invaluable to me.

It was a great honor to have Nancy—my very first scientific mentor, role model and inspiration—as my committee chair. As an undergraduate student, it was from her that I first learned to question dogma and to embrace the complexities of biological systems. Ever since—for over a decade—she has been a positive guiding force in my life. I will be eternally grateful for her tireless support, encouragement, advice, and the countless times she helped me through obstacles I faced both in and out of lab. I have the most sincere respect and affection for her and am deeply indebted to her for my growth not only as a scientist but also as a person. I hope that this work will make her proud.

My family gave me endless love and support. My parents and in-laws made sure that I was well fed even when I had to tell them “I can’t, I have to go run an experiment” for so many holidays, birthdays and other celebrations. My sister Jennie was always there to lend me an ear whenever I needed help or advice. My little baby niece Lily distracted me from my concerns with her smile and boundless energy. Thank you, everyone, for your love and care.

Finally, to my wife, Ana—thank you for waiting for me to come home even at completely unreasonable hours and spending long nights with me in the lab just so I wouldn’t be alone. When my work threatened to overwhelm me, you helped me keep a level head and not lose perspective of the big picture. Thank you for all your love, patience and inspiration—and most of all, thank you for believing in me.
CHAPTER 1

INTRODUCTION

Clonal populations of bacteria are traditionally considered to be homogeneous. However, growing evidence suggests that many bacteria, even under identical growth conditions, have subpopulations that execute markedly distinct gene expression programs from the rest of the populations (“phenotypic heterogeneity”) (1-6). A particularly well known example is the spontaneous formation of a small subset of “persisters”—cells that stop growing and become highly resistant to antibiotics—in *Escherichia coli* (1, 2). Another example of phenotypic heterogeneity is natural competence, i.e. the ability of bacteria to be uptake extracellular DNA. In *Bacillus subtilis*, most cells undergo sporulation upon starvation but a small subset of cells become naturally competent (4, 5).

A common theme is that phenotypes such as persistence or competence confer protective or adaptive advantages but potentially at a large cost (for example reduced growth). Thus, it has been hypothesized that phenotypic heterogeneity is a “bet-hedging” strategy that allows cells to prepare a subset of its population for unexpected fluctuations in the environment (6-8). Consistent with this hypothesis, a second common theme is that these heterogeneous phenotypes appear to be generated independently of environmental signals (1, 9). Many studies have demonstrated that the heterogeneity is driven by amplifying the stochastic fluctuations inherent to gene expression (“noise”) (4, 10). This amplification has often been shown to be modulated by gene-regulatory networks via feedback loops (10, 11). However, theoretical and experimental studies also suggest that other mechanisms can help amplify stochastic fluctuations (12, 13). These include mechanisms that rely on molecular cooperativity (12, 14), zero-order kinetics of
enzymes working at saturation (15, 16), or sequestration by titratable inhibitors (17, 18).

Bacterial genetic networks that generate heterogeneous phenotypes are often fascinating “real-life” examples that implement these different mechanisms.

One emerging class of heterogeneously regulated phenotypes is flagellar motility. Although flagella allow cells to forage for nutrients and avoid harmful environments (19, 20), it is a bioenergetically expensive structure (21). Motility can also lead to the dispersal of cells which can, in turn, prevent efficient collective behavior such as the formation of biofilms (22). Finally, flagella can trigger immune responses in many host-organisms which can affect both commensal and pathogenic strains seeking to colonize these environments (23-25). Consequently, some bacteria have been observed to heterogeneously regulate the expression of flagella. One prominent example comes again from *B. subtilis*. During exponential growth, *B. subtilis* cells have been shown to exist in two alternate forms, one motile and expressing flagella and one sessile (26). The sessile state is particularly visually striking because multiple cells are “chained” together. The molecular mechanisms underlying the stochastic decisions between motile and chained states are now well-characterized. A recent study also examined the temporal dynamics underlying this process by monitoring *B. subtilis* cells in microfluidics devices over numerous cell divisions (9). This study revealed that the time spent in the motile state was “memoryless”—i.e. there was a constant probability of switching to the chained state, regardless of how long cells had already been motile—the sessile chained state was more tightly timed—i.e. once initiated, it lasted a stereotypical amount of time.

A second example comes from *Salmonella*, one of the classic model systems for the study of flagellar synthesis. Like *B. subtilis*, during exponential growth, only a subset of *Salmonella* cells were found to express flagella (27, 28). However, the regulatory network governing flagellar
synthesis in *Salmonella* is considerably different from that found in *B. subtilis* (29-31). Thus, the specific molecular players and mechanisms involved in *Salmonella* are completely distinct from *B. subtilis*, although certain common regulatory “themes” such as positive feedback appear to be preserved (32-34). Due to the clinical relevance of *Salmonella*, several studies have investigated how heterogeneous regulation of motility might allow *Salmonella* to reduce or even evade host immune responses (28, 35). However, to date, the temporal dynamics underlying the switch between motile (i.e. flagella-“on”) and non-motile (i.e. flagella-“off”) cells remain unknown.

Alongside *Salmonella*, the gram-negative bacterium *Escherichia coli* has been a classic model system for flagellar biosynthesis. The flagella of these two bacteria are highly similar and many of their flagellar proteins are directly interchangeable (36-38). Moreover, the basic regulatory organization of these genes appears to be the same (39). In these two bacteria, more than 50 genes involved in the synthesis of flagella are expressed from 14 different operons. These operons (or the promoters driving these operons) are categorized into three classes. Class 1 consists of the transcription factor FlhDC (more precisely, FlhD_C2 (40)) which acts as the master regulator for flagella production (41). FlhDC directly activates the expression of Class 2 genes which encode the hook-basal body (i.e. the “base” of the flagellum) as well as the alternative sigma factor FliA (σ28) (42). FliA, in turn, complexes with RNA polymerase (RNAP) to transcribe Class 3 genes encoding flagellin—the main component of the flagellar filament—as well as proteins necessary for powering flagellar rotation and chemotaxis (43).

Despite these similarities to *Salmonella*, however, flagellar synthesis in *E. coli* is believed to occur homogenously throughout exponential growth. This view has traditionally been supported by physiological observations of *E. coli*. Using different staining techniques or electron microscopy, cells during exponential growth were each found to possess 2-3 flagella on average.
(38, 44, 45). Based on this assumption of homogeneity, studies of flagellar gene expression in *E. coli* have largely been carried out using bulk population measurements (20, 46-48).

However, in the last decade, it has been recognized that many laboratory strains of *E. coli*, including strains traditionally used for studies of flagella, harbor an insertion (IS) element in the regulatory region of FlhDC (49). The IS element is found upstream of the RNA polymerase binding site and is thought to disrupt the binding sites for the transcriptional repressors of FlhDC. Consequently expression of flagellar genes in these strains (henceforth abbreviated as “insertion element mutants” or “IS mutants”) is significantly upregulated compared to laboratory strains with wild-type regulation of FlhDC (49-51). Interestingly, the identity of the insertion element varies between many strains (IS1 in some and IS5 in others) and even for strains with the same insertion element, the precise locus of insertion appears to be different (49). These variations suggest that the insertion element has been acquired independently several times during laboratory culture of *E. coli*.

While studies have compared flagellar gene expression in these insertion elements to strains with wild-type FlhDC regulation (henceforth, simply called “wild-type strains”), they have only measured differences at the population level using bulk culture measurements (49, 51). In these studies, the average expression of flagellar genes in wild-type populations was found to be low but not fully repressed. Moreover, bulk population studies of motility found that wild-type strains could (as a population) migrate through soft-agar, again, albeit less effectively than insertion element strains (49, 51, 52). Thus, a common interpretation has been that each wild-type cell simply produces fewer flagella per cell than their insertion element counterparts. However, one potentially puzzling observation is that some flagellar genes are downregulated as much as 30-50 fold in wild-type populations. If “hypermotile” insertion elements strains can
only produce 2-3 flagella per cell, it seems unlikely that wild-type cells would be able to produce any flagella at all.

In this dissertation, I address this discrepancy by studying the regulation of flagellar genes in *E. coli* at the single cell level. I present my novel discovery that during exponential growth, individual *E. coli* cells regulate flagellar genes using a stochastic “pulsing” strategy. As a consequence, wild-type populations consist of cells with strikingly heterogeneous levels of flagellar expression.

In Chapter 2, I first describe my discovery and characterization of this stochastic mode of flagellar regulation that has been previously obscured by the insertion element mutation. I use fluorescent protein fusions to establish that each layer of flagellar regulatory cascade has a distinct pattern of heterogeneity. I further use time-lapse microscopy to demonstrate that heterogeneity at each level of the regulatory cascade arises from stochastic pulses in flagellar gene expression. Strikingly, genes within the same class pulse synchronously but each class has distinct pulsing dynamics, which explains the distinct patterns of heterogeneity in each class.

Finally, I investigate how this pulsing strategy might influence the fitness of *E. coli*.

In Chapter 3, I follow the observations in chapter 2 to try to elucidate (i) how these stochastic pulses are generated from fluctuations in Class 1 gene expression and (ii) how the resulting pulses propagate from Class 2 to Class 3 genes. I demonstrate that, at both transitions, only a subset of upstream fluctuations (or pulses) leads to an activation of downstream genes. Analysis of the “dose-response” relationships between upstream input and downstream output reveals that, at both levels, there is a tri-phasic response: a “repressed” phase, an “ultrasensitive” phase and a “proportional” phase. This type of pattern is well-explained by a “molecular titration model”
(18). In this context, I propose YdiV, a post-translational regulator of FlhDC (53, 54) and FlgM, a post-translational regulator of FliA (55), to be critical regulators of the two transitions.

In Chapter 4, I examine many different molecules known to be involved in regulation of flagellar gene expression. This work excludes some proposed molecules as regulators and suggests others as potential secondary modulators of basic regulatory cascade.

Finally, in Chapter 5, I further discuss the implications of my results and discuss potential future studies.
References


CHAPTER 2

FLAGELLAR GENES OF E. COLI ARE REGULATED IN STOCHASTIC PULSES DURING EXPONENTIAL GROWTH

2.1 INTRODUCTION

Clonal populations of bacteria can exhibit significant phenotypic heterogeneity (1-6). These variations can arise from inhomogeneous nutrient depletion or spatial distributions within structures such as biofilms (7, 8). However, even in the absence of such extracellular cues, stochastic differences in gene expression can drive large changes in cellular phenotypes (1, 5, 9, 10). In Chapter 1, I reviewed some notable examples of such phenotypic heterogeneity, including, in some bacteria, the regulation of motility (11-13).

Here, I report a new example of heterogeneous regulation in exponentially growing Escherichia coli. In this classic model system, flagellar motility requires the expression of more than 50 genes that are organized in a three-tiered transcriptional cascade, driven by the master regulator FlhDC (see Chapter 1). Traditionally, the expression of these genes has been implicitly assumed to be homogeneous during exponential phase. However, it has now been recognized that many laboratory strains of E. coli, including strains commonly used for studies of flagella, harbor an insertion (IS) element in the regulatory region of FlhDC (14). This mutation disrupts the binding of transcriptional repressors, causing FlhDC expression to be upregulated. In this study, I will show that in strains with wild-type regulation of FlhDC, flagellar genes are only expressed in a subset of cells during exponential growth. By contrast, strains with the insertion element mutation (from here on, I will call them “IS mutants” for simplicity) express flagellar genes homogenously.
I will further demonstrate that this heterogeneity arises during exponential growth because flagellar promoters stochastically switch on and off in slow pulses that span several generations. These pulses occur synchronously for genes within the same transcriptional class. However, each class has a distinct pulsing pattern, ranging from more to less frequent in Classes 1, 2 and 3 respectively.

Finally, I will show that in contrast to IS mutants, which grow more slowly due to flagellar expression, wild type strains minimize the growth cost of flagella expression even in the subset of cells producing flagella. On the basis of these observations, I will briefly speculate on the potential biological significance of stochastic pulses.

2.2 RESULTS

2.2.1. Expression of flagellin is highly heterogenous in *E. coli* cells with wild-type regulation of FlhDC

A common transcriptional marker for flagellar synthesis is the expression of the gene FliC (8, 15-17). Its gene product, flagellin, is the basic subunit that polymerizes to form the flagellar filament (18). Moreover, FliC is a Class 3 gene (19), i.e. the final tier of the flagellar transcriptional cascade—thus, the expression of FliC is also used to infer the expression of the previous tier genes, i.e. Class 2 and Class 1 genes.

To visualize the expression of flagellin, I constructed a fluorescent transcriptional reporter by fusing the regulatory region of FliC gene (P_{fliC}) to YFP. This reporter was inserted into the *E. coli* chromosome to avoid artificial cell-to-cell variations that can arise from fluctuations in plasmid segregation and replication.
Initially, I examined MG1655, a strain that lacks any insertion element mutations in FlhDC (CGSC 6300, see Appendix for further discussion) (14). Surprisingly, only a subset of cells were fluorescent during exponential phase (Fig. 2.1A). The difference between individual cells was dramatic: a majority of cells had no detectable fluorescence above background but a subset of cells was highly fluorescent (Fig. 2.1A, B). This behavior was observed for a range of optical densities from early (OD$_{600}$ = 0.1) to mid/late (OD$_{600}$=0.6) exponential phase.

The heterogeneity in flagellar gene expression appeared to contradict previous physiological observations that mid-exponential phase cells have on average 2-3 flagella each (20-22). An obvious difference between traditional strains used in these studies (such as RP437 or W3110) and the experimental strain MG1655 used here is the insertion element mutation in FlhDC. However, these strains have several other notable genetic differences (23, 24). To determine whether the apparent difference in flagellar gene regulation arose specifically due to the insertion element mutation, I replaced the native FlhDC regulatory region with a homologous region from W3110 which carries an IS5 insertion. This “IS5” strain of MG1655 began expressing flagellin homogenously at a high level (Fig. 2.1B).

Additionally, I examined the distribution of flagella in MG1655 cells. To facilitate direct comparison of flagellin expression and presence of flagella, I fused a fluorescent protein YFP to FliM, a protein which assembles into the basal body of the flagellar apparatus. This fusion protein is functional and has previously been used to monitor the assembly and inheritance of flagella (25). Consistent with my measurements of gene expression, basal bodies were detected in only a subset of cells (Fig. 2.1C). Moreover, cells with basal bodies matched cells with high flagellin gene expression (Fig. 2.1C). Together, these results confirm that both flagellin and flagella are expressed only in a subset of cells under wild-type regulation of FlhDC.
Figure 2.1 Flagellin Expression Is Heterogeneous During Exponential Growth

(A) Images of E. coli cells with flagellin promoter (P\textsubscript{fliC}) fused to YFP. Cells were harvested during early exponential phase (OD = 0.2). Left: Cells imaged on agarose pad via phase contrast. Right: YFP fluorescence in same set of cells.

(B) Distribution of fluorescence in cells with P\textsubscript{fliC}-YFP measured via flow cytometry. Wild type strains (WT, light blue) show large heterogeneity while insertion element mutants (IS5, dark blue, see main text) express flagellin homogeneously. The dashed red line indicates 2σ threshold for cellular auto-fluorescence (measured from cells without any fluorescent reporters).

(C) Expression of flagellin reporter correlates with presence of flagella. Cells harboring both a flagellin transcriptional reporter (P\textsubscript{fliC}-CFP, middle) and a flagellar basal body protein fused to YFP (FliM-YFP, right) were harvested during early exponential phase (OD = 0.2). Cells were placed on patterned agarose pads which have linear grooves (phase contrast, left, see Appendix). By improving the spatial separation between cells, the grooves make it easier to visualize to which cell the basal bodies belong.
2.2.2. Each class of flagellar genes shows distinct patterns of heterogeneity

Next, I investigated whether other flagellar genes (Fig. 2.2A) were also expressed heterogeneously during exponential growth. For each of the 14 flagellar gene promoters, strains harboring promoter-YFP fusions similar to that of FliC were constructed. Strikingly, three distinct patterns of gene expression emerged. Cells with reporters for class 3 genes such as MotA displayed the same expression pattern as FliC (Fig. 2.2B). Cells with class 2 transcriptional reporters also showed extremely large variations in fluorescence (coefficient of variation ≥~100%) with over a 100-fold difference in fluorescence between bright and dim cells. However, unlike cells with Class 3 promoters where distinct “on” and “off” populations could be identified, cells with Class 2 promoters had a more “smeared” unimodal distribution (Fig. 2.2C). Finally, cells with the sole Class 1 gene promoter, P_{flhD}, displayed a narrow fluorescence distribution with a coefficient of variation comparable to that of an unregulated constitutive promoter of similar transcriptional strength (CV=∼44%) (Fig. 2.2D, E).

Having observed similar patterns of expression for genes within the same class, I proceeded to measure the activity of two Class 2 or two Class 3 promoters simultaneously within the same cell. For the comparison of Class 2 promoters, a “reference” reporter was constructed fusing the promoter of the Class 2 gene FliF to CFP. This reporter, along with a Class 2 promoter fused to YFP was integrated into the chromosome (see Appendix), allowing for simultaneous measurement of FliF and another Class 2 gene expression. The expression of each Class 2 gene strongly correlated with the expression of the Class 2 reference promoter (Fig. 2.3A, C).
Figure 2.2 Patterns of Expression for Different Classes of Flagellar Genes During Exponential Growth

(A) Schematic of flagellar transcription network.

(B) Typical fluorescence distribution of Class 3 gene promoters. Shown is the fluorescence distribution of cells harboring the motA promoter fused to YFP, measured via flow cytometry. Like flagellin (FliC), motA expression is highly heterogeneous.

(C) Typical fluorescence distribution of Class 2 gene promoters. Shown is the fluorescence distribution of cells harboring the fliF promoter fused to YFP, measured via flow cytometry. The fluorescence distribution is still considerably wide (compared to (E)) but is no longer distinctly bimodal.

(D) Fluorescence distribution of the sole Class 1 gene promoter flhD.

(E) Fluorescence distribution of a constitutive promoter (Pro4) with similar mean expression to the wild-type flhD promoter.
Figure 2.3 Flagellar Genes Within The Same Class Show Correlated Expression

(A) Typical distribution of fluorescence in cells harboring two Class 2 promoter reporters, measured via flow cytometry. Shown is a 2D density plot of YFP and CFP fluorescence in cells harboring the fliA promoter fused to YFP and the flfF promoter fused to CFP.

(B) Typical distribution of fluorescence in cells harboring two Class 3 promoter reporters, measured via flow cytometry. Shown is a 2D density plot of YFP and CFP fluorescence in cells harboring the motA promoter fused to YFP and the fliC promoter fused to CFP.

(C) Table of correlation values for YFP and CFP fluorescence in cells with two Class 2 (red) or two Class 3 (blue) promoters. Fluorescence in each strain was measured via flow cytometry.

A similar strategy was used for the comparison of Class 3 promoters, where a FliC promoter fused to CFP served as the reference reporter. Again, the expression of each Class 3 promoter strongly correlated with the expression of the Class 3 reference promoter (Fig. 2.3B, C). These results established that genes within the same class are regulated coordinately within the same cell.
2.2.3. Heterogeneity in flagellar gene expression arises from stochastic switches in flagellar gene expression (“pulses”) during exponential growth

In principle, heterogeneity in flagellar gene expression could arise from cells transiently differentiating into subpopulations that stably express or repress flagella. Alternately, cells could “sample” different modes of gene expression over time. To better understand the temporal dynamics of flagellar gene expression, I decided to track changes in flagellar gene expression in single cells.

Time-lapse microscopy is a powerful method for such measurements (26). A major challenge arises, however, when studying fast dividing microbes such as E. coli over multiple division cycles: due to their exponential rate of growth, cells quickly overwhelm the imaging field of view and make accurate segmentation and fluorescence quantitation difficult (27). Moreover, due to nutrient consumption, the environment quickly becomes inhomogeneous. This effect can potentially confound whether observed heterogeneities in gene expression truly arise from internal processes or are artifacts of external environmental changes.

To address these concerns, I used a microfluidic device, the “mother machine”, which has previously been used to monitor gene expression in bacteria for long time periods (9, 28, 29). Briefly, this device consists of an array of ~1µm wide channels. These channels are connected to a larger “flow channel” at one end but are closed at the other end (Fig. 2.4A). The micron-wide channels are just wide enough to accommodate a single linear microcolony of E. coli. As cells grow and divide, new-born cells are pushed uni-directionally out of the channel. However, the cell at the bottom of the channel (the “mother” cell) remains trapped—this allows a single continuous lineage to be monitored for numerous generations (Fig. 2.4A). Microfluidic flow of
Figure 2.4 Flagellar Genes Are Expressed in Stochastic Pulses

(A) Schematic of microfluidic device for single cell studies. Left: Cells are grown in narrow tracks that constrain cells to grow in linear colonies. Excess cells are washed away by microfluidic flow which also provides fresh growth media. Right: at each time point, we follow the cell at the bottom of the channel (the “mother” cell, orange box). The segmentation channel defines the location of the cell while fluorescence from the flagellar promoter is obtained from the measurement channel.

(B) A kymograph of cells growing in the microfluidic device (20 mins per frame). Constitutive promoters such as the promoter used for cell segmentation (RNAIp from ColE1) show only small variations in fluorescence over time (Gray, top panel in each pair). By contrast, Class 3 (Top pair, bottom panel) and Class 2 (Middle pair, bottom panel) flagellar promoters switched between active and inactive states over multiple cell divisions. The only exception was Class 1 (Bottom pair, bottom panel) which, though fluctuating in intensity, remained continuously fluorescent.
growth media in the “flow channel” washes away excess cells and provides a constant source of fresh nutrients.

Cells harboring the flagellar promoter-YFP fusions were grown in this device for numerous generations. Strikingly, flagellar promoters randomly switched between being brightly fluorescent and having nearly no fluorescence over the course of multiple cell generations (Fig. 2.4B). The “pulses” of gene expression were completely uncorrelated in cells in adjacent channels. It was also visually apparent that different classes of flagellar promoters had distinct temporal dynamics with progressively decreasing frequency of switching events from Class 3 to 2 to 1 (Fig. 2.4B)—in fact, Class 1 genes remained fluorescent for the entire duration of the experiment although I could detect subtle fluctuations in fluorescence that also lasted multiple generations.

To better quantify the dynamics of pulsing in different promoters, I analyzed the various time traces of fluorescence. To characterize the frequency of switching events, I defined a heuristic threshold by adapting Otsu’s method, a technique commonly used for discriminating foreground and background pixels in images (30). However, qualitatively similar results were observed for a range of threshold values. Time periods when fluorescence was continuously above the threshold were defined as “on” periods while contiguous time periods with fluorescence values below the threshold were defined as “off” periods (Fig. 2.5A). The distribution of “on” and “off” periods for different class genes confirmed the visual observation: Class 2 genes pulsed more frequently than Class 3 genes while Class 1 genes fluctuated even more frequently than the other two gene classes (Fig. 2.5A, B).
Figure 2.5 Quantitative Characterization of Pulses

(A) Top row: Typical dynamics of pulsing in flagellar promoters. Shown are time traces of CFP and YFP fluorescence in cells harboring two Class 3 promoters (fliC and motA, left), two Class 2 promoters (fliF and fliA, middle), or the sole Class 1 promoter (flhD, right, only YFP trace shown). Also shown are thresholds used to define “on” and “off” (or “high” and “low”) periods. Bottom row: distribution of “on” and “off” periods. The “on” duration of Class 3 could be well-approximated by a gamma distribution with shape parameter 2. All other durations were well-approximated by an exponential distribution.

(B) Table summarizing mean duration and coefficient of variation of the various “on” and “off” phases.

(C) Cross-correlation analysis of CFP and YFP fluorescence in cells harboring two Class 3 (blue) or Class 2 (red) promoters. The average cross-correlation peaks at time offset 0, suggesting that promoters within the same class pulse synchronously.
The duration of the “on” and “off” phases was approximately exponentially distributed. The only significant exception was the “on” duration of Class 3 genes, which was better described by a gamma distribution with shape parameter \( \sim 2 \) (Fig. 2.5A, B). This observation suggests that with the exception of Class 3 gene firing, the duration of “on” and “off” states is “memoryless”—that is flagellar genes have the same probability of turning “off” (or “on”) regardless of how long it has already spent in the “off” (or “on”) phase. Additionally, the duration of “on” states was completely uncorrelated with the duration of previous or subsequent “off” states (Fig. 2.5D). This result implies that the amount of time that flagellar genes remain in the “on” or “off” phase does not affect the duration of the subsequent “off” or “on” phase.

I also compared the temporal dynamics of two Class 2 or two Class 3 genes in the same cell using the “two-color” strains with one flagellar promoter driving CFP and a second driving YFP (see above). In each case, two promoters in the same class fluctuated synchronously (Fig. 2.5B). Cross-correlation analysis further confirmed that there was no systematic temporal delay (lag) in gene expression between genes of the same class (Fig. 2.5C).

Taken together, these results explain many of the features of heterogeneous gene expression observed via flow cytometry. If flagellar genes pulse infrequently like Class 3 genes, most cells in the population would not have any fluorescence while few cells would be extremely bright. However, as the switching frequency becomes higher as in the case of Class 2 genes, more cells become caught “in transition”. Therefore, the distribution starts to become more unimodal. This
effect has previously been described in mathematical models of stochastically switching phenotypes (31). It is also easy to see how synchronous pulsing of genes within the same class gives rise to strong correlations in gene expression levels. Thus, we conclude that stochastic pulsing directly underlies the heterogeneous pattern of flagellar gene expression in mid-exponential populations.

2.2.4. Flagellar gene pulses and cell growth

Flagellar synthesis can have a large bio-energetic cost (32). Indeed, when I compared the growth rate of MG1655 and the “IS5” strain I engineered, I found that the IS5 strain grew noticeably slower (Fig. 2.6A). In turn, I wondered whether even in wild-type strains, the growth rate of the cell was reduced during time periods when flagellar gene expression was upregulated. From the time-lapse measurements, I compared the elongation rate of cells with high and low levels of flagellin (FliC) expression—cells with low flagellin expression were defined as cells with fluorescence below the heuristic threshold described above while cells with high flagellin expression were defined as cells with the top 20% fluorescence values. Surprisingly, no significant difference in growth rate could be observed between these two subpopulations (Fig. 2.6B).

I also explored whether the heterogeneous flagellar regulation was specific to the growth conditions used. To reduce background fluorescence from media, my experiments were performed in a MOPS based medium supplemented with a mixture of 20 amino acids and glycerol (Teknova, see Appendix). However, heterogeneous expression of flagellin was also observed when cells were grown in traditional complex media such as LB or tryptone-broth (Fig. 2.6C). In particular, tryptone-broth is a commonly used growth medium in the study of E. coli
motility. Thus, pulsing does not appear to be caused by a specific ingredient in the defined media I have used.

Figure 2.6 Relationship Between Flagellar Gene Expression and Cell Growth

(A) Growth cost of continuous flagellar expression. Cells with wild-type flagellar gene regulation, i.e. pulsing (“WT”, light blue), were compared an otherwise identical strain, engineered to harbor the typical insertion element mutation in FlhD regulation found in many laboratory strains (“IS5”, dark blue). Growth rate was estimated via linear fit to log-transformed optical density data. The IS5 strain grows roughly 18% slower than the wild type strain.

(B) Growth cost of flagellar expression during pulsing. Cells with P_{fliF}-YFP were grown in the microfluidics device. For each time trace, the elongation rate was measured for time periods when cells had low fluorescence vs high fluorescence. The distribution of the elongation rates for the low (gray) and high (red) fluorescence periods show virtually no difference.

(D) Effect of different growth media on flagellin expression. Cells with P_{fliC}-YFP were grown in different media and the fluorescence was measured via flow cytometry. Heterogeneity in flagellar expression is observed for both defined and complex media. The presence of amino acids in the media appear to promote flagellar gene expression.
I was unable to identify specific inducers that significantly biased MG1655 towards expression of flagellin. However, I found that the flagellin expressing population was considerably diminished when amino acids were removed from the growth media (Fig. 2.6D). Currently, it is unclear whether this effect is due to the specific absence of amino acids or the overall reduced growth rate. However, this observation hints at the possibility that nutrient availability can affect the frequency of flagellar gene expression within the population.

2.3 DISCUSSION

In this work, I have identified and characterized a stochastic pulsing strategy for flagellar gene regulation that results in a striking heterogeneity in flagellar production in exponentially growing E. coli. My work also suggests this heterogeneity may have been previously obscured by the FlhDC-upregulating insertion element mutation found in many laboratory strains. Additionally, I note that most previous studies of flagellar gene expression have used bulk population measurements. Thus, even in studies where strains with wild-type regulation of FlhDC were examined, this single-cell heterogeneity would likely have been missed.

One notable feature of the stochastic pulses is that genes within the same class fire synchronously. This observation contradicts a traditional model of flagellar gene expression which suggested that flagellar genes within the same class are activated in sequence over a period of multiple cell divisions (33). However, this model was inferred from bulk population measurements on an insertion element mutant strain and therefore likely describes the behavior of flagellar promoters at artificially saturated expression levels. It is easy to see that in wild-type strains, synchronous pulsing would be beneficial since all the proteins that are necessary for flagellar synthesis are produced at the same time while the cell has decided to sample a “motile”
lifestyle. Synchronous pulsing also suggests that the pulsing dynamics are mediated by FlhDC, the common transcription factor for Class 2 genes, and FliA, the common sigma factor required for the transcription of Class 3 genes. However, I note that regulation by a common transcription factor does not in itself guarantee a synchronous response—for example, if downstream genes had large differences in binding affinities for the transcription factor, potentially only a subset of downstream genes would be activated. It would be interesting to analyze how individual Class 2 and Class 3 promoters are tuned to ensure synchronous firing of all Class 2 or all Class 3 genes during the stochastic pulses.

In addition to establishing that flagellar genes are expressed in stochastic pulses, my work also establishes that each class of genes have distinct pulsing dynamics. The implication of this observation is that at each stage of the transcriptional cascade, fluctuations in the production of FlhDC and FliA are modified. I will examine this process more in detail in the next chapter (Chapter 3).

Qualitatively, the heterogeneity expression of flagellar genes in E. coli is reminiscent of heterogeneous regulation of motility in Bacillus subtilis and Salmonella (9, 11-13). As described in Chapter 1, B. subtilis cells during exponential growth exist either in a flagella expressing “motile” state or a morphologically distinct sessile state where multiple cells are “chained” together (9, 12). Similarly, in Salmonella, cells in exponential growth were found to exist in two subpopulations, one repressing and the other upregulating the production of flagella (11, 13).

The comparison with B. subtilis is particularly intriguing because the temporal dynamics underlying the heterogeneity have also been investigated. Norman et al. used a modified version of the mother machine and observed switching between “motile” and “chained” states over
numerous cell divisions under constant environmental conditions (9). Interestingly, they found that the time spent in the “motile” state was exponentially distributed but that the time spent in the “chained” state was tightly regulated. This observation implies that *B. subtilis* spends a random amount of time in the “motile” state but then executes a stereotypical “chained” program of fixed duration. In *E. coli*, I found that the flagellar gene “on” states—functionally analogous to the “motile” states—were similarly exponentially distributed but the “off” phases—the analogy to “chained” states, by extension—were also exponentially distributed. Thus, unlike *Bacillus*, *E. coli* appears to sample both motile and nonmotile lifestyles for a random amount of time. These differences are likely a reflection of the considerably different genetic organization of flagellar genes in these two organisms.

Unlike *Bacillus*, *Salmonella* shares many regulatory features with *E. coli*, including the master regulator FlhDC and the three-tiered arrangement of flagellar genes (34, 35). These similarities have often led to observations from *Salmonella* being used interchangeably with observations from *E. coli*. As such, a comparison between flagellar gene heterogeneity in *E. coli* and *Salmonella* would be highly interesting. However, to date, the temporal dynamics of flagellar gene heterogeneity in *Salmonella* are unknown. Nonetheless, I note that in recent studies, flow cytometry measurements of Class 2 gene expression in *Salmonella* showed a clear bimodal distribution (36) which contrasts my observations that Class 2 genes in *E. coli*, unlike Class 3 genes, show a broad but unimodal distribution. I will provide a more detailed comparison with *Salmonella* in subsequent chapters where I analyze the molecular mechanisms that might regulate or affect the stochastic pulses.

What might be the biological significance of the pulsatile regulation of flagellar synthesis? Although bacterial flagella and motility are highly useful for bacteria, one well-known cost of
flagellar synthesis is the large amount of cellular resources required. Synthesis of a single flagellum requires the expression of over 10,000 molecules of flagellin in addition to the over 40 different gene products required for the assembly of the motor—previously, in *E. coli*, the cost has been estimated to be at least 2% of the cell’s total energetic budget (37-39). Indeed, my observations confirm that IS element mutants that are continuously producing flagella grow noticeably slower even in relatively rich growth conditions.

This cost presents a potential dilemma for *E. coli*. When growth conditions are poor, motility would allow cells to seek out better environments. However, it would also be difficult for cells to devote precious resources to a highly expensive process. Conversely, when growth conditions are good, cells can afford to produce flagella but may not need it since they are already in a favorable environment. In fact, flagellar production in *E. coli* and *Salmonella* is repressed by glucose (40-42), suggesting that when growth conditions are optimal, motility is deemed “unnecessary”. Thus, one possibility is that the stochastic expression of flagella during exponential growth might be a mechanism to periodically “anticipate” potentially harmful changes to the environment (such as nutrient depletion or sudden introduction of stress).

My observation also suggests that during stochastic pulses, the growth of cells producing flagella does not appear to be noticeably different from the growth of cells in which flagellar synthesis is off. Thus, an alternate possibility is that the cost of flagellar synthesis might depend on the duration of flagellar gene expression—that is, only prolonged expression of flagella starts to incur a growth cost. If so, stochastic pulses might be a mechanism to ensure that cells can periodically “recover” from the strain of flagellar synthesis before its ability to buffer the stress is significantly compromised.
Finally, I note that flagella can also efficiently provoke immune responses in many host-organisms, including humans (43-46). Thus, for enteric bacteria such as *E. coli*, there may be an additional cost to continuously producing flagella. In fact, in *Salmonella*, it has been suggested that the heterogeneity in flagellar gene expression minimizes or reduces the inflammatory response from human cells (13, 47). As such, stochastic pulsing in *E. coli* might similarly be motivated by an effort to periodically “hide” flagella from host immune systems.
References


CHAPTER 3

GENERATION AND PROPAGATION OF STOCHASTIC PULSES ACROSS THE FLAGELLAR TRANSCRIPTIONAL CASCADE

3.1 INTRODUCTION

Gene expression in individual cells inevitably undergoes random fluctuations even in the absence of explicit environmental variations (1-3). In transcriptional cascades, such fluctuations in upstream genes can propagate to downstream genes (4-6). Even in relatively simple synthetic systems, the propagation can be highly complex because different sources of fluctuations in gene expression can combine in non-intuitive ways (4). Conversely, similar patterns of propagation can potentially arise from a diverse set of mechanisms (2, 7-9). A question of general interest is how different biological networks have evolved regulatory strategies to address or even exploit these fluctuations.

Previously, I demonstrated that genes in the flagellar synthesis network, one of the classic examples of a transcriptional cascade in *E. coli*, stochastically switch on and off in slow pulses that span several generations. I observed that genes within the same class pulsed synchronously, suggesting that the pulsing behavior of Class 2 and Class 3 genes is coordinated (and presumably mediated) by the transcription factor for these genes, namely, FlhDC (for Class 2) and FliA (for Class 3) respectively.

Interestingly, I also found that the dynamics of pulses were markedly different at each step of the transcriptional cascade: Class 1 genes fluctuated only a small amount but at a higher frequency (on average every 2-3 generations) while Class 2 genes underwent more dramatic fluctuations at
reduced frequencies (on average, every 5 generations). Class 3 genes also showed dramatic fluctuations but at even lower frequencies (on average, every 10 generations). This observation raises the question of how fluctuations or pulses in transcription of FlhDC (Class 1) and FliA (Class 2) result in different pulsatile dynamics in the downstream Class 2 and Class 3 genes, respectively (Fig. 3.1).

**Figure 3.1 Dynamics of Pulsing Are Modified At Each Step of The Transcriptional Cascade**

Each class of flagellar genes has a different characteristic pulsing dynamic. Class 1 gene expression fluctuates frequently but with limited dynamic range (left). Class 2 gene pulses become less frequent but during each pulse, gene expression is upregulated by a considerably larger magnitude (middle). Class 3 gene pulses become even less frequent so that periods in which flagellar genes are expressed become more sharply punctuated (right).

Here, using time-lapse microscopy, I have directly measured how transcription of two genes of different classes fluctuate or pulse in the same cell. My studies indicate that pulses in downstream genes correlate with fluctuations in upstream genes but also that not all upstream fluctuations result in downstream gene activation. To understand the underlying basis for this
process, I have used a combination of time-lapse microscopy and synthetic expression systems to determine the input-output relationship between FlhDC and Class 2 as well as FliA and Class 3 genes. Interestingly, for both steps of the transcriptional cascade, I observe a triphasic response: a “repressed” phase where the output genes are insensitive to variations in input concentration, an “ultrasensitive” phase where small changes in input variations give large changes in downstream gene activity and finally a “proportional” phase where the output varies approximately linearly with the input. I also find that for both FlhDC and FliA, mutating a post-translational repressor YdiV (10, 11) and FlgM (12), respectively, appears to “linearize” the input-output relationship and cause fluctuations across the classes to be more synchronized. These behaviors are reminiscent of features of a “molecular titration” model (13) where a small “titratable” concentration of a tight-binding negative regulator acts as a threshold for output activation. I hypothesize how the observed patterns of pulse propagation in flagellar gene expression might be explained by this model.

3.2 RESULTS

3.2.1. Only a Subset of Upstream Gene Pulses Activate Downstream Genes

To study how fluctuations in Class 1 gene expression correlate with pulses of Class 2 genes in the same cell, I constructed a strain with an YFP reporter for FlhDC transcription and a CFP reporter for Class 2 transcription (Fig. 3.2A, also see Appendix). These cells were grown in the “mother machine” under the same conditions as before (see Chapter 2 and Appendix). For each channel, I monitored YFP and CFP fluorescence in the “mother cell” at the bottom of the channel for over 40 generations.
Figure 3.2 Only a Subset of Upstream Gene Pulses Activate Downstream Genes

(A) System for simultaneous measurement of Class 1 and 2 genes. Class 1 transcription was measured via YFP inserted at the end of the FlhDC transcript. Class 2 expression was measured via a chromosomally inserted copy of a Class 2 promoter ($P_{fliF}$) fused to CFP.

(B) An example time trace of Class 1 (green) and Class 2 (red) gene expression in the same cell. Fluctuations in Class 1 can be accompanied by pulses in Class 2 ("overlapping", green) but fluctuations of similar magnitude in Class 1 can also result in no Class 2 expression (“non-overlapping”, gray).

(C) Cross-correlation analysis of Class 1 and Class 2 pulses. Class 1 “pulses” occur on average 50 mins (~1 cell division) ahead of Class 2 pulses.

(D) System for simultaneous measurement of Class 2 and 3 genes. Transcription from the Class 2 FliF promoter (monitored with CFP) was used as a proxy for FliA transcription based on the previously observed synchrony between the two promoters (see Chapter 2). Class 3 expression was measured using the FliC promoter fused to YFP. Both reporters were chromosomally inserted.

(E) An example time trace of Class 2 (red) and Class 3 (blue) gene expression in the same cell. Some Class 2 pulses are accompanied by Class 3 pulses (“overlapping”, red) whereas during other Class 2 pulses do not activate Class 3 (“non-overlapping”, gray).

(F) Cross-correlation analysis of Class 2 and Class 3 pulses. Class 2 pulses occur on average 30 mins ahead of Class 3 gene pulses.
Analysis of the time traces revealed that fluctuations in Class 1 genes only partially correlated with Class 2 gene pulses: some Class 1 fluctuations (which for simplicity, I will henceforth also call “pulses”) clearly correlated with Class 2 pulses but other Class 1 pulses of similar magnitude failed to activate Class 2 genes (Fig. 3.2B). I could also observe a difference in relative timing of Class 1 and Class 2 pulses. Cross-correlation analysis confirmed that Class 1 gene pulses occurred on average 50 minutes (approximately 1 generation under experimental conditions) prior to Class 2 gene pulses (Fig. 3.2C). It was also apparent that when Class 1 fluctuations activated Class 2 genes, only small changes in Class 1 gene activity led to dramatic changes in Class 2 activity.

I also constructed a strain with a CFP reporter for Class 2 transcription and YFP reporter for Class 3 transcription (Fig. 3.2D). Again, these cells were examined in the “mother machine”. Like the Class 1 and 2 gene pairs, Class 2 gene pulses only partially correlated with Class 3 gene pulses—that is, only a subset of Class 2 pulses activated Class 3 pulses (Fig. 3.2E). Unlike Class 1 gene pulses, however, Class 2 gene pulses did not appear to be further magnified in relative amplitude. Cross-correlation analysis indicated that Class 3 gene pulses were delayed on average 30 minutes after Class 2 gene pulses (Fig. 3.2F).

3.2.2. Probing the “dose-response” relationship (gene regulatory function) at each stage of the flagellar transcriptional cascade

To better understand the patterns of correlations that I observed, I first investigated the “dose-response” relationship between Class 1 and 2 gene expression. In the case of transcriptional networks, such input-output relationships are also called the “gene regulatory function” or GRF.
Because the natural fluctuations in Class 1 gene expression were relatively small in magnitude, I decided to use a synthetic promoter to artificially adjust the levels of Class 1 expression.

Typically, for such studies, inducible promoters such as the IPTG-inducible lac promoter are used to tune different levels of expression (4, 14, 15). However, many inducible promoters rapidly transition from being maximally-repressed to maximally-induced over a small range of inducer concentrations (16). This “all-or-none” response makes it difficult to probe the behavior of the system at intermediate levels of transcription. I instead opted to use a set of variable-strength synthetic constitutive promoters whose strengths were tuned to be ~2-3 fold of each other (17).

I replaced the endogenous FlhDC promoter in the dual Class 1 and Class 2-reporter strain with the synthetic promoters from this set. In the process, I also replaced the endogenous 5’ untranslated region (UTR) with the 5’-UTR of the synthetic promoter. This additional modification was made so that the sequences around the synthetic promoter’s transcriptional start site would not be substantially altered—otherwise, the relative transcriptional strengths of the different promoters might not be preserved (17). In the subsequent chapter, I will describe a further consequence of this modification in more detail.

I identified a promoter Pro4 which gave rise to Class 2 gene expression levels that were similar to levels found in wild-type cells. In turn, I chose Pro3, a promoter approximately 2-fold weaker than Pro4, and ProB, a promoter approximately 3.5-fold stronger than ProB. I grew these three strains in the “mother machine” and measured the natural fluctuations in FlhDC transcription and the resultant Class 2 activity in the same cell for multiple generations. Strains with Pro3 driving FlhDC had virtually no Class 2 transcription even though the transcription strength was only half
Figure 3.3 Probing the “dose-response” relationship (gene regulatory function) at each stage of the flagellar transcriptional cascade

(A) Relationship between Class 1 and Class 2 expression in strains with a synthetic promoter driving FlhDC. Class 2 gene expression ($P_{fli}$-YFP) was measured in cells where the native Class 1 promoter was replaced by a weak (Pro3, dark green), intermediate (Pro4, green) or strong (ProB, light green) synthetic promoter. Left: scatter plot of FlhDC transcriptional reporter concentration vs. Class 2 transcriptional reporter concentration obtained from time-lapse microscopy. Middle: example time traces from these strains. Right: the scatterplot shows a triphasic response.

(B) Scatter (density) plot of Class 1 and Class 2 reporter expression in WT strains (left) and WT strains superimposed with insertion element mutant (IS5, black circle) (middle). These distributions roughly correspond to two of the three phases in (A).

(C) Scatter (density) plot of Class 2 and Class 3 reporter expression in WT strains. Interestingly, a similar triphasic response can be found between Class 2 and 3.
of Pro4 (Fig. 3.3A, “weak” promoter). In strains with ProB, on the other hand, the Class 2 gene was expressed continuously and fluctuations in its expression mirrored the fluctuations in Class 1 transcription reporter (Fig. 3.3A, “strong” promoter). The most striking dynamic was observed in the Pro4 strain where even small variations in the Class 1 transcription led to large swings in Class 2 expression, a behavior highly reminiscent of endogenous Class 2 gene pulses in wild-type strains (Fig. 3.3A, “intermediate” promoter).

These differences became even more apparent when the Class 1 reporter fluorescence from each experiment was plotted against the Class 2 reporter fluorescence to map out a “dose-response” (Fig. 3.3A, left). The resulting scatterplot revealed three distinct phases. In the first “repressed” phase, Class 2 genes remained inactive regardless of Class 1 expression. In the second “ultrasensitive” phase, however, even small changes in Class 1 expression resulted in extremely large variations in Class 2 activity. Finally, in the third “proportional” phase, changes in Class 1 expression resulted in a more linear response in Class 2 expression (Fig. 3.3A, right).

Having mapped the full “dose-response” (which I will henceforth call the Class 1-GRF) with synthetic promoters, I decided to compare it to the input-output relationship between Class 1 and Class 2 genes in wild-type strains. When I plotted the Class 1 fluorescence against the Class 2 fluorescence from the time-lapse data (see above, Fig. 3.2A), the resulting distribution mostly resembled the second phase of the Class 1-GRF (Fig. 3.3B, left). This observation was further reinforced by examining Class 1 and Class 2 gene expression in a strain engineered to have an insertion element mutation in the regulatory region of FlhDC (see Chapter 2). When the Class 1 and Class 2 fluorescence from these strains were mapped together with fluorescence measurements from the wild-type strain, the resulting distribution resembled even more strongly the full Class 1-GRF (Fig. 3.3B, middle and right).
I also examined the analogous “dose-response” between Class 2 and Class 3 genes (which I will call the Class 2-GRF). Unlike Class 1 genes, Class 2 genes naturally fluctuate over a large range of expression levels. As a result, I found that the time-lapse data of Class 2 and Class 3 genes in wild-type strains was sufficient to map a full “dose-response” (Fig. 3.3C, left). Interestingly, I found that the full Class 2-GRF also had three phases that resembled the full Class 1-GRF (Fig. 3.3C, right).

3.2.3. Role of Post-Translational Regulators of FlhDC and FliA

My experiments with the synthetic promoters indicated that the triphasic input-output response, and specifically the “ultrasensitive” behavior seen in wild-type strains, occurred independently of transcriptional or translational regulation of FlhDC (I will provide a more detailed discussion in Chapter 4). In turn, I decided to investigate how post-translational regulation of FlhDC might affect the input-output relationship between Class 1 and Class 2 genes.

YdiV is a protein which has been shown to bind FlhDC and promote its degradation and/or prevent its binding to DNA (10, 11, 18). Although it is highly effective at inhibiting the activation of Class 2 genes by FlhDC, a previous study indicated that YdiV is only expressed in low quantities in \textit{E. coli} (11). The same study also concluded that YdiV had negligible impact on flagellar gene expression (18). However, the strain used in the study, W3110, carries the previously described insertion element mutation (19). Therefore, I wondered whether YdiV might have a more noticeable impact in a strain with wild-type regulation of FlhDC.

To examine the impact of YdiV on Class 1 and Class 2 expression, I engineered an in-frame deletion of the ydiV gene in the dual Class 1 and Class 2-reporter strain. In this strain, the
Figure 3.4 Mutations in YdiV and FlgM Improve Correlations Between Class 1 and 2, and Class 2 and 3, Respectively

(A) Effect of knocking out YdiV, a post-translational repressor of FlhDC. Left: Example time trace of Class 1 and Class 2 gene expression in a YdiV knockout mutant. Fluctuations in Class 1 and Class 2 gene expression become well-correlated. Right: scatter (density) plot of Class 1 and Class 2 reporter expression in a ΔydiV strain.

(B) Effect of knocking out FlgM, a post-translational repressor of FliA. Left: Example time trace of Class 2 and Class 3 gene expression in a FlgM knockout mutant. Pulses in Class 2 and Class 3 genes become synchronized. Right: Scatter (density) plot of Class 2 and Class 3 reporter expression in a ΔflgM strain.

(C) Fraction of cells with basal bodies as a function of Class 2 gene expression. Cells were “binned” according to the Class 2 reporter (P_fliF-CFP) fluorescence. Basal bodies were detected by the chromosomally inserted fusion protein FliM-YFP (see Chapter 2).
correlation between Class 1 and Class 2 gene expression was markedly improved: pulses in Class 2 gene expression mostly matched fluctuations in Class 1 gene expression (Fig. 3.4A, left). The input-output relationship also become more linear (Fig. 3.4A, right) although I note that there was still a modest amplification of Class 1 fluctuations.

I also decided to examine whether the propagation of pulses from Class 2 to Class 3 is influenced by post-translational regulation of FliA. A well-known mechanism for this process involves the anti-FliA protein FlgM (20). FlgM, like FliA is a Class 2 flagellar gene. When both proteins are present in the cell, FlgM binds tightly to FliA and prevents its interaction with RNA polymerase (12, 20). This repression is relieved when flagellar basal bodies are assembled (from other Class 2 genes) because the basal bodies export FlgM from the cytoplasm (12). Traditionally, cells were thought to produce flagella continuously and as a consequence FlgM was essentially absent from the cell during mid-exponential growth (21). However, my discovery of pulses raised the possibility that since Class 2 gene expression levels fluctuate, FlgM might play a non-trivial role.

I engineered an in-frame deletion of the flgM gene in the dual Class 2- and Class 3-reporter strain. This deletion caused Class 2 and Class 3 gene expression to be almost perfectly correlated (Fig. 3.4B). In light of the important role that basal bodies play in the regulation of FlgM, I also decided to examine the correlation between Class 2 gene expression and formation of basal bodies in wild-type cells (Fig. 3.4C). To visualize basal bodies, I used a fusion protein FliM-YFP (see Chapter 2). I discovered that there was a threshold concentration of Class 2 gene expression below which no basal bodies could be observed. This observation led me to hypothesize that FlgM represses Class 3 pulses when Class 2 expression levels are low (Fig. 3.3C) due to the absence of functional basal bodies.
Taken together, these results suggested that YdiV and FlgM modulate the propagation of pulses from Class 1 to 2 and Class 2 to 3, respectively.

3.2.4. Molecular Titration Model

The propagation of pulses from Class 1 to 2, and from Class 2 to 3 shared common features. First, both processes showed a triphasic “dose-response” or GRF for both transitions (Fig. 3.3). A particularly pronounced feature of these two GRFs was the suppression of outputs (Class 2 or 3, respectively) below a threshold concentration of inputs (Class 1 or 2, respectively), i.e. the “repressed” phase. Tantalizingly, I also found that for both processes, a post-translational negative inhibitor played a major role. These observations led me to hypothesize that a “molecular titration” model might describe both steps of the flagellar transcriptional cascade.

Briefly, the “molecular titration” model describes a scenario where an active protein (TF) is sequestered by a high affinity repressor (anti-TF) into an inactive complex (13). For simplicity, let’s consider a situation where the anti-TF and TF bind 1:1. If the concentration of TF falls below that of the anti-TF, nearly of the TF would be sequestered. However, when the concentration of TF exceeds that of the anti-TF, the number of active or free TF starts to increase proportionally to the increase in total TF concentration (Fig. 3.5A). An interesting feature of this mechanism is that when concentrations of the TF are approximately equal to the concentration of the anti-TF (i.e. near the “threshold” concentration), even small variations in the concentration of TF (or anti-TF) can lead to large variations in the amount of free TF. This “ultrasensitive” behavior can readily be visualized by plotting the dose response in a log-log plot (Fig.3.5B).
Figure 3.5 Molecular Titration Model

(A) Sequestration of a transcription factor (“TF”) by a strong binding repressor (“anti-TF”) results in a threshold response. The blue curve depicts an idealized amount of TF available as a function of total TF produced based on a simple kinetic model (see text).

(B) The same model also predicts an “ultrasensitive” response near the threshold—that is small fold changes in output result in large fold changes in output. This effect can be visualized by plotting the input-output response in log-log space. The “ultrasensitivity” near the threshold is comparable to that of a cooperative-process approximated by a high Hill coefficient (red dotted line).

(C) Cartoon depicting how small fluctuations in Class 1 might result in large pulsatile Class 2 response. If Class 1 is produced in similar concentrations to a negative inhibitor (possibly YdiV), the model predicts a large amplification in Class 2 activity fluctuations. If the negative inhibitor concentration also fluctuates, similar levels of Class 1 expression could result in noticeably different levels of Class 2 expression.

(D) Class 2 gene expression fluctuates over a wider range of expression levels. Therefore, the dominant effect that is observed is the threshold response where Class 2 pulses of low magnitude are “filtered-out”.


Here, a clear tri-phasic behavior emerges: at low concentrations, the output is insensitive to changes in the input while at high concentrations, changes in the output are approximately proportional to changes in the input. These two phases are bridged by a phase where, small fold changes in input are amplified into larger changes in output.

The amplification of input fluctuations is highly reminiscent of sigmoidal Hill-function dose-responses that can arise from TFs regulated by cooperative multimerization. However, a key difference is that the output response of cooperative mechanisms saturate at high concentrations whereas in molecular titration models, the output response can continue to increase with increasing inputs (Fig. 3.5B). Of note, models analogous to “molecular titration” have been shown to be relevant in a variety of systems including RNA-RNA and toxin-antitoxin interactions (22-25).

How might this model explain the propagation of flagellar gene pulses? Let’s imagine that the production of Class 1 genes is tuned such that the average concentration of FlhDC is close to the stoichiometric concentration of a negative inhibitor (possibly YdiV). If so, the molecular titration model predicts that small variations in Class 1 expression would lead to large changes in the amount of free FlhDC and in turn, large fluctuations in Class 2 activity (Fig. 3.5C).

Moreover, if the concentration of the negative inhibitor also fluctuates (i.e. is “noisy”), similar levels of Class 1 gene production could yield considerably different amounts of Class 2 activity. The net result would be that relatively small magnitude fluctuations in Class 1 transcription become stochastically amplified as large magnitude pulses in Class 2 transcription.

In the next step of the flagellar cascade, Class 2 gene expression naturally fluctuates over a large range. At low levels of Class 2 gene expression, flagellar basal bodies cannot be assembled (Fig. 3.4C). Thus FlgM effectively sequesters all FliA molecules. However, when Class 2 gene
expression is sufficiently high, basal bodies can be assembled, FlgM is exported, and FliA can now trigger Class 3 gene expression. I note that both FliA and basal body genes are expressed in a correlated manner (see Chapter 2). As a consequence, FlgM functions analogously to a titratable repressor in the “molecular titration” model—FlgM can only repress FliA at low levels of FliA expression. The thresholding behavior results in only a subset of Class 2 pulses being propagated to Class 3.

3.3 DISCUSSION

In this chapter, I investigated how fluctuations in Class 1 gene expression propagate across the flagellar transcriptional cascade. In principle, pulses in flagellar gene expression could have been coordinated simply by large changes in the transcription of FlhDC with Class 2 and Class 3 genes passively mirroring this change. If so, flagellar gene in all three classes would have been highly correlated. Instead, I find that at each stage of the flagellar cascade, the dynamics of flagellar gene transcription undergo important modifications. This observation suggests that the function of the three-tiered cascade is not simply to temporally offset flagellar gene expression as has been traditionally assumed. Rather, the flagellar cascade acts as a series of circuits to convert a relatively low noise signal (Class 1) into discrete and large pulses (Class 3).

The most striking feature in the transition from Class 1 to Class 2 genes is an apparent “ultrasensitivity”. In principle, this behavior could have arisen from cooperativity in the master regulator FlhDC, which exists as the hexamer FlhD₄C₂. However, the apparent Hill coefficient of DNA binding for this multimer has been estimated to be only around 1.8-2.2 (26), which would only result in a modest amplification of Class 1 pulses. (I briefly note that in ΔydiV mutants, I still observed a modest amplification of Class 1 fluctuations in Class 2. This residual
amplification could potentially be due to this cooperativity in FlhDC multimerization).

Moreover, as discussed above, a cooperative mechanism would result in a saturated response at high levels of Class 1. Although presumably there is a level of Class 1 expression above which Class 2 gene expression is saturated, my measurements suggest that even the hypermotile insertion element mutants do not reach this point.

In this work, I have hypothesized that a molecular titration model could explain this ultrasensitive response. YdiV is a particularly attractive candidate for the titratable negative repressor in this model because it is likely expressed in low quantities and has been shown to be a highly effective repressor of FlhDC. However, the production and activity of FlhDC can be influenced by a variety of post-transcriptional mechanisms including small RNAs (27). Although I will discuss in the subsequent chapter why these other mechanisms are unlikely to be “essential” for the ultrasensitive behavior, they nonetheless may play a role in shaping the final dose-response between Class 1 and 2 genes.

I also note that molecular titration is by no means the only mechanism that can give rise to the observed behavior. For instance, special synthetic circuits have been engineered to take advantage of the ultrasensitivity of positive feedback circuits, but avoid the “saturating” effect by using “shunt” plasmids to titrate away some of the transcription factors (28). The dose-response from such circuits can also qualitatively approximate the behavior seen here. Even more complex mechanisms involving complex protein-protein or protein-DNA interactions could certainly be possible.

The most pronounced feature of the Class 2 to Class 3 transition is the “thresholding” effect. This observation fits well with the classic model that Class 3 expression is suppressed until
sufficient Class 2 proteins are produced (more specifically, enough to produce a basal body).

The observation that threshold levels of Class 2 expression are required for the formation of basal bodies as well as the effect of FlgM on the correlation between Class 2 and Class 3 provide support for this interpretation. An interesting observation, however, is that when Class 2 pulses are close to the threshold for basal body formation, pulses with similar Class 2 expression only occasionally give rise to Class 3 pulses (Fig. 3.2B, Fig. 3.3C). It is currently unclear whether this is because at that concentration, basal bodies occasionally fail to form or if basal bodies occasionally are unable to pump out sufficient amounts of FlgM before the Class 2 pulses subside. Nonetheless, this observation matches the prediction of the molecular titration model that around the threshold, the output response is highly sensitive to small variations in the input (and repressor concentration).

Finally, I note that the molecular titration model described here only provides a basic “qualitative” description of the observed behavior. The simplest descriptions of the model assume that the system is at steady-state (22). By contrast, flagellar gene expression is dynamic. However, because the typical pulses typically span several generations, to a first order approximation, each “on” and “off” phase could be thought of as a transient steady state. In addition, more complex models of molecular titration have demonstrated that even at faster time scales, the key features of the model, namely thresholding and ultrasensitivity, are preserved (13). It would nonetheless be interesting to develop a more specific model for the flagellar network with physiological parameters and compare it directly to the observed gene expression dynamics.
References


CHAPTER 4

MUTATIONS THAT AFFECT FLAGELLAR GENE EXPRESSION DURING EXPONENTIAL GROWTH OF E. COLI

4.1 INTRODUCTION

In *Escherichia coli*, the master regulator of flagellar synthesis FlhDC is regulated at multiple levels. The production of FlhDC is regulated both at the transcriptional (1-4) and translational level (5). After translation, additional regulators can affect the ability of FlhDC to activate transcription of Class 2 genes (6, 7). There is also some evidence that downstream flagellar genes can potentially provide feedback regulation on FlhDC (8-11) (Fig. 4.1).

In the previous two chapters, I described the effect of a few mutants that had the most direct influence on the process that was being studied. Here, I will provide a more comprehensive view of FlhDC regulation by examining various regulators previously suggested or shown to affect FlhDC transcription, translation or activity (Fig. 4.1). Specifically, I will describe how mutants of these regulators affect Class 2 gene expression during exponential growth. I will also briefly examine the role of two proteins, CpxR and CsgD, which have recent been reported to play a role in regulating Class 3 gene expression (12). Many of these results will contribute to my comparison of stochastic pulsing in *E. coli* to previously known examples of heterogeneous motility in other bacteria in the final chapter.
Figure 4.1 Production and Activity of FlhDC is Regulated At Multiple Levels

(A) Cartoon of FlhDC regulation at various levels of FlhDC production. Concentration of FlhDC is regulated by a diverse range of transcription factors (blue), small RNAs (green), and regulatory proteins (orange). Several other transcriptional and translational regulators have been proposed to regulate FlhDC expression but here are omitted for simplicity. Note that flhD transcription requires CRP and is therefore sensitive to catabolite repression.

(B) Potential feedback interaction from downstream flagellar genes. In Salmonella, FliZ is shown to repress transcription of YdiV. However, the YdiV regulatory region in E coli lacks the FliZ binding site. FliT has been shown to promote degradation of FlhC but is unable to do so when FlhDC is bound to DNA. Finally, it is unknown whether the flagellar apparatus itself can directly or indirectly affect FlhDC production or activity.
4.2 RESULTS

4.2.1. Measuring Flagellar Gene Expression via Flow Cytometry

In many previous studies of FlhDC-regulator mutants, changes in flagellar gene expression were monitored by bulk population measurements such as microarrays or bulk liquid culture fluorescence. However, based on my discovery of phenotypic heterogeneity in flagellar gene expression during exponential growth, I decided to use flow cytometry to monitor the distribution of fluorescence in cells with flagellar promoter-fluorescent protein fusions. A secondary benefit of flow cytometry was that it permitted measurement of a large number of cells even at low cellular densities. Thus, I could monitor cells before they had depleted substantial amounts of nutrients to better approximate the conditions of the microfluidic devices.

![Figure 4.2 Flagellar Gene Pulses Result in a Characteristic Flow Cytometry Profile](image)

This diagram illustrates how pulses in Class 2 expression result in a broad unimodal distribution. This distribution can be readily distinguished from narrow unimodal distributions observed for constitutive promoters. The time traces and histograms were constructed from data acquired in Chapter 2 and 3 for P_{flif}-YFP (red) and P_{proB}-YFP.
As noted before, in wild-type strains, slow stochastic pulses in Class 2 gene expression gave rise to a broad unimodal distribution (CV\textasciitilde100\%) (Fig. 4.2). Flow cytometry allowed me to compare changes in the fluorescence distribution in the various mutants against this “WT” distribution. I note that it is difficult to directly reconstruct the underlying temporal dynamics from this dataset because similar distributions can arise from a variety of mechanisms (13). Moreover, my previous studies indicated that Class 1 to Class 2 relationship can be complex and makes such reconstructions even more challenging. For simplicity, I will limit my comparisons to how various mutants affect the static “snapshot” distribution of Class 2 gene expression within the population.

### 4.2.2. Transcriptional Regulators of FlhDC

Transcription of FlhDC is governed by a variety of positive and negative regulators. The positive regulators include CRP (14, 15) and H-NS (15). Because they are global regulators, mutants of these genes display widespread pleiotropic effects that are difficult to interpret. Therefore, they were omitted from this study. The major negative regulators of FlhDC are LrhA (2), HdfR (3), OmpR (1) and RcsB (4) (Fig. 4.1A). Mutants of these four regulators were the primary subjects of my study.

LrhA and HdfR are LysR type regulators that have been shown to directly repress FlhDC (2, 3). To date, their primary function appears to be regulation of FlhDC, although LrhA has also been suggested to indirectly regulate translation of the sigma factor RpoS (16). Consistent with previous results, mutations in either protein resulted in significant up-regulation of Class 2 activity (Fig. 4.3A). Additionally, this up-regulation resulted in a narrower distribution of Class 2 gene activity, similar to the effect of the insertion element mutation (17) (see Chapter 2).
Figure 4.3 Effect of Various Mutants on Class 2 Gene Expression.

(A) Effect of deleting transcriptional regulators of FlhDC. Class 2 gene expression was measured in WT (red, line) and mutants (olive, bars) via flow cytometry. LrhA and HdfR have the largest effect on Class 2 expression. OmpR had only a negligible effect while RcsB had an intermediate effect. The positive effect of the knockouts on Class 2 gene expression is consistent with their previously reported role as a repressor.

(B) Effect of deleting small RNA regulators of FlhDC. Class 2 gene expression was measured in WT (red, line) and mutants (purple, bars) via flow cytometry. OxyS and OmrAB had only a negligible effect. Deletion of ArcZ had a positive effect while deletion of McaS had a negative effect on Class 2 gene expression, consistent with their previously reported role as a translational repressor and activator, respectively.

(C) Effect of deleting YdiV. Class 2 gene expression in WT (red, line) and YdiV knockout (gold, bars). Contrary to previous reports, deletion of YdiV has a measurable effect on flagellar synthesis.
The other major negative regulators, OmpR and RcsB, are two-component system response regulators that are activated primarily in response to environmental cues (18, 19). Although they respond to explicit environmental perturbations, these two regulators have also been shown to play an important role regulating flagellar gene expression during the growth of biofilms (20, 21). However, during exponential growth, deletion of OmpR had a negligible effect on Class 2 transcription while deletion of RcsB only resulted in modest up-regulation of Class 2 activity (Fig. 4.3A). Moreover, the distribution of Class 2 activity in the RcsB mutant was still broad, suggesting that much of the underlying pulsing behavior was preserved.

From these observations, I propose LrhA and HdfR to be the two main regulators of FlhDC transcription during exponential growth.

4.2.3. Translational Regulation of FlhDC

Small RNAs (sRNAs) mediate the second layer of FlhDC regulation. The 5’ untranslated region (UTR) of the FlhDC mRNA has been shown to be a hub for a variety of sRNAs including ArcZ, OxyS, OmrA/B and McaS (Fig. 4.1A) (5, 22). Among these sRNAs, McaS uniquely promotes the translation efficiency of FlhDC, while the other three sRNAs downregulate translation. It is still debated whether these sRNAs act by modifying the secondary structure around the ribosomal binding site or by altering the mRNA lifetime.

Deletion of OxyS and OmrA/B had no effect on Class 2 activity (Fig. 4.3B). By contrast, deletion of ArcZ increased Class 2 activity while deletion of McaS decreased Class 2 activity, consistent with their known mode of action (Fig. 4.3B). Thus, ArcZ and McaS are proposed to be the two main regulators of FlhDC translation during exponential growth. Unlike mutations in LrhA or HdfR, however, the fluorescence distributions in ArcZ and McaS mutants still exhibited
a significant “tail” suggesting that some cells had significantly lower/higher activity than most of the population. This observation suggests that heterogeneity in gene expression is not completely abolished in these mutants.

4.2.4. Post-Translational Regulation of FlhDC by YdiV

The protein YdiV was first characterized in Salmonella as a post-translational inhibitor of FlhDC (7). Subsequently, its homolog in E. coli (also called YdiV) was also shown to be an effective anti-FlhDC factor (6, 23). In both cases, YdiV can bind both the FlhD monomer and the FlhD₄C₂ complex (which I will simply call FlhDC) (6, 7). The Salmonella YdiV was shown to decrease the lifetime of both the FlhD monomer and FlhDC (24). Both Salmonella and E. coli homologs of YdiV also caused FlhDC to dissociate from its binding site on DNA (6, 7).

In Salmonella, YdiV is part of an important positive feedback regulation of FlhDC (25-27)—the precise details of this mechanism will be further discussed in the next chapter where I compare Salmonella and E. coli. By contrast, YdiV in E. coli previously was found to play a negligible role in flagellar gene expression. However, as previously discussed in Chapter 3, this result was inferred from studies in W3110, a strain of E. coli with the hypermotile insertion element mutation (23). Thus, I hypothesized that this previous result could have arisen from excessive production of FlhDC “overriding” the regulation by low quantities of YdiV.

Indeed, I discovered that contrary to the previous observation in W3110, deletion of YdiV in MG1655 (CGSC6300)—my experimental strain with wild-type regulation of FlhDC—resulted in up-regulation of Class 2 activity (Fig. 4.3C). Like mutants of ArcZ, mutants of YdiV had a narrower distribution of Class 2 expression but still retained broader tails. For this mutant, I was able to interpret this “snapshot” distribution from time-lapse measurements (see Chapter 3). As
previously described, in this strain, fluctuations in Class 2 genes became more “proportional” to fluctuations in Class 1, likely due to the disruption of ultrasensitivity. Consequently Class 2 gene pulses in ΔydiV strains more closely resembled Class 1 gene fluctuations, albeit with still a modest amplification in magnitude. These effects narrow the distribution of Class 2 gene activity.

4.2.5. Effect of Various Mutants on Strain With Synthetic Expression of FlhDC

Previously, I engineered strains to express FlhDC from a synthetic constitutive promoter (28) to probe the gene regulatory function (GRF) between Class 1 and Class 2 genes (see Chapter 3). Interestingly, these strains showed a triphasic “dose-response” or GRF that resembled features of the “dose-response” measured in wild-type strains. When constructing these strains, I replaced the native promoter of FlhDC as well as the 5’ untranslated region (UTR) with a synthetic sequence. This modification was done because previously it had been shown that the relative difference in transcriptional strength between the different synthetic promoters was better preserved when the synthetic UTR was retained as an “insulating” sequence (28). A consequence of this modification was that this synthetic FlhDC transcript no longer contained sequences to which the sRNAs ArcZ and McaS could bind.

I confirmed that this synthetic expression strain was no longer sensitive to transcriptional or translational regulation as expected. Deletion of LrhA, HdfR, ArcZ or McaS had no effect on Class 2 expression from this strain (Fig. 4.4). By contrast, deletion of YdiV resulted in an up-regulation of Class 2 expression, similar to the effect observed in the wild-type strain. Since the synthetic expression strain retained the key features of the wild-type GRF even in the absence of influence from transcriptional or translational regulators, I decided to focus my investigation on
YdiV (see Chapter 3). However, it should be noted that although ArcZ and McaS are apparently not necessary for observed ultra-sensitivity between Class 1 and Class 2 genes, they might nonetheless help shape or modify the GRF in the wild-type strain.

**Figure 4.4 Effect of Various Mutants on Class 2 Gene Expression in Strain with Synthetic FlhDC Promoter.**

Flow cytometry measurements of Class 2 gene expression in various mutants in a background strain with synthetic constitutive expression of FlhDC (P<sup>Pro4</sup>-FlhDC). In these strains, the 5’ UTR is also replaced by a synthetic sequence. As a result, neither transcriptional (olive, bars) nor translational regulators (purple, bars), have any effect on Class 2 gene expression compared to the background strain (red, line). However, YdiV, a post-translational regulator, still plays a role as evidenced by the increased Class 2 expression in a ∆ydiV mutant (gold, bars).

**4.2.6. Influence of Downstream Flagellar Genes on Class 2 Gene Expression by FlhDC**

In *Salmonella*, two Class 2 genes have been shown to participate in feedback regulation of FlhDC. The Class 2 gene product FliZ in *Salmonella* acts as a transcriptional repressor for YdiV (8). Hence, expression of FliZ leads to increased activation of FlhDC via positive feedback (Fig 4.1B). This positive feedback in *Salmonella* was shown to be crucial for generating the heterogeneity in flagellar gene expression (25, 26).
A second Class 2 protein FliT, again in *Salmonella*, was found to directly bind the FlhC subunit of FlhDC and prevent the complex from associating with DNA (9, 10). Thus FliT was proposed to be a post-translational inhibitor of FlhDC like YdiV. Interestingly, however, FliT was unable to interact with FlhDC that was pre-bound to DNA (9).

![Figure 4.5 Mutations That Do Not Affect Flagellar Gene Regulation](image)

(A) Effect of deleting flagellar genes. Class 2 gene expression was measured in WT (red, line) and mutants (gray, bars) via flow cytometry. FliZ and FliT, two proteins previously implicated in *Salmonella* for their role in post-translational regulation of FlhDC, have a negligible effect on Class 2 gene expression in E. coli. Deletion of FliA also has a negligible effect, which by extension suggests that Class 3 gene products do not affect Class 2 gene expression. Finally, deletion of FliF, which impairs the assembly of flagella also has no effect on Class 2 expression suggesting that Class 2 expression is insensitive to the presence/absence of flagella.

(B) Effect of ΔcpxR or ΔcsgD on Class 3 gene expression (P_{fliC}-YFP) measured via flow cytometry. Wild type expression is shown in blue (line) while the mutant expression profile is shown in pink (bars). Recently, CpxR and CsgD were implicated as negative regulators of FliA transcription (and consequently, negative regulators of Class 3 gene expression) during exponential growth. However, under my experimental conditions, deletion of CpxR or CsgD have no effect on Class 3 gene expression.
Surprisingly, I found that neither FliZ nor FliT had a significant effect on Class 2 gene expression in *E. coli* (Fig. 4.5A). For FliZ, I note that the 5’ untranslated region of YdiV in *Salmonella* and *E. coli* show considerable divergence. Transcription of YdiV in *Salmonella* is thought to be regulated by the neighboring nlpC gene promoter, while its *E. coli* counterpart is thought to be expressed from its own promoter within this region (23). Thus, I hypothesize that the FliZ-YdiV positive feedback may be a *Salmonella* specific feature. It is less clear why FliT does not appear to affect Class 2 gene expression in *E. coli*. I speculate that perhaps during exponential growth, most available FlhDC complexes are bound by DNA during pulses and therefore are insensitive to the action of FliT—meanwhile, the “off” phases might be sufficiently long for any FliT proteins produced to be diluted away.

I also decided to investigate whether FliA, the sigma factor responsible for Class 3 activation, could potentially affect FlhDC expression or activity (and as a consequence Class 2 gene expression). However, deletion of FliA did not alter the distribution of Class 2 gene expression (Fig. 4.5A). Since Class 3 genes cannot be expressed in these mutants, by inference, I could also conclude that Class 3 gene products such as the chemotaxis signaling pathway did not influence Class 2 gene expression.

Finally, I decided to investigate whether the flagellar apparatus itself could potentially influence FlhDC activity. Although no specific mechanisms are known for how this might occur, in principle, a regulator similar to FlgM that is exported by the basal body could allow FlhDC expression to be coupled to the presence of basal bodies in the cell. To test this hypothesis, I deleted the Class 2 protein FliF which is necessary for the formation of flagellar basal bodies. However, I observed no difference in Class 2 gene expression (Fig. 4.5A). Thus Class 2 gene pulses appear to occur independently of basal body assembly.
4.2.7. Influence of CpxR and CsgD on Class 3 Gene Expression

Recently, a study of flagellar gene regulation in *E. coli* proposed that during exponential growth, the regulators CpxR and CsgD repress the expression of Class 3 genes by repression of Class 2 transcription (12). This study measured the bulk fluorescence from flagellar promoter-GFP reporters in the background strain BW25113. Notably, BW25113 does not harbor the insertion element mutation upstream of FlhDC (29). Thus, I decided to investigate whether these two regulators played a role under my experimental conditions. Surprisingly, I found that deletion of these two regulators did not affect Class 3 gene expression (Fig. 4.5B). It is currently unclear whether this difference reflects other genetic variations in the background strains (namely BW25113 vs MG1655) or the specific growth conditions tested.

4.3 DISCUSSION

In this chapter, I examined the effect of a variety of regulators previously shown to regulate FlhDC. Mutants that altered Class 2 expression, with the notable exception of ΔrcsB, all demonstrated a narrower fluorescence distribution. As previously noted, it is difficult to infer directly from these distributions how the underlying dynamics of pulses have changed. Consequently, direct measurement of the temporal dynamics of flagellar gene expression in these mutants would be a useful follow-up to the observations presented here. Nonetheless, it appears that each of these regulators can bias the population towards “off” or “on” states of flagellar regulation. Therefore, it is tempting to speculate that different environmental cues might be able to tune the degree of flagellar heterogeneity via these regulators.

I also investigated whether other regulators whose functions were inferred from studies in *Salmonella* played a similar role in *E. coli*. My studies revealed many interesting similarities but
also striking differences which were discussed above. In the subsequent chapter, I will explore some of the broader implications of these observations as well as some future experiments they motivate.
References


CHAPTER 5

CONCLUSIONS AND PROSPECTS

In this work, I described my discovery of a previously unknown heterogeneous mode of flagellar gene regulation in *E. coli*. This heterogeneity was found to arise from stochastic “pulses” of flagellar gene expression that spanned multiple cell divisions. These pulses were synchronous for genes within the same class. However, the dynamics of pulsing were sequentially modified at each step of the transcriptional cascade so that small “fluctuations” (or mini-“pulses”) of Class 1 activity first become amplified into large pulses of Class 2 genes. These pulses were then additionally filtered during the propagation to Class 3 genes, resulting in rare and discrete pulses. My studies also suggested that both of these transitions might be regulated by a “titratable” inhibitor, which I currently hypothesize to be the post-translational regulator YdiV and FlgM.

Throughout this work, I have endeavored to discuss the significance of many of my observations within each chapter. Here, I will discuss some of the broader implications of my work as well as some future experiments that they motivate.

5.1 Comparison with *Salmonella*

Based on my studies, flagellar motility in *E. coli* now joins the growing list of bacterial phenotypes that are shown to be regulated heterogeneously even in the absence of explicit environmental variations. Perhaps the most interesting comparison lies with the heterogeneous regulation of flagellar motility in *Salmonella*, in part because they directly share many of the genes and regulatory motifs. Both bacteria have a three-tiered organization of flagellar genes (Class 1, 2, and 3) which are regulated by FlhDC and FliA (1). Many of the regulatory proteins I
have discussed here such as the transcriptional regulators LrhA and RcsB and post-translational regulators YdiV (and FlgM for FliA) are also shared (2). Therefore, it is tempting to consider the heterogeneity in flagellar gene expression in both bacteria to be identically regulated.

However, a few crucial details are different. In *Salmonella*, a protein RflM which is transcribed by FlhDC, acts as a negative regulator of FlhDC (3). However, no homologs of RflM have been found in *E. coli*. This difference is interesting because negative feedback by RflM has been found to be important during exponential growth regulation of flagellar gene regulation (2).

Perhaps a more important difference, which I have discovered in this work, is that the FliZ-YdiV positive feedback, which was found to be crucial for maintaining and fine-tuning heterogeneity in *Salmonella* (4-6), is apparently absent/non-functional in *E. coli*. As previously discussed, FliZ in *Salmonella* represses YdiV transcription by binding to the neighboring nlpC gene promoter (7). It has therefore been postulated that YdiV is primarily transcribed by read-through transcription. Although the *E. coli* homolog of NlpC also lies upstream of the YdiV coding sequence, the intergenic region between nlpC and ydiV is considerably different from that of *Salmonella* (8). The intergenic region in *E. coli* is postulated to contain a σ24 dependent promoter (8). Thus, rather than being produced by read-through transcription, YdiV in *E. coli* might be primarily transcribed from its own promoter. This difference might account for the absence of the FliZ mediated feedback.

One consequence of positive feedback regulation is that it allows cells to more stably exist in “on” or “off” states. As previously noted, in *Salmonella*, flow cytometry measurements reveal a clear bimodal distribution of Class 2 gene expression (4). By contrast, I observed for all Class 2 genes, a broad unimodal distribution. Bimodal distributions are predicted to arise from relatively
long residence times in the “on” and “off” states, i.e. slow switching frequencies. Thus, one hypothesis is that flagellar gene heterogeneity in *Salmonella* is more “stable”. It would be interesting to use similar time-lapse studies in microfluidics devices (9-12) to investigate the temporal dynamics of flagellar gene expression in *Salmonella*.

5.2 Environmental Cues that Modulate Stochastic Pulsing

Another interesting line of inquiry comes from my observations that the relative distribution of flagellar “on” populations appears to be sensitive to the presence of amino acids in the media (see Chapter 2). Previously, in insertion element mutants, poor nutrient conditions were inferred to stimulate flagellar motility (8, 13). By contrast, in *Salmonella*, motility was higher in rich nutrient conditions (4, 8). My observations in wild-type strains of *E. coli* appear to qualitatively agree with the *Salmonella* model. However, in *Salmonella*, increased concentrations of yeast extract were found to bias populations towards the flagella-“on” state (4). By contrast, I observed a modest decrease in Class 2 gene expression in media supplemented with yeast extract (e.g. compare LB vs tryptone broth, Chapter 2, Fig. 6). Thus, while the general trend of nutrients favoring flagella expression may be preserved between the two strains, the specific “inducer” to which the two bacteria are sensitive may be different. It would be interesting to investigate whether certain nutritional cues, e.g. specific amino acids, or even average growth/cell division rates can tune the dynamics of pulsing in *E. coli*.

One “natural” environmental cue comes from the different growth phases of *E. coli*. Traditional physiological studies, albeit on “insertion element strains”, have found changes in mean flagellar number per cell during early-, mid-, and late-exponential growth (14, 15). In principle, one relatively straight-forward variation of the experiments presented here would be to observe the
temporal dynamics of flagellar gene expression in microfluidic devices without applying any flow. However, a potential challenge with this strategy is that to characterize the “typical” timescale of gene expression fluctuations, cells need to be observed for numerous generations—in the absence of flow, the growth phase would change too quickly to permit such observations. A different strategy might be to grow batch cultures to different optical densities and then use a constant flow of “conditioned” media. A limitation of this strategy is that over the course of the microfluidics experiments, fast degrading molecules might be steadily depleted from the conditioned media. Nonetheless, it would be a useful first order “approximation” of the different growth phases.

These studies could in turn be coupled with the different mutant strains I have examined in this study. As previously remarked, OmpR and RcsB have previously been shown to play an important role in biofilms (16). Thus, it is possible that as cells are grown in media that are increasingly closer to stationary conditions, these two regulators might play an enhanced role.

5.3. Probing the Molecular and Dynamical Details of Flagellar Gene Pulses

In this work, I observed a “triphasic” input-output response for each step of the flagellar gene cascade that motivated a “molecular titration” model. Although my work currently points to YdiV and FlgM as two possible molecular candidates for “titratable” inhibitors that can generate this behavior, further studies would be required to test this hypothesis. In the case of YdiV, one possible experiment would be to simultaneously measure the transcription of FlhDC, YdiV and Class 2 genes. This experiment could help determine whether the combined fluctuations of YdiV and FlhDC better predict Class 2 gene activity, in support of the molecular titration model. An obvious challenge, currently, is the lack of a spectrally compatible third fluorescent protein
which has a maturation time comparable to that of YFP and CFP. However, since pulses in
Class 2 gene expression are slow and large in magnitude, it may nonetheless be possible to
obtain reliable estimates of the underlying transcriptional dynamics. A second set of
experiments, for both YdiV and FlgM, would be to artificially tune the expression of these
proteins and observe whether the input-output dynamics change as predicted by a molecular
titration model. One difficulty may be tuning the artificial expression level of YdiV and FlgM to
appropriate levels so as not to require excessive amounts of FlhDC or FliA. This challenge
could be overcome either by the use of different strength promoters to drive artificial expression
(17) or even by tuning the translational rate by mutating the ribosomal binding site of the
synthetic expression system (18). Finally, these experiments could be expanded to additionally
investigate the role of small RNAs, such as ArcZ and McaS, both in the presence and absence of
YdiV. Together, these additional studies will provide the basis for a more predictive and
quantitative model of how pulses are generated and propagated across the flagellar network.

5.4. Consequences of Flagellar Gene Pulsing on Other Genetic Networks of E. coli

Could the pulsatile regulation of flagellar genes have consequences beyond the production of
flagella?

My studies here demonstrate that pulsatile synthesis of FliA, albeit with modification by FlgM,
results in a coordinated pulsatile synthesis of Class 3 genes. In addition to being the alternative
sigma factor for Class 3 genes, however, FliA regulates the expression of several non-flagellar
genes (19). The most prominent of these are cellular genes related to the synthesis and
degradation of secondary messenger c-di-GMP (20, 21), which is in turn a key signaling
molecule for the coordination of cell adhesion and biofilm formation (22). Additionally, the
Class 2 gene FliZ was shown to antagonize action of the stationary phase sigma factor RpoS (σ38) by binding to −10 sites recognized by this sigma factor (23). A future line of study, therefore, would be to examine whether the expression of other non-flagellar genes, in particular, those related to the c-di-GMP or RpoS, are also coordinated by the pulses of flagellar genes.

5.5. Conclusion

The flagellar biosynthesis network in *E. coli* is one of the classic “model” transcriptional systems. Although at first glance it appears to be a relatively simple linear cascade, a closer examination reveals many complex regulatory details. Many of these features presumably ensure the proper coordinated expression of the 14 different core operons as well as potentially other regulatory genes. In this work, I have demonstrated that some of these regulatory features may also contribute to a complex dynamic regulation that results in a striking phenotypic heterogeneity. As such, this classic network continues to yield new fascinating lines of research.
References


APPENDIX

MATERIALS AND METHODS

Many of the materials and methods were common for the work contained in each chapter. As such, they are presented here in a single appendix.

A. STRAINS AND PLASMIDS

a. Strain and Plasmid Table

Table A1. List of Plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Source</th>
<th>Antibiotic Marker</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMK4</td>
<td>This work</td>
<td>Kan</td>
<td>Empty vector with YFP</td>
</tr>
<tr>
<td>pMK7</td>
<td>This work</td>
<td>Amp</td>
<td>Empty vector with CFP</td>
</tr>
<tr>
<td>pMK4-FlgA</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flgA}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlgB</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flgB}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlhB</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flhB}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlIA</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flIA}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlID</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flID}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlIE</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flIE}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlIF</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flIF}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlIL</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flIL}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlgK</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flgK}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlgM</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flgM}$-YFP</td>
</tr>
<tr>
<td>pMK4-FlIC</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{flIC}$-YFP</td>
</tr>
<tr>
<td>pMK4-MotA</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{motA}$-YFP</td>
</tr>
<tr>
<td>pMK4-Tar</td>
<td>This work</td>
<td>Kan</td>
<td>$P_{tar}$-YFP</td>
</tr>
<tr>
<td>pMK7-FlIF</td>
<td>This work</td>
<td>Amp</td>
<td>$P_{flIF}$-CFP</td>
</tr>
<tr>
<td>pMK7-FlIC</td>
<td>This work</td>
<td>Amp</td>
<td>$P_{flIC}$-CFP</td>
</tr>
<tr>
<td>pProB-YFP</td>
<td>This work</td>
<td>Gent</td>
<td>$P_{ProB}$-YFP</td>
</tr>
<tr>
<td>pPro4-YFP</td>
<td>This work</td>
<td>Gent</td>
<td>$P_{Pro4}$-YFP</td>
</tr>
<tr>
<td>pPro3-YFP</td>
<td>This work</td>
<td>Gent</td>
<td>$P_{Pro3}$-YFP</td>
</tr>
<tr>
<td>pSIM5</td>
<td>Gift of D. Court</td>
<td>Cm</td>
<td>Helper plasmid encoding red recombinase proteins</td>
</tr>
<tr>
<td>Table A1 (Continued)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCP20</td>
<td>CGSC</td>
<td>Amp, Cm</td>
<td>Helper plasmid encoding FLP recombinase</td>
</tr>
</tbody>
</table>

Table A2. List of Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG1655 (CGSC 6300)</td>
<td>CGSC</td>
<td>Background Strain</td>
</tr>
<tr>
<td>MG1655+IS5</td>
<td>This work</td>
<td>See below</td>
</tr>
<tr>
<td>MGR</td>
<td>This work</td>
<td>MG1655 IntS::ZeoR-P_RNA1-mCherry</td>
</tr>
<tr>
<td>MGR</td>
<td>This work</td>
<td>MGR <em>motA(E98K)</em></td>
</tr>
<tr>
<td>MGR-E98K</td>
<td>This work</td>
<td>MGR-E98K galK::AmpR-P_fliF-CFP flhDC-YFP</td>
</tr>
<tr>
<td>MGR-E98K FD</td>
<td>This work</td>
<td>MGR-E98K (see below)</td>
</tr>
<tr>
<td>MGR-E98K FliF FliG</td>
<td>This work</td>
<td>MGR-E98K P_fliG-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliB</td>
<td>This work</td>
<td>MGR-E98K P_fliB-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliB</td>
<td>This work</td>
<td>MGR-E98K P_fliB-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliA</td>
<td>This work</td>
<td>MGR-E98K P_fliA-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliD</td>
<td>This work</td>
<td>MGR-E98K P_fliD-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliE</td>
<td>This work</td>
<td>MGR-E98K P_fliE-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliF</td>
<td>This work</td>
<td>MGR-E98K P_fliF-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliF FliL</td>
<td>This work</td>
<td>MGR-E98K P_fliL-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FC</td>
<td>This work</td>
<td>MGR-E98K P_fliC-YFP galK::AmpR-P_fliF-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliC FliGK</td>
<td>This work</td>
<td>MGR-E98K P_fliGK-YFP galK::AmpR-P_fliC-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliC FliM</td>
<td>This work</td>
<td>MGR-E98K P_fliM-YFP galK::AmpR-P_fliC-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliC FliC</td>
<td>This work</td>
<td>MGR-E98K P_fliC-YFP galK::AmpR-P_fliC-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliC MotA</td>
<td>This work</td>
<td>MGR-E98K P_motA-YFP galK::AmpR-P_fliC-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K FliC Tar</td>
<td>This work</td>
<td>MGR-E98K P_tar-YFP galK::AmpR-P_fliC-CFP attB::KmR-</td>
</tr>
<tr>
<td>MGR-E98K Pro3 FD</td>
<td>This work</td>
<td>MGR-E98K galK::AmpR-P_fliF-CFP flhDC-YFP P_fliD::GentR-Pro3</td>
</tr>
</tbody>
</table>
Table A2 (Continued)

<table>
<thead>
<tr>
<th>Strain Description</th>
<th>Source</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGR-E98K Pro4 FD</td>
<td>This work</td>
<td>MGR-E98K galK::AmpR-PfliCFP flhDC-YFP PflhD::GentR-Pro4</td>
</tr>
<tr>
<td>MGR-E98K ProB FD</td>
<td>This work</td>
<td>MGR-E98K galK::AmpR-PfliCFP flhDC-YFP PflhD::GentR-ProB</td>
</tr>
<tr>
<td>MGR-E98K FD ΔydiV</td>
<td>This work</td>
<td>MGR-E98K galK::AmpR-PfliCFP flhDC-YFP ydiV::FRT</td>
</tr>
<tr>
<td>MGR-E98K FC ΔflgM</td>
<td>This work</td>
<td>MGR-E98K galK::AmpR-PfliCFP flhDC-YFP flgM::FRT</td>
</tr>
<tr>
<td>MGR-E98K+IS5 FliF FliC</td>
<td>This work</td>
<td>MGR-E98K+IS5 galK::AmpR-PfliCFP FliM(GSGGG)YFP</td>
</tr>
<tr>
<td>MGR F FliM-YFP</td>
<td>This work</td>
<td>MGR galK::AmpR-PfliCFP FliM(GSGGG)YFP (see below)</td>
</tr>
<tr>
<td>MGR C FliM-YFP</td>
<td>This work</td>
<td>MGR galK::AmpR-PfliCFP FliM(GSGGG)YFP (see below)</td>
</tr>
<tr>
<td>MGR FC ΔLrhA</td>
<td>This work</td>
<td>MGR galK::AmpR-PfliCFP attB::KmR-PfliCF-YFP</td>
</tr>
<tr>
<td>MGR FC ΔHdfR</td>
<td>This work</td>
<td>MGR FC hdfR::FRT</td>
</tr>
<tr>
<td>MGR FC ΔOmpR</td>
<td>This work</td>
<td>MGR FC ompR::FRT</td>
</tr>
<tr>
<td>MGR FC ΔRcsB</td>
<td>This work</td>
<td>MGR FC resB::FRT</td>
</tr>
<tr>
<td>MGR FC ΔOxyS</td>
<td>This work</td>
<td>MGR FC oxyS::FRT</td>
</tr>
<tr>
<td>MGR FC ΔOmrAB</td>
<td>This work</td>
<td>MGR FC omrAB::FRT</td>
</tr>
<tr>
<td>MGR FC ΔArcZ</td>
<td>This work</td>
<td>MGR FC arcZ::FRT</td>
</tr>
<tr>
<td>MGR FC ΔMcaS</td>
<td>This work</td>
<td>MGR FC mcaS::FRT</td>
</tr>
<tr>
<td>MGR FC ΔYdiV</td>
<td>This work</td>
<td>MGR FC ydiV::FRT</td>
</tr>
<tr>
<td>MGR FC ΔFlZ</td>
<td>This work</td>
<td>MGR FC flZ::FRT</td>
</tr>
<tr>
<td>MGR FC ΔFlT</td>
<td>This work</td>
<td>MGR FC flT::FRT</td>
</tr>
<tr>
<td>MGR FC ΔFlA</td>
<td>This work</td>
<td>MGR FC flA::FRT</td>
</tr>
<tr>
<td>MGR FC ΔFlF</td>
<td>This work</td>
<td>MGR FC flF::FRT</td>
</tr>
</tbody>
</table>

b. Background Strain

The background strain for this work was MG1655 (Coli Genetic Stock Center (CGSC) 6300).

There are two stocks of MG1655 that are commonly available. MG1655 (seq) (CGSC 7740) was the subculture used by the Blattner lab for the complete genomic sequencing of *E. coli* K-12.

This stock harbors an IS1 element in the regulatory region of FlhDC (1). By contrast, CGSC
6300, the isolate of MG1655 first submitted to the Coli Genetic Stock Center (CGSC), does not have this insertion element sequence.

c. Chromosomal Engineering

Strains used in this study were constructed using red recombination (2). Parental strains were transformed with the helper plasmid pSIM5 and red recombination was performed using standard protocols (3). During strain construction, cells were grown at 30°C to ensure maintenance of pSIM5, which has a temperature sensitive origin. All strains used in experiments were cured of the help plasmid prior to the experiment.

(i) Construction of In-frame Deletion Mutants

To generate in-frame deletion mutants, I designed a custom cassette similar to the cassette used to generate mutants in the Keio Collection (4). This cassette contained a gentamicin resistance marker (5) flanked by FLP recognition target (FRT) sites and 20bp sequences optimized for binding of PCR primers. An initial “template” cassette flanked by 40bp of sequences homologous to the GalK gene was ordered as a gBlock® (IDT). This cassette was chromosomally integrated into the GalK locus and sequenced. For subsequent in-frame deletions, strains harboring this cassette were used as templates for PCR amplification. The PCR primers were designed to contain 40bp overhangs homologous to the deletion target and 20bp specific to my deletion cassette. The homologous overhangs were designed to insert the cassette between the first 10th and last 10th codon of the coding sequence.
Following successful chromosomal insertion of the cassette, the gentamicin resistance marker was removed by Flp-mediated recombination using the helper plasmid pCP20 following previously described protocols (6).

(ii) “Scarless” Chromosomal Engineering

Some chromosomal manipulations such as the generation of point-mutations required “scarless” chromosomal engineering—that is changes to the chromosome that do not leave behind selective markers or other extraneous sequences such as FRT sites. I adopted a two-step “selection-counterselection” strategy (7). I modified a dual selection/counter-selection cassette in pKD45 (8) consisting of a kanamycin resistance marker and a toxin ccdB driven by a rhamnose inducible promoter (P_{RhaB}). This original cassette required counter-selection to be done on minimal M9 plates containing rhamnose because presence of other carbon sources allowed cells to avoid activation of the rhamnose promoter and escape counter selection. Cell growth was extremely slow on these plates, requiring almost two full days of incubation before colonies became visible. To overcome this limitation, I replaced the rhamnose inducible promoter with the arabinose inducible P_{araB} using standard molecular biology techniques and isothermal assembly (9) using the NEB HiFi DNA Assembly Master Mix (New England Biolabs). The resulting kmR-araBp-ccdB (“KAC”) cassette allowed for efficient counter-selection on LB plates containing 10% arabinose. I also constructed a variant of this cassette with the gentamicin resistance (gmR-araBp-ccdB or “GAC”) again using standard molecular biology techniques and isothermal assembly, again using the NEB HiFi DNA Assembly Master Mix (New England Biolabs).

d. Construction of Background Strain with Insertion Element Mutation in FlhDC
The effect of the insertion element mutation in FlhDC could directly be analyzed by comparing MG1655 (seq) (CGSC 7740) and MG1655 (CGSC 6300). However, I decided to investigate whether different insertion element mutations found in other laboratory strains could also play the same role. Unlike MG1655 (seq) which carries an IS1 element in the FlhDC regulatory region, W3110 carries an IS5 insertion element (1). These two insertion elements are also oriented in the opposite direction. Thus, I decided to replace the wild-type FlhDC regulatory region in MG1655 (CGSC6300) with the regulatory region from W3110. First, a KAC cassette was inserted into the promoter region of FlhDC in MG1655 (CGSC) using the techniques described above. The FlhDC regulatory region in W3110 was then amplified by PCR with primers in the FlhD and UspC genes that flank this region. This linear DNA was used to “knock-out” the KAC cassette which resulted in the strain “MG1655+IS5” (or “IS5” in various figures).

e. **FliM-YFP**

The functional fusion protein FliM-GSGGG-YFP has previously been used to visualize flagellar motor assembly (10). In this work, I generated this fusion protein directly in the native chromosomal locus to ensure that this protein is produced with the same dynamics as the wild-type FliM protein. To reduce the impact of the fluorescent protein insertion on the translation of the neighboring gene FliN, I designed the following custom DNA sequence: a fragment of the FliM coding sequence with the last 21bp replaced with synonymous codons was fused to a sequence encoding the GSGGG linker and a coding sequence of YFP. To the end of the YFP coding sequence, the original 21bp encoding the C-terminus of FliM was attached. This hybrid sequence was ordered as a gBlock® (IDT). To insert this sequence into the chromosome, a
KAC cassette was first inserted 21 bp upstream of the FliM stop codon in the chromosome. The
KAC cassette was subsequently “knocked-out” using the gBlock® DNA. The resulting strain
showed slightly reduced but mostly functional motility consistent with original characterizations
of this fusion protein (10).

f. **Constitutive RFP Marker**

To improve identification of cells in flow cytometry experiments and facilitate segmentation of
cells in time-lapse experiments, the background strain was first engineered to constitutively
express mCherry. An expression cassette was designed from previously published sequences:
this cassette contained (i) a Zeocin resistance marker, (ii) a RNA1 promoter from ColE1 driving
expression of mCherry and (iii) 50 bp sequences homologous to the IntS gene at each end. A
linear DNA encoding cassette was ordered as a gBlock® (IDT). This linear fragment was
chromosomally inserted into the IntS locus. I confirmed that this insertion did not affect the
growth rate of our strains under our experimental conditions or flagellar promoter activity. The
strain harboring the constitutive red marker was designated “MGR” (*MG1655 “Red”*).

g. **Point Mutation of MotA**

Motile cells can swim out of the channels in the “mother machine”. To prevent the loss of cells
expressing flagella from the channels, I engineered a point mutant of MotA, MotA(E98K) (11),
directly in the chromosome using the KAC cassette described above. This point mutation
prevents MotA from powering the rotation of the flagellar motor. I confirmed via flow
cytometry that this point mutation did not affect flagellar gene expression. For time-lapse
studies in the “mother machine”, a MGR strain harboring the MotAE98K mutation (“MGR-E98K”) was used as the background strain.

h. Construction of Transcriptional Reporters

To generate the library of transcriptional reporters, I first generated the cloning vectors pMK4 and pMK7 which were generated from pUA66 (12). In pMK4, the GFP coding sequence and ribosomal binding site (RBS) were replaced by a custom high translation efficiency YFP cassette developed in collaboration with Enrique Balleza. In pMK7, the GFP coding sequence and RBS were replaced by a custom high translational efficiency CFP cassette, also developed in collaboration with Enrique Balleza. In pMK7, the kanamycin resistance cassette and the lambda T0 terminator were also replaced with an ampicillin resistance cassette and a high efficiency terminator from the ilvGEDA regulatory region. In addition to these changes, the original restriction cloning site was replaced with a region flanked by two BsaI recognition sites to allow for insertion of promoter sequences via Golden Gate assembly (13, 14). Both pMK4 and pMK7 were constructed by assembling PCR amplified cassettes into a circular plasmid via isothermal assembly using the NEB HiFi DNA Assembly Master Mix (New England Biolabs)

For each transcriptional reporter, I amplified via PCR a region of the chromosome encompassing the flagellar promoter of interest as well as 30bp of the coding sequences flanking the promoter. This strategy was implemented in case any previously unknown regulator binding sites existed in that region. The PCR primers were designed to contain overhangs that contained BsaI sites as well as target sequences to pMK4 or pMK7 that allowed for the assembly of the PCR product with pMK4 or pMK7 via Golden Gate assembly. The Golden Gate assembly reaction was
performed as previously described, using the higher efficiency protocol of cycling between 3min incubation at 37°C and 4 min incubation at 16°C for 25 cycles (13).

The transcriptional reporters were inserted into one of two chromosomal loci. The first locus was the lambda attB site in the chromosome. The second locus was the GalK. Both loci have previously been used for the insertion of chromosomal transcriptional reporters (6, 15, 16). The two sites are approximately 20kb apart, which is sufficiently far to prevent read-through transcription (this read-through is already low due to the strong rrnB T1 terminator at the end of the YFP cassette) from interfering with the measurements. Simultaneously, the two sites are sufficiently close to each other to limit significant variations in chromosomal copy number during replication. Integration into both sites was performed by amplifying the transcriptional reporter from the plasmid template with primers each containing a 40bp overhang homologous to the target locus. This PCR fragment was inserted chromosomally using the techniques described above.

i. Construction of Class 1 Transcriptional Reporter

Unlike other flagellar promoters, the regulatory region of the Class 1 gene FlhDC is highly complex. Moreover, some evidence suggests that transcription from this promoter may be sensitive to long range interactions (1, 17). Therefore, I decided to examine transcription of Class 1 genes by directly integrating an YFP cassette (with its own RBS) into the 3’ region after the FlhC coding sequence. One complication was the promoter of the neighboring gene MotA which resides within the FlhC coding sequence. To avoid interference from this promoter, I additionally designed synonymous mutations in the C-terminus region of FlhC that would eliminate this promoter. A gBlock® (IDT) containing these mutations, the YFP expression
cassette and homologies to the chromosomal target was ordered. I first inserted the KAC cassette (see above) into the C terminus region of FlhC being mutated. In the second step, I inserted the linear fragment encoding the final desired sequence and used counter-selection to identify successful recombinants. The resulting strain was assayed with a Class 2 transcriptional reporter to confirm that the mutations and insertion of YFP did not significantly alter the underlying flagellar gene expression.

j. Synthetic Promoters Driving FlhDC Expression

The synthetic constitutive promoters Pro3, Pro4 and ProB (18) were constructed by site-directed mutagenesis (Q5® Site-Directed Mutagenesis Kit, New England Biolabs) of a plasmid created by E. Balleza harboring the related promoter ProD (18). The resulting plasmids were used as PCR templates to amplify linear DNA for insertion into the native FlhDC locus. This insertion was mediated by the “scarless” technique described above.

B. GROWTH CONDITIONS

For all experiments, except where otherwise indicated, I used a modified version of the Neidhardt EZ rich media (Teknova), an optically clear rich defined media based on a MOPS buffer (19). I replaced the 0.2% glucose in the original formulation with 0.4% glycerol to prevent catabolite repression of flagellar synthesis.

The MOPS minimal media was identical to this media except for the absence of the amino acids (Supplement EZ, Teknova). LB (Difco) and Tryptone-Broth (Difco) were prepared using standard protocols.
For liquid culture experiments, bacterial cultures were first grown overnight in the growth medium being studied. Subsequently, the overnight cultures were diluted at least 1000 fold in fresh media (see below for specific variations). Cells were grown at 30°C with shaking at 250rpm.

C. DATA ACQUISITION AND ANALYSIS

a. Imaging of Cells on Agarose Pads

Cells were grown as described above to various optical densities corresponding to early through late exponential phase ($OD_{600} = 0.1-0.6$). For cells harvested at low optical densities, cultures were gently centrifuged at $2000 \times g$ for 10 minutes and resuspended in fresh media $1/5^{\text{th}}$ of the original volume. In this work, two different forms of agarose pads were used. The first, without any structural features, was prepared by dissolving 2% (w/v) low melting point agarose (BP165, Fisher Scientific) in fresh growth media and pipetting the mixture onto a glass cover slide (25mm×40mm, No. 1.5, VWR). A second cover slide was added to the top of the agarose which upon cooling resulted in a uniform thin agarose pad. One cover slide was removed and a small volume of cells (typically 1µl) was pipetted onto the pad. Subsequently a new cover slide (25mm×40mm, No. 1.5, VWR) was placed on top and cells were imaged by phase contrast and fluorescence microscopy (see below).

A second type of agarose pad had patterned linear grooves. This feature allowed cells to be better separated from each other facilitating cell segmentation and other image analyses. This patterned agarose pad was prepared by dissolving 3% (w/v) melting point agarose (BP165, Fisher Scientific) in fresh growth media and pipetting the mixture onto a previously designed
PDMS template (20). Upon cooling, the agarose pad was peeled off from the template and 2-3µl of cells was added to the patterned surface. A glass cover slide (25mm×40mm, No. 1.5, VWR) was placed on top and cells were again imaged by phase contrast and fluorescence microscopy (see below).

Phase contrast and fluorescence images were acquired with a Zeiss Axiovert 200M microscope equipped with a Plan-Apochromat 40x/1.3 Oil Ph3 Objective and a CCD camera (Hamamatsu C4742-98-24ERG). For fluorescence excitation I used a solid state LED illumination source, SOLA SE II (Lumencor). Filters with the following specifications were used: for YFP (Ex. 500/24, Di. 520, Em.542/27), CFP (Ex. 438/24, Di. 458, Em. 483/32) and RFP (Ex. 586/20, Di. 605, Em. 647/57). All filters were from Semrock. The variation in fluorescence intensity illumination across the field-of-view was less than 10% in all channels. The microscope setup was controlled using custom software on MATLAB 2013a (Mathworks) interfacing with µManager 1.4.

b. Flow Cytometry

For flow cytometry experiments, an overnight culture was diluted at least 1:3000 into 500µl of fresh growth media. Cells were incubated at 30°C with shaking at 250rpm for at least 6 hours prior to measurement. At this point, the cell density was typically around or below OD=0.1.

Cells were analyzed on a BD Fortessa flow cytometer (Becton Dickinson). The side-scatter (SSC-A) profile was used to first discriminate cells from background particles. Where possible, the constitutive red fluorescence from the mCherry marker was used as a second gate. The raw
flow cytometry data was imported into MATLAB 2015b (Mathworks) and analyzed via custom software.

c. **Time-Lapse Experiments with the Mother Machine**

   (i) **Microfluidic Device Preparation**

   I used an epoxy-based replica of the silicon wafer described in Potvin-Trottier et al (21). The replica mold was a generous gift from Dr. Matthew Cabeen (Harvard University). To prepare a new device, dimethyl siloxane monomer (Sylgard 184, Dow Corning) was mixed with the curing agent at a 10:1 ratio, degassed and poured over the epoxy mold. This mixture was degassed for an additional 1 hour and cured overnight at 65°C before being removed from the mold. Individual devices were first cut from the cured PDMS. Subsequently, inlets and outlets for each flow channel (“lane”) were created using a 0.75 mm biopsy punch. The device was treated with oxygen plasma in a plasma cleaner along with a 25mm×40mm No. 1.5 coverglass (VWR) for 15s at 30W and oxygen pressure 200mTorr. Following plasma treatment, the device and glass were bonded and incubated at least 1 hour at 65°C prior to use.

   (ii) **Cell and growth media preparation**

   For time-lapse experiments, I added a passivating agent Pluronic F-108 (Sigma Aldrich) at a final concentration of 0.85g/L to the growth media as previously described. *E. coli* strains were first grown overnight in media without the passivating agent. Cells were diluted the morning of the experiment 1:100 fold in fresh media, containing the passivating agent and allowed to grow to late exponential or early stationary phase. The reduced cells at this growth phase improved the efficiency with which cells loaded in the device.
The cell culture was loaded into the inlet of the device by pipetting. The device was then centrifuged on a custom adaptor fit into a standard table-top centrifuge at 6000g for 10 mins. The inlets were connected to syringes filled with the growth media (+passivating agent) via Tygon® tubing (VWR, ID 0.02”×OD 0.06”). The flow-through was collected from the outlet via a second Tygon® tubing into an empty beaker. The growth media was first pumped at a rate of 35µl/min for at least 1 hour to allow for the inlets and outlets to be cleared. Afterwards, the flow rate was reduced to 4-5µl/min for the duration of the experiment.

(iii) Imaging Protocol

The cells were allowed to adapt to growth in the device for at least 2 additional hours before imaging. Time-lapse images were acquired on a Nikon Eclipse Ti inverted microscope equipped with a 60× Plan Apo oil objective (numerical aperture (NA) 1.4, Nikon), an Orca R2 CCD camera (Hamamatsu), an automated xy-stage (Ludl) and a SOLA light engine LED excitation source (Lumencor). The microscope was also surrounded by a temperature controlled enclosure. The following filter sets were used for acquisition: YFP (Semrock YFP-2427A), CFP (Semrock CFP-2432A), and RFP (Semrock mCherry-A).

All experiments were performed at 30 °C. Automated time-lapse acquisition was controlled using custom MATLAB (Mathworks) scripts interfacing with μManager. Images were taken every 10 minutes and focal drift was corrected via the Nikon PerfectFocus system and periodically recalibrated using z-stacks on a sacrificial position. Images in the RFP (the cell segmentation marker) were acquired at full camera resolution (1344×1024 pixels) to improve segmentation while CFP and YFP images were acquired using 2×2 binning to reduce
measurement noise. Short exposure times (typically 200-300ms) and low illumination intensities (<30% of maximum illumination power) were used to minimize the effects of photobleaching.

(iv) Image Quantification and Lineage Tracking

For each experiment, individual cells were identified and segmented using custom software implementing previously described algorithms (21-23). Parameters for each segmentation step were optimized based on morphological characteristics of the cells and the typical fluorescence intensities in the segmentation channel. Once “masks” defining the boundaries of individual cells were determined, images from the YFP and CFP data channels were first aligned to the RFP segmentation channel to correct for any mis-registration between the data channels.

The cells at the bottom of the growth channels (i.e. the “mother cell”) remained virtually in the same position from frame-to-frame. In some cases, there was a small drift in the field-of-view over the course of the experiment. However, because this drift was relatively small, it could easily be corrected by re-aligning each frame based on the location of the linear colony. Subsequently, the mother cell lineage could be easily identified by following cells closest to the “mother cell” in the previous frame. Potential “cell divisions” events were first identified by sudden decreases in cell area, i.e. if a cell’s area dropped to less than 60% of its current value in the next frame. Manual review revealed that most lineages were properly constructed. Occasionally, however, due to mis-segmentation, the cell area increased or decreased precipitously. A “voting mechanism” similar to the algorithm described in (24) was used to correct these errors. Interestingly, measurements of mean fluorescence were surprisingly robust to these errors.
Each lineage was tracked to the end of the experiment or until the mother cells were “lost” from the channels. For simplicity, only lineages that survived to the end of the experiment were analyzed—the inclusion of shorter lineages in select cases did not substantially alter my results. Occasionally, cells in the tracks grew filamentously or stopped growing altogether. These lineages were discarded from my analysis.

d. **Time-Lapse Data Analysis**

All time-lapse data were analyzed using custom MATLAB software.

(i) **Measuring Duration of “On” and “Off” States**

To define “on” and “off” states for each promoter, I defined a heuristic threshold by applying the MATLAB function `multithresh` which implements Otsu’s method (25), a technique commonly used for discriminating foreground and background pixels in images. This method gave thresholds which were similar to “visually intuitive” thresholds. Varying the threshold several fold above and below this method gave qualitatively similar results. The only exception was Class 1 genes which had a narrow range of fluctuations and no “intuitively” obvious “on” or “off” period. As a consequence, the `multithresh` algorithm gave a threshold value similar to the mean fluorescence value. For simplicity, in this work, I use the “high” and “low” fluorescence periods of Class 1 expression interchangeably with “on” and “off” periods.

Time periods when fluorescence was continuously above the threshold was defined as “on” periods. Occasionally, fluorescence values transiently peaked just above the threshold before dropping back down. If both the duration was sufficiently short (less than 2 time points, i.e. 20 mins) and the peak magnitude of fluorescence during this time was low (within 10% of the
threshold value), the “on” period was ignored. Similarly, “off” periods were defined as contiguous time periods with fluorescence values below the threshold. Again, if the duration was sufficiently short (less than 2 time points) and the minimum fluorescence value was within 10% of the threshold value, the “off” period was ignored. Once these two corrections were made, the duration of the remaining “on” and “off” periods were re-computed.

(ii) Cross-correlation Analysis Between Fluorescence Time-Traces

The cross-correlation between two promoters was estimated by averaging the cross-correlation from individual time traces. Two alternate methods were explored. In the first method, for each lineage, the cross-correlation between the YFP and CFP time traces was computed with the MATLAB function \textit{xcov} using the ‘coeff’ option which first subtracts the mean of each time trace and then normalizes the cross correlation by the standard deviations of each time trace. These normalized cross-correlation values were then averaged together. In the second method, I first determined the mean YFP and CFP values for all the traces within the same experiment. For each lineage, I subtracted the global mean YFP value from the YFP time trace and the global mean CFP value from the CFP time trace. I then applied the MATLAB function \textit{xcorr} to these two time traces (which omits the subtraction of the local time trace mean) using the ‘unbiased’ option. The resulting outputs for different lineages were averaged and then normalized by the global standard deviations of YFP and CFP. The two methods gave largely very similar cross-correlation functions.

(iii) Plotting the “Dose-Response” Relationship
The cross-correlation analysis between Class 1 and Class 2 genes indicated that the peak correlation occurred when the two time traces were offset by a “lag” of 50 minutes. To account for this lag, the “dose-response” relationship was constructed by comparing the fluorescence value of the Class 1 time trace at time $t$ with the fluorescence value of the Class 2 trace at a time point 50 minutes afterwards (i.e., at $t+50$ mins). Similarly, to determine the dose-response relationship between Class 2 and Class 3 genes, I compared the fluorescence value of Class 2 genes at each timepoint $t$ with the fluorescence value of Class 3 genes 30 minutes afterwards (i.e. at $t+30$ mins). Interestingly, scatterplots without this “lag-correction”, i.e. a direct comparison of fluorescence values from the same timepoint, gave very similar results, likely because each “on” or “off” state persisted for several generations. The density of the scatter plot was estimated by the MATLAB $histogram2$ function.
References


