Multiple Unnecessary Protein Sources and Cost to Growth Rate in E.coli

Multiple Unnecessary Protein Sources and Cost to Growth Rate in *E.coli*

A dissertation presented

by

Luke Julien Bruneaux

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

May 2013

Thesis advisor | Author

Professor Melissa Franklin | **Luke Julien Bruneaux**

Multiple Unnecessary Protein Sources and Cost to Growth Rate in *E.coli*

Abstract

The fitness and macromolecular composition of the gram-negative bacterium *E.coli* are governed by a seemingly insurmountable level of complexity. However, simple phenomenological measures may be found that describe its systems-level response to a variety of inputs. This thesis explores phenomenological approaches providing accurate quantitative descriptions of complex systems in *E.coli*.

Chapter 1 examines the relationship between unnecessary protein production and growth rate in *E.coli*. It was previously unknown whether the negative effects on growth rate due to multiple unnecessary protein fractions would add linearly or collectively to produce a nonlinear response. Within the regime of this thesis, it appears that the interplay between growth rate and protein is consistent with a non-interacting model. We do not need to account for complex interaction between system components.

Appendix A describes a novel technique for real-time measurement of messenger RNA in single living *E.coli* cells. Using this technique, one may accurately describe the transcriptional response of gene networks in single cells.

Contents

Citations to Previously Published Work

Appendix A is largely reproduced from the following paper:

> "Minimally invasive determination of mRNA concentration in single living bacteria", Guet, C. C., Bruneaux, L., Min, T. L., Siegal-Gaskins, D., Figueroa, I., Emonet, T. and Cluzel, P.; Nucleic Acids Research, 36, 12 (2008);

Acknowledgments

I have had the great fortune to work with a number of excellent scientists in the lab of Philippe Cluzel and at the Bauer Center for Systems Biology, including Arvind (Rasi) Subramaniam, J. Mark Kim, Enrique Balleza, Lisa Marshall, Calin Guet, Panos Oikonomou, Israel Figueroa, and Beverly Neugeboren and Bodo Stern. This work would not have been possible without the determined support of my thesis committee of Professors Mara Prentiss, Erel Levine, David Nelson and Melissa Franklin. I greatly appreciate all the time and effort they devoted to the process and the education they provided.

I owe a debt of gratitude to Bruce Ennis for introducing me to physics, Prof. Phil Nelson for his biological physics course at Penn, Prof. Mark Devlin and Jeff Klein for showing me how experiments are done. Thanks to Prof. John Huth, Dr. David Morin, Dr. Margo Levine and Anne Marie Calareso for giving me the chance to serve the Harvard community, and Ellen Fox and Janet Sand for guiding me through.

I would like to thank my loving mother and father, Cynthia and Gilles Bruneaux, for their encouragement and support over the many years of school, as well as my siblings David, Mathieu, Judith and Bess for taking me under their wings. Many thanks to Bob, Rose and Betsy Shor for spiritual guidance and a bottomless well of kindness. Thanks to Christopher Semisch, Rafael Jaramillo, Sarah Wood, Ben Costa and Conor Allen for the many helpful discussions. Kevin Wood will have my eternal gratitude for the countless conversations and unwavering support over

the last five years. And finally, infinite gratitude to my wife Rebecca, whose love, dedication and inspiring example I appreciate more than anything in this world.

Dedicated to Rebecca

List of Figures

Chapter 1

Multiple Unnecessary Protein Sources Combine Additively to Reduce Growth Rate in *E.coli*

1.1 Abstract

The relationship between unnecessary protein production and growth rate in bacteria has been of interest since Jacob and Monod's classic studies inducing the production of β-galactosidase in *E.coli* [43]. Multiple sources of growth rate reduction have been suggested in connection to protein production, including ribosome destruction [15], diversion of metabolites [58], competition among mRNA for ribosomes [63], and protein specific toxicity [16]. The significance of reduced

growth rate in an evolutionary framework has led to cost-benefit models of protein production [13, 16]. Recently, *Scott et al* proposed a phenomenological model describing the linear relationship between fraction of the proteome devoted to unnecessary protein and the reduction of growth rate[56]. Entering this study, it was unknown whether the negative effect of unnecessary protein from multiple sources would combine to produce a nonlinear reduction of growth rate. By inducing production of YFP, flagellin and β-galactosidase in combination, we present evidence that the addition of two unnecessary protein burdens combine additively in their negative effect on growth rate. Within the regime studied here, we find no evidence that the effects of multiple sources add nonlinearly. However, there may be regimes outside of the scope of this study in which a nonlinear relationship exists.

1.2 Introduction

The relationship between cell composition and growth rate of *E.coli* has been examined since the mid-20th century in studies widely ranging in focus and approach. The components of the cell have been probed experimentally in many studies: monitoring growth [43], protein production[44], DNA replication [38], translation inhibition [24], gratuitous overexpression [15], evolutionary tuning of expression [13], titration of ribosomes [58], bistability as a result of growth reduction [60] and the entire biochemical composition of the cell as a function of growth rate [8]. Following the development of a mechanistic models [29, 54, 59],

the Hwa group presented a coarse-grained model relating the growth of cultures to the macromolecular makeup of the cell [56] reducing the number of relevant parameters to a handful of easily measurable quantities. They also propose a model suggesting that the production of unnecessary protein reduces growth rate in a linear fashion, describing the effect of production of β-galactosidase in *E.coli*.

While this model accurately predicts the effect of a single source of unnecessary proteins, it is not clear how multiple sources of unnecessary protein will affect the growth rate of cells. Multiple inputs produce strong nonlinearities in many systems of *E.coli*: antagonistic and suppressive drug interactions [7], autoregulation in gene expression[1], and epistasis of deleterious mutations [18]. However, efforts to simplify these interactions with linear models such as Bliss Independence[6] have shown that in many cases two inputs can combine additively. In order to properly explain the phenomenological model explored in this study, we will briefly delineate the historical basis for parameter values of the macromolecular composition of the cell.

1.2.1 Historical Background

In 1958, Moselio Schaechter, Ole Maaløe, and Neils Kjeldgaard demonstrated that the chemical composition of the cell depended on the overall growth rate of *E.coli B/r*, independent of media composition. They showed that at a given temperature the average mass, RNA and DNA per cell were exponentially re-

lated to the growth rate. In their estimation, for doubling times from 20 to 120 minutes, the protein mass per cell is roughly proportional to the RNA mass per cell, irrespective of growth rate. They studied two kinds of exponential growth: "balanced" growth and continuous culture growth.

In balanced growth, cells grew in large, aerated liquid cultures and were diluted 1:2 with fresh media with a period corresponding average generation time. The optical density at a wavelength of 420 nm (OD_{420}) was kept between 0.2 and 0.4, and no more than 0.8, which corresponds to $1.2 - 6.4 \cdot 10^8$ bacterial/mL and 140 μg /mL bacterial dry weight. The researchers also measured the composition of the cells at two temperatures: 25°C and 37°C . They found that at the lower temperature, the relationship between growth rate and mass, RNA and DNA per cell remained exponential but with different parameters. For a given nutrient medium, the growth rate at 37°C is close to double of that at 25°C .

They concluded that the composition of the cell is not dependent on the components of the media in which it grows, but rather on the overall effect on growth rate. However, it is not clear that cells growing at a specific division time in rich media at 25°C are of the same composition as those growing at the same division time in poor media at 37°C . Of course, the exponential relationship of each quantity is of major significance at well. As Cooper says in [11], the work was the first to demonstrate that there is no obligatory life cycle, but that cell size is determined by growth rate.

Ten years after the publication of the Schaechter-Maaløe-Kjedgaard experiments, Stephen Cooper and Charles Helmstetter described the C and D periods of the cell cycle in order to relate chromosome replication to DNA content in the cell. During the C period, replication begins at the origin of replication and stops at the terminus. The D period simply described the period of time between the end of replication and cell division. The startling thing about the Cooper-Helmstetter experiments was that C and D are relatively fixed quantities across a wide range of growth rates ($20 < \tau < 60$ minutes), rather than a function of division time.

In order to study chromosome replication, the two developed a method to synchronize bacterial cultures by cell division. Bacterial cells grow on a membrane such that when they divide, daughter cells are eluted off the membrane into a flask below. Cell culture is sampled over time, yielding measurements of cell composition as a function of cell cycle. Later, this device would be referred to informally as the "baby-machine".

Helmstetter developed the technique to measure DNA concentration by pulse-labeling exponentially growing cultures with thymidine and measuring the radioactivity as a function of dilution. They saw distinct initiation and termination moments as a function of cell cycle. The period of chromosomal replication was essentially constant ($C = 40$ minutes) for all observed doubling times. Similarly D was relatively constant at about 20 minutes. They found the point of initiation,

rather than being decoupled from the cell cycle, was reliable for a given doubling time. Since the doubling time of fast-growing cultures is shorter than 40 minutes, cells will carry chromosomes with multiple initiation forks, and hence, an average of more than two copies of the chromosome.

Cooper and Helmstetter released a model alongside their experimental findings in [12]. In the model, replication activity is fixed for the cell cycle of a cell growing at doubling time τ. For simplicity, the C period is 40 minutes, and D 20 minutes. A slow-growing cell (doubling time $\tau = 60$ minutes) begins initiation at cell division, terminates replication at 40 minutes, and has a 20 minute period without activity. As a result, the cell has one chromosome at age= 0 and two chromosomes at age= 1, when it divides again.

The story is more complicated for faster-growing cells. For example, when $\tau = 40$ minutes, the cell is in the middle of replication at age= 0 (thus containing 1.5 chromosomes), terminates at minute 20 and immediately initiates the next round of replication, such that the cell contain two chromosomes and two half-chromosomes at age= 1. The constraint that cell division at age= 1 must reproduce the initial chromosome copy number at age= 0 fixes, in addition to the constant C and D periods, leaves only one solution per τ. Cells that are growing at $\tau = 20$ minutes will begin the cycle with three chromosomal equivalents and divide again with six. On the heels of the Cooper-Helmstetter Model, William Donachie published a simple model describing overall cell mass as a function of di-

vision time [14]. In an exponentially growing population, the rate of mass increase is given by Equation (1.1),

$$\frac{d\,M}{dt} = kM = \frac{M}{T} \tag{1.1}$$

where M is the cell mass, k is the growth rate and T is the doubling time of the cell. The total mass of the population at time t is given by (1.2). Here we've modified the exponential base in Donachie's original equations to 2 to reflect doubling as a marker for growth.

$$M(t) = M(t=0) \cdot 2^{t/T} \tag{1.2}$$

Single cells growing exponentially have a mass at division, M_d, given by (1.3). This depends on the number of origins of replication in the cell, as each origin represents a cell fraction doubling a the given rate.

$$M_d = \frac{M_i}{N_i} \cdot 2^{(C\,+\,D)/T} \tag{1.3}$$

where M_i is the mass at initiation, and N_i is the number of origins at initiation.

$$\log_2 M_d = \log_2 \frac{M_i}{N_i} + \frac{(C+D)}{T} \tag{1.4}$$

Thus, the log of the mass at division will vary as the growth rate provided the mass per origin at initiation is constant. The model predicts that cells dividing at $\tau = 20$ minutes will be four times as massive as those with $\tau = 60$ minutes.

In their chapter of *E.coli and Salmonella* [8], Hans Bremer and Patrick Dennis expand the Cooper-Helmstetter model to describe the cell's biomolecular composition. Using [55], they define the growth rate of the cell as depending on two

parameters, c_p (the rate of peptide chain elongation) and β_r (the fraction of total ribosomes actively engaged in peptide chain elongation) 1.5.

$$\mu = ({}^{60}/_{ln(2)}) \cdot ({}^{N_r}/_{P}) \cdot \beta_r \cdot c_p \tag{1.5}$$

The model that follows then predicts the levels of DNA, mRNA, protein, RNA polymerase, tRNA, etc. for doubling times $\tau < 100$ minutes. The DNA per cell(G_p), RNA per cell(R_p) and protein per cell(P_c) are given by 1.6c.

$$P_C = P_o \cdot 2^{(C+D)/\tau} \tag{1.6a}$$

$$R_C = K'({}^{P_0}/_{c_p})({}^{1}/_{\tau})2^{(C+D)/\tau} \tag{1.6b}$$

$$G_C = \frac{\tau}{C \cdot ln(2)} \cdot (2^{(C+D)/\tau} - 2^{D/\tau}) \tag{1.6c}$$

In this equation $K' = {}^{nucl.}/_{rib.} \cdot ln(2)/[f_s \cdot (1 - f_t) \cdot \beta_r \cdot 60]$, where f_s is the fraction of total RNA that is stable RNA and f_t is the fraction of stable of RNA that is tRNA. The total mass of the cell is simply the linear addition of measured optical density values per amino acid residue, RNA nucleotide residue, and genome equivalent.

$$M_C = k_1 \cdot P_C + k_2 \cdot R_C + k_3 \cdot G_C. \tag{1.7}$$

The model predicts that faster growing cells will have a higher proportion of RNA relative to protein than slower growing cells, given by a linear relationship between the ratio and ${}^{1}/_{\tau}$. This is largely tied to ribosome number since at fast doubling rates the majority of the RNA in the cell is ribosomal rRNA. Not surprisingly, the level of mRNA to protein is relatively constant across many doubling times, and the relative amount of DNA per protein decreases as well (see Figure 1.1).

Bremer and Dennis assert that these general relationships hold for cells growing at different temperatures. Under varying temperature, the C period and doubling time τ change proportionally to one another. For E.coli B/r, the relationship between temperature and growth rate is linear.

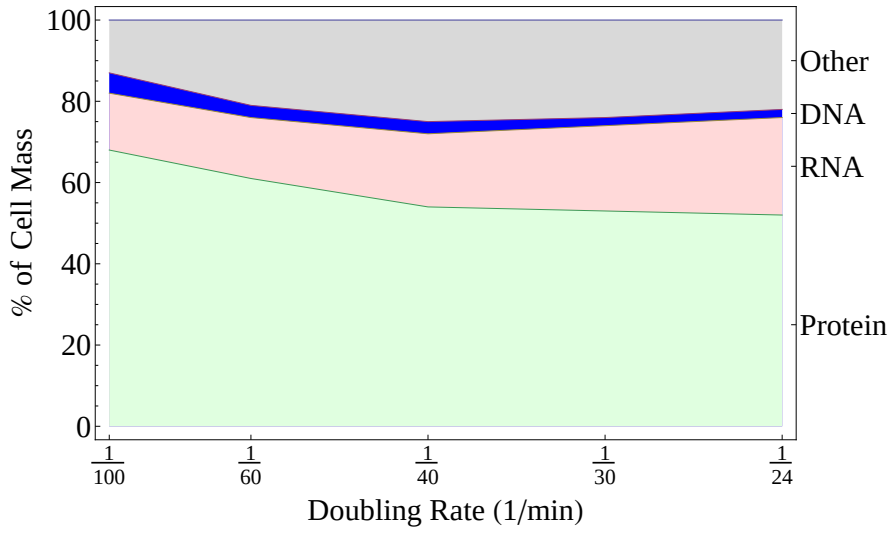


Figure 1.1: Cell composition as a function of growth-rate. As growth rate increases, the relative amount of protein and DNA per cell decreases and RNA increases, mostly due to increased ribosome count: as much as four-fifths of the RNA corresponds to ribosomal rRNA. The remaining section (Other) includes translation, initiation and elongation factors, subunits of RNA polymerase, and the aminoacyl-tRNA synthesases. Figure is modified from [8].

1.2.2 Growth Laws

The Hwa model for cell growth presented in [56] simplifies complex molecular models by using the strong correlation between growth rate and RNA content (and therefore ribosome content). They base this assumption upon the results of the Schaechter-Maaløe-Kjedgaard experiments, as well those by Mikkola and Kurland [42]. The model assumes for doubling times $20 < \tau < 120$ minutes that all ribosomes are actively translating. The proteome of the bacteria is divided into

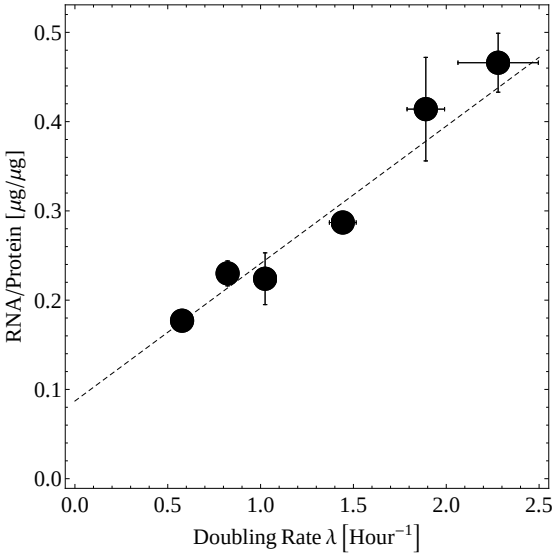


Figure 1.2: RNA/protein is linearly related to growth rate under varying nutrient levels. Data points reflect cells growing in a variety of nutrient conditions ranging from minimal media (low growth rate) to rich defined media (high growth rate), and varying carbon source (glycerol or glucose). RNA/protein is also negatively related to growth rate under translation inhibition (not shown). Figure is modifed from [56].

three classes of protein labeled P, Q, and R. Roughly speaking, R corresponds to the "extended ribosome", i.e. mass of ribosomal proteins and affiliated proteins, Q to the fixed proteins necessary for reproduction (metabolism, cell membrane, etc.) and P to the rest of the proteome. Experimentally the proteome devotes a fixed sector Φ_{Q} of about one half, and that the maximum ribosomal component Φ_{R}^{max} is also roughly half the proteome, which Φ_{R} approaches with decreasing growth rate. As growth rate increases,the ribosomal fraction decreases, and the third section Φ_{P} increases.

$$\Phi_{\mathrm{R}}^{max} \approx 55\% \tag{1.8a}$$

$$\Phi_{\mathrm{Q}} = 1 - \Phi_{\mathrm{R}}^{max} \tag{1.8b}$$

$$\Phi_{\mathrm{P}} = \Phi_{\mathrm{R}}^{max} - \Phi_{\mathrm{R}} \tag{1.8c}$$

At most growth rates, $\Phi_{\mathrm{R}} < \Phi_{\mathrm{R}}^{max}$, while Φ_{Q} remains constant, leaving a third proteome fraction Φ_{P}. The mass of each of these proteome pieces is given by M_i such that $M_{\mathrm{P}} + M_{\mathrm{Q}} + M_{\mathrm{R}} = M$ where M is the total mass of the proteome. Likewise the fraction of ribosomes devoted to synthesizing each class given by f_i, such that $f_{\mathrm{P}} + f_{\mathrm{Q}} + f_{\mathrm{R}} = 1$. Cells growing at an exponential rate, such that $M_i(t) = M_i(0) \cdot e^{\lambda t} = M_i(0) \cdot 2^{\mu t}$, resulting in relationships 1.9, which describe equalities between proteome fractions and fraction of ribosomes devoted to their synthesis.

$$\lambda = f_{\mathrm{R}} \cdot \frac{k}{m_{\mathrm{R}}} \tag{1.9}$$

Figure 1.3: The proteome partitioning model wherein R represents ribosome-affiliated proteins, Q the fixed protein fraction unaffected by translation inhibition, and P the remaining useful proteins. The relative proportion of P-protein increases with growth rate. Figure is modifed from [56].

The ribosomal mass fraction $\Phi_{\mathrm{R}} = M_{\mathrm{R}}/M$, using equation 1.9, is equal to two quantities

$$\Phi = \frac{\lambda}{k/m_{\mathrm{R}}} = \rho \cdot r \tag{1.10}$$

where the conversion factor $\rho = 0.76^{\mu\mathrm{g\ extended\ ribosome\ protein}/\mu\mathrm{g\ RNA}}$. The correlation between RNA/protein ratio r and growth rate λ includes an extrapolated intercept at zero growth rate, $r_0 > 0$ since at very slow growth not all ribosomes are actively translating.

$$r = r_0 + \frac{\lambda}{\kappa_t} \tag{1.11}$$

where κ_t is defined as the "translational capacity", $\kappa_t = \rho \cdot k/m_{\mathrm{R}}$, which can be rewritten to estimate the elongation rate $k = 20.1^{\mathrm{a.a.}/\mathrm{sec.}}$. Using equation 1.8c, a

second linear relationship between growth rate and RNA/protein mass results:

$$\Phi_{\mathrm{R}} = \Phi_{\mathrm{R}}^{max} - \Phi_{\mathrm{P}} = \Phi_{\mathrm{R}}^{max} - \frac{\rho\lambda}{\kappa_n} \tag{1.12a}$$

$$r = r_0 - \frac{\lambda}{\kappa_n} \tag{1.12b}$$

The combination of two linear equations amounts to an analogy of Ohm's law applied to two resistors in a series. One could imagine the cell's proteome as an electric circuit with a battery of voltage $\Phi_R^{max} - \Phi_0$ producing a current λ. Two two resistors correspond to the P section and the R section with respective resistance values ρ/κ_n and ρ/κ_t. The resulting growth rate is given by

$$\lambda = \frac{\Phi_{\mathrm{R}}^{max} - \Phi_0}{\rho} \frac{\kappa_n \kappa_t}{\kappa_n + \kappa_t} \tag{1.13}$$

With this relationship, the Hwa model predicts the growth rate of any liquid culture depending on two quantities: the nutritional capacitance (a function of the nutrient concentration of the media) and the translational capacitance (affected by the presence of translation-inhibiting antibiotics, for example).

1.2.3 Unnecessary protein expression in the Hwa Model

In order to explain the relationship between growth cost and gene expression, the Hwa model introduces a fourth mass fraction Φ_U representing the unnecessary protein mass fraction such that $\Phi_U = \Phi_R^{max} - \Phi_R - \Phi_P$. The production of unnecessary protein mass fraction Φ_U causes an effective reduction of r_{max} to

$r_{max} - \Phi_U/\rho$, since the new growth rate

$$\lambda = \frac{\Phi_R^{max} - \Phi_0 - \Phi_U}{\rho} \frac{\kappa_n \kappa_t}{\kappa_n + \kappa_t} \tag{1.14}$$

can be effectively rewritten as

$$\lambda = \lambda_0 \left(1 - \frac{\Phi_U}{\Phi_R^{max} - \Phi_0}\right) \tag{1.15}$$

This yields a negative linear relationship between unnecessary protein expression and growth rate, which can be more simply rewritten as

$$\lambda = \lambda_0 \left(1 - \frac{\Phi_U}{\Phi_C}\right) \tag{1.16}$$

where $\Phi_C = \Phi(r_{max} - r_0) \approx 0.48$. This relationship fits reasonably well to classic β-gal experiments, such as [15](see Figure 1.4).

1.2.4 Multiple sources of unnecessary protein

The relationship between unnecessary protein and growth rate has been tested for single protein sources (e.g. β-galactosidase or β-lactamase). The source of the unnecessary protein is not significant to the model, which implies that all extraneous protein may be described in a generalized "U" section of the proteome. The model does not explicitly predict the cumulative effect on growth rate of multiple sources of unknown protein (simultaneous production of β-galactosidase and β-lactamase, for example). The null hypothesis of this study is that the effect of multiple protein sources will be additive, such that

$$\lambda(\Phi_{\mathrm{U,i}}) = \lambda_0 \left(1 - \frac{\sum_i^n \Phi_{\mathrm{U,i}}}{\Phi_{\mathrm{C}}}\right) \tag{1.17}$$

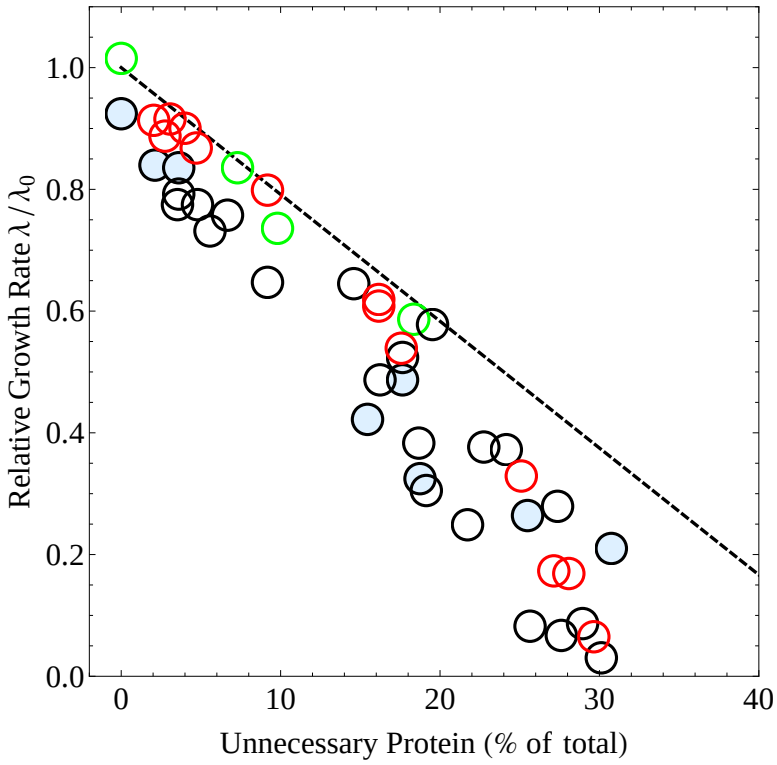


Figure 1.4: Relative growth rate vs. unnecessary protein fraction in *Scott et. al.* The light blue circles represent *E.coli* growth with β-galactosidase, clear circles represent historical data: β-galactosidase (red circles) and Δ-EF-Tu (black circles)[15] and β-lactamase [2]. The dashed line is the linear model for growth rate under the load of unnecessary protein synthesis, with the empirical estimate$\Phi_C = 0.48$ (modified from[56]).

In order to test this hypothesis, we will introduce first yellow fluorescent protein (YFP) as an unnecessary protein, which to our knowledge has not been directly tested with respect to the linear Hwa model. As a second unnecessary protein, we will modify our wild-type cells in order to produce additional flagella.

In low-nutrient environments *E.coli* cells use flagella to propel themselves towards more nutrient-rich environments using a combination of swarming and tumbling (depending on the rotation direction), controlled by the chemotaxis network [50]. The flagellum is also used in processes such as adherence to host cells and biofilm formation [25]. The number of flagella per cell typically depends on the environment, which is not surprising given the significant energy demands of assembling and operating the device [50]. The flagellum is composed of three primary pieces (the basal body, hook and filament) and uses a rotational motor and type III secretion system to export proteins [25]. The assembly of the rotational motor is modular in nature [3] and is regulated by a cascade of fourteen related operons [30, 32, 53]: a master regulator pflhD, a second class included in the assembly of the motor, and a third class III, which initiates chemotaxis $motA$ and tar. In this study, we use the pfliL promoter as a reporter for overall promoter activity, and measure flagellin directly using protein immunoblotting technique. We anticipate that the resulting unnecessary protein fraction due to both YFP and extraneous flagellin will reduce growth rate in the linear fashion described by 1.24.

1.3 Results

1.3.1 Relative growth rate as a function of unnecessary yellow fluorescent protein

We first tested the hypothesis that an extrinsic reporter protein YFP would produce a reduction in growth rate in a fashion similar to β-galactosidase in [56]. We inserted YFP-reporter plasmids into the WT strain MG1655 (referred to in this study as the *sequencing* strain) and the hyper-motile but otherwise identical MG1655+IS5 (referred to in this study as the *insertion* strain). In each strain, a promoterless reporter plasmid was used as a control for the pfliL reporter plasmid reporting on the flagellum assembly cascade. The cells were grown in a robustly shaking flask in media with either glucose or glycerol as a carbon source, and at either 30°C or 37°C , and their growth rates were recorded. In each case, the growth rate of the promoterless *sequencing* strain was treated as the optimal growth rate to which the remaining strains-plasmid combination samples were compared.

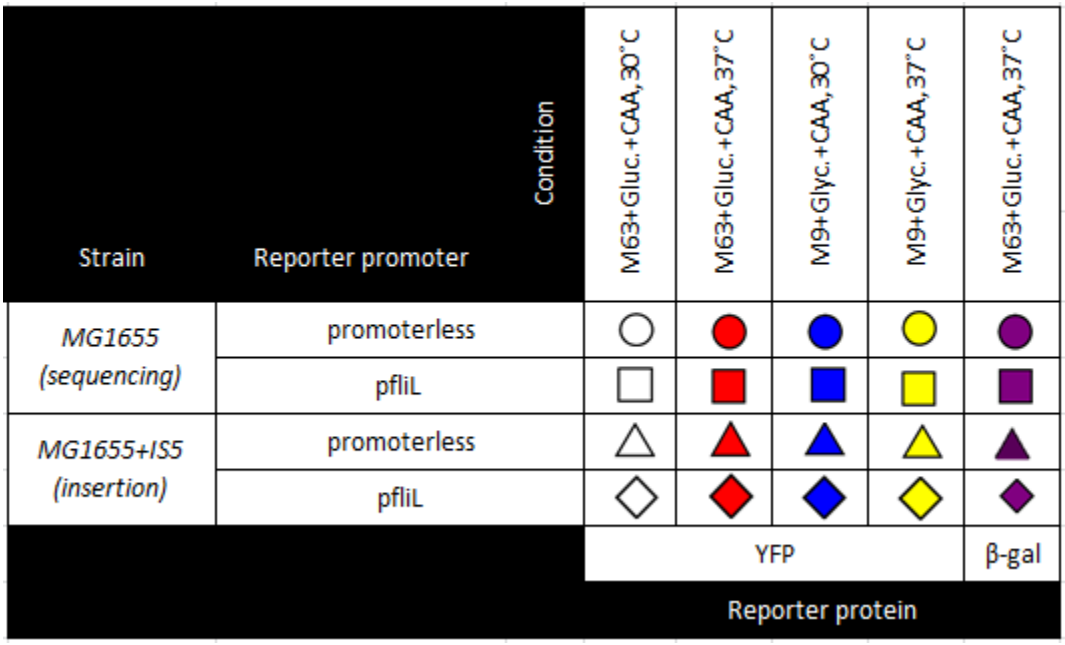


Figure 1.5: Legend of samples taken in this study. For each media and temperature, the *sequencing* strain MG1655 and *insertion* strain MG+IS5 were grown in a large flask of rapidly shaking culture. The strains were tested reporting on flagellum assembly promoter pfliL and a promoterless control using YFP as a reporter protein.

The relative YFP amount in a particular sample varied with temperature, carbon source, and strain. In the wild-type *sequencing* strain MG1655, the plasmid reporting the pfliL reporter produced comparable amounts of YFP to the promoterless control ($\approx 10-60\mu M$ and $\approx 10-20\mu M$ respectively), indicating that transcriptional activity of the flagellum assembly network was minimal in WT cells. However, in the *insertion* strain MG+IS5, YFP reporting on pfliL ranged

from 270μM (30°C , glucose, CAA) to 800μM (37°C , glycerol, CAA), indicating strong activity at the transcriptional level of the flagellum assembly cascade. Overall,*insertion* strains exhibit 20 to 40 times more activity than the corresponding WT cells.

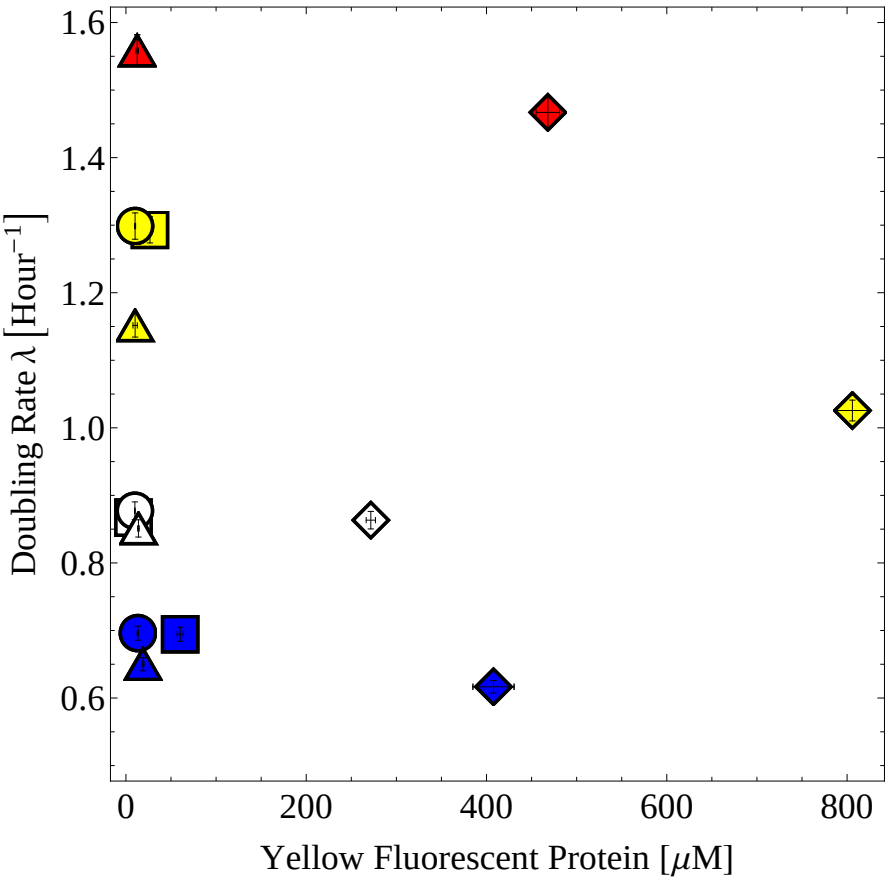


Figure 1.6: Doubling time vs. Yellow Fluorescent Protein concentration. Within a given condition, the promoterless *sequencing* strain doubled at the fastest rate, producing very little YFP. In some conditions, the *insertion* strain doubled at rates similar to those of the *sequencing* strain for both reporter plasmids. However, in some conditions, the *sequencing* strain demonstrated a marked decrease in growth rate corresponding to high YFP concentrations. In general cells grew faster at 37°C than 30°C , and faster with glucose as a carbon source than glycerol. The amount of YFP reporting was affected by condition as well, reflecting regulation of the flagellum cascade.

Within a given condition, the growth rate of the WT *sequencing* strain was similar for both the pfliL and promoterless variants, suggesting that the relatively low amount of YFP was not significant enough to reduce growth rate to a significant amount. It also suggests that the act of reporting at the transcriptional level does not significantly perturb the cells, which has been of concern using reporter plasmids [2]. However in several conditions, the pfliL variant of the *insertion* strain grew significantly slower than its promoterless counterpart, suggesting that YFP indeed affects growth rate at high concentrations (Figure 1.6).

We normalize the growth rates by the optimal growth rate of the promoterless variant within each strain in order to compare relative growth rates across conditions. We also convert the absolute YFP concentration into a fraction of the proteome using total protein content (see section 1.3.2). Figure 1.7 demonstrates that the growth rate is reduced in proportion to the fractional amount of YFP, even across temperature and carbon-source. At first glance, the relationship appears consistent with the linear model described in subsection 1.2.3,

$$\lambda\left(\Phi_{\mathrm{U}}\right)=\lambda_0\left(1-\frac{\Phi_{\mathrm{U}}}{\Phi_{\mathrm{C}}}\right) \tag{1.18}$$

treating YFP as "unnecessary" protein fraction Φ_{U} and using the established empirical parameter $\Phi_{\mathrm{C}}=0.48$.

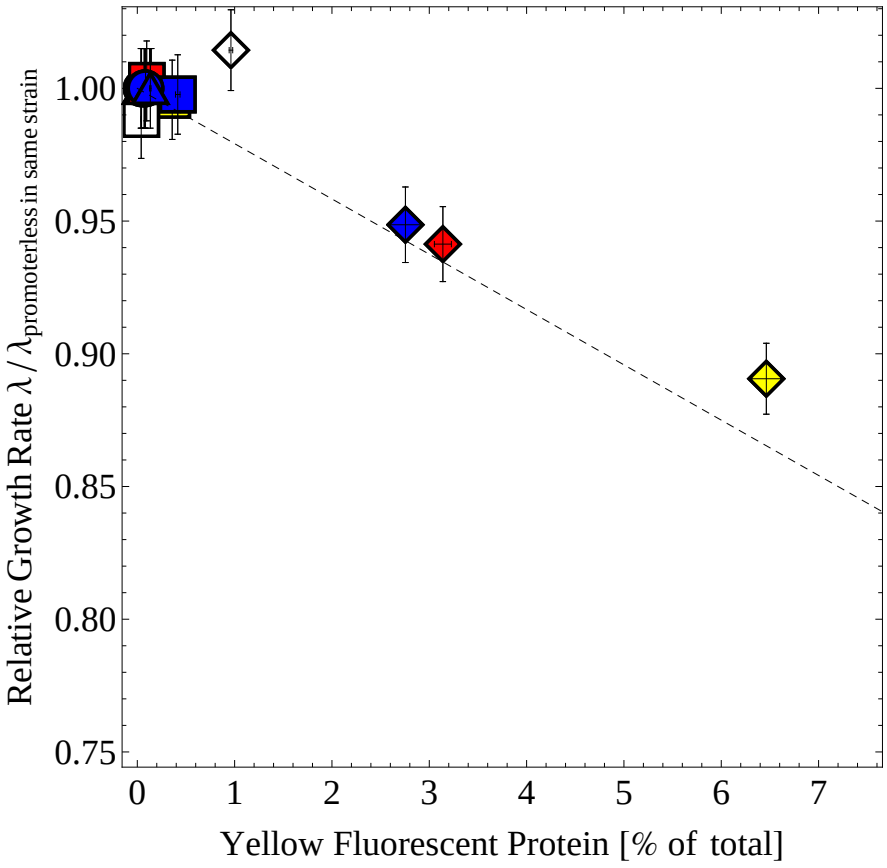


Figure 1.7: Relative Growth Rate (within strain) vs. YFP proteome fraction Φ_F. The fraction of the proteome devoted to YFP and the growth rate varies in the pfliL variant of the *insertion* strain with condition. Comparing the growth rates of the pfliL variant to the promoterless variant of either the *sequencing* or the *insertion* strain, the relative cost to growth does not appear to deviate significantly from the linear model (equation 1.18,dashed line).

However, if we compare the *insertion* strain directly to the *sequencing* strain, normalizing all growth rates to the promoterless variant in the *sequencing* strain, and plot with respect to YFP proteome fraction Φ_F, we find that the reduction in

growth rate cannot be explained by the model using $\Phi_U = \Phi_F$ alone (Figure 1.8).

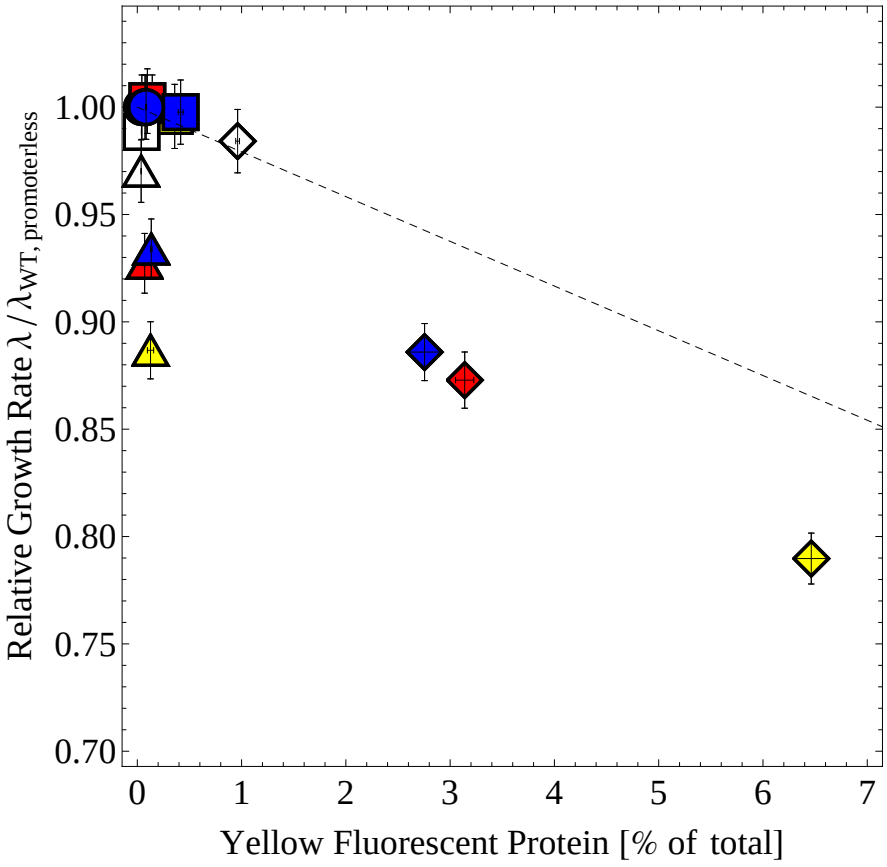


Figure 1.8: Relative growth rate (across strains) vs. YFP proteome fraction Φ_F. If we compare the *insertion* strain to the *sequencing* strain, the difference in relative growth rate cannot be accounted for by YFP proteome fraction Φ_F alone. The *sequencing* and *insertion* strains are identical except for an insertion in the flagellum assembly cascade, for which activity is reflected in YFP reporting of the pfliL promoter. Higher flagellum count is suspected as the additional burden in the *insertion* strain.

The comparison suggests than another protein may be considered unnecessary

in the effect on growth rate. To make the direct comparison, we argue that the *insertion* strain is identical to the *sequencing* strain in macromolecular composition except for the insertion of an element into the regulation of the flagellum assembly cascade. Therefore the difference in flagellum count between the two strains may account for an additional source of unnecessary protein. In order to validate the direct comparison between strains, we measured the RNA and protein composition and compared the results to those described in past empirical studies.

1.3.2 Macromolecular composition of the cell

In order to properly compare the *sequencing* and *insertion* strain in the context of the model outlined in subsection 1.2.3, we verified that the macromolecular composition of strains used in this study were similar to one another and to those in previous studies. RNA and protein were extracted and measured from identical samples, and their ratio plotted as a function of growth rate. The RNA/protein ratio of strains growing in multiple conditions is consistent with those presented both in [56] and [8]. In general, faster growing cells demonstrated a lower protein content as a function of the optical density OD_{600}, reflecting the increase in ribosomal RNA at faster growth rates, reducing the relative amount of protein per cell (Figure 1.10). The slope of RNA/protein generally increased with growth rate (Figure 1.9) in the manner described by the [8, 56]. The amounts of RNA/protein

can be found in TableB.2.

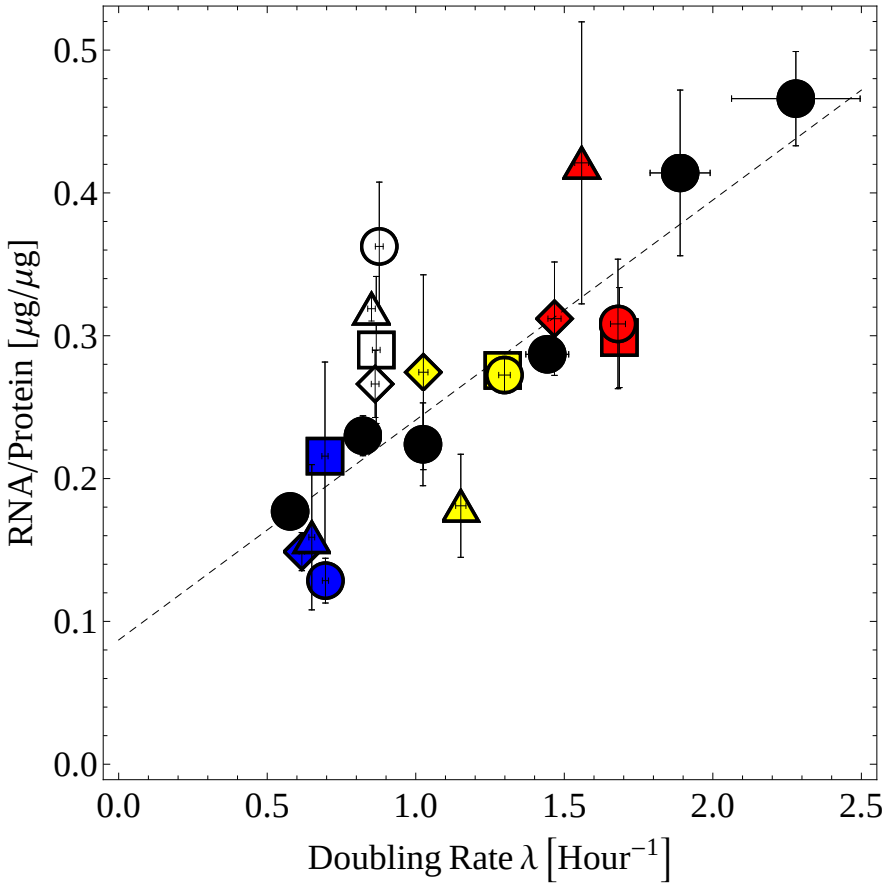


Figure 1.9: RNA/protein ratio vs Doubling Rate. The relative abundance of RNA increases with doubling rate, reflecting the increase in ribosome number needed to sustain faster doubling. The samples in this study do not deviate significantly from the data (black dots) and relationship (dashed line) seen in [56]. The value of the ratio r at a given doubling rate determines the relative amounts of ribosomal protein in the proteome model. The intercept on the y-axis reflects the absolute minimal RNA/protein ratio required for cell growth.

Growth rates of the wild-type,promoterless controls strain increased with tem-

perature and the presence of glucose over glycerol. The absolute growth rates were comparable to cells grown in equivalent conditions in past studies [56, 8] (Figure 1.9). The growth rate of the promoterless *insertion* strain was in general less than or equal to that of the promoterless *sequencing* strain, ranging from nearly identical (30°C ,glucose,CAA) to 11 %slower (37°C ,glycerol,CAA). Absolute measurements of growth rate and doubling times may also be found in Table B.2. Cell counts of the *sequencing* and *insertion* strains were found to be identical within error bars and consistent for the conditions tested.

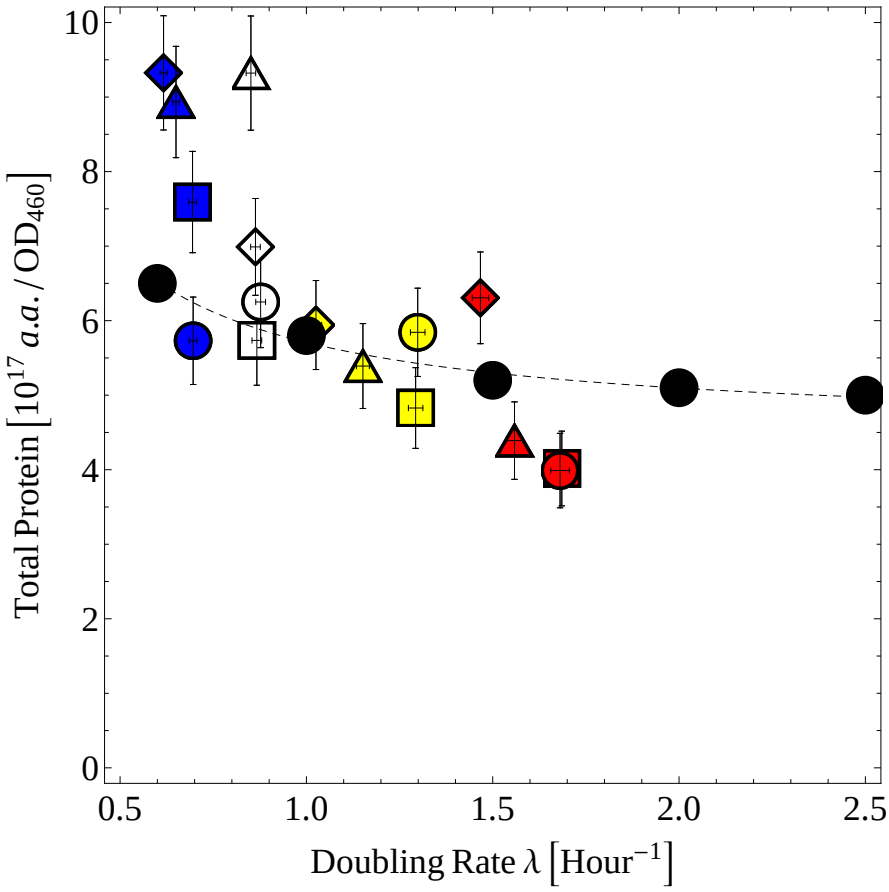


Figure 1.10: Protein/OD ratio vs. Doubling Rate. The absolute amount of protein per optical density decreases with growth rate, as cells grow denser with ribosomal RNA. Samples in this study are compared to absolute protein content of equivalent cells grown at 37°C in a survey by Bremer & Dennis [8] (Black dots, dashed line is guide to the eye). At low growth rate, the ratio appears to be significantly higher for cells in this study. However, these samples were grown at 30°C , as opposed to 37°C in [8].

1.3.3 Additional flagellin as unnecessary protein

Once we were able to determine that the RNA/protein ratio of the *sequencing* and *insertion* strains were not significantly different, we hypothesized that an additional source of unnecessary protein in the *insertion* strain may be additional flagella that does not appear in the *sequencing* strain. The flagellum assembly cascade responds differently in both *sequencing* and *insertion* strains depending on condition. According to reporting at the transcriptional level, the pfliL reporter is 20 to 40 times more active in the *insertion* strain than in the *sequencing* strain, (Figure 1.6). Based on blotting of flagellin protein, however, the actual count of additional flagella in the *insertion* strain is only 2-6 times higher than in the *sequencing* strain, depending on condition (Figure 1.18). The discrepancy is not surprising, as transcriptional reporting may deviate from translational reporting due to the presence of control systems in the assembly of the flagellum ([32]). Based on the relationship between the two promoters, it appears that the transcription signal is saturated (Figure 1.11).

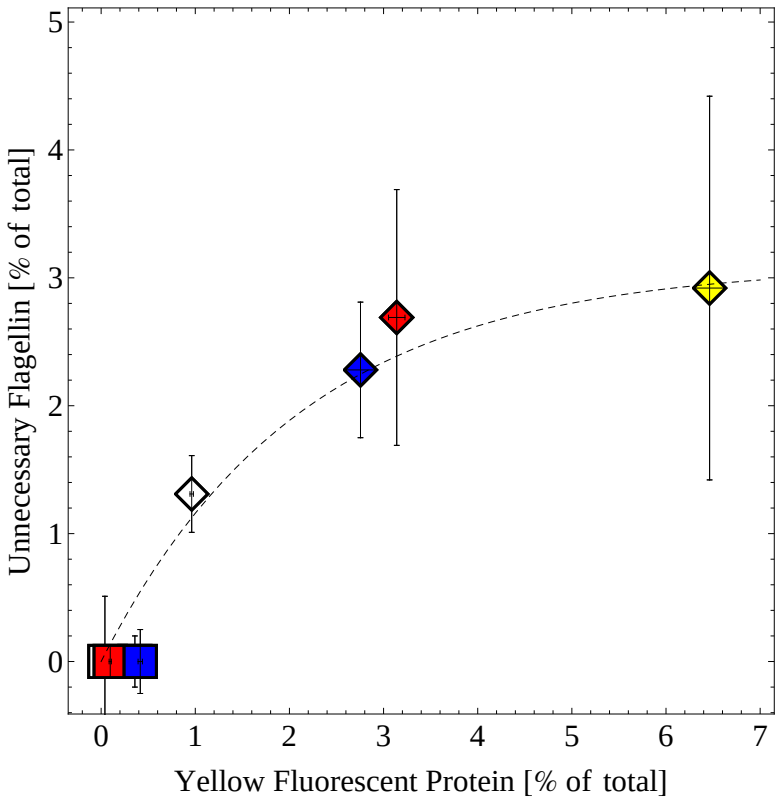


Figure 1.11: Additional flagellin vs. fluorescent protein. The x-axis represents the total YPF proteome fraction reporting on the flagellum assembly cascade at the transcriptional level via the pfliL promoter. The y-axis represents the amount of flagellin determined via Western Blot. The flagellin content saturates with increasing pfliL activity, although uncertainty of the Western Blot measurement is large.

The presence of additional flagella is further corroborated by swarming assays in which the *insertion* strain travels much further on a plate of thick agar in a limited nutrient environment (see 1.5.1). While the existence of additional flagella

benefit cells in an agar swarming plate, where the spread of nutrients is limited to diffusion, it is unlikely to benefit cells in the case of a well-mixed liquid media in a large beaker shaking at high intensity. For the purposes of this study, we treat the additional flagellin proteins (over wild-type levels) as unnecessary, where the total protein incorporated in a single flagellum corresponds to roughly 0.5% of the proteome [50].

1.3.4 Burden due to unnecessary protein increases with multiple sources of unnecessary protein

Treating both additional flagellin and YFP as unnecessary protein, we plot the relative growth rates of each strain as a function of unnecessary protein fraction, which ranges from $1 - 8\%$, depending on condition. (Figure 1.12). Comparing these growth rates of these two strains directly, we find that the linear relation 1.18 continues to hold, approximately.

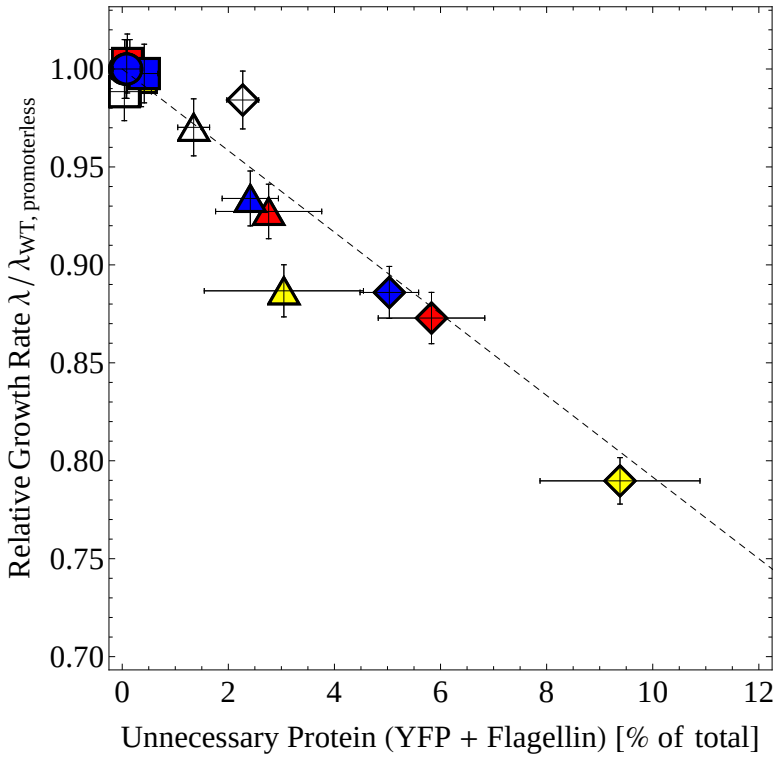


Figure 1.12: Relative Growth Rate vs. Unnecessary Protein Fraction. The growth rate of all strains is reduced as a linear function of the unnecessary protein fraction, which includes both YFP and additional flagellin. Growth rates in a given condition are relative to the WT promoterless *sequencing* strain. Relative growth rate does not seem to deviate significantly from the linear cost model (dashed line).

We hypothesized that this relationship would hold for not only these two species of protein YFP and flagellin, but for two-protein combinations featuring for β-galactosidase as well, which was the original protein used in [56]. By replacing the YFP with β-galactosidase on the reporter plasmid, we find that the

linear relationship holds as well (see Figure 1.13).

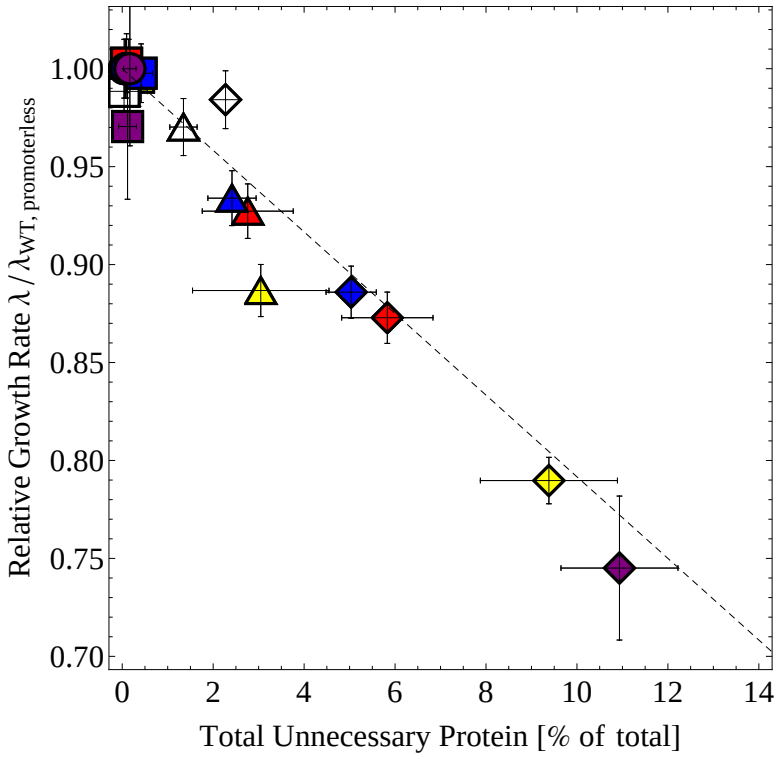


Figure 1.13: Relative Growth Rate vs. Total Unnecessary Protein. The growth rate of all the strains are compared to the WT promoterless *sequencing* strain as a function of total unnecessary protein, which in this case includes all three species included in this study: YFP, flagellin and β-galactosidase. The cost to growth rate of multiple protein species does not significantly deviate from the linear model.

1.3.5 The effect on growth rate from multiple protein sources is additive

In [56], the authors do not explicity predict the cumulative effect of unnecessary proteins from multiple sources. Based on the derivation of their model, there exist two completing naïve approaches: the additive and the multiplicative models.

In the additive model, we would be to assume the total unnecessary protein Φ_{U} is simply the addition of proteome fractions of separate origins $\Phi_{\mathrm{U,i}}$, such that $\Phi_{\mathrm{U}} = \sum \Phi_{\mathrm{U,i}}$. Therefore the model would give the relative growth rate of a strain as

$$\lambda(\Phi_{\mathrm{U,i}}) = \lambda_0 \left(1 - \frac{\sum \Phi_{\mathrm{U,i}}}{\Phi_{\mathrm{C}}}\right) \tag{1.19}$$

For example, in the case of the *insertion* strain under both large YFP production and extra flagellin production, $\Phi_{\mathrm{U}} = \Phi_{\mathrm{YFP}} + \Phi_{\mathrm{flagellin}}$ the growth rate $\lambda_{ins,\mathrm{pfliL}}$ should grow at a rate relative to the promoterless *sequencing* strain $\lambda_{seq,\mathrm{promoterless}}$ with the function

$$\lambda\left(\Phi_{\mathrm{YFP}}, \Phi_{\mathrm{flagellin}}\right)_{ins,\mathrm{pfliL}} = \lambda_{seq,\mathrm{promoterless}} \left(1 - \frac{\Phi_{\mathrm{YFP}} + \Phi_{\mathrm{flagellin}}}{\Phi_{\mathrm{C}}}\right) \tag{1.20}$$

This however, presents, a potential contradiction when comparing multiple sources of protein. As we saw above in section 1.3.4, the linear relationship holds within the *insertion* strain as well as across the two strains. The growth rate of the high-YFP*sequencing* strain $\lambda_{ins,fliL}$ relative to its promoterless variant $\lambda_{ins,\mathrm{promoterless}}$

is given by the relationship

$$\lambda(\Phi_{\mathrm{U,i}})_{ins,\mathrm{pfliL}} = \lambda_{seq,\mathrm{promoterless}} \left(1 - \frac{\Phi_{\mathrm{YFP}}}{\Phi_{\mathrm{C}}}\right) \tag{1.21}$$

We do not include the additional flagellin here, because it is shared by both *insertion* strains. However, the relative growth rate $\lambda_{seq,\mathrm{promoterless}}$ is in turn given by

$$\lambda(\Phi_{\mathrm{U,i}})_{ins,\mathrm{promoterless}} = \lambda_{seq,\mathrm{promoterless}} \left(1 - \frac{\Phi_{\mathrm{flagellin}}}{\Phi_{\mathrm{C}}}\right) \tag{1.22}$$

Therefore the effect of both the addition flagellin and YFP on the growth rate $\lambda_{ins,\mathrm{pfliL}}$ is

$$\begin{aligned}\lambda(\Phi_{\mathrm{U,i}})_{ins,\mathrm{pfliL}} &= \lambda_{seq,\mathrm{promoterless}} \left(1 - \frac{\Phi_{\mathrm{flagellin}}}{\Phi_{\mathrm{C}}}\right)\left(1 - \frac{\Phi_{\mathrm{YFP}}}{\Phi_{\mathrm{C}}}\right) \\ &= \lambda_{seq,\mathrm{promoterless}} \left(1 - \frac{\Phi_{\mathrm{flagellin}} + \Phi_{\mathrm{YFP}}}{\Phi_{\mathrm{C}}} + \frac{\Phi_{\mathrm{flagellin}}\Phi_{\mathrm{YFP}}}{\Phi_{\mathrm{C}}^2}\right)\end{aligned}$$

which differs from the relationship in 1.20 by a second-order term. This multiplicative model is similar to the Bliss independence model describing fitness reduction due to non-interacting drugs[6]. The general form is

$$\lambda(\Phi_{\mathrm{U,i}}) = \lambda_0 \left(1 - \frac{\sum_i^n \Phi_{\mathrm{U,i}}}{\Phi_{\mathrm{C}}} + \frac{\sum_{U,j\neq i}^n \Phi_{\mathrm{U,i}}\Phi_{\mathrm{U,j}}}{\Phi_{\mathrm{C}}^2} - \cdots + \frac{\prod_i^n \Phi_{\mathrm{U,i}}}{\Phi_{\mathrm{C}}^n}\right) \tag{1.23}$$

Within the regime studied here $\Phi_{\mathrm{U}} < 12\%$, the second-order term is likely to be small relative to the first-order term, and the two relationships will be equivalent, reducing to

$$\lambda(\Phi_{\mathrm{U,i}}) = \lambda_0 \left(1 - \frac{\sum_i^n \Phi_{\mathrm{U,i}}}{\Phi_{\mathrm{C}}}\right) \tag{1.24}$$

which for a single protein species is identical to the original model.

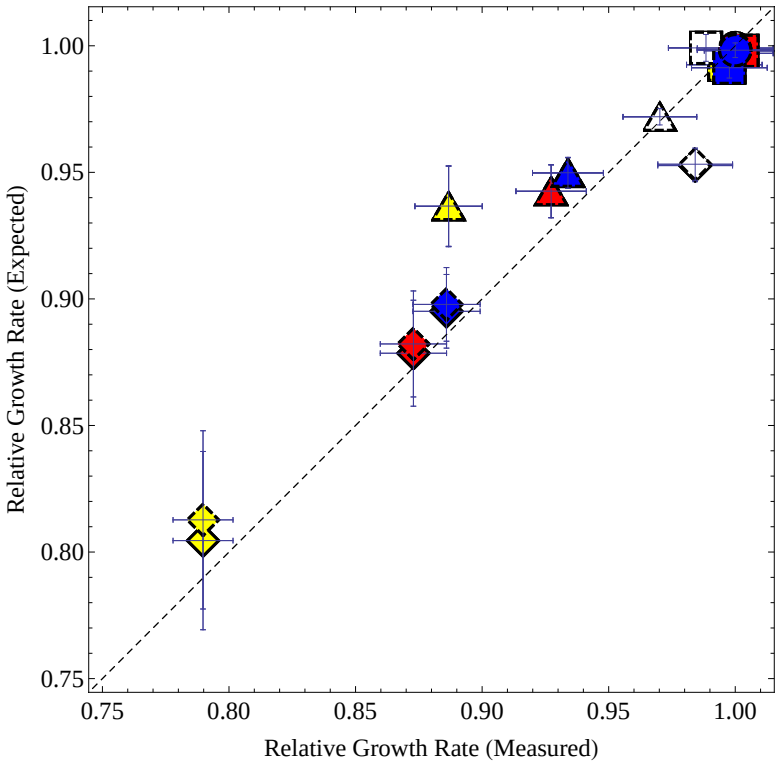


Figure 1.14: Comparison of first-order and second-order models to measured growth rates. Relative growth rate as predicted by the first-order model (solid borders, y-axis), using the measured unnecessary proteome percentages as variables, are slightly closer to actual values of the growth rate (x-axis) than those predicted the second-order model (dashed border, y-axis). The second-order model predicts relatively faster growth rates than the first-order model, especially at larger unnecessary proteome fractions.

We can the two models using Akaike's Information Criterion [9, p. 69], which for finite data sets is given by

$$AIC = 2K + n\log(\frac{rss}{n}) + \frac{2K(K+1)}{n-K-1}$$

where K is the number of estimable parameters (including 1 for the variance inflation factor), n is the number of data points, and rss is the residual sum of squares, and errors are assumed Gaussian and independent. The second-order model (AIC=-49.9) is 1.5 times as likely to describe the data as the first-order model (AIC = -49.1). We can conclude that for $\sum \Phi_U < 10\%$, unnecessary protein from multiple sources can be considered one single source. The pie chart in Figure 1.15 demonstrates a typical proteome partitioning under this model.

(a) Sequencing strain (b) Insertion strain (c) Insertion strain +YFP

Figure 1.15: Proteome partitioning under multiple unnecessary protein burdens. (a) The *sequencing* strain exhibits the three partitions as in original model[56]: fixed partition Q, ribosomal partition R, and "other" partition P (b) The *insertion* strain exhibits a burden due to unnecessary flagellin, denoted F. (c) Cells exhibit an addition burden due to the expression of reporter proteins YFP or β-galactosidase (U).

Based on the similarity of the models within this regime, we may compare the

data from this study to that of previous studies, we find that the effect of two unnecessary proteins may be treated as a single protein (see Figure 1.16).

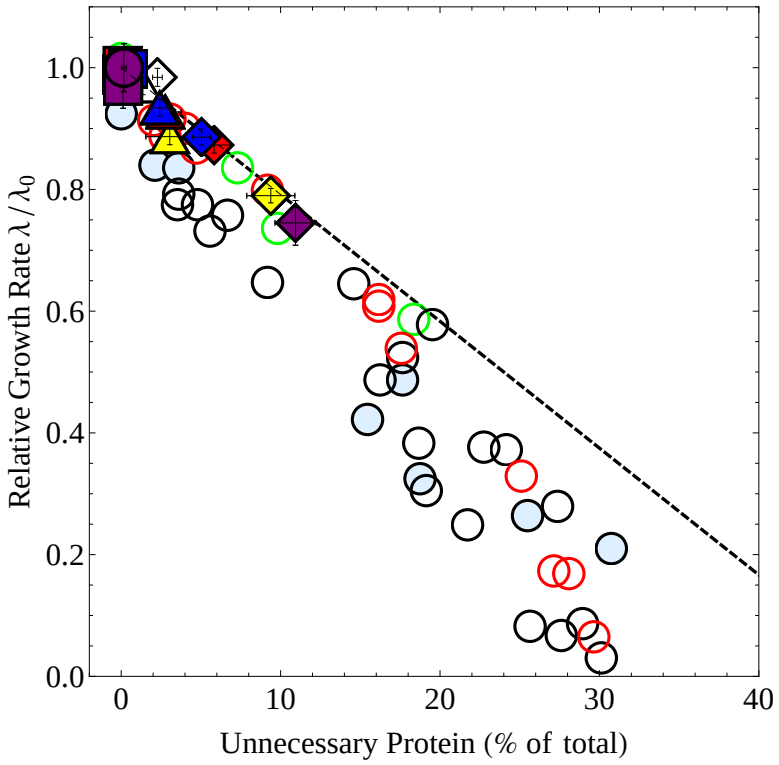


Figure 1.16: Comparison of this study and *Scott et. al.* Unnecessary protein production from multiple sources of up to $\approx 10\%$ of the proteome may be treated as a single source with regard to the growth rate of *E.coli*. The linear model is the additive model using the same value for $\Phi_C = 0.48$ [56].

1.4 Discussion

Prior to this study, it was not known whether multiple sources of unnecessary protein may be treated as a single source with respect to the effect on growth

rate in *E.coli*, or if the combination of multiple inputs would produce a nonlinear effect.The null hypothesis predicted that unnecessary proteins from multiple sources were simply additive in their affect on growth rate and described by the linear relationship. We tested this by inducing unnecessary flagellin production in addition to report yellow fluorescent protein, and then confirmed with the production of β-galactosidase and flagellin. For the regime encompassed by this study, for which the unnecessary protein does not exceed 12% of total protein, we did not see significant deviation from the linear relationship. As a result, in most cases the relative growth rates of separate strains of *E.coli* may be compared as a function of unnecessary protein fraction, provided their protein profiles are well understood.

The effect of multiple unnecessary protein sources on growth rate is still not understood, however, at protein fractions greater than 12%. The linear model presents a potential contradiction in that regime (as described in section 1.3.5). Further steps to test this regime would include sythetic induction of extrinsic proteins and environmental induction of intrinsic proteins, such as heat-induced chaperones[57] to increase total unnecessary protein in multiple linear combinations while monitoring growth rate. The simple addition of multiple sources may break down under cooperative burdens[2], complicating existing models of cell fitness [49]. However, as most proteins occur in relatively small amounts [61], the linear model proposed in [56] and further validated in this study should hold in

most cases.

1.5 Materials & Methods

1.5.1 Strains

The strain MG1655 was used as wild-type (WT) (referred to as the *sequencing* strain). This strain is the widely-used variant 6300 from Coli Genetic Stock Center (http://cgsc.biology.yale.edu/). The strain has low motility as compared to the nearly identical CGSC variant 7740, which is hypermotile. A key genetic difference is the existence of insertion element $insB$ at, labelled IS1, upstream from the $flhD$ gene in the hypermotile strain, which is absent in the sequencing strain. A slightly different insertion IS5 exists in the widely-used WT strain W3110. In order to create a hypermotile strain otherwise identical to the WT strain, IS5 was inserted at the IS1 location. Hypermotility was confirmed via swarming assay.

Cells were transformed with vectors modifed from the Alon plasmid library [65]. Each plasmid is low-copy (3-4 per cell) and confers resistance to the antibiotic Kanamycin for selection purposes. Each plasmid in the library also carries a specific promoter fused to the Green Fluorescent Protein (GFPmut2), Yellow Fluorescence Protein (YFP-Venus)[45], or β-galactosidase, depending on the strain and assay. Each strain was stored at -80°C in 30% glycerol, and plated on agarose plates prior to experiment.

1.5.2 Growth Conditions

Cells were grown from a single colony overnight in media (∼16 hours) shaking at 200 rpm at 30°C . Two media were used, M63+gluc+CAA and M9+glyc+CAA. M63+gluc was taken from [17]: Miller's M63 salts (in 1 liter of water: 2g$(NH_4)_2SO_4$, 13.6 g KH_2PO_4, 0.5 mg $FeSO_4 \cdot 7H_20$, pH adjusted to 7 with KOH and sterilized, to which was added 1mL 1 M $MgSO_4 \cdot 7H_2O$, 10 mL 20% glucose, 5mL 20% Casamino Acids (CAA) and antibiotics (25 μg/mL Kanamycin). M9+gly+CAA was taken from [27]. In 1 L, M9 minimal salts (Fisher): 6 g Na_2HPO_4, 3 g KH_2PO_4, 1g NH_4Cl, 0.5g NaCL, 3 mg $CaCl_2$, 2mM $MgSO_4$, 0.4% glycerol, 0.1% Casamino Acids, and antibiotics (25 μg /mL Kanacymin][17]. Overnight cultures were grown at 30°C or 37°C depending on the experiment. Overnight cultures were diluted 1:2000 (10μL in 20mL of fresh media), shaking at 250 rpm in a water bath at 30°C or 37°C . To measure cell density in the media, the optical density at wavelength of light 600 nm (OD_{600}) was taken on a spectrophotometer at a regular period (20-40 minutes, depending on the doubling rate). When the OD_{600} reached ∼0.01, samples were extracted for a series of measurements: fluorescence, protein concentration, RNA concentration, Western Blotting, and colony forming units. Doubling rates were calculated from fitting $Log_2(OD_{600})$ versus time to a line and taking the slope as the doubling rate. Uncertainties were taken from best-fit uncertainties.

1.5.3 RNA Measurement

RNA was isolated from 1.4mL samples taken during exponential growth, chilled in an ice slurry, and spun down at 3000g for 10 minutes then flash-frozen on dry ice (protocol by Arvind Subramaniam). All steps until final resuspension were performed at 4°C . A 1.5mL microcentrifuge tube was pre-chilled with 500μL of acetate-saturated phenol-chloroform, 50μL of 10% Sodium Dodecyl Sulfate (SDS) and 250μL of 300μm acid-washed glass beads. Cells were resuspended in 500μL 0.3 M Sodium Acetate (NaOAc)-10 mM Ethlyenediaminetetracetic Acid (EDTA), and the mixture deposited into the tube containing the glass beads, phenol chloroform, and SDS. Cells were lysed by vortexing at maximum speed for 10 minutes. The tubes were centrifuged at max speed to produce three layers: aqueous layer on top containing RNA, middle interphase layer with white precipitated DNA, and a pink organic layer on the bottom. The aqueous phase was removed and mixed with 500 μL of phenol-chloroform in a fresh tube. The tube was vortexed and spun for 2 minutes, and the aqueous layer was removed again and the extraction was repeated with an equal volume of chloroform. The RNA was precipiated with an equal volume of isopropanol, incubated at -80°C and spun at max rcf for 10 minutes. The isoproanol was pipete out carefully, and the pellet was washed with 150μL of 70 % Ethanol, spun for 5 minutes, and then the pellet was air-dried. The RNA was resuspended in DNase-free water, and quantified at OD_{260} in the Nanodrop. The samples were then run in a 1% agarose gel using TBE buffer to

see sharp rRNA and tRNA bands.

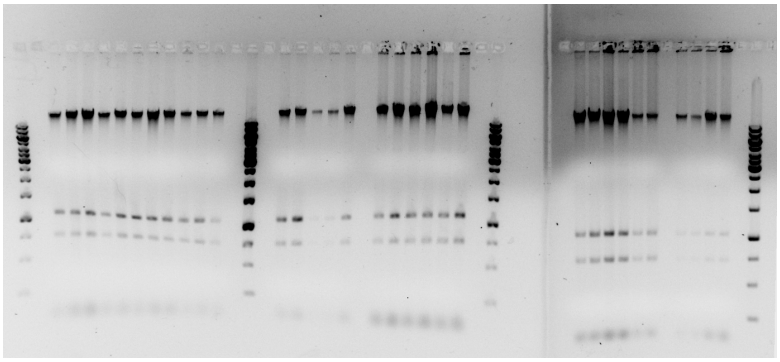


Figure 1.17: Example of RNA and DNA bands in samples. After RNA extraction, samples are run in an electrophoresis gel to seperate nucleic acids of differing lengths. In each sample, the largest nucleicc acids (top) correspond to chromosomal DNA. Ribosomal RNA can be seen in the weaker middle bands (23s and 16s, respectively). Faint bands of very short tRNA can be seen at the bottom of gel. To calculate the ribosomal RNA, the intensity of each band was measured and converted using the calibration of the DNA ladder, whose absolute mass per band is known.

1.5.4 Protein Measurement

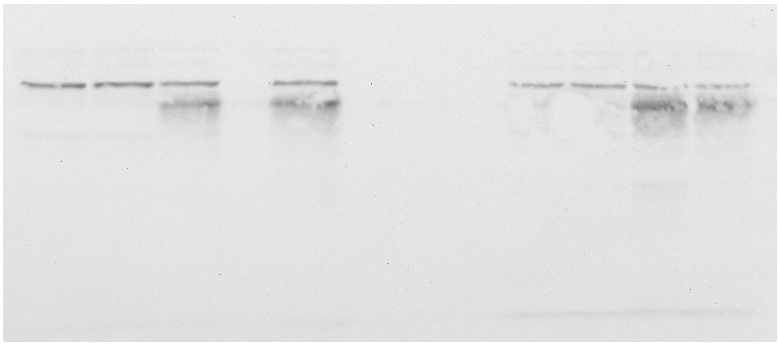
Protein extraction was carried out in 200μL of exponentially-growing samples which were flash-frozen in ethanol and dry ice after the method described in [56], using the Total Protein Kit (Sigma TP0300) and bovine serum albumin (BSA) as a standard. Samples were thawed in 800 μL of 0.015% (w/v) sodium deoxycholate. Cell samples and BSA standards ranging from 10-400μg /mL to were diluted to

1mL in 1.5 mL microcentrifuge tubes. 100 μL of 0.15 % deoxycholate were added to each sample, which were well-mixed and incubated for 10 minutes at room temperature. 100μL of 72% trichloroacetic acid were added to each microtube and well-mixed, incubated briefly and centrifured at 14,000rpm in a table-top microcentrifuge. The pellet was resupended in 1 mL of 50%Lowry reagent solution and incubated for 20 minutes. 400μL of Folin & Ciocalteu's Phenol Reagent Working Solution (provided in kit). After 30 minutes of incubation, the optical density at 720nm was measured. The BSA standards were used to calibrate the amount of protein in each sample.

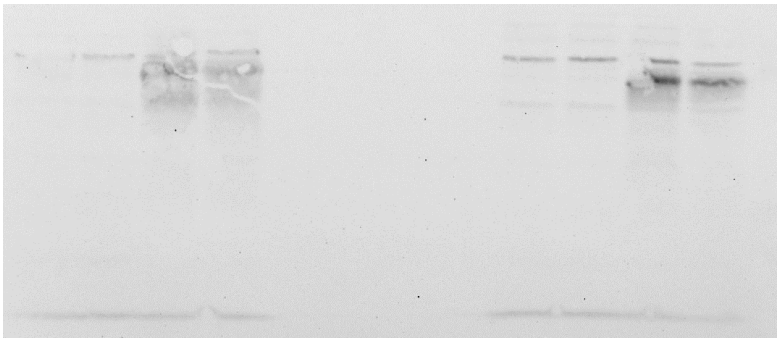
1.5.5 Estimate of Flagellin amount

Flagellin concentration was measured using the Western Blot protocol (from Vikram Vijayan). 5mL of samples were cooled in an ice-slurry for 5 minutes, then spun down at 5000rpm at 4°C . The cells were resuspended in lysis buffer (7.5M Urea, 20 mM HEPES pH 8.0) and transfered to pre-chilled 1.5mL microcentrifuge tubes with 300 μL of acid-washed glass beats. Cells were alternately vortexed and chilled for five cycles at 4°C , then spun at 20,000g 4°C for 10 minutes to settle the beads. The suspension was transfered to a fresh 1.5mL, and the process repeated. The final suspension was frozen at -80°C until all samples were collected. A bicinchoninic acid assay (BCA) was used to determine the protein concentration of each sample (similar to Lowry method outlined above). Samples were equal-

ized to roughly 10μg protein in a 15μL volume of lysate, whch was mixed with 15μL Laemmli buffer prepared fresh with β-metacptoethanol. The tubes were heated at 95°C for 5 minutes and then spun at 20,000g for 1 minute. Samples were loaded onto Biorad 12+2 well precast SDS-page gel, which was run 125V for 100 minutes to resolve protein bands. Protein was transferred to a nitrocellulose membrane in Novoblot transfer buffer at 180mA for 1 hour. The membrane was blocked in 2% casein milk TBST at 4°C overnight. 10μL of $1^{\mu g}/_{\mu L}$ anti-fliC primary antibody (Abcam) was added to 10mL of 2% milk TBST, which was shaken for 1.5 hours. The membrane was washed four times with TBST, then incubated with a 1:2000 dilution of anti-goat secondary antibody (Abcam), and washed again. Finally, 1:1 chemiluminescence reagents were added to the membrane, and the chemiluminescence was measured. The amount of flagellin in each sample was measured by image analysis in ImageJ. To estimate the fractional amount of flagellin to the proteome, the estimate was used that WT *sequencing* cells possess roughly one flagellum per cell, whose $\approx$ 6000 proteins correspond to 10^6 amino acids, or $\approx 0.5\%$ of the proteome.



(a) M63+glucose+cAA, Left: 30 °C , Right: 37 °C



(b) M9+glycerol+cAA, Left: 30 °C , Right: 37 °C

Figure 1.18: (a,b) Western Blots of flagellin protein. In each blot, the left two lanes represent the *sequencing* strain and the right two lanes the *insertion* strain, (pfliL , promoterless samples). Samples are equilibrated to $\approx 10\mu g$ total cell protein. The bands at 70 kDa correspond to flagellin, which is roughly equivalent in each strain *insertion* but varies depending on condition.

1.5.6 Estimate of fluorescent protein

To measure the concentration of fluorescent protein, absolute measure, 150 μL samples were taken from exponentially-growing culture applied them to a

slide and in a 96-well plate. individual cells were placed at the focus of the microscope (x100, numerical aperture=1.3, Olympus, Melville, NY). A confocal volume within the cell is illuminated with a laser (Sapphire 488nm, 20mW, Coherent, Santa Clara, CA), The resulting beams are each filtered and focused on an avalanche photo-diode (spcmaqr-16fc, PerkinElmer). The photon detector's signal is analyzed in real time with a fast correlator (5000EPP, ALV, Langen, Germany) [33].

For the convenience of the reader when considering the experimental results in this thesis, important theoretical underpinnings presented in previously published work by Oleg Krichevsky and Grgoire Bonnet (reference [31]) are summarized in the following two pages.The molecules we study move through the cell via diffusion, described by

$$\frac{\delta C}{\delta t} = D\nabla^2 C$$

where the concentration is given by $C(\vec{r},t) = \bar{C}(\vec{r},t) + \delta(C(\vec{r},t))$. The distribution of the excitation light in the sample is denoted by $I(\vec{r})$. We assume that the number of photons emitted and collected from each molecule is proportion to this intensity, so that the fluctuation $\delta n(t)$ of the photon counts from the mean $\bar{n} =< n(t) >$ is

$$\delta n(t) = n(t) - \bar{n} = \triangle t Q \int d^3 r I(r) \delta C(\vec{r},t)$$

where Q is some proportionality factor. We consider $G(t)$ as the normalized time

average of products of the intensity fluctuations

$$G(t) = \frac{1}{\bar{n}^2 T}\sum_{i=0}^{T-1}\delta n(t')\delta n(t'+t) = \frac{1}{\bar{n}^2} < \delta n(t')\delta n(t'+t) >_{time}$$

This is the normalized autocorrelation function, where $\bar{n}^2$ is the square average intensity, T is the total number of accumulated sampling intervals and $T\triangle t$ is the total integration time of the experiment. For the purpose of the derivation we will use the ergodicity of the system to write $G(t)$ as an ensemble average

$$G(t) = \frac{1}{\bar{n}^2} < \delta n(0)\delta n(t) >_{ens} = \frac{1}{\bar{n}^2}\triangle t^2 Q^2 \int\int d^3r d^3r' I(r)I(r') < \delta C(\vec{r},0)\delta C(\vec{r'},t) >_{ens}$$

Using Fourier Analysis we define

$$\delta\tilde{C}(\vec{q},t) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int d^3\vec{r}e^{i\vec{q}\cdot\vec{r}}\delta C(\vec{r},t)$$

We then can calculate the correlation of the concentration in space and time

$$< \delta C(\vec{r},0)\delta C(\vec{r'},t) >= \frac{1}{(2\pi)^{\frac{3}{2}}}\int d^3\vec{q}e^{-i\vec{q}\cdot\vec{r}} < \delta C(\vec{r},0)\delta\tilde{C}(\vec{q},t) >$$

Substituting $\delta\tilde{C}(\vec{q},t) = \delta\tilde{C}(\vec{q},0)e^{-Dq^2t}$ leads us to

$$< \delta C(\vec{r},0)\delta C(\vec{r'},t) >= \frac{\bar{C}}{(2\pi)^3}\int d^3\vec{q}e^{-i\vec{q}\cdot(\vec{r'}-\vec{r})}e^{-Dq^2t}$$

Using $\tilde{I}(\vec{q}) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int d\vec{r}^3 e^{-i\vec{q}\cdot\vec{r}}I(\vec{r})$ we find

$$G(t) = \frac{\bar{C}}{\bar{n}^2}\triangle t^2 Q^2\int d^3\vec{q}|I^2(q)|e^{-Dq^2t}$$

The confocal volume is a Gaussian cylinder with intensity profile

$$I_{x,y,z} = I_0 e^{\frac{-2(x^2+y^2)}{\omega_{x,y}^2}}e^{\frac{-2z^2}{\omega_z^2}} \Rightarrow \tilde{I}(\vec{q}) = I_0\frac{\omega_{x,y}^2\omega_z}{8}e^{\frac{\omega_{x,y}^2}{4}(q_x^2+q_y^2)-\frac{\omega_{x,y}^2}{8}(q_z^2)}$$

where $\omega_{x,y}$ is the radius and ω_z the length of the detection volume cylinder and $I(0)$ is the maximum intensity. If we insert this definition into $G(t)$ and perform a Gaussian integration we obtain

$$G(t) = \frac{1}{N}\left(\frac{1}{1+\frac{4Dt}{\omega_{x,y}^2}}\right)\left(\frac{1}{1+\frac{4Dt}{\omega_z^2}}\right)^{\frac{1}{2}} = \frac{1}{N}\left(\frac{1}{1+\frac{t}{\tau_{x,y}}}\right)\left(\frac{1}{1+\frac{t}{\tau_z}}\right)^{\frac{1}{2}}$$

where $\tau_{xy} = \frac{\omega_{xy}^2}{4D}$, $\tau_z = \frac{\omega_z^2}{4D}$ and N is the number of fluorescent particles in the volume. In the limit $t \to 0$, $G \to 1/N$, reflecting the Poisson Statistics of this sample size.

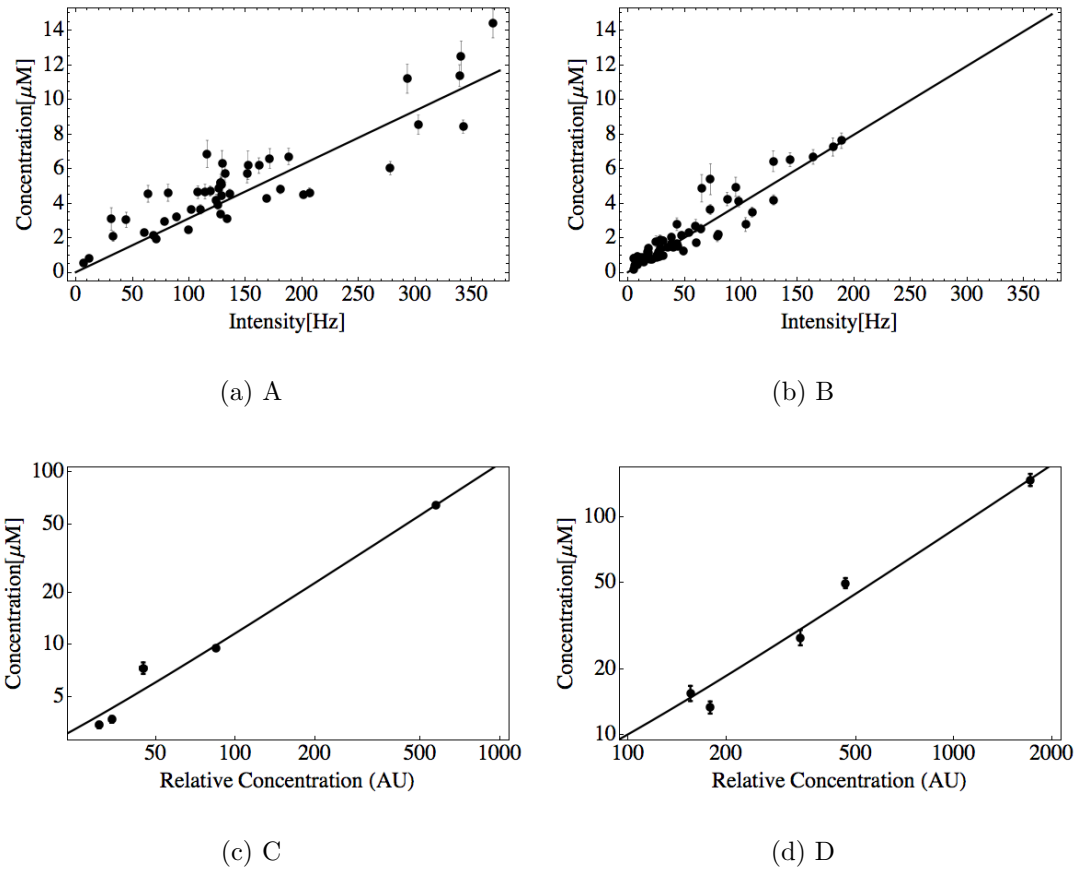
Notice that $\tau_z = \frac{\omega_z^2}{\omega_{xy}^2}\tau_{xy}$, and in our case, the soft focus of the laser constitutes a volume that is much longer than it is wide. We are more interested in smaller time scales, so the element of the correlation function relating to z can be assumed to be ~ 1. We therefore have

$$G(t) = \frac{1}{N}\left(\frac{1}{1+\frac{t}{\tau_D}}\right)$$

where τ_D is the characteristic time of diffusion.

Using equation [FCS] I demonstrated a proportional relationship between photon incidence (kHz) and fluorescent protein concentration (uM), which varied depending on GFP and YFP (see Figure 1.19). I took a sample from each of the five flagellar promoters in both GFP and YFP, and using the respective proportionality constants, established a calibration curve between the concentration obtained from the platereader and the absolute concentration found within single cells (see Figure 1.19c,1.19d). Fluorescence intensity per YFP-venus molecule was

approximately 28% greater than GFPmut2.



(a) A

(b) B

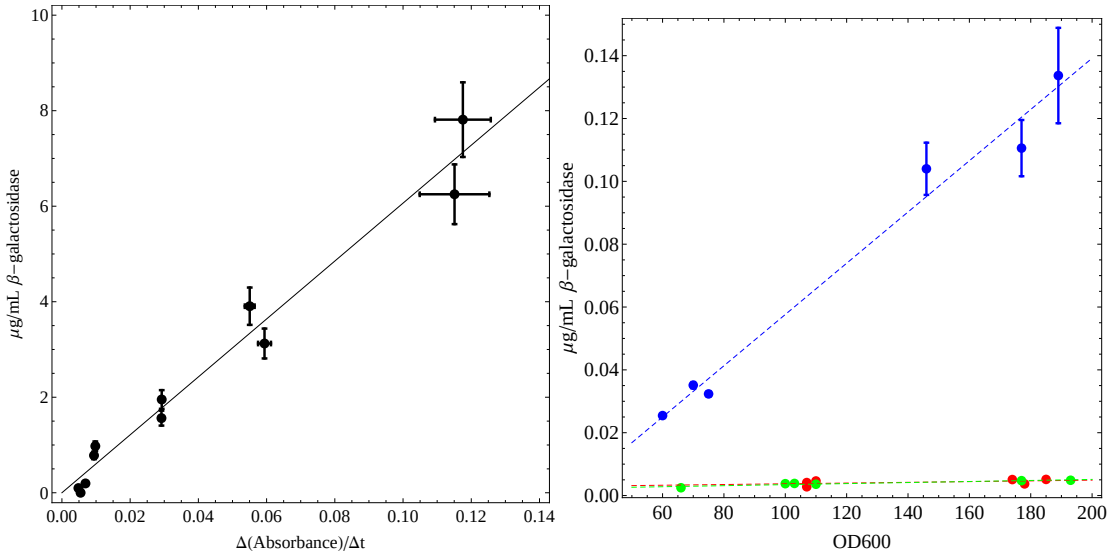
(c) C

(d) D

Figure 1.19: Calibration of Fluorescence Intensity to Protein Concentration. Cells are transformed with (a,c) GFP (b,d) YFP reporter plasmids. Each point above represents a single FCS measurement corresponding to midlog growth. N is a fit parameter representing the average number of proteins within the FCS volume. For each flagellar promoter, we measured 50 cells taken from culture. We fit a linear calibration curve through the axis (solid line). From this, we convert readings on the plate reader to absolute concentrations of fluorescence protein. Error bars are standard deviations.

1.5.7 Measurement of β-galactosidase

Measurement of β-galactosidase was modifed from [56]. During exponential growth between OD_{600} of 0.1 and 0.2, samples of 500 μL were vortex and places on ice. Z-buffer was prepared (in 1 Liter: 8.52g Na_2HPO_4, 5.5g $NaH_2PO_4 \cdot H_2O$, 0.75g KCL, 0.25g $MgSO_4 \cdot 7H_20$. The pH was adjusted to 7.0, then 0.004% SDS (w/v),40mMβ-mercaptoethanol(BME) and 100 μL chloroform was added. Lysate was diluted with 50% Z-buffer solution (in media) into two separate samples 1:10 and 1:5 to test for linearity of measurement. 200 μL was added to 96-well plate (with three repeats), and 40 μL of 4 μg /mL ortho-Nitrophenyl--galactoside (ONβG) in 0.1M Phosphate buffer (pH 7.0) was added. The absorbance at 420 nm and 600 nm was taken every minute for 30 to 60 minutes at 30°C . The slope at 420 nm indicates the concentration of β-galactosidase, using a similarly diluted sample of purified β-galactosidase (Sigma) as a control.



(a) Standards

(b) Samples

Figure 1.20: β-galactosidase content in samples. (a) A set of controls was used to calibrate the increase in optical density (absorbance) over time to an absolute concentration of β-galactosidase (b) The absolute concentration of β-galactosidase vs. OD_{600} of cell samples.

1.5.8 Colony Forming Units

To verify that OD_{600} accurately represents the number of living cells in media, 100 μL samples were extracted and serially diluted in media such that cell concentration ranged over 7 orders of magnitude. Using a multi-channel pipettor, 5μL of each concentration were dripped and spread on a large agar plate (Teknova). After an overnight incubation at room temperature, colonies were counted using a Zeiss Lumar.V12 SteREO microscope and Axiovision microscopy software. Uni-

formity of fluorescence across the colonies was verified using a Zeiss HBO 100 Mercury lamp with an excitation filter at 500nm (24nm bandwidth) and dichromatic mirror cutoff at 520nm. Colonies were counted across two-three orders of magnitude, depending on cell density. The number of colony forming units (cfu) was determined by taking the weighted average of these measurements.

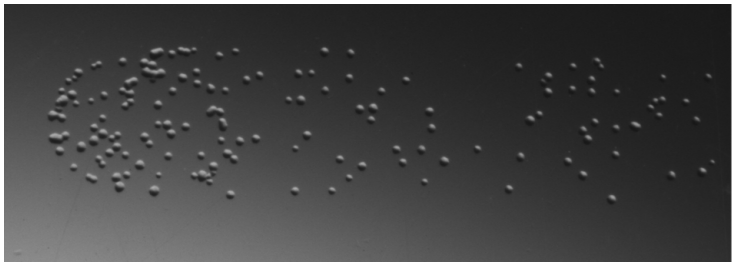


Figure 1.21: Micro-colonies growing on agar plates. The colonies were counting using a stereo-microscope and compared across strains as a function of optical density OD_{600}. Uniformity of fluorescence was also confirmed.

Appendix A

Simultaneous determination of mRNA and protein concentration in single living bacteria

A.1 Abstract

Recent imaging technologies have opened a route to the quantitative visualization of low numbers of mRNA transcripts in eukaryotes and prokaryotes. Such imaging technologies are powerful tools for monitoring the localization of sparse RNA transcripts, however they may be limited when mRNA is abundant and not localized. Alternatively, Fluorescence Correlation Spectroscopy (FCS) permits the characterization of high concentrations of non-coding RNAs in a single living

bacterium. Here we extend the use of FCS to measuring concentration of coding RNAs in single living cells. We genetically fuse a red fluorescent protein (RFP) gene and two binding sites for an RNA binding protein, whose translated product is the protein RFP alone. Using this construct, we determine in single cells both the [mRNA] concentration and the associated [RFP] expressed from an inducible plasmid. We find that this method allows us to reliably monitor in real-time [mRNA] down to 40 nM (i.e. 2 transcripts per volume of detection). To validate these measurements, we show that [mRNA] is proportional to the associated expression of the RFP protein. This FCS-based technique establishes a framework for simultaneous detection of transcription and translation in individual living bacteria.

A.2 Introduction

Recently, a series of innovative image-based assays have demonstrated the possibility for determining the localization [4, 23, 46, 62],as well as the concentration of non-coding RNA and mRNA in single living cells[21, 22]. Although these approaches are powerful tools to detect and visualize a very small number of RNA molecules, they also have significant limitations regarding higher levels of [mRNA][21, 22]. Moreover, imaging techniques use a large number of fluorescent probes to tag the individual RNA molecule, which form a molecular complex often as large as the endogenous mRNA itself[21, 22]. Most of these imaging-based as-

says rely on the visualization of well-localized fluorescent spots each associated to individual tagged mRNA transcripts. However, in recent years Fluorescence Correlation Spectroscopy (FCS)[39, 40, 20] has been increasingly applied to the study of biological systems[26, 28, 19] and furthermore to determine protein concentration and RNA concentration in real time in living bacteria[10, 41, 41]. Assays using FCS rely on freely diffusing RNA molecules that are tagged with few fluorescent probes of relatively small size compared to the size of the native RNA transcripts[35, 33]. Le et al[35] have recently used an FCS-based assay to measure the concentration of a non-coding RNA in single living bacteria. In [35], an RNA binding fusion protein, MS2-GFP, binds specifically to a tandem of 23-nucleotide RNA binding sites. FCS is sensitive to the diffusion of fluorescent molecules. When MS2-GFP molecules are free, they diffuse fast through the volume of detection with a typical time of 1ms. When the MS2-GFP molecules bind to the tandem of RNA binding sites, the fluorescent molecules diffuse slower, revealing the presence of the specific RNA molecules. Additionally, the fusion of a ribosomal binding site to the tandem of RNA binding sites drastically increases the sensitivity of detection by causing the RNA/MS2-GFP/ribosome complex to diffuse 30-fold slower (30ms) than free MS2-GFP. Due to the strong binding of MS2-GFP to its RNA binding sites, all target RNA molecules in the cell are practically labeled at all times. Using this molecular labeling scheme, fluorescent correlation spectroscopy has allowed us to measure the relative concentrations of slow dif-

fusing RNA/MS2-GFP complexes and fast diffusing MS2-GFP proteins. In our previous study[35], a single ribosome binds to each non-coding RNA transcript. MS2-GFP, RNA, and the bound ribosome form a well defined molecular system that unequivocally determines the diffusion of this complex. In this report, we extend the use of the FCS technique from non-coding RNA detection to that of mRNA in living single bacteria of E. coli. Measuring [mRNA] is potentially more challenging than our previous study because each mRNA transcript is actively translated in the cell and thus can have more than one bound ribosome at any given time. Therefore it is not clear that FCS can be easily used for mRNA detection. However, here we show that our FCS assay is sensitive enough to measure reliably mRNA levels in living single cells. We demonstrate that the lower limit of [mRNA] detection is a few tens of nM, which corresponds to about two mRNA transcripts per volume of detection. We also measure simultaneously in each cell the concentration of a red fluorescent protein [RFP] expressed from the measured [mRNA]. Finally, we use the independent measurements of [RFP] to validate the accuracy of [mRNA] measured with FCS.

A.3 Materials and Methods

A.3.1 Genetic assay for mRNA/protein in vivo measurements

We construct a synthetic gene in order to measure simultaneously mRNA levels and the associated protein levels. The gene sequence carries at the level of DNA a coding region for the red fluorescent protein dsRed followed by a tandem of two ms2 binding sites, ms2x2, which ends with a transcriptional terminator (Figure A.1).

Figure A.1: Genetic assay for mRNA/protein measurements in living single cells. Transcriptional fusion of the coding region of the red fluorescent protein dsRed fused to two ms2 binding sites. The dsred-ms2x2 gene was placed under the control of an inducible Tet promoter, pLtetO-1. rbs: ribosomal binding site, taa: in frame stop codon; terminator: transcriptional terminator.

Finally, a stop codon in the dsred gene ensures that translation does not proceed beyond the dsred coding region into the non-coding ms2x2 binding sites. The resulting synthetic gene, which we call dsred-ms2x2, is very tightly controlled by an inducible TetR regulated promoter [37]. The coding region of dsRed is at the 5 end of the ms2 binding sites. Therefore, the dsred gene is transcribed first

and the ms2x2 binding sites last (Figure A.2). This order ensures that MS2-GFP molecules bind the mRNA transcript only once the gene has been fully transcribed. Therefore we can detect only fully transcribed mRNA transcripts. MS2-GFP fusion proteins are pre-expressed in the cytoplasm from a low copy plasmid and are immediately available to bind the mRNA coded by the dsred-ms2x2 gene.

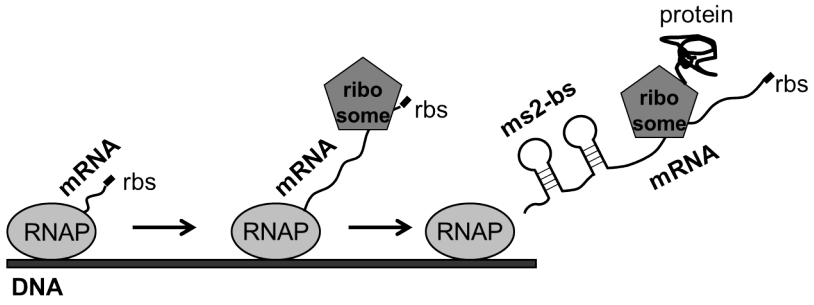


Figure A.2: A molecule of RNA polymerase transcribes the dsred-ms2x2 gene. The transcription of the ms2 binding sites occurs only at the end of the dsred gene. The ribosome bind the nascent mRNA transcript (middle cartoon), and the MS2-GFP proteins cannot bind the mRNA transcript unless the ms2 binding sites are fully transcribed. This genetic design guarantees that our FCS detection method only measures mRNA transcripts that are free to diffuse and that are not in complex with the RNA polymerase and the DNA.

A.3.2 Strains and plasmids

We use strain Frag1B, which is wildtype for the multi-drug efflux pump acrAB and strain Frag1A, which is a deletion mutant of the efflux pumps genes acrAB

[35]. We modified the MS2-GFP fusion protein initially developed by Bertrand et al [4] and described in [35]. We introduce a five amino acid linker (Ser, Gly, Gly, Gly, Gly) between the C-terminal of the ΔlFG mutant MS2 coat protein [47] and the N-terminal of the GFP coding part of the fusion gene. Plasmid pZS*12-MS2-GFP. A fusion-PCR reaction using the five amino acid linker as an internal primer produced the MS2-GFP fusion. The ms2-gfp fusion gene was cloned between the KpnI and HindIII sites of a pZS*12 vector [37] to produce pZS*12-MS2-GFP. Plasmid pZE31-dsRed-ms2x2. We used plasmid pQE31-dsRed.T3f (a gift from the Glick Lab, The University of Chicago), which carries a variant of the fast folding dsRed.T3 gene [5], to PCR amplify the dsred gene and clone it into a pZE21 vector resulting in plasmid pZE21-dsRed. The two ms2 binding sites were PCR amplified from pZE31-ms2[35] flanked by HindIII restriction sites with the following primers (5 ttaagcttgatatcgaattccga and 3 aagcttccgctctagaactagtggatcc). This fragment was introduced into the HindIII site of plasmid pZE21-dsRed, producing plasmid pZE21-dsRed-ms2x2. Subsequently, the kanamycin resistance cassette was replaced with a chloramphenicol resistance cassette from the pZE series of vectors [37], resulting in the plasmid pZE31-dsRed-ms2x2. Electro-competent cells of both Frag1B and Frag1A were transformed with plasmids pZS*12-MS2-GFP and pZE31-dsRed-ms2x2.

A.3.3 Growth conditions

Cells were grown from a single colony overnight for 14 hours in Luria Broth (LB) medium, at 30∘ C and 300 rpm, in the presence of IPTG (200 μM) and the antibiotics ampicillin (50μg/ml) and chloramphenicol (20μg/ml). Cells were diluted 1:100 into fresh LB medium, and induced with anhydrotetracycline (aTc) at various concentrations. Cells were grown on average for 2.5-3 hours, and harvested at an OD of 0.150-0.200. The same overnight culture was used for all different aTc induction levels, by growing distinct cultures. A 0.3 μl drop of cell culture was placed on a glass cover slip then covered with a 3% low melting-point agarose LB gel containing equal concentrations of aTc as the liquid culture. The gel padding did not contain any antibiotics or IPTG. The sample was sealed within the cavity of an aluminum slide. The prepared sample was then placed on a heated microscope stage at 30∘ C for measurements.

A.3.4 Single cell measurements

FCS measurements were performed on individual cells within the first hour after harvest. On average 70-80 different cells were measured for each inducer concentration. The fluorescent signal was acquired over a 3s interval.

A.3.5 Fluorescence correlation spectroscopy (FCS) apparatus

A collimated laser beam (Sapphire 488 nm, 20 mW, Coherent, Santa Clara, CA) is expanded ten-fold via an achromatic diverging lens (f2=25mm) and a converging lens (f1=250mm) to a diameter of 8mm. Two neutral density filters (OD=1 and OD=3, NE10A and NE30A, Thor Labs) attenuate the expanded laser beam before it enters the microscope (OlympusX71, Melville, NY). A dichroic mirror, DM1 in A.3, reflects the incoming laser beam, (492dclp Chroma Technologies) and is focused with a 100X microscope objective lens (NA=1.3) to a diffraction limited spot onto a single bacterium. The fluorescent signals from the green and red fluorescent proteins are transmitted by DM1 and rejected by a second dichroic mirror, DM2 in Figure 2, (700 dcxr, Chroma Technologies) outside the microscope.

The out coming fluorescent light is focused using an achromatic convex lens (f=250mm, ThorLabs). A high-pass filter (RazorEdge LP02-488RU, Semrock) removes residual 488nm laser light from the fluorescent signal. The fluorescent light is separated into two components, green and red, with a third dichroic mirror, DM3 (Figure A.3), (575dcxr, Chroma Technologies). A band-pass filter (FF01-520/35, Semrock) selects the green signal and a band-pass filter (FF01-618/50, Semrock) selects the red signal. The cross contribution from the GFP and dsRed fluorescent signal is 2.4% in the green channel and 8.6% in the red channel. Finally, each signal is collected in a confocal geometry using a multimode optical

fiber (AFS50/125Y, ThorLabs) whose 50 μm core acts as a pinhole. Photons are detected at the other end of the fiber with a photon counting module (avalanche photodiode spcmaqr-16fc, Perkin Elmer). A sub-micrometer positioning system allows us to adjust the x,y,z position of the pinhole, which is critical to FCS measurements. A fast correlator (5000EPP, ALV, Langen, Germany) computes in real-time the auto-correlation function from the fluorescent signal detected by the photon-counting module.

A.3.6 Analysis of FCS measurements

The autocorrelation function of the green fluorescent intensity signal is fitted with the following function.

$$G(t) = \frac{1}{N}\frac{1}{[1+y]^2}\left(\frac{1-y}{1+4Dt/\zeta free} + \frac{4y}{1+t/\zeta_{bound}}\right) \tag{A.1}$$

This function models the free and bound MS2-GFP as two species with fast and slow diffusion [31]. This model also accounts for the two-fold increase in brightness per bound species since per mRNA we have two ms2 binding sites each binding a MS2-GFP homodimer [35]. N represents the number of individual fluorescent particles in the observation volume, y is the fraction of MS2-GFP molecules bound to the mRNA-ribosome complex, ζ_{free} and ζ_{bound} represent the diffusion times of free and bound MS2-GFP respectively. N and y are the fitting parameters. ζ_{free} is 1.55$\pm$0.07 ms in Frag1B cells and 1.59$\pm$0.08 ms in Frag1A

cells. We obtain zfree by performing FCS measurements on cells carrying only the pZS*12-MS2-GFP plasmid and by fitting the autocorrelation curves with the following function:

$$G(t) = \frac{1}{N}\left(\frac{1}{1 + t/\zeta free}\right) \tag{A.2}$$

A fitting procedure using the two-component function determines the value of ζ_{bound} for each strain (Frag1B and Frag1A) induced with 10ng/ml aTc. This procedure uses zbound as a free parameter to minimize simultaneously the total residual sum of squares for all the autocorrelation functions. We find that zbound is 28 ms for Frag1B cells, and 29.4 ms for Frag1A cells. Several factors affect the quality of FCS measurements: excessive bleaching of the GFP molecules and bad alignment of the focused laser beam with the cell. A set of selection criteria allows us to discard poor FCS measurements based on poor fitting of the auto-correlation curves. These criteria emerge from the noise analysis of each auto-correlation curve[64, 52], as well as a set of empirical observations that relate the number of fluorescent MS2-GFP measured in the detection volume and the associated mean fluorescence intensity.

A.3.7 Auto-correlation curve noise analysis

At long timescales (τ_jj longer than 4.5 ms), low-quality auto-correlation data exhibits large deviations from the fit. Several phenomena can cause these devi-

ations: poor positioning of the confocal volume on cells, bleaching of the GFP, which artificially introduces a long-time component in the autocorrelation function, and large aggregates of fluorescent proteins. First, poor positioning produces deviations at timescales larger than 30 ms (Figure A.4).

This behavior was readily apparent during data acquisition and we discarded immediately these curves without further analysis. Second, we use Fourier analysis of the normalized residuals to characterize the effect of bleaching and the presence of aggregates in order to reject low-quality curves. We use two parameters: 1) the power spectrum of the normalized residuals at frequency zero indicates the presence of a constant component in the autocorrelation curve. We reject curves where the value of this parameter exceeds 0.03. 2) The variance of the residuals over time scales longer than 4.5 ms, accounts for large the deviations from the fit of the auto correlation function. We reject curves where the variance is above 0.2. In Figure S2 we show typical examples of auto-correlation curves that passed or failed this analysis.

At short times scales (dwell times $\Delta\tau_j$ shorter or equal to 256 μs), the auto-correlation function is computed from linearly distributed dwell times $\Delta\tau_j$. Consequently, the variance of the residuals is proportional to $(\Delta\tau_j)^{-1}$. Individual auto-correlation curves are rejected if the linear fit of the variance of the residuals is poor. We discard curves when the variance at a given timescale deviates from the linear fit by more than 0.4 (Figure A.5).

A.3.8 Monomer number and fluorescence intensity

Under working conditions, the total number of MS2-GFP molecules, N(1+y), in the detection volume is proportional to the measured mean fluorescent intensity. We reject FCS measurements that deviate by more than 30% from the linear relationship between N(1+y) and mean fluorescence intensity in the detection volume.

A.3.9 Bleaching during measurements

Bleaching of MS2-GFP reduces as a function of time the mean fluorescence intensity during signal acquisition. To characterize this decay, we divided the acquisition time (3 Sec.) into 16 intervals. We take the mean of the fluorescence intensity over each interval and fit the 16 means with a line. We rejected measurements that give linear fits with a slope steeper than -6% or +8%.

A.3.10 Determining the number of dsRed molecules

The low quantum efficiency of dsRed yields sometimes a poor signal to noise ratio in our FCS measurements. Nevertheless, we used measurements on cells for which FCS measurements were accurate enough to determine a linear relationship between the number of dsRed molecules and the measured fluorescence intensity (Figure A.6). During our measurements, we use this linear relationship to infer directly the number of dsRed molecules and we determine simultaneously [mRNA]

using FCS.

A.3.11 Volume of detection of FCS setup

In order to determine the detection volume of the FCS setup, we used fluorescent polystyrene beads of known size (44nm in diameter, Duke Scientific). To measure the diffusion coefficient of the beads, we fit the measured autocorrelation function to A.3.

$$G(t) = \frac{1}{N}\left(\frac{1}{1+4Dt/\omega^2}\right) \tag{A.3}$$

This function describes a two-dimensional translational diffusion process for a fluorescent molecule. N is the number of diffusing particles in the confocal volume, D is the two dimensional diffusion constant, t is time, and ω is the radius of the detection volume. Using the Stokes-Einstein relationship we can determine the value of $D = k_B T/6\Pi\nu R$, where k_B is the Boltzman constant, T is temperature in Kelvin (300K), ν is the viscosity of water (1 mPa·s), and R is the radius of the beads (22nm). We find $D = 10\mu m^2/s$. Using $D = \omega^2/4\zeta$, where ω is the radius of the detection volume and ζ the measured diffusion time of the beads (1.01 ms, FigureA.7), we determined that $\omega = 0.2\mu m$.

Next we estimate the width of the E.coli cells as follows: we image polystyrene beads of known size (diameter = 1.0 μm) as well as individual E.coli cells and compare the width of the cells to the size of the beads. We find that the width of

the cells is 0.73μm. Since the diameter of the detection volume is smaller than the width of the cell, we assume the volume of detection to be a cylinder and compute the volume of detection as: $V_{detection} = \Pi\omega^2 h$, where h is the volume height determined by the width of the cell in this case h=0.73μm. Therefore we find that the volume of detection is 0.09 fl. The presence of one molecule in this volume corresponds to a concentration of 18.5 nM.

A.3.12 Rate of translation

Based on both mRNA concentration measurements and dsRed concentration measurements we evaluate the mean translation rate per mRNA. Experimentally we find that the relationship between transcription and translation is linear (Figure A.8).

Therefore we can use the slope of the fitting function in FigureA.8 to infer the rate of translation per mRNA transcript, where a=0.11 for Frag1B and a=0.10 for Frag1A. Under this steady-state condition, one mRNA transcript yields about 100 dsRed tetramers (400 monomers of protein). The division time, T, of Frag1B cells growing with 10ng/ml [aTc] is T=42 min. Since $T = \frac{ln2}{\mu}$, the resulting growth rate is μ=0.016 min-1. The rate of production of dsRed monomers, P, is $\frac{dP}{dt} = k[mRNA] - \mu[P]$ where k is the protein production rate. At steady state,$k = \frac{\mu[P]}{[mRNA]}$. We find k=6.7 $dsRed$ monomers per mRNA transcript per minute. The dsRed gene codes for 220 amino acids. If we assume that only one

ribosome is loaded per mRNA message, the average transcription rate is about 24 amino acids per second. This rate is line with the transcription rates of previous studies[22, 48].

A.4 Results

A.4.1 Using FCS to monitor mRNA concentration

FCS is a technique sensitive to the diffusion coefficient of fluorescent molecules[40]. We detect dsred-ms2x2 mRNA transcripts during translation because ribosomes decrease drastically the diffusion of mRNA transcripts bound by MS2-GFP molecules. Using FCS we determine the fraction, y, of bound MS2-GFP molecules, and we infer the number of mRNA transcripts present in the detection volume. In this inducible Tet promoter system, the average [mRNA] measured across a population of individual cells increases with [aTc] (FigureA.9).

We use two different strains, a wildtype (Frag1B) and the Frag1A mutant ($\Delta acrAB$) null for multi-drug resistance efflux pumps[36]. We find that in both wild-type and mutant strains, [mRNA] increases with [aTc] in sigmoidal fashion (Figure A.9). The observed sigmoidal curves are indicative of a cooperative mechanism in the regulation of the transcription process (Hill coefficients 4). The fact that [mRNA] induction curves for strains Frag1B and Frag1A have sigmoidal curves with similar Hill coefficients but exhibit different levels of expression con-

tributes to the validation of our method. In these two strains the presence or absence of the acrAB efflux pumps affects the internal aTc concentration. The absence of the acrAB efflux pumps in strain Frag1A results in higher levels of mRNA than in Frag1B (2 fold higher for the highest aTc induction level). However, this difference between the two strains should not affect the cooperative nature of the TetR regulation system as seen in FigureA.9.

A.4.2 Simultaneous detection of protein concentration by fluorescence intensity levels

As the main validation of our [mRNA] measurements, we monitor the fluorescent intensity from the translated dsred gene. Our inducible genetic system is regulated solely at the transcriptional level, and uses TetR as repressor and aTc as inducer. Consequently, if the FCS measurements of [mRNA] are robust, the induction curve of [mRNA] and [dsRed] should exhibit a similar expression profile. As expected, the levels of mRNA (Figure 3A) and dsRed (FigureA.10) as a function of [aTc] display induction curves with comparable Hill coefficients.

Moreover, we also find that the induction curves [dsRed] in wild-type and mutant cells have similar Hill coefficients of 3 (FigureA.10, inset). The similarity between the Hill coefficients across strains, mRNA, and protein measurements constitutes an additional control for our method. FigureA.10 shows the direct relationship between mRNA and protein levels in wild-type Frag1B and mutant

Frag1A cells. Both strains display a similar linear relationship between transcription and translation, whose slope allows us to estimate a translation rate of 24 amino acids per second. This rate is in agreement with that of the visualization method for mRNA determination [22]. We previously showed that the mutant Frag1A cells expressed more proteins than wild-type Frag1B cells [34]. However, this effect was observed at much higher levels of the inducer aTc (400ng/ml) than the inducer levels used in the present study. The two-fold difference in expression between the two strains observed at the level of mRNA is also present in the levels of [dsRed] in FigureA.10.

A.5 Discussion

The combination of a genetic fusion with the use of the FCS technique demonstrates that simultaneous detection of a protein and its associated mRNA is technically possible. In this report we have extended our approach from the detection of non-coding RNA to that of a coding RNA. Detecting mRNA with FCS is potentially more difficult than the detection of non-coding RNAs. mRNA molecules are actively translated by the attached ribosomes. The number of translating ribosomes will not necessarily be limited to one per mRNA transcript. The difficulty arises from the fact that there will be a distribution of heterogeneously diffusing mRNA/MS2-GFP/ribosome complexes. We coarse-grained the complexity of this molecular system by using a single diffusion time for the mRNA in complex.

This simplification implicitly attributes to the measured fraction y the meaning of an effective translational activity. The validity of this simplification is supported by the similarity of the Hill coefficients of mRNA and protein induction curves. Presently, our ability to detect in real-time protein concentration is limited by the dsRed protein maturation time of about 80 min.[5]. The use of a faster non-fluorescent reporter, such as a luciferase, should circumvent such limitations. Imaging methods [22] are superior to FCS approaches when mRNA is localized and does not diffuse freely as it is often the case in eukaryotes. Imaging is also better suited for the detection of very low numbers of mRNA molecules per cell [22]. Conversely, in bacteria FCS-based methods for mRNA detection have two important advantages over imaging methods: FCS works over a wide range of concentrations (from nM to μM) of both mRNA and protein. Importantly, we label the mRNA transcript with only two MS2-GFP molecules, whereas imaging methods employ a large array of 96 binding sites for MS2-GFP [22]. The large size of this ms2x96 probe can greatly affect the tertiary structure as well as the degradation and diffusion characteristics of mRNAs. Therefore, the FCS based detection of mRNA is potentially less invasive than imaging based techniques. The small ms2x2 tag can be cloned in principle at the 3 end of any bacterial gene, which now makes single cell transcriptome FCS analysis technically attainable[51].

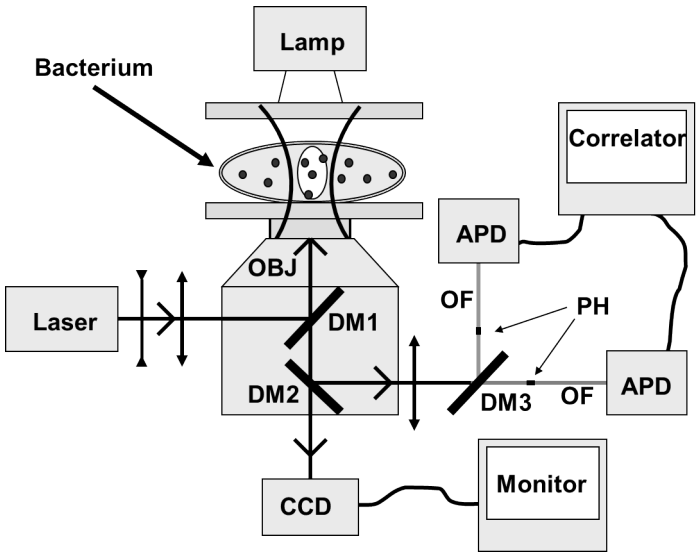


Figure A.3: Experimental FCS setup. A 488 nm laser beam is expanded using a divergent and a convergent lens. The expanded beam feeds a Olympus X71 microscope, and a dichroic mirror (DM1) reflects the laser blue light on the bacterium. The emitted green and red light from the fluorescent proteins expressed in the bacterium are transmitted through the dichroic mirror (DM1) and are reflected by a second dichroic mirror (DM2). A third dichroic mirror (DM3) splits it into a green component and a red component. The light is focused with an achromatic convergent lens (CL) onto the cores of two optical fibers (OF), which act as pinholes (ph). The fibers feed two avalanche photodiodes (APD) that produce photon counting time series. An ALV correlator connected to a computer records the time series and computes in real-time the associated autocorrelation functions. A red light (lamp) illuminates the sample from above, and a CCD camera is connected to a monitor (more details in Materials and Methods).

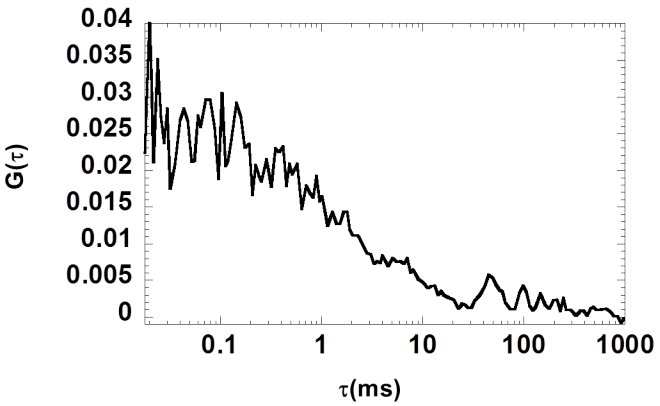
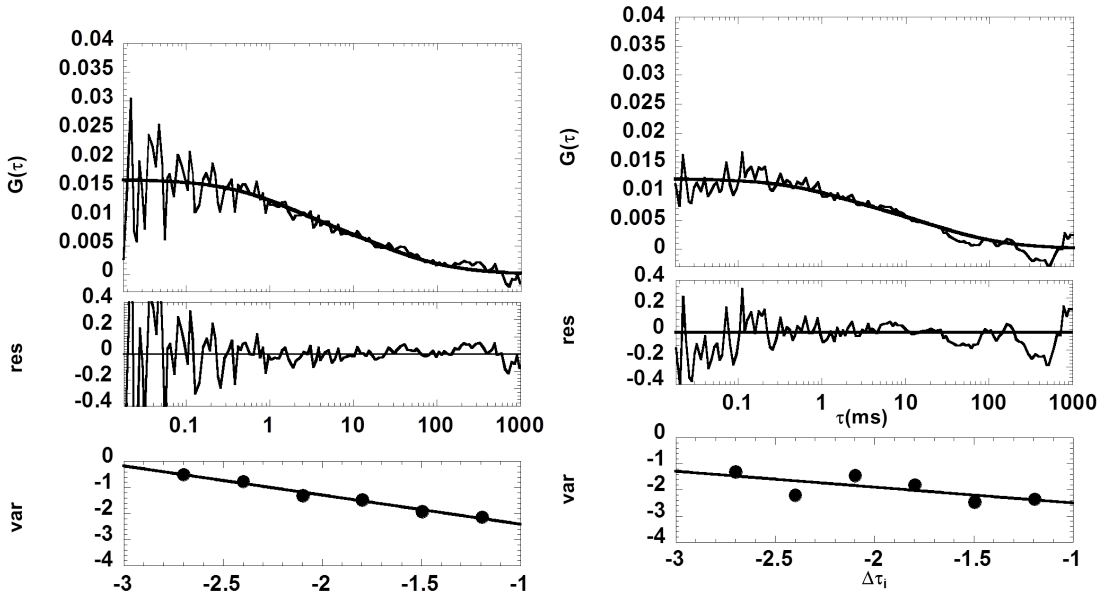


Figure A.4: Example of a typical autocorrelation curve when the positioning of the detection volume on the cell is poor. The circle on the right of the graph shows the characteristic shape of the G(t) curve when cells were not properly positioned.



(a) A (b) B

Figure A.5: Noise analysis of autocorrelation curves. Typical FCS autocorrelation curves that passed (A) or failed (B) the selection criteria discussed in Materials and Methods. On each figure, the top panel shows an autocorrelation curve. The data is fitted with the two-component function G(τ). The middle panel represents a plot of the residuals of the fit. The bottom panel depicts the value for the variance of the signal dwell time as a function of log of the dwell time $\Delta\tau(i)$. The noise analysis of each autocorrelation curve relies on three parameters, whose values determine if a curve is kept or rejected. The value of the residuals displayed in the middle panels define two parameters for the behavior at long time scales: 1) power spectrum at frequency zero, psf_0, 2) variance of residuals at times higher than 4.5 ms,psf_var. The third parameter determines the distance to the linear fit of the variance of residuals at short time scales, max_d.

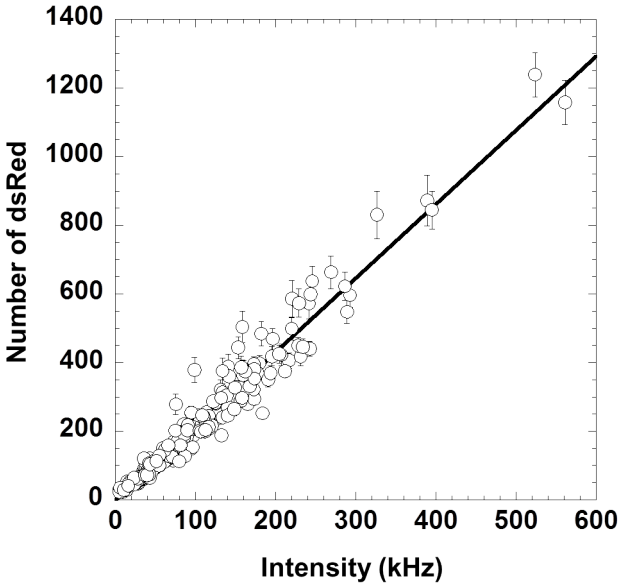


Figure A.6: Number of dsRed molecules versus intensity calibration curve. The red channel data for all the cells were analyzed using , and the ones which resulted in good quality fits were plotted against the mean intensity measured across the 3s acquisition interval. The linear fit (forced through 0) of the data gives a slope of 2.15 dsRed tetramers per 1kHz of measured signal intensity. The error bars represent uncertainties in the fit parameters of G(t).

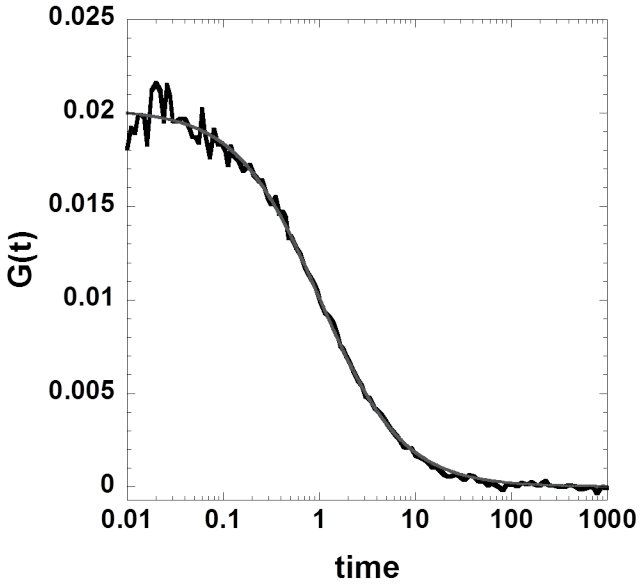


Figure A.7: FCS determination of the diffusion of polystyrene beads. The autocorrelation function $G(t)$ for 44nm diameter polystyrene beads. Fluorescence data was acquired during an interval of 30s. We focus the laser beam close to the surface of the glass coverslip as is the case for FCS measurements on bacteria. Grey line represents the fit function A.1. We perform ten distinct measurements and determine the diffusion constant of the beads to be ζ 1.01$\pm$0.05 ms.

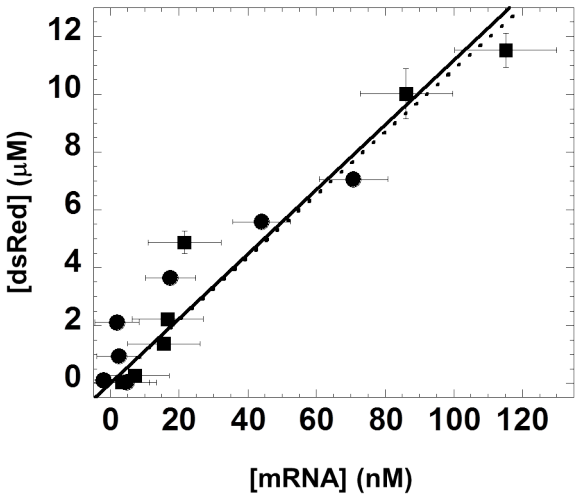


Figure A.8: Protein and mRNA concentrations within the detection volume. Frag1B (circles), Frag1A (squares). Linear fit (forced through 0): dashed lines (Frag1B) with R factor of 0.94, continuous line (Frag1A) with R factor of 0.97.

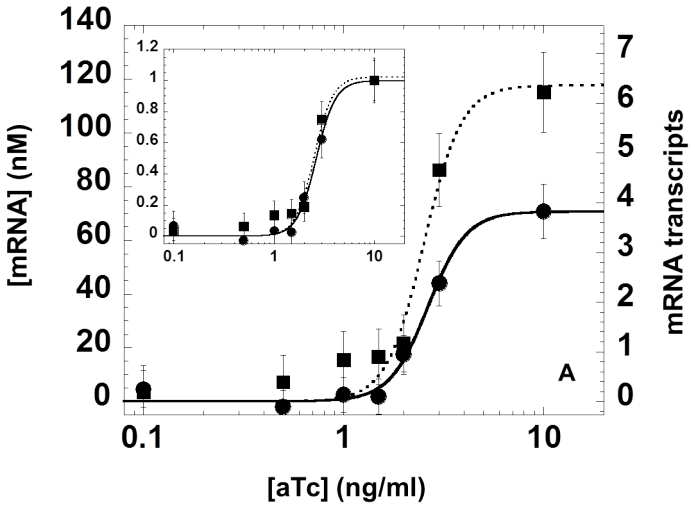


Figure A.9: [mRNA] (nM) as a function of inducer concentration. Each point represents the average mRNA concentration measured across 30-50 individual cells. Frag1B (circles and solid line), Frag1A (squares and interrupted line). The number of mRNA transcripts present in the detection volume is converted into concentration (nM). The farthest point on the left of the graph represents [mRNA] in the absence of inducer (0 [aTc]). The inset represents a Hill fit of the normalized data with a Hill coefficient 4.5±1.6 for Frag1B and 4.6±1.5 for Frag1A. Error bars represent the combination of the standard error and the systematic error. The systematic error is determined from the measured [mRNA] in a cell with no pZE31-dsRed-ms2x2 plasmid. This value was 6nM for Frag1B and 10nM for Frag1A and was subtracted from all data points.

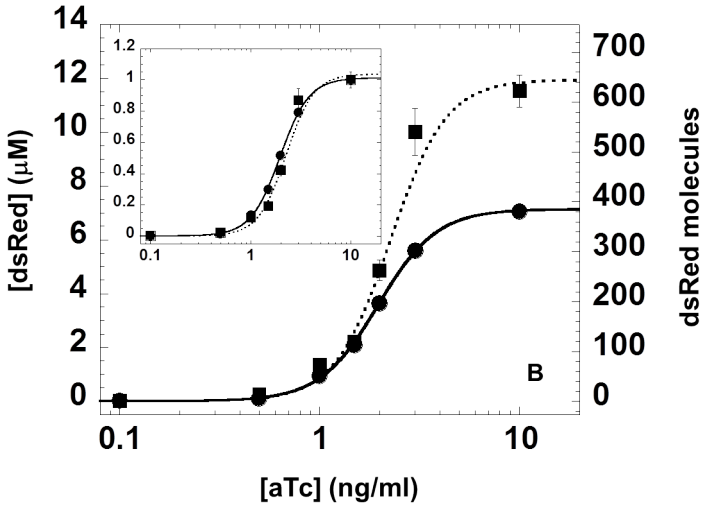


Figure A.10: dsRed tetramer concentration (μM) as a function of inducer level. Each point represents the average protein concentrations measured across the same individual cells as in A. Error bars represent the standard error. The number of dsRed tetramers in the detection volume is converted into concentration (μM) on the left side of the graph. The farthest point on the left of the graph is [dsRed] in the absence of inducer (0 aTc). The inset represents a Hill fit of the normalized data with Hill coefficient of 3$\pm$0.2 for Frag1B and 3.1$\pm$0.3 for Frag1A. Frag1B (circles and solid line), Frag1A (squares and interrupted line).

Appendix B

Supplemental Data

Table B.1: List of Conditions and Strains Used in this Study

ref	Condition	Strain	Plasmid	Marker
1	M63+Gluc.+CAA,30°C	seq.	promoterless	White Circle
2	M63+Gluc.+CAA,30°C	seq.	pfliL - YFP	White Square
3	M63+Gluc.+CAA,30°C	ins.	promoterless	White Triangle
4	M63+Gluc.+CAA,30°C	ins.	pfliL - YFP	White Diamond
5	M63+Gluc.+CAA,37°C	seq.	promoterless	Red Circle
6	M63+Gluc.+CAA,37°C	seq.	pfliL - YFP	Red Square
7	M63+Gluc.+CAA,37°C	ins.	promoterless	Red Triangle
8	M63+Gluc.+CAA,37°C	ins.	pfliL - YFP	Red Diamond
9	M9+Glyc.+CAA,30°C	seq.	promoterless	Blue Circle
10	M9+Glyc.+CAA,30°C	seq.	pfliL - YFP	Blue Square
11	M9+Glyc.+CAA,30°C	ins.	promoterless	Blue Triangle
12	M9+Glyc.+CAA,30°C	ins.	pfliL - YFP	Blue Diamond
13	M9+Glyc.+CAA,37°C	seq.	promoterless	Yellow Circle
14	M9+Glyc.+CAA,37°C	seq.	pfliL - YFP	Yellow Square
15	M9+Glyc.+CAA,37°C	ins.	promoterless	Yellow Triangle
16	M9+Glyc.+CAA,37°C	ins.	pfliL - YFP	Yellow Diamond

Table B.2: Macromolecular Values for Strains in this Study

ref	Doubling Rate(/hr)	μg protein/OD_{600})	RNA/protein
1	0.877±0.013	79.7±7.8	0.363±0.045
2	0.867±0.013	73.2±7.7	0.290±0.051
3	0.851±0.013	118.9±9.8	0.319±0.009
4	0.863±0.013	89.2±8.3	0.266±0.023
5	1.681±0.025	50.9±6.4	0.308±0.045
6	1.685±0.025	51.2±6.4	0.299±0.035
7	1.558±0.023	56.0±6.6	0.421±0.099
8	1.467±0.022	80.4±7.8	0.312±0.040
9	0.696±0.010	73.1±7.5	0.129±0.016
10	0.694±0.010	96.8±8.7	0.216±0.066
11	0.650±0.010	114.0±9.5	0.159±0.051
12	0.617±0.009	118.9±9.8	0.149±0.013
13	1.299±0.019	74.5±7.6	0.272±0.011
14	1.293±0.019	61.6±6.9	0.276±0.006
15	1.152±0.017	68.8±7.3	0.181±0.036
16	1.026±0.015	75.8±7.6	0.274±0.068

Table B.3: Unnecessary Protein Fractions for Strains in this Study

ref	YFP	Flagellin	Total Unnecessary Protein
1	0.04±0.00%	0.00±0.51%	0.04±0.51%
2	0.04±0.00%	0.00±0.51%	0.04±0.51%
3	0.04±0.00%	1.31±0.30%	1.35±0.30%
4	0.96±0.02%	1.31±0.30%	2.27±0.30%
5	0.07±0.00%	0.00±0.13%	0.07±0.13%
6	0.10±0.01%	0.00±0.13%	0.10±0.13%
7	0.07±0.01%	2.69±1.00%	2.76±1.00%
8	3.14±0.09%	2.69±1.00%	5.83±1.00%
9	0.08±0.00%	0.00±0.25%	0.08±0.25%
10	0.42±0.02%	0.00±0.25%	0.42±0.25%
11	0.13±0.01%	2.28±0.53%	2.41±0.53%
12	2.75±0.15%	2.28±0.53%	5.03±0.55%
13	0.14±0.01%	0.00±0.20%	0.14±0.20%
14	0.36±0.02%	0.00±0.20%	0.36±0.20%
15	0.13±0.03%	2.92±1.50%	3.05±1.50%
16	6.46±0.15%	2.92±1.50%	9.38±1.51%

Bibliography

[1] Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, 2000.

[2] W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol. Bioeng.*, 35(7):668–81, 1990.

[3] Howard C. Berg. The rotary motor of bacterial flagella. *Annual Review of Biochemistry*, 72(1):19–54, 2003.

[4] E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long. Localization of ash1 mrna particles in living yeast. *Molecular Cell*, 2(4):437–445, 1998.

[5] B. J. Bevis and B. S. Glick. Rapidly maturing variants of the discosoma red fluorescent protein (dsred). *Nature Biotechnology*, 20(1):83–87, 2002.

[6] C. I. BLISS. The toxicity of poisons applied jointly1. *Annals of Applied Biology*, 26(3):585–615, 1939.

[7] Tobias Bollenbach, Selwyn Quan, Remy Chait, and Roy Kishony. Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. *Cell*, 139(4):707 – 718, 2009.

[8] Dennis PP Bremer H. Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella*. American Society for Microbiology, 1996.

[9] Kenneth P. Burnham and David. R Anderson. *Model Selection and Multimodel Inference*. Springer, 1998.

[10] P. Cluzel, M. Surette, and S. Leibler. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science*, 287(5458):1652–1655, 2000.

[11] S. Cooper. The origins and meaning of the Schaechter-Maaløe-Kjeldgaard experiments. *Journal of General Microbiology*, 139:1117–1124, 1993.

[12] S. Cooper and C. E. Helmstetter. Chromosome replication and division cycle of *Escherichia Coli B/R*. *Journal of Molecular Biology*, 31(3):519–+, 1968.

[13] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–92, 2005.

[14] W. D. Donachie. Relationship between cell size and time of initiation of DNA replication. *Nature*, 219(5158):1077–+, 1968.

[15] H. Dong, L. Nilsson, and C. G. Kurland. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J Bacteriol*, 177(6):1497–504, 1995.

[16] D.A. Drummond and C.O. Wilke. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet*, 10:715–724, 2009.

[17] K. L. Elbing and R. Brent. Media preparation and bacteriological tools. *Current Protocols in Protein Science*, page A.4A.1A.4A.5., 2001.

[18] Santiago F. Elena and Richard E. Lenski. Test of synergistic interactions among deleterious mutations in bacteria. *Nature*, 390:395–398, 1997.

[19] E. L. Elson. Fluorescence correlation spectroscopy measures molecular transport in cells. *Traffic*, 2(11):789–796, 2001.

[20] E. L. Elson and D. Magde. Fluorescence correlation spectroscopy .1. conceptual basis and theory. *Biopolymers*, 13(1):1–27, 1974.

[21] I. Golding and E. C. Cox. Rna dynamics in live escherichia coli cells. *Proceedings of the National Academy of Sciences of the United States of America*, 101(31):11310–11315, 2004.

[22] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, 2005.

[23] L. Haim, G. Zipor, S. Aronov, and J. E. Gerst. A genomic integration method to visualize localization of endogenous mrnas in living yeast. *Nature Methods*, 4(5):409–412, 2007.

[24] R. J. Harvey and A. L. Koch. How partially inhibitory concentrations of chloramphenicol affect the growth of escherichia coli. *Antimicrob Agents Chemother*, 18(2):323–37, 1980.

[25] Todd G. Smith Jennifer K. Anderson and Timothy R. Hoover. Sense and sensibility: flagellum-mediated gene regulation. *Trends in Microbiology*, 18(1):30 – 37, 2010.

[26] N. Kahya, D. A. Brown, and P. Schwille. Raft partitioning and dynamic behavior of human placental alkaline phosphatase in giant unilamellar vesicles. *Biochemistry*, 44(20):7479–7489, 2005.

[27] S. Kalir, J. McClure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M. G. Surette, and U. Alon. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science*, 292(5524):2080–2083, 2001.

[28] S. A. Kim, K. G. Heinze, and P. Schwille. Fluorescence correlation spectroscopy in living cells. *Nature Methods*, 4(11):963–973, 2007.

[29] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–75, 2009.

[30] Y Komeda. Fusions of flagellar operons to lactose genes on a mu lac bacteriophage. *Journal of Bacteriology*, 150(1):16–26, 1982.

[31] O. Krichevsky and G. Bonnet. Fluorescence correlation spectroscopy: the technique and its applications. *Reports on Progress in Physics*, 65(2):251–297, 2002.

[32] K Kutsukake, Y Ohya, and T Iino. Transcriptional analysis of the flagellar regulon of salmonella typhimurium. *Journal of Bacteriology*, 172(2):741–747, 1990.

[33] T. T. Le, T. Emonet, S. Harlepp, C. C. Guet, and P. Cluzel. Dynamical determinants of drug-inducible gene expression in a single bacterium. *Biophysical Journal*, 90(9):3315–3321, 2006.

[34] T. T. Le, O. C. Guet, and P. Cluzel. Protein expression enhancement in efflux-deleted mutant bacteria. *Protein Expression and Purification*, 48(1):28–31, 2006.

[35] T. T. Le, S. Harlepp, C. C. Guet, K. Dittmar, T. Emonet, T. Pan, and P. Cluzel. Real-time rna profiling within a single bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 102(26):9160–9164, 2005.

[36] X. Z. Li and H. Nikaido. Efflux-mediated drug resistance in bacteria. *Drugs*, 64(2):159–204, 2004.

[37] R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in escherichia coli via the lacr/o, the tetr/o and arac/i-1-i-2 regulatory elements. *Nucleic Acids Research*, 25(6):1203–1210, 1997.

[38] Ole Maaløe. Role of Protein Synthesis in DNA Replication Cycle in Bacteria. *J Cell Physiol*, 62:SUPPL1:31–44, 1963.

[39] D. Magde, E. L. Elson, and W. W. Webb. Fluorescence correlation spectroscopy .2. experimental realization. *Biopolymers*, 13(1):29–61, 1974.

[40] D. Magde, W. W. Webb, and E. Elson. Thermodynamic fluctuations in a reacting system - measurement by fluorescence correlation spectroscopy. *Physical Review Letters*, 29(11):705–706, 1972.

[41] G. Meacci, J. Ries, E. Fischer-Friedrich, N. Kahya, P. Schwille, and K. Kruse. Mobility of min-proteins in escherichia coli measured by fluorescence correlation spectroscopy. *Physical Biology*, 3(4):255–263, 2006.

[42] R. Mikkola and C.G. Kurland. Evidence for demand-regulation of ribosome accumulation in *E. coli*. *Biochimie*, 73(12):1551 – 1556, 1991.

[43] J. Monod. The growth of bacterial cultures. *Ann. Rev. Microbiol*, 3:371–394, 1949.

[44] J. Monod, Jr. Pappenheimer, A. M., and G. Cohen-Bazire. The kinetics of the biosynthesis of beta-galactosidase in *Escherichia coli* as a function of growth. *Biochim. Biophys. Acta.*, 9(6):648–60, 1952.

[45] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology*, 20(1):87–90, 2002.

[46] T. Ozawa, Y. Natori, M. Sato, and Y. Umezawa. Imaging dynamics of endogenous mitochondrial rna in single living cells. *Nature Methods*, 4(5):413–419, 2007.

[47] D. S. Peabody and K. R. Ely. Control of translational repression by protein - protein interactions. *Nucleic Acids Research*, 20(7):1649–1655, 1992.

[48] S. Pedersen. Escherichia-coli ribosomes translate invivo with variable-rate. *Embo Journal*, 3(12):2895–2898, 1984.

[49] Lilia Perfeito, Stphane Ghozzi, Johannes Berg, Karin Schnetz, and Michael Lssig. Nonlinear fitness landscape of a molecular pathway. *PLoS Genet*, 7(7):e1002160, 07 2011.

[50] Macnab R. Flagella and motility. In *Escherichia coli and Salmonella*. American Society for Microbiology, 1996.

[51] Y. J. Ruan, P. Le Ber, H. H. Ng, and E. T. Liu. Interrogating the transcriptome. *Trends in Biotechnology*, 22(1):23–30, 2004.

[52] S. Saffarian and E. L. Elson. Statistical analysis of fluorescence correlation spectroscopy: The standard deviation and bias. *Biophysical Journal*, 84(3):2030–2042, 2003.

[53] Floess E Mears PJ Chemla YR Golding I Aldridge C Aldridge PD Rao CV Saini S, Koirala S. "fliz induces a kinetic switch in flagellar gene expression.".

[54] M. A. Savageau. Demand theory of gene regulation: I. Quantitative development of the theory. *Genetics*, 149(4):1665–76, 1998.

[55] R. Schleif. Control of production of ribosomal protein. *Journal of Molecular Biology*, 27(1):41–&, 1967.

[56] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–102, 2010.

[57] Hans Sorensen and Kim Mortensen. Soluble expression of recombinant proteins in the cytoplasm of escherichia coli. *Microbial Cell Factories*, 4(1):1, 2005.

[58] D. M. Stoebel, A. M. Dean, and D. E. Dykhuizen. The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics*, 178(3):1653–60, 2008.

[59] A. D. Tadmor and T. Tlusty. A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLoS Comput Biol*, 4(4):e1000038, 2008.

[60] C. Tan, P. Marguet, and L. You. Emergent bistability by a growth-modulating positive feedback circuit. *Nat Chem Biol*, 5(11):842–8, 2009.

[61] Y. Taniguchi, P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–8.

[62] M. Valencia-Burton, R. M. McCullough, C. R. Cantor, and N. E. Broude. Rna visualization in live bacterial cells using fluorescent protein complementation. *Nature Methods*, 4(5):421–427, 2007.

[63] Jesper Vind, Michael A. Srensen, Michael D. Rasmussen, and Steen Pedersen. Synthesis of proteins in escherichia coli is limited by the concentration of free ribosomes: Expression from reporter genes does not always reflect functional mrna levels. *Journal of Molecular Biology*, 231(3):678 – 688, 1993.

[64] T. Wohland, R. Rigler, and H. Vogel. The standard deviation in fluorescence correlation spectroscopy. *Biophysical Journal*, 80(6):2987–2999, 2001.

[65] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette, and U. Alon. A comprehensive library of fluorescent transcriptional reporters for escherichia coli. *Nat Methods*, 3(8):623–8, 2006.