MODELS OF NOISE PROPAGATION IN GROWING AND REGULATING BACTERIAL CELLS

Modellen van Ruispropagatie in Groeiende en Regulerende Bacteriële Cellen
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

woensdag 20 april 2022

des middags te 4.15 uur

door

Laurens Hubert Jozef Krah

geboren op 13 augustus 1992
te 's-Hertogenbosch

Promotor:

Prof. dr. R.J. de Boer

Copromotor:

Dr. R. Hermsen

Laurens H.J. Krah
Models of noise propagation in growing and regulating bacterial cells

CONTENTS

INTRODUCTION

1.1 GRASPING THE MICROBIAL WORLD

Intuitions about how the world around us works are shaped and fundamentally limited by the things we observe. However, the vast majority of organisms in the world are invisible to the naked eye. It was not until Antoni van Leeuwenhoek (1676) peered through his microscope that we first observed new, small and tumbling organisms [206]. Such organisms (viruses, bacteria, archea and certain algae species), together called microorganisms, often exist out of only a single cell and live in any thinkable (and unthinkable!) habitat on earth — deep sea [140], volcanoes, human skin [30], even space satellites [80, 105]. Microorganisms have an astonishing diversity and multitude, and, interestingly, are quite far beyond the grasp of our basic intuitions. For example, bacteria are not really affected by gravity, since the random, tiny collisions with molecules in their environment (*e.g.*, water molecules) already have a larger effect on the bacterium's movement than gravity [186]. Due to these frequent collisions with molecules, bacteria experience water as a very thick, viscous liquid. A seemingly easy task like swimming, then suddenly becomes an enormous challenge[1].

Most microorganisms have such a small volume that diffusion (the 'slow' process by which a drop of ink randomly spreads out in a glass of water) inside the cell is actually very fast. Concentrations of molecules quickly equilibrate throughout the cell's tiny volume ($\sim 1\ \mu m^3$), such that diffusion is the main driver of intracellular molecular transport [92, 93]. However, the random motion of molecules through the cell's volume still has an important impact on biochemical reactions happening inside the cell. If a reaction takes place only when two molecules meet, the reaction can happen quickly, or, by chance, happen only after a while. Reactions happening inside cells are therefore unavoidably stochastic (*i.e.*, the timing of reaction event cannot be completely controlled, but is instead based on probability). The unavoidable uncertainty in the timing of cellular processes in microorganisms is a key focus point of this thesis, and we will discuss stochasticity in more detail later.

Microorganisms live everywhere in rich diversity, often performing important and complex ecological tasks. Indeed, although they are commonly associated with a broad range of diseases, many microorganisms are actually mutualistically symbiotic with humans [8, 35]. In our gut, bacteria help with our metabolism by pre-processing certain sugars into smaller metabolites that our gut can absorb. Many microorganisms are moreover used to produce a range of useful biochemical compounds, such as penicillin, vinegar, insulin, etc., which are still hard, or expensive to synthesise in a lab or factory [33, 74, 202]. The industrial value of understanding the growth, metabolism, and biosynthesis of microorganisms is therefore immense [34, 152].

---

1 Imagine a swimming through syrup! Flapping with something like an arm will not get you anywhere: moving your 'arms' forward to make a swimming stroke will push you backwards! Instead, bacteria use an ingenious rotary mechanism to swim [14].

Still, partly due to the fact that the bacterial world is so vastly different from ours, humans lack a strong intuition of its functioning. Therefore, concise, quantitative experiments are essential, alongside (mathematical) models that can help us gain a deeper understanding of the microbial world[2].

1.2 BACTERIA: INGENIOUS TINY FACTORIES

This thesis will focus on bacteria, an immensely diverse clade of organisms, of which only < 10% can be cultured in a lab [197], meaning that the shape and functioning of a vast majority of bacterial species is completely unknown. Bacteria are small (about one-tenth the size of human cells), but are incredibly abundant —more bacteria live inside each human, than humans live on earth. Bacteria have no nucleus or other evident compartmentalisation. Instead, genetic material, as well as proteins (the cell's working units), metabolites (small building blocks) and even the ribosomes (relatively large complexes that synthesise proteins) float around in the cell. This has classically lead to the view of bacteria as 'bags of proteins', a view that is now debated [121, 142], but is a useful first approach to describing bacteria [76]. Bacteria have a surprisingly constant density (the number of proteins and metabolites, relative to the size of the bag is seemingly fixed under many external conditions, possibly even at the single-cell level [24, 50, 58, 101, 119, 201]), a feature that will return many times throughout this thesis.

Due to the complexity and overwhelming diversity of species, the workflow biologists have so far been following is to study only a few species extensively, in the hope to learn general features of all organisms. Countless biological experiments, and therefore almost all theoretical models, concern the model bacterium *Escherichia coli*. This bacteria primarily lives in the guts of animals, such as humans. There, it provides us with certain vitamins[3] (K_2 [28], B_{12} [150]). Since by far the most experiments have been done with *E. coli*, experimental tools for this particular bacterium are widely available, and new experiments are therefore again performed in *E. coli*. In turn, thanks to the overwhelming amount of *E. coli* data, most quantitative models are tailored to fit this specific bacterium as well. Still, our goal is to understand the functioning of bacteria in general, so our models will include as little biochemical detail as possible, only to be fitted to (coarse-grained) *E. coli* data in a later stage. General concepts that we learn, and intuitions that we build, hopefully apply to other bacterial species as well.

1.3 ENVIRONMENT AND GROWTH

The growth rate, the rate at which cells expand their volume and eventually divide, is a crucial physiological variable. The fitness of bacteria is to a certain extent determined by their rate of clonal expansion —genotypes that result in a higher growth rate increase in frequency in the population, increasing their chances of survival[4]. Since bacteria divide according to the adder principle (division is initiated after a fixed amount of mass is added

---

2 Such an approach to science has also been used in modern physics, where any hope of gaining intuitions seems to be abandoned and mathematics is to be followed instead.

3 Sadly, *E. coli* is still mostly known for its rare role in food poisoning, and other very rare diseases.

4 Fitness is here defined as the expected, long term number of offspring. To increase the chance of having a large number of offspring, bacteria of course need to perform other tasks as well: survive stress, heat shocks, antibiotics, etc. However, growing as fast as possible while still performing the other tasks can never hurt.
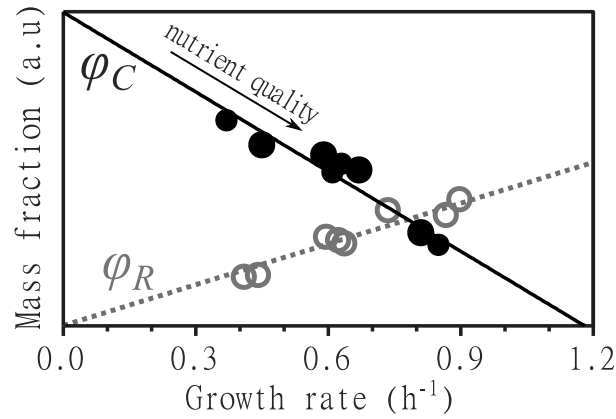
Figure 1.1: Example of two growth laws in *E. coli*. Approximately linear relationships between the growth rate in a particular environment, and the proteome fraction allocated to catabolic protein (ϕ_C, black line and dots) or to ribosomes (ϕ_R, dashed lines and grey circles). Nutrients (carbon sources) of higher quality allow for a faster growth rate, because a smaller catabolic sector is required and the concentration of ribosomes can be increased. Data from You *et al* [198].

[103, 170, 194]), bacteria that have a higher growth rate, also have a higher replication rate. Therefore, it is of crucial importance for bacteria that they are able to grow as fast as their environment and their biophysical constraints allow.

In the different environments encountered, bacteria produce different proteins. For instance, depending on the current nutrients in the environment, specific proteins are expressed to import, and metabolise the specific carbon source(s) available [65, 75, 155, 198]. However, not every carbon source provides, per imported molecule, an equal number of carbon molecules or an equal amount of energy. In any environment, bacteria thus also have to carefully tune the concentration of each protein species, to try and increase their growth rate [159, 198]. *E. coli* mainly regulates this by measuring the internal concentration of certain key metabolites[5]. The concentrations of internal metabolites inform the cell about the richness of the medium, such that regulatory metabolites, by binding to transcription factors, can tune the expression of many genes in consort [94, 136, 198]. Interestingly, *E. coli* makes use of a few very important transcription factors, so-called master regulators, that regulate enormous parts of the proteome at the same time [94, 95, 155, 198]. Indeed, the many *E. coli* genes ($\sim$ 4500) are regulated by a still large, but much smaller, number of transcription factors ($\sim$ 250) [53, 82, 95, 114].

In the light of the roughly constant density of the cell, adjusting protein concentrations with the goal to increase the cell's growth rate poses new challenges to the cell: only a limited number of proteins fit inside the 'bag of proteins'. An increase in the concentrations of particular proteins ultimately comes at the expense of others. Experiments show that this is indeed the case in populations of cells: when, in response to a poorer carbon source, the cells up-regulate proteins responsible for catabolism, the expression of other parts of the proteome (such as ribosomal proteins) becomes lower [91, 95, 155, 158, 159].

---

5 Instead, another strategy would be to have specific, external receptors for every possible environment, but this might crowd the cell wall too much, and it would cost a lot of resources to make many different receptors.

Under variation of the environment, the expression of many proteins show a simple, approximately linear, correspondence to the realised growth rate, called a growth law[6] (Fig. 1.1) [91, 155, 159]. As mentioned above, in reaction to a certain change in growth condition some protein species increase in abundance, while others decline. Classes of proteins whose concentrations show a similar relationship with the growth rate are called proteome sectors. So far, the following sectors have been observed, each showing a qualitatively different (linear) relationship with the growth rate in response to different limitations: a ribosomal sector (R), a catabolic sector (C), a nitrogen-fixating sector (N), an anabolic sector (A) and a part of the proteome that seems to have a fixed size in every environment (called Q or sometimes H sector) [68]. In Box 1, we discuss the regulatory mechanism behind the C-sector growth law: the master-regulator CRP that activates the expression of catabolic proteins by allosterically binding the metabolite cAMP. Growth laws as found in *E. coli* have also been found in other species, such as yeast [19, 88, 124], showing that, in this case, lessons learned from *E. coli* apply more generally.

Box 1. cAMP-CRP regulation. The cAMP Receptory Protein (*CRP*)[7] is a regulatory protein that, when bound to the internal metabolite cyclic adenosine monophosphate (cAMP), promotes the expression of over 100 proteins [36, 46, 54, 68, 83, 94, 95, 203]. The cAMP-CRP complex activates many important enzymes that have a role in central carbon metabolism, including all enzymes from the TCA circle, and certain steps in the glycolysis pathway. CRP-activated proteins therefore belong to the catabolic C-sector (Fig. 1.1), and sometimes the term 'C-sector' is used to describe the set of proteins that are regulated by cAMP-CRP.

The molecule cAMP itself is synthesised from the energy carrier ATP by the enzyme adenylate cyclase (*cyaA*), and actively degraded by cAMP phosphodiesterase (*cpdA*). Multiple central metabolites (mainly α-ketoacids, including oxaloacetate (OAA) and pyruvate (PYR)), themselves synthesized by enzymes that are activated by cAMP-CRP, inhibit the synthesis of cAMP [55, 198]. The cAMP-CRP network is therefore a closed, negative feedback loop: C-sector proteins are responsible for the import of carbohydrates and their conversion into smaller metabolites via the TCA cycle and glycolysis, and those metabolites inhibit the production of the C-sector (Fig. 1.2). It is this negative feedback that is thought to be responsible for the negative slope in the C-line [181, 198] (Fig. 1.1).
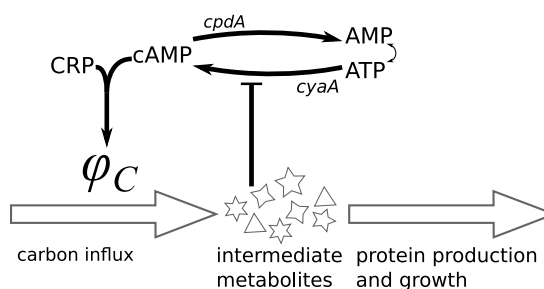
Figure 1.2: Cartoon of the cAMP-CRP regulatory network.

---

6 Not to be confused with '*the* growth law' which usually refers to the positive relationship between cell size and growth rate.

The growth laws can be understood from the point of view of resource allocation [116, 128, 159]. 'Richer' nutrient conditions require less investment in catabolic proteins (that import and degrade external carbon), because per imported molecule more energy or carbon is present, or simply because fewer enzymatic steps are required to degrade the nutrient to basic building blocks. Instead, more cellular resources can be allocated towards making new biomass (ribosomes). For each specific environment, the cell must thus balance the amount of resources it allocates towards each proteomic sector, such as the C-sector (nutrient import and catabolism) and the R-sector (ribosomes). Expressing, for example, higher concentrations of catabolic proteins might increase the import rate and the internal concentrations of metabolites, but would also result in fewer ribosomes due to cellular constraints. At some point, the ribosomes present can not match the flux generated by the C-sector, resulting in high internal metabolite concentrations that inhibit the working of the C-sector proteins. As a result, the efficiency of those C-sector proteins (flux catalysed per invested protein) decreases, and the growth rate would not be optimal. On the other hand, expressing too many ribosomes at the expense of catabolic proteins would deplete internal metabolites, causing translation to slow down and therefore an inefficient use of ribosomes.

Interestingly, multiple studies, both theoretical and experimental, suggest that for many carbon sources the particular division of resources (what part of the cell's internal metabolic flux is invested in which protein sectors) *E. coli* chooses results in the highest possible growth rate [32, 98, 128, 157, 181, 204].

1.4 STOCHASTICITY IN GENE EXPRESSION: NOISE

The seemingly perfect regulation at the population level is in sharp contrast with the inevitable randomness involved in the expression of proteins. As mentioned above, the microscopic world in general, including bacteria, is prone to stochasticity [42, 174]. The random thermal motion of all molecules causes an unavoidable uncertainty in the timing of all biochemical processes happening inside the cell: in order to react, the trajectories of two randomly diffusing molecules have to intersect in exactly the right way for the reaction to occur. Due to the small volume of bacteria, and the low copy number of many internal molecules, the stochasticity of reaction events in a cell do not simply average out, causing for example protein copy numbers to fluctuate over time [40, 86, 189]. The stochasticity in gene expression, referred to as 'noise in gene expression', has been intensely studied in the last decades, ever since the first single-cell experiments became possible [4, 25, 40, 42, 100, 168, 173, 189].

Generally, noise in protein species i is quantified in terms of the *coefficient of variation-squared* (often written as CV^2 or η^2):

$$\eta_i^2 = \frac{\text{Var}\{x_i\}}{\mathbb{E}[x_i]^2},\tag{1}$$

where x_i is the protein's concentrations and $\mathbb{E}[x_i]$ the protein's mean concentration. Context should make clear what kind of mean and variance one is currently talking about: an average over time, over a population, or over a lineage.

---

7 CRP was earlier named *CAP*: catabolite gene activator.

1.4.1   *Two ways of describing a stochastic process*

Theoretical definitions and descriptions of noise, in increasing levels of detail, have been useful to guide follow-up experiments, and to unravel the mechanisms underlying protein stochasticity. For example, during the process of creating new proteins, a distinction can be made between the stochastic waiting times between transcription events (to create mRNA), and the (exponentially distributed [25]) number of times a single mRNA molecule is translated during its lifetime (order of 3-8 minutes [11]). These processes have been neatly summarised by Friedman and Cai [47], in the following set of reactions and equations, that are explicitly solvable:

$$DNA \xrightarrow{k_1} mRNA \xrightarrow{k_2} protein, \quad mRNA \xrightarrow{\gamma_1} \emptyset, \quad protein \xrightarrow{\gamma_2} \emptyset, \tag{2}$$

$$\frac{\partial p(x)}{\partial t} = \frac{\partial [\gamma_2 x p(x)]}{\partial x} + k_1 \int_0^x w(x, x') p(x') dx'. \tag{3}$$

Here, k_1 and k_2 are the mRNA and protein synthesis rates, γ_1 and γ_2 are their decay rates[8]. The probability distribution for the protein concentration ($x = n/V$, the protein's copy number n divided by the cell's volume V) is denoted with $p(x)$, and $w(x, x')$ is the rate at which, due to translation from a single transcript, the protein concentrations jumps from x' to x, determined by the parameters γ_1 and k_2. These equations can be solved to give the steady-state probability of finding a certain protein concentration. The solution predicts a Gamma-distributed probability distribution. For many protein species in *E. coli*, the measured protein distributions —determined by measuring, via a fluorescent marker, the protein's concentration in many single cells— match the predicted distributions strikingly well [25, 173].

The above description of stochastic processes is common and intuitive, where the process is described using rates of the stochastic process (in this case k_1, γ_1, etc) that are themselves fully, deterministically, determined given the system's current state. Stochasticity arises only from the timing between (and therewith also the order of) events, and from the exact realisation of each event (*e.g.* the number of translated proteins from one transcript). If the rate is constant (*i.e.* does not depend on the system's state), the rate can straightforwardly be interpreted as the number of events per unit of time, averaged over a long period that includes many events. An equation that describes a stochastic process in terms of the time-evolution of its probability distribution is called a *Master Equation* (ME)[9] [78], a differential equation that describes how the probability distribution (deterministically) changes over time. Note that indeed the equations themselves are deterministic, although they describe a stochastic process. Of course, the specific example mentioned can be extended by allowing rates to change as some function of time and/or the system's state, or by adding cellular process in more details. Sadly, the mathematics involves becomes complicated very quickly.

A different framework and a conceptually different approach of describing (mostly continuous) stochastic systems is the use of *Stochastic Differential Equations* (SDE)[10] [147]. Here,

---

8 Note that since most proteins are stable for a long time, the protein's decay rate, γ_2, is not much higher, and thus approximately equal to, μ, the (mean) growth rate (will be made more explicit later).

9 Master equations are continuous in time, with, for biochemical processes, often discrete states, although the example presented here is a continuous ME. A (discrete) ME can be approximated by a continuous Fokker-Planck equation (FPE) that is sometimes slightly easier to solve.

10 SDE are not completely different from MEs, since for every SDE, one can write a corresponding FPE that can be seen, in turn, as an approximation of a ME.

instead of tracking individual events, the concept of an ordinary differential equation (ODE) is followed. However, rates of a process are now no longer fully determined by the system's current state, but are instead stochastic —constantly and randomly changing over time. Below is a general example of a SDE:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t. \tag{4}$$

Here, X_t is the current state of the process, which each 'timestep' (dt) is updated with a normal (deterministic) rate of change, $f(X_t, t)dt$. However, each timestep the process is also affected by a stochastic variable (scaled by $g(X_t, t)$), dW_t, called a Wiener increment. Over each time interval dt, the Wiener increment, dW_t, is a stochastic normally distributed variable with mean 0 and variance $\sqrt{dt}$, *i.e.* $dW_t \propto \mathcal{N}(0, \sqrt{dt})$. The sum of all the increments is called the Wiener process, $W(\tau) = \sum^\tau dW_t$. Box 2 describes a famous example of a SDE: the Ornstein-Uhlenbeck (OU) process. Ornstein-Uhlenbeck processes will be used often in this thesis.

Box 2. The Ornstein-Uhlenbeck Process[183]. A common Stochastic Differential Equation, used often in this thesis, describes the motion of a particle that diffuses, while it is constantly pushed back to the origin with a linear force:

$$dN_t = -\beta N_t \, dt + \theta \, dW_t. \tag{5}$$

Here, N_t is the current state of the process, which each 'time step' (dt) decays with a deterministic rate β representing the restoring force. However, each time step the process is also again disturbed by a Wiener increment. This Wiener increment, dW_t is an independent, stochastic variable that follows a normal distribution with mean 0 and variance $\sqrt{dt}$. The parameter θ sets the variance of the stochastic kicks. Still, for all values of θ, the Ornstein-Uhlenbeck process itself has mean zero and finite variance (the probability that N reaches infinity is zero, see also Fig. 1.3).

The OU Process is often used to describe a phenomenological noise source: a stochastic time series that, when added to a deterministic rate, results in a stochastic rate that randomly fluctuates over time. Note that no biochemical details are taken into account. The OU process instead simulates the net result of many stochastic process that together cause the rate of a certain process to be stochastic. Many stochastic signals have a timescale at which its fluctuations die out. The parameter β in the OU process sets this timescale.
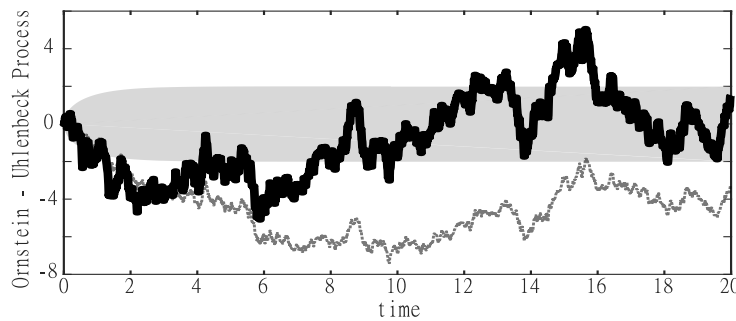
Figure 1.3: Example path of an Ornstein-Uhlenbeck Process, with $\beta = \frac{1}{2}$, $\theta = 2$, $dt = 10^{-2}$. Shaded area represents the standard deviation of many OU processes that all started at $N_{t=0} = 0$. The dashed line is the corresponding Wiener Process (W_t).

Another crucial difference between the ME-approach and the SDE-approach, is that the solution to a SDE is always a single realisation of the process (one of infinitely many trajectories that the process could possible take). Different realisations of the Wiener process, $W(t)$, will result in different values of $X(t)$. Only by then averaging over all possible trajectories can one calculate the mean $\mathbb{E}[X(t)]$, the variance $\text{Var}\{X(t)\}$, or the probability distribution $p(X, t)$. Averaging over many paths may seem a daunting task, but mathematical tricks to do this luckily exist. In the Appendix, we work out an example: given an Ornstein-Uhlenbeck noise source (Box 2) that influences the rate of protein production, what is the resulting variance and distribution of the concentration of that protein, and how does this relate to (the mean and variance of) the same system modelled according to the ME approach?

1.4.2 *Noise propagation*

Since protein content is stochastic, and proteins perform countless important tasks for the cell, noise in gene expression results in a wide range of phenotypical behaviour [113, 145]. Genetically identical cells in the same, controlled, environment, can still grow at different rates, become locked into different metabolic states, or differ in their sensitivity to antibiotics [144, 145, 171, 176, 192]. The intrinsic stochasticity of cellular reactions thus seems to have a large impact on the cell's physiology. However the opposite is also true: growth of cell volume, while protein copy numbers remain constant, results in lower (diluted) protein concentrations. Since most proteins are stable for times (much) longer than the cell cycle, most proteins leave the cell due to dilution by growth and division rather than by degradation [133]. Thus, while noise in protein concentrations can influence the cell's growth rate, growth in turn adds additional feedback to the noise in protein concentrations. Complications due to feedbacks such as this 'dilution-by-growth-feedback', has sparked the interest of many biologists, physicists and mathematicians that aim to quantify and understand cellular noise properties.

Although the mathematical definition and description of gene expression noise is quite clear, it remains difficult to understand how noise in gene expression influences (and is influenced by) downstream process such as the expression of (other) proteins, the growth rate, fitness, population fitness, etc. Note that the examples mentioned so far ignored any effect of the modelled protein on the growth rate! Until recently, models of stochasticity in gene expressions have often assumed the growth rate to be constant, although time-lapse experiments show that the growth rate of individual *E. coli* cells differs significantly between cells [61, 170], and fluctuates over time, on timescales faster than a single cell cycle [86, 189]. Additionally, noise in the expression of proteins is known to propagate through gene regulatory networks [21, 40, 139] and, via the proteins' effect on metabolism, even propagate to the cellular growth rate directly [48, 86, 171, 185]. Models that try to elucidate not only the origin of observed noise, but also the effect of noise —how noise propagates and reverberates through the cell to affect its functioning [86, 160, 178]— must thus include the biological wiring of the cell, and, ultimately a biological level of complexity.
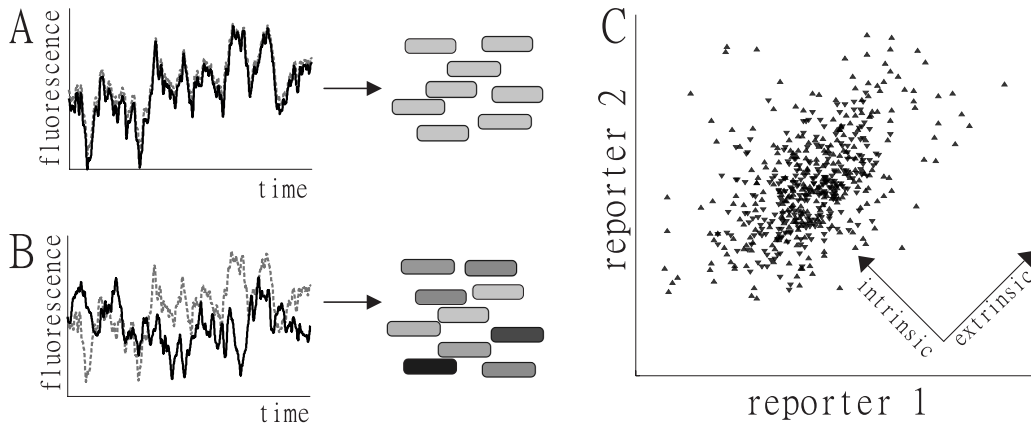
Figure 1.4: Fluorescence over time of two reporters with identical promoters whose stochastic dynamics are dominated by extrinsic noise (A), or both extrinsic and intrinsic noise (B). (C) In a scatter plot of the expression of the two reporters in a snapshot of a population of bacteria, the shape of the cloud yields information about the relative importance of extrinsic and intrinsic noise. Figure is adapted with permission from Elowitz *et al* [42].

1.4.3 *Intrinsic and extrinsic noise*

Early work decomposed the noise in the concentration of a particular protein into an intrinsic and extrinsic component. As discussed above, not all noise observed in a single protein species stems only from the stochasticity that is intrinsic to the processes of transcription and translation (intrinsic noise) of that specific protein species. Rates of transcription and translation themselves may be influenced by (noisy) factors determined by the rest of the cell, or even the cell's local environment. Additionally, imperfect (asymmetrical) cell division adds noise to the copy number of each protein. Such noise contributions are called extrinsic noise sources. Extrinsic noise acts similarly on multiple protein species, and therefore causes the noise in two similar reporters to be, to some extent, correlated. If the stochastic dynamics are mainly determined by the same, extrinsic noise sources, two fluorescent reporters with identical promoters but different colours, should still give cells of similar hue only (Fig. 1.4A). If instead noise in transcription and translation would be dominant, individual cells should display different colours, changing over time, displaying the colour of the reporter that happens to be most abundant at that time (Fig. 1.4B). This realisation has lead to the operational definition of intrinsic and extrinsic noise as respectively the decorrelated and the correlated part of the expression of two equal reporters (Fig. 1.4C). It turns out that the intrinsic noise component scales as $\eta_{in}^2 = 1/\mathbb{E}[x]$, with $\mathbb{E}[x]$ the protein's mean concentration[11] [173]. The extrinsic noise component however, adds a baseline noise level to all proteins and is dominant when the mean expression is large (and the intrinsic component therefore small). Which processes contribute to extrinsic noise, and how strong those contributions are, is a topic of ongoing discussion and study [18, 42, 87, 97, 109, 168, 177].

---

11 This scaling is reminiscent of a Poisson process, where events with a two times higher rate (in this case more transcription/translation events per unit of time) have a two times lower coefficient of variation.

1.4.4 *Cross-correlations*

When noise in the concentration of a particular protein propagates to other protein species and to the growth rate, and growth itself also affects protein expression, the cell's stochastic dynamics very quickly become complicated. One way to further dissect the noisy dynamics of gene expression and growth as they are happening inside living cells is by using time-lapse microscopy. Experimental advances have made it possible to follow a single bacterial lineage over an extended period of time while measuring reporter concentrations and the single-cell growth rate at a sub-cell-cycle time resolution [86, 167, 189]. The resulting stochastic time-traces can be analysed by using a mathematical concept called the cross-correlation function [16, 40, 59, 86, 132]. Cross-correlations capture the similarity between two time-dependent signals of which one has been shifted by a certain delay time τ. If, for example, stochastic fluctuations that first appear in the expression of protein A propagate through the cell and at a later time arrive at protein B, the cross-correlation between the concentrations $A(t)$ and $B(t)$ will have a positive value at certain positive delay times (Fig. 1.5). Mathematically, the cross-correlation between two (real-valued) signals $A(t)$ and $B(t)$ at delay time τ, $R_{A,B}(\tau)$ is similar to their convolution:

$$R_{A,B}(\tau) := \frac{1}{\sqrt{\mathrm{Var}\{A\}\,\mathrm{Var}\{B\}}} \int \left(A(t) - \mathbb{E}\left[A\right]\right)\left(B(t+\tau) - \mathbb{E}\left[B\right]\right) dt, \qquad (6)$$

where the means and variances of the signals are also calculated over time. For discrete measurements the integral turns to a summation. For finite time-series, practical problems appear when the overlap over which the correlation is calculated becomes too narrow (Fig. 1.5B). Throughout this thesis we will revisit cross-correlation functions multiple times, in order to dissect how their shapes are determined by the propagation of noise through bacterial cells.

1.5 ABOUT THIS THESIS

Up to here, this introduction has given an overview of the microbial world, of regulation in the bacterial cells at the population level, and of the inherent stochasticity at the single cell level. So far, limited by the experimental possibilities, earlier work mainly focused on noise that propagates from a single protein species. This thesis started with the idea that although the expression of only a small number of protein species can be measured at the same time, the expression of *all* protein species in the cell is stochastic. Together, all these proteins, all deviating from their regulated average, determine the current metabolic flux and growth rate. In other words, the intrinsic noise of all unmeasured proteins contributes to the extrinsic noise of a particular protein of interest. To make things more complex, all proteins again receive feedback via global (density) constraints and growth, while individual protein species might receive additional, specific regulatory feedback. To gain intuition about such a complex, noisy system with countless feedback loops, a mathematical framework can be of great use. The goal is to calculate how the noise in all individual protein species can transfer through the cell to influence the stochastic dynamics of a particular protein of interest. In this thesis we try to derive such frameworks and compare, where possible, with experimental data.

Chapter 2: Noise propagation in a holistic cell model. First, we build a general framework that allows noise propagation properties to be calculated for an arbitrary number
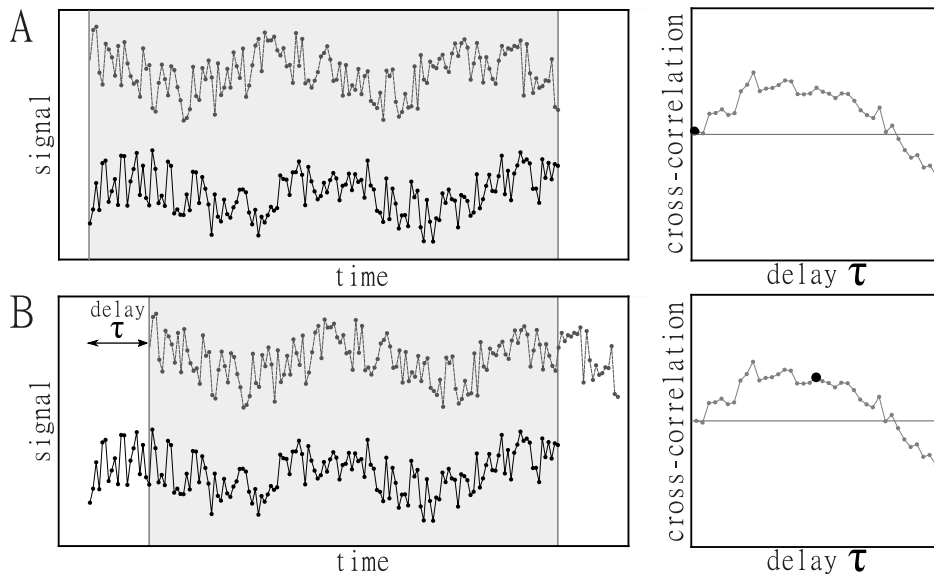
Figure 1.5: Visual explanation of the cross-correlation function between two finite, stochastic, discrete time series. Although both time series are noisy, on average the signals can correlate, or anti-correlate, depending on the chosen delay. (A) The cross-correlation without any delay ($\tau = 0$) equals the correlation between the two signals. (B) For non-zero delay times, a completely different correlation is measured between the signals (calculated over the grey area). Each possible delay time results in a correlation-value as plotted in the panels on the right. Here, the boundary effects also become tangible: for a large delay time, the correlation is calculated over very little data and can therefore become dominated by noise.

of protein species. The resulting analytical expressions predict cross-correlations between the expression of a reporter gene and the growth rate. We found important quantities that shape the cross-correlations: "Growth Control Coefficients". Growth Control Coefficients (GCCs) quantify to what extent a particular protein exerts control over the growth rate. (If a protein's GCC is large, it means that the growth rate is sensitive to the fluctuations in the concentration of that particular protein.) Further study of the GCCs and their role in the cross-correlations, shows how the noise in all proteins can affect the measured cross-correlation of the single protein that is currently tracked. Using an extensive library study, we distil noise parameters of the model from the measured noise levels of more than 1000 *E. coli* genes. Resulting cross-correlations qualitatively fit published cross-correlations in three conditions.

Chapter 3: Growth Control Coefficients are subjected to evolution. This chapter continues the study of Growth Control Coefficients. In Chapter 2 we show that the GCCs shape noise propagation properties in growing bacterial cells. Here, we argue that the GCCs themselves are subject to evolution. If a cell could grow (on average) faster by increasing the expression of a particular protein (*i.e.*, that protein has a (too) high GCC!), natural selection for fast growth rates might well result in an increased average expression, as long as no other constraints are hit. For cells that, by means of natural selection, optimised a part of their proteome for faster growth, the GCCs of all proteins can be derived formally. In such an optimum, a protein's GCC scales linearly with its average abundance. In other words: the higher the expression level of the protein, the stronger its fluctuations affect the growth

rate! We show that abundant proteins, despite their low noise levels, are then responsible for most of the variance in the cellular growth rate. This result is counter-intuitive, because a common intuition was that proteins with a *low* expression level, due to their high noise levels, are the important drivers of stochasticity at the cellular level.

Chapter 4: Noise propagation and cAMP-CRP regulation. So far, regulatory networks, and the effect they possibly have on the single-cell stochasticity, have been ignored in this thesis. This chapter therefore studies the interplay between the cAMP-CRP regulatory network and stochasticity at the single-cell level, in a fixed environment. Measurements of the stochastic dynamics of a CRP-regulated reporter, together with measurements of the growth rate, show that removal of the cAMP-CRP feedback changes the stochastic dynamics of *E. coli* cells. To explain this behaviour, a mathematical model is constructed and analysed. The model matches the experiments and confirms that the changes in stochastic dynamics are indeed caused by removing the regulation. Although the model itself describes only a small number of protein species (in contrast to the earlier, holistic cell model from Chapter 2), intuitions learned about growth control are used in this study. Again, we show an important role for fluctuations in other, not explicitly measured or modelled proteins, in this case via their effect on the metabolic flux: due to (intrinsic) stochasticity of all the metabolic proteins, the metabolic flux becomes stochastic, further transmitting the noise to other parts of the cell.

Chapter 5: Dynamical cell model that predicts noise propagation in each condition. The model presented in Chapter 4 was independently fitted to data from several conditions: to describe cells growing in different conditions, different sets of a parameters had to be found. A more logical approach would be to start with a model of gene expression and growth that includes non-linear regulation functions with which (population of) cells react to changes in their external environment. Hopefully, a stochastic extension of such a model will then also be able to describe, in a fixed environment, the stochastic fluctuations observed in individual cells. In this chapter we set the first steps in deriving a stochastic mathematical model of a growing cell that sets the correct average expression upon changing environments, but also reproduces the measured, single-cell stochastic dynamics within each fixed environment. We stay as close to (population-level) data as possible, while trying to explain the multiple experiments from Chapter 4 with a single set of parameters.

NOISE PROPAGATION IN AN INTEGRATED MODEL OF GENE EXPRESSION AND GROWTH

Istvan Kleijn*, Laurens H.J.Krah*, Rutger Hermsen

* These authors contributed equally.

ABSTRACT

In bacterial cells, gene expression, metabolism, and growth are highly interdependent and tightly coordinated. As a result, stochastic fluctuations in expression levels and instantaneous growth rate show intricate cross-correlations. These correlations are shaped by feedback loops, trade-offs and constraints acting at the cellular level; therefore a quantitative understanding requires an integrated approach. To that end, we here present a mathematical model describing a cell that contains multiple proteins that are each expressed stochastically and jointly limit the growth rate. Conversely, metabolism and growth affect protein synthesis and dilution. Thus, expression noise originating in one gene propagates to metabolism, growth, and the expression of all other genes. Nevertheless, under a small-noise approximation many statistical quantities can be calculated analytically. We identify several routes of noise propagation, illustrate their origins and scaling, and establish important connections between noise propagation and the field of metabolic control analysis. We then present a many-protein model containing > 1000 proteins parameterized by previously measured abundance data and demonstrate that the predicted cross-correlations between gene expression and growth rate are in broad agreement with published measurements.

2.1 INTRODUCTION

Few processes are more fundamental to life than the growth and proliferation of cells. Bacterial cells in particular are highly adapted to grow rapidly and reliably in diverse habitats [70]. Yet, the composition of individual bacteria grown in a constant environment is known to fluctuate vigorously, in part due to the stochastic nature of gene expression [42, 86, 123, 174]. Many experimental and theoretical studies have shed light on the origins, characteristics and consequences of this "noisy" expression [21, 25, 42, 47, 86, 104, 123, 137, 139, 149, 161, 168, 174, 176, 195, 200]. Still, it remains unknown to what extent, and by what routes, noise in gene expression propagates through the cell and affects the rate of growth [61, 86, 178], which is often considered a proxy for its fitness [17, 178].

Recently, important progress towards understanding noise propagation in single cells has been made through experiments in which the instantaneous growth of individual *Escherichia coli* cells was monitored in real time under fixed growth conditions [86, 170]. Such experiments have revealed large fluctuations in the growth rate, with coefficients of variation of the order of 25%, which in part result from noise in the concentrations of metabolic enzymes [86]. Conversely, growth-rate fluctuations affect the concentrations of individual enzymes, because the cell's constituents are diluted whenever the cell grows [182]. Such results emphasize that a clear understanding of these processes is complicated by the fact that gene expression, metabolism, and growth are highly interdependent, involving multiple layers of feedback and cellular constraints.

This interdependence is also central to a series of recent studies that characterize the *average* composition and growth rate of *Escherichia coli* cultures in balanced exponential growth under variation of the growth medium [68, 89, 91, 112, 155, 172, 198]. In particular, these experiments have revealed striking linear relations between their mean proteomic composition and their mean growth rate [65, 68, 89, 155, 158, 198]. Phenomenological models have demonstrated how such "growth laws" can be understood as near-optimal solutions to constrained allocation problems [17, 116, 128, 157]. These results also stress that global physiological variables and constraints strongly affect the expression of individual genes. As such, both these experiments and the single-cell experiments mentioned above suggest a "holistic" perspective: the behavior of individual components cannot be understood without some knowledge of the cell's global physiological state [12, 160].

Here, we present a model of bacterial cells growing under fixed external growth conditions, in which gene expression, metabolism and growth are fully integrated. We offer a highly simplified description that nevertheless imposes several essential global cellular constraints. Both gene expression and growth rate fluctuate due to the stochastic synthesis of many protein species that together control the rates of metabolism and growth. Conversely, the rate of metabolism constrains the protein synthesis rates and the growth rate sets the dilution rate of all proteins. As a result, noise in the expression of each gene propagates and affects the expression of every other gene as well as the growth rate—and *vice versa*.

Below, we first introduce the generic modeling framework and its assumptions. We then make an excursion to the theory of growth control, in order to define growth-control coefficients and establish connections between the propagation of noise and the field of Metabolic Control Analysis. Next, we discuss how the concentration of each protein is affected by the synthesis noise in all other proteins; this exposes a hidden assumption in a standard operational definition of intrinsic and extrinsic expression noise. We subsequently
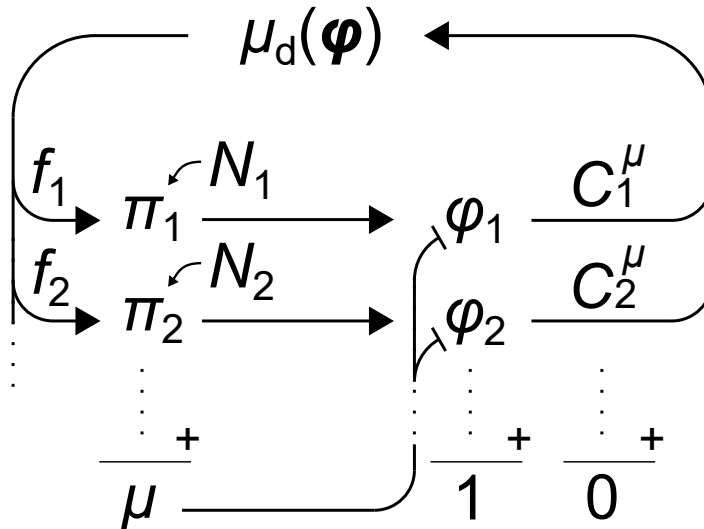
Figure 2.1: Integrated model of stochastic gene expression and cell growth. The cell contains many protein species, with proteome mass fractions ϕ_i that sum to 1. Mass fractions are increased by protein synthesis but diluted by growth. The synthesis rate π_i of each species i is modulated by a noise source N_i. The instantaneous growth rate μ reflects the total rate of protein synthesis. Proteins affect metabolism and thus the deterministic growth rate $\mu_d(\vec{\phi})$, as quantified by growth-control coefficients C_i^μ. A fraction f_i of the total metabolic flux is allotted to the synthesis of protein i. The inherent noise in the expression of each gene reverberates through the cell, affecting cell growth and the expression of every other gene.

explain the noise modes that characterize the noise propagation between gene expression and growth in the context of a toy model with just two proteins. Lastly, we present a many-protein model that includes 1021 protein species with experimentally measured parameters. We demonstrate that the cross-correlations functions between expression and growth rate predicted by this model capture the main features of published measurements.

2.2 RESULTS

2.2.1 *Modeling framework*

We here discuss the key assumptions of the modeling framework (Fig. 2.1); see SI, section S2.1 for details. We consider a culture of bacterial cells that has reached steady-state exponential growth under fixed external growth conditions. We study fluctuations of gene expression within individual cells in this steady state, and in particular how these fluctuations reverberate through the growing cell. Similar assumptions connecting the increase in biomass, the cellular growth rate, protein synthesis, and growth-mediated dilution were explored in a recent review article [76].

The mass density of *E. coli* cells is dominated by protein content [20] and under tight homeostatic control [101]. We assume that this homeostasis also eliminates long-lived protein-density fluctuations in single cells. Then, the volume of a cell is proportional to its protein mass $M := \sum_i n_i$, where n_i is the abundance (copy number) of protein i. (We ignore that different proteins have different molecular weights.) The instantaneous growth

rate is then defined by $\mu := \dot{M}/M$, and the proteome fraction $\phi_i := n_i/M$ of enzyme i measures its concentration. Differentiation of ϕ_i with respect to time then yields

$$\dot{\phi}_i = \pi_i - \mu\phi_i, \tag{2.1}$$

where π_i is the synthesis rate per protein mass. (Here we neglect active protein degradation, which on average amounts to about 2% of the dilution rate [122].) By definition, proteome fractions obey the constraint $\sum_i \phi_i = 1$. Combined with equation 2.1 this results in

$$\mu = \sum_i \pi_i. \tag{2.2}$$

That is, the growth rate equals the total rate of protein synthesis.

Another key assumption of our model is that the cellular growth rate is an *intensive* quantity. That is: given fixed mass fractions, the growth rate does not depend on the cell size, as suggested by the observation that individual *E. coli* cells grow approximately exponentially within their cell cycle [71, 86]. Based on this, we express the synthesis rate of protein i as:

$$\pi_i = f_i\,\mu_d(\vec{\phi}) + N_i, \tag{2.3}$$

in which

$$\mu_d(\vec{\phi}) := J/M. \tag{2.4}$$

The first term in equation 2.3 is an intensive function; it captures the deterministic effect of the cellular composition $\vec{\phi} = (\phi_1, \phi_2, \ldots)$ on the metabolic flux J that quantifies the rate of biomass production, normalized by the protein mass M. (Note that, here and below, we use the term metabolism in a broad sense; it is intended to encompass all catabolic and anabolic processes required for biomass production and cell growth, including protein synthesis.) The coefficients f_i specify which fraction of this flux is allocated towards the synthesis of protein species i. Because the f_i are fractions, $\sum_i f_i = 1$.

The second term of equation 2.3 couples each synthesis rate π_i to a zero-mean Ornstein–Uhlenbeck noise source N_i that represents the stochasticity of both transcription and translation [40]. Each noise source is characterized by an amplitude θ_i and a rate of reversion to the mean β_i; the latter's inverse β_i^{-1} characterizes the time scale of intrinsic fluctuations in π_i. The variance of N_i is given by $\mathrm{Var}(N_i) = \theta_i^2/(2\beta_i)$. All noise sources are mutually independent, and we neglect other sources of noise, such as the unequal distribution of molecules over daughter cells during cell division (see Discussion).

Combining equation 2.2 and 2.3 reveals that

$$\mu = \mu_d(\vec{\phi}) + \sum_i N_i, \tag{2.5}$$

which identifies $\mu_d(\vec{\phi})$ as the growth rate afforded by a given proteome composition $\vec{\phi}$ in the zero-noise limit. Given a function $\mu_d(\vec{\phi})$, equations 2.1–2.4 fully define the dynamics of the cell.

Below, we focus on the simplest case where, under given environmental conditions, the allocation coefficients f_i are constant. This means that the cell does not dynamically adjusts its allocation in response to fluctuations in expression levels. We note, however, that such dynamical effects of gene regulation could be included by allowing the f_i to depend on intra- and extra-cellular conditions, and in particular on the cellular composition $\vec{\phi}$. (See SI,

sections S2.1.6 and S2.4.2) We also stress that the allocation coefficients may differ strongly between growth conditions, as demonstrated by the growth laws mentioned above. For example, the f_i's of ribosomal proteins must be considerably larger in media that support a fast growth rate than in media with strong nutrient limitation, because the mean mass fraction of ribosomal proteins increases with the growth rate [158]. Here, however, we describe stochastic cell growth under fixed environmental conditions, so that the (mean) allocation of resources is well-defined and knowable in principle—for example through proteomics data.

Fig. 2.1 is an illustration of the modeling framework. Noise in the synthesis of a protein species induces fluctuations in its mass fraction (equation 2.1). Through their effect on metabolism, these fluctuations propagate to the deterministic growth rate μ_d, which modulates the synthesis of all protein species (equation 2.3). In parallel, all noise sources directly impact the growth rate μ (equation 2.5) and thus the dilution of all proteins (equation 2.1).

Linearization under a small-noise approximation

The results below rely on the assumption that equations 2.1–2.5 may be linearized around the time-averaged composition $\vec{\phi_0}$. This transforms equation 2.5 to

$$\frac{\delta\mu}{\mu_0} = \sum_i C_i^\mu \frac{\delta\phi_i}{\phi_{0,i}} + \sum_i \frac{N_i}{\mu_0}, \tag{2.6}$$

where $\delta\phi_i$ is the deviation of ϕ_i from its time average $\phi_{0,i}$ and $\delta\mu$ the deviation of μ from $\mu_0 := \mu_d(\vec{\phi_0})$. (See SI, section S2.1 for derivations.) The coefficients C_i^μ are defined as

$$C_i^\mu := \left[\frac{\phi_i}{\mu_d} \frac{\partial\mu_d}{\partial\phi_i} \right]_{\vec{\phi_0}}. \tag{2.7}$$

In the terminology of linear noise models, the C_i^μ are transfer coefficients: they quantify to what extent fluctuations in ϕ_i transmit to μ_d. Equation 2.6 demonstrates that the growth rate is affected by all noise sources, both directly (second term on the right-hand side) and indirectly through fluctuations in the protein mass fractions.

2.2.2 *Transfer coefficients are growth-control coefficients*

The transfer coefficients C_i^μ are reminiscent of the logarithmic gains defined in biochemical systems theory, which relate enzyme abundances to the metabolic flux in a given pathway [154]. It has previously been shown that these gains are relevant in the context of noise propagation [138]. Here, however, we consider the growth rate of the cell rather than the flux through a distinct pathway. In this section, we connect the transfer coefficients C_i^μ to the control of cellular growth and the field of Metabolic Control Analysis (MCA) [63, 77].

In MCA, flux-control coefficients (FCCs) C_i^J are defined that quantify to what extent an enzyme concentration ϕ_i limits (controls) a metabolic flux J [63, 77]:

$$C_i^J := \left[\frac{\phi_i}{J} \frac{\partial J}{\partial\phi_i} \right]_{\vec{\phi_0}}. \tag{2.8}$$

In direct analogy to this definition of FCCs, the transfer coefficients of equation 2.7 can be interpreted as growth-control coefficients (GCCs) that quantify each enzyme's control of

the growth rate. From equation 2.4 a direct link between FCCs and GCCs can be derived (see also [196], SI section S2.3.1, and Fig. S2.1):

$$C_i^\mu = C_i^J - \phi_i. \tag{2.9}$$

The GCCs are specified by the sensitivity of the growth rate $\mu_d(\vec{\phi})$ to changes in the proteome composition $\vec{\phi}$, evaluated in the steady-state mean, $\vec{\phi}_0$. Both the mean composition $\vec{\phi}_0$ and the function μ_d clearly differ between growth conditions; therefore, the GCCs depend on the growth conditions as well.

As mentioned, studies on the resource allocation of cells grown under different growth conditions have revealed striking empirical relations between the mean proteome composition and the mean cellular growth rate [68, 155, 158, 198]. Even though these growth laws describe relations between growth rate and composition, they should not be confused with μ_d. The growth laws describe correlations between the mean composition and the mean growth rate under variation of the growth conditions, whereas μ_d describes the deterministic effect of the instantaneous composition on the instantaneous growth rate under a particular, fixed growth condition. There is no direct relation between the two. By extension, the growth laws do not directly translate into knowledge on the GCCs.

Growth-control coefficients and their sum rule

An important difference between metabolic flux and cellular growth rate lies in their behavior under a scaling of the system size. It is routinely assumed that metabolic fluxes scale linearly with the system size, meaning that an increase in the abundances of all enzymes by a factor α increases the metabolic flux J by the same factor α. That is, fluxes are *extensive* variables. Based on this assumption, a famous sum rule has been derived for FCCs [63, 77]:

$$\sum_i C_i^J = 1. \tag{2.10}$$

In contrast, we assumed the growth rate to be invariant under scaling of the system size, *i.e*, that the growth rate is an *intensive* variable. (Indeed, as equation 2.4 directly shows, if J is extensive, μ_d must be intensive, and *vice versa*.) Under this assumption, GCCs obey a markedly different sum rule:

$$\sum_i C_i^\mu = 0. \tag{2.11}$$

This sum rules articulate a delicate trade-off: the excess of one protein implies the lack of another.

Both sum rules are special cases of Euler's homogeneous function theorem. Specific derivations are presented in SI, section S2.3.2. In general, for an arbitrary function f with a scaling relation $f(\alpha\vec{\phi}) = \alpha^k f(\vec{\phi})$, a sum rule can be derived by differentiating this equation with respect to α and evaluating the result in $\alpha = 1$. The particular cases $k = 1$ (for the flux J), and $k = 0$ (for the growth rate μ_d) lead to equations 2.10 and 2.11.

In theory, all expression levels could be regulated such that $C_i^\mu = 0$ for all protein species i. In reality, however, many protein species do not have a function within metabolism or biomass growth. By definition, the metabolic flux J does not depend on the expression levels of these proteins; therefore, their FCCs are zero. The GCC of such a protein, with mass fraction ϕ_h, then follows from equation 2.9:

$$C_h^\mu = -\phi_h. \tag{2.12}$$

That is, the control of all non-metabolic enzymes on the growth rate is negative. The sum rule then implies that the sum of GCCs of all proteins that do contribute to biomass growth must be positive and equal to

$$\sum_{i \notin H} C_i^\mu = - \sum_{h \in H} C_h^\mu = \sum_{h \in H} \phi_h = \phi_H. \tag{2.13}$$

where H denotes the set of non-metabolic proteins. This goes to show that any system that bears the cost of producing non-metabolic proteins must contain other proteins that have positive growth control.

This conclusion has implications for the propagation of noise. We saw that the the noise transfer coefficients appearing in the linear noise model are in fact GCCs. The analysis in the previous paragraph demonstrates that these GCCs cannot all vanish; it then follows that there must be linear-order noise transfer from protein levels to the growth rate in all cells that maintain non-metabolic proteins.

Non-metabolic proteins are common, both in wild-type cells and in engineered constructs. In wild-type *E. coli*, the expression level of proteins that do not contribute to biomass growth were estimated recently in a study that combined a genome-scale allocation model with proteomics data sets [135]. Direct estimates of ϕ_H ranged from 25% to 40%, depending on the precise growth conditions. Although not directly beneficial to the growth of the cell in constant environments, the non-contributing proteome fraction is thought to provide fitness benefits to cells that encounter frequent changes in growth conditions [135]. Furthermore, synthetic biologists commonly study systems with a large expression burden [15].

2.2.3 *Separating in- and extrinsic noise components*

Within the above framework, many statistical properties can be calculated analytically [40, 86]. In particular, the noise level of the concentration of protein i, quantified by the coefficient of variation η_i, can be expressed as:

$$\eta_i^2 = \frac{(1 - \phi_{0,i})^2}{\phi_{0,i}^2} \frac{\text{Var}(N_i)}{\mu_0(\mu_0 + \beta_i)} + \sum_{j \neq i} \frac{\text{Var}(N_j)}{\mu_0(\mu_0 + \beta_j)}. \tag{2.14}$$

The derivation is provided in the SI, section S2.2.2. Equation 2.14 shows that the coefficient of variation has two components: the first term results from the noise in the synthesis of the protein itself, the second from the noise in the synthesis of all other proteins. Each term is proportional to the variance of the corresponding noise source, but weighted by a factor that decreases with the mean growth rate μ_0 and the reversion rate β_i of that noise source. This analysis confirms that the inherent noise in the synthesis of one protein affects all other proteins.

A fundamental distinction is commonly made between intrinsic and extrinsic noise in gene expression [138]. Intrinsic noise results from the inherently stochastic behavior of the molecular machinery involved in gene expression; extrinsic noise from fluctuations in the intra- and extracellular environment of this machinery. In this sense, the two terms in equation 2.14 can be identified as intrinsic and extrinsic contributions.

Complications arise, however, if the standard operational definition of these terms is applied [42, 168]. This definition considers two identical reporter constructs R and G expressed in the same cell (Fig. 2.2A). Noise sources extrinsic to both reporters affect both
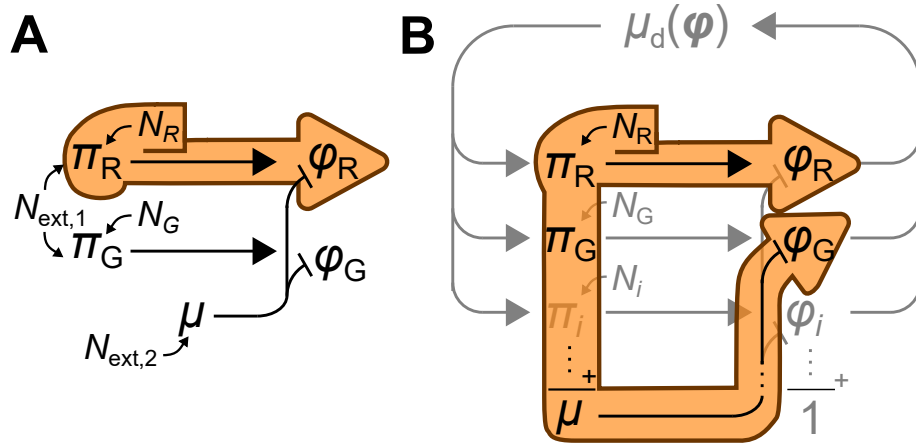
Figure 2.2: Limitations of the operational definition of in- and extrinsic expression noise.(*A*) Extrinsic noise is measured by the covariance between the expression levels of two identical reporter systems R and G. This presupposes that the intrinsic noise N_R of system R affects concentration ϕ_R but not ϕ_G (orange outline), so that the covariance between ϕ_R and ϕ_G quantifies the contribution of extrinsic sources $N_{ext,i}$. (*B*) But in our model, N_R affects the growth rate and thus the dilution of ϕ_G. This adds a negative term to the covariance, which no longer measures just the extrinsic noise.

reporters identically, inducing positively correlated fluctuations in the concentrations of the reporter proteins. Intrinsic noise sources instead produce independent fluctuations in each concentration. Extrinsic noise is therefore measured by the covariance between both expression levels; intrinsic noise by their expected squared difference. This operationalization, however, implicitly assumes that intrinsic noise does not propagate between the reporters. This assumption is violated in our model because the synthesis of reporter R directly contributes to the dilution of protein G (Fig. 2.2B). Consequently, the covariance between the expression levels has two contributions:

$$\frac{\text{Cov}(\phi_R, \phi_G)}{\phi_{0,b}^2} = \underbrace{-\frac{2(1-\phi_{0,b})}{\phi_{o,b}}\frac{\text{Var}(N_b)}{\mu_0(\mu_0 + \beta_b)}}_{\text{transmission between R and G}} + \underbrace{\sum_{j \neq R,G} \frac{\text{Var}(N_j)}{\mu_0(\mu_0 + \beta_j)}}_{\text{other sources}}, \qquad (2.15)$$

where the label "b" indicates quantities that are by definition identical for both expression systems. The second term on the right-hand side is positive and stems from noise sources that affect both reporters identically. The first term, however, is negative; it reflects the transmission of noise between reporters R and G. It would be misleading to identify equation 2.15 as the extrinsic component of the noise—it is not even guaranteed to be positive. We conclude that the operational definition is not suitable when noise propagates between arbitrary genes.

2.2.4 *Expression–growth correlations in a two-protein toy model*

The circulation of noise in the cell can be studied by measuring cross-correlations between expression and growth rate in single-cell experiments [86]. Interpreting measured cross-correlations, however, is non-trivial. To dissect them, we now discuss a toy version of the

model with just two protein species, X and Y. Despite its simplicity, it displays many features seen in more realistic models.

Within the linear noise framework, ϕ_Y–μ and π_Y–μ cross-correlations, respectively denoted $R_{\phi_Y\mu}(\tau)$ and $R_{\pi_Y\mu}(\tau)$, can be calculated analytically [40]. Up to a normalization, the results can be written as:

$$R_{\phi_Y\mu}(\tau) \propto \underbrace{C_Y^\mu S_Y(\tau)}_{\text{Control}} + \underbrace{\phi_{o,Y}A_Y(\tau)}_{\text{Autogenic}} - \underbrace{\sum_{j=X,Y} \phi_{0,j}\left[C_j^\mu S_j(\tau) + \phi_{0,j}A_j(\tau)\right]}_{\text{Dilution}}; \qquad (2.16)$$

$$R_{\pi_Y\mu}(\tau) \propto \underbrace{C_Y^\mu A_Y(-\tau)}_{\text{Control}} + \underbrace{\phi_{o,Y}B_Y(\tau)}_{\text{Autogenic}} + \underbrace{\sum_{j=X,Y} C_j^\mu\left[C_j^\mu S_j(\tau) + \phi_{0,j}A_j(\tau)\right]}_{\text{Transmission}}. \qquad (2.17)$$

(For a full derivation, not limited to the two-protein case, see SI, sections S2.1-S2.2. The two-protein case is discussed further in SI section S2.4) These equations are plotted in Fig. 2.3A and B (see caption for parameters). As the equations show, the cross-correlation functions are linear combinations of three functions $S_i(\tau)$, $A_i(\tau)$, and $B_i(\tau)$, which are also illustrated in the figure.

To aid interpretation, the cross-correlations can be decomposed into four noise modes, as indicated in equations 2.16 and 2.17.

The **control mode** (Fig. 2.3C) reflects the control of enzyme Y on the growth rate. Noise N_Y in the synthesis of Y causes fluctuations in ϕ_Y, which transfer to the growth rate in proportion with the GCC C_Y^μ. Because the effect of ϕ_Y on μ is instantaneous, the contribution to the ϕ_Y–μ cross-correlation is proportional to the *symmetric* function $S_Y(\tau)$. In contrast, the effect of π_Y on μ involves a delay; hence the contribution to the π_Y–μ cross-correlation is proportional to the *asymmetric* function $A_Y(\tau)$. In both cases, the amplitude scales with C_Y^μ.

The **autogenic mode** (Fig. 2.3D) is a consequence of equation 2.2. Because the growth rate matches the total rate of protein synthesis, noise in the synthesis of Y instantly affects the growth rate, resulting in a noise mode in the π_Y–μ cross-correlation that is proportional to the symmetric function $B_Y(\tau)$. With a delay, this noise also affects ϕ_Y, adding an asymmetric mode to the ϕ_Y–μ cross-correlation. This mode does not depend on the control of Y; instead, its amplitude is proportional to the mean concentration $\phi_{o,Y}$.

The **dilution mode** (Fig. 2.3E) pertains only to the ϕ_Y–μ cross-correlation. It reflects that the growth rate of the cell is also the dilution rate of protein Y (equation 2.1). With a delay, upward fluctuations in μ therefore cause downward fluctuations in ϕ_Y. A subtle complication is that noise in the synthesis rate of both proteins reaches μ via two routes: through the immediate effect of π_Y on μ, and through the delayed effect of π_Y on ϕ_Y, which in turn affects μ in proportion with C_Y^μ (see in equation 2.6). Together, these routes result in a mode towards which each protein contributes both a symmetric and an asymmetric function.

Lastly, the **transmission mode** (Fig. 2.3F) is unique to the π_Y–μ cross-correlation. It reflects that all noise sources affect the cell's composition $\vec{\phi}$ and therefore μ_d; this in turn induces fluctuations in the synthesis rate π_Y. The noise sources again affect the growth rate via the two routes explained above, causing a symmetric and an asymmetric component to the π_Y–μ cross-correlation for each protein.
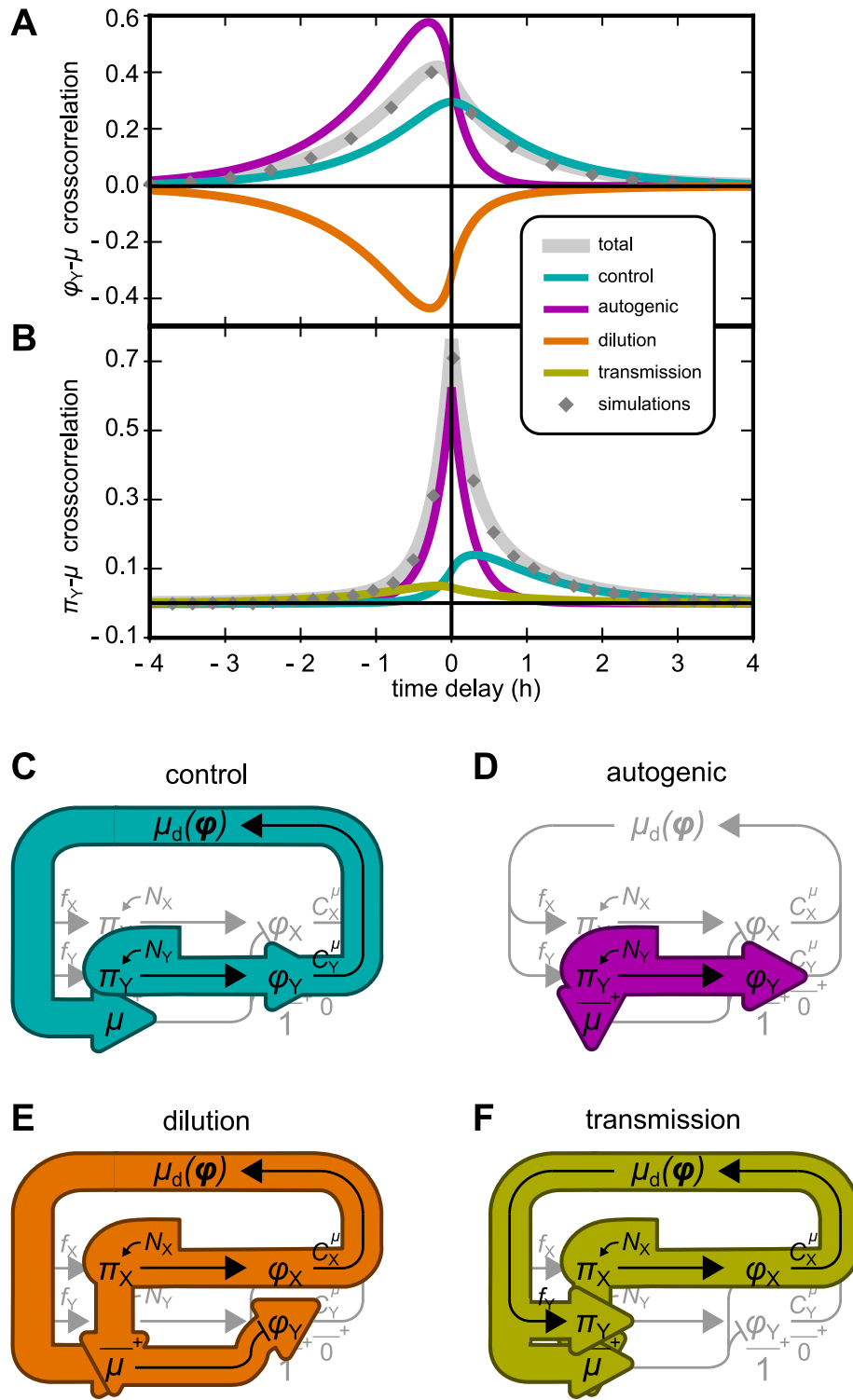
Figure 2.3: Noise modes in a toy model containing only two protein species, X and Y. (*A*): Analytical solution for the cross-correlation between protein Y's proteome fraction ϕ_Y and growth rate μ (gray curve), verified by simulations (gray diamonds, details in SI, section S2.4). The contributing noise modes are indicated (colored curves). (*B*): Same as (A), but for the synthesis rate π_Y. The cross-correlation functions are linear combinations of three classes of functions, called $A_i(\tau)$, $B_i(\tau)$, and $S_i(\tau)$ (see SI, section S2.2, equations S2.47–S2.49 for their definitions). In panels (A) and (B), noise modes that are proportional to just one of these functions are annotated accordingly. (*C*)–(*F*): Noise propagation routes underlying the noise modes. The control mode and the autogenic mode arise from noise source N_Y alone. Both noise sources N_X and N_Y contribute to the dilution and transmission modes, but only the contribution of N_X is illustrated in Fig. (D) and (F). Parameters for (A) and (B): $C_Y^\mu = 0.25$; $\phi_{o,Y} = 0.33$; mean growth rate $\mu_0 = 1\ \mathrm{h}^{-1}$; noise sources of N_Y and N_X have amplitudes $\theta_Y = 0.5$ and $\theta_X = 0.5$ and reversion rates $\beta_Y = \beta_X = 4\mu_0$.

The above analysis shows that, even in a highly simplified linear model, the cross-correlations are superpositions of several non-trivial contributions. The intuitions gained from this exercise will be used below when we present the results of a more complex model.

The effects of gene regulation

Above, we assumed that the cell allocates a fixed fraction f_i of its metabolic flux towards the synthesis of protein i. Within this two-protein model all cross-correlations can still be computed if the f_i are linear(ized) functions of the concentrations $\vec{\phi}$ (see SI, sections S2.1.6 and S2.4.2, and Fig. SS2.2). The resulting feedback regulation affects the decay of fluctuations: a negative feedback shortens the correlation time scales and reduces variance, whereas positive feedback lengthens them and increases variance (*cf.* [39, 47, 174]).

2.2.5   *Expression–growth correlations in a many-protein model*

In single *E. coli* cells, the cross-correlations between gene expression and growth rate have been measured by Kiviet *et al*[86]. To test whether the above framework can reproduce their results, we constructed a model that includes 1021 protein species with realistic parameters.

In the experiments, micro-colonies of cells were grown on lactulose (a chemical analog of lactose) and expression of the *lac* operon was monitored using a green fluorescent protein (GFP) reporter inserted in the operon. Because intrinsic fluctuations in GFP expression affect the cross-correlations directly as well as indirectly, through their impact on the growth rate and the expression of other genes, we modeled this reporter construct explicitly (see Fig. 2.4A, and SI, section S2.5). Specifically, the *lac* operon O was represented as a collection of three proteins Y, Z, and G (for LacY, LacZ, and GFP) affected by a shared noise source N_O in addition to their private sources N_Y, N_Z, and N_G. The GCC of the operon as a whole is the sum of the GCCs of its genes.

By varying the mean expression of the *lac* operon with a synthetic inducer, Kiviet *et al* measured cross-correlations in three growth states with different macroscopic growth rates: "slow", "intermediate", and "fast" [86]. Empirically, the macroscopic growth rate obeyed a Monod law [129] as a function of the mean *lac* expression. We therefore mimicked the three growth states by choosing their mean *lac* expression levels and growth rates according to three points on a Monod curve that approximates the empirical one (Fig. 2.4B, labels D, E, and F). Via equation 2.7, the same curve also is also used to estimate the GCC of the *lac* operon in each condition. Under "slow" growth conditions, the *lac* enzymes limit growth considerably (large GCC); under "fast" conditions, *lac* activity is almost saturated (small GCC).

To choose realistic parameter values for all other proteins, we used a published dataset of measured means and variances of *E. coli* protein abundances [173]. For each of the 1018 proteins in the dataset, the model included a protein with the exact same mean and variance (see Fig. 2.4C). This uniquely fixed the amplitudes of all noise sources. The GCCs of all proteins were randomly sampled from a probability distribution that obeyed the sum rule of equation 2.11.[1] (See Materials and Methods, and SI, section S2.5.2)

---

1 More about this distribution in Chapter 3.
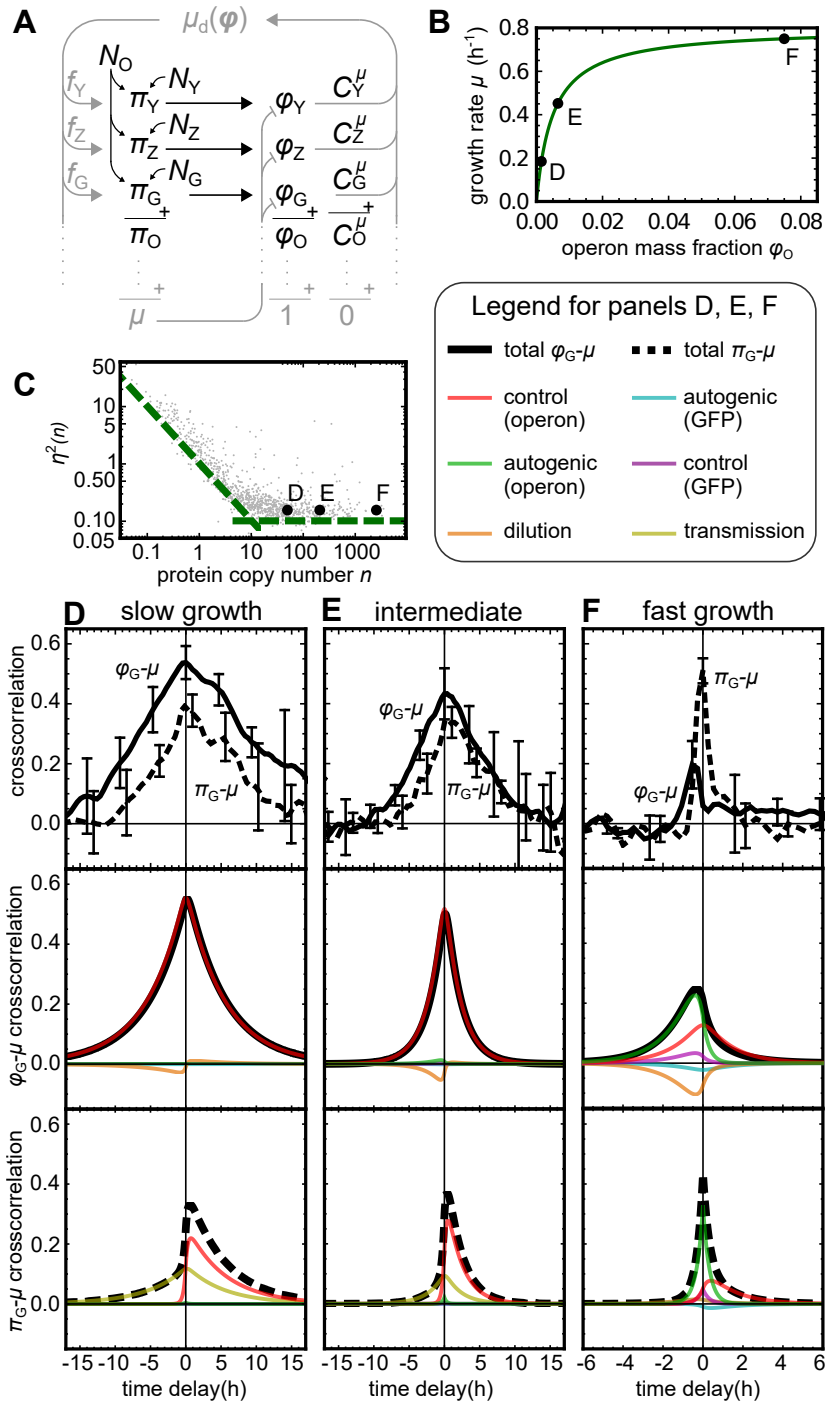
Figure 2.4: (*A*) Cartoon of the noise propagation network. (*B*) Monod curve describing the mean growth rate as a function of *lac* expression. Black dots indicate the operon mass fractions and growth rate used to calculate the cross-correlations in (D)-(F). (*C*) Noise distribution of the proteome (gray cloud) taken from Ref. [173], and the values chosen for proteins on the *lac* operon (black dots). Green dashed lines are guides for the eye. (*D*)-(*F*) Experimental [86] (top panels) and theoretical (middle and bottom panels) cross-correlations for three growth conditions. Proteome fraction–growth and production–growth cross-correlations are plotted as solid and dashed black lines, respectively. As in Fig. 2.3A and B, coloured lines show the contributing noise modes.

Comparison with measured cross-correlations

The experimental results on the cross-correlations between GFP synthesis π_G, GFP expression ϕ_G, and growth rate μ [86] are reproduced in Fig. 2.4D–F (top panels), together with the model predictions (middle and bottom panels).

The predicted cross-correlations are linear superpositions of the same noise modes as described for the two-protein model. However, the dilution and transmission modes are now driven by all 1022 noise sources, and there are two instances of the control and autogenic modes: one associated with the expression and GCC of the operon as a whole, and one with the expression and GCC of GFP separately. (See SI, equations S2.89–S2.94, section S2.5.)

At slow growth, the $\phi_G - \mu$ cross-correlation is almost symmetrical (Fig. 2.4D, middle panel). Here the control mode of the operon dominates due to its large GCC. At higher growth rates, the autogenic modes become more prominent because their amplitudes are proportional to the expression level of the *lac* genes; at the same time, the amplitudes of the control modes decrease with the GCCs (Fig. 2.4F). As a result, the cross-correlation becomes weaker and more positively skewed.

At slow growth, the π_G–μ cross-correlation is negatively skewed because the operon control mode is dominant (Fig. 2.4D, bottom panel). It also shows a notable transmission mode. With increasing growth rate, the autogenic modes increase in importance, which narrows the peak, increases its height, and reduces its asymmetry. The patterns seen in both cross-correlations are in good qualitative agreement with the experimental data (top panels).

Alternative dataset, similar results

In the dataset that we used to parameterize protein expression, the abundances are consistently low compared with other studies [1, 155]. However, an alternative analysis based on different abundance data [1] and sampled variances [195] yielded similar results (SI section S2.5.3, and Fig. S2.3). We conclude that the qualitative trends are insensitive to the precise dataset used.

2.3 DISCUSSION

We have presented a model of stochastic cell growth in which the growth rate and the expression of all genes mutually affect each other. Systems in which all variables communicate to create interlocked feedback loops are generally hard to analyze. Analytical results were obtained by virtue of stark simplifying assumptions. Nevertheless, the predicted and measured cross-correlations have similar shapes and show similar trends under variation of the growth rate.

That said, a few differences are observed. Chiefly, at slow and intermediate growth rates the model consistently underestimates the decorrelation timescales (peak widths). In the model, the longest timescale is the doubling time; this timescale is exceeded in the experimental data. This suggests a positive feedback that is not included in the model, possibly as a result of gene regulation (also see Fig. S2), or else a noise source with a very long auto-correlation time.

Alongside their measurements, Kiviet *et al* published their own linear noise model, which fits their data well. In fact, the shapes of the noise modes emerging in that model are mathematically identical to those presented above [40]. Yet, the models differ strongly in their setup and interpretation. Kiviet *et al* model a single enzyme E that is produced and diluted by growth. It features only three noise sources: one directly affects the production of E ("production noise"), one the growth rate μ ("growth noise"), and one affects both simultaneously ("common noise"). While these ingredients are sufficient to fit the data, the interpretation and molecular origins of the common and growth noise are left unspecified. In our model, which includes many proteins, similar noise modes emerge without explicit growth or common noise sources. Each enzyme perceives fluctuations in the expression of *all* genes as noise in the growth rate; this results in a dilution mode similar to that of Kiviet *et al*. Furthermore, noise in the synthesis of each enzyme instantaneously affects the growth rate (equation 2.2) due to the assumed homeostatic control of protein density. Hence, this noise behaves as a common noise source, which explains why the autogenic mode is mathematically identical to the common-noise mode of Kiviet *et al*. We conclude that noise in the expression of many enzymes, combined with homeostatic control of protein density, can contribute to the observed but unexplained common- and growth-noise modes.
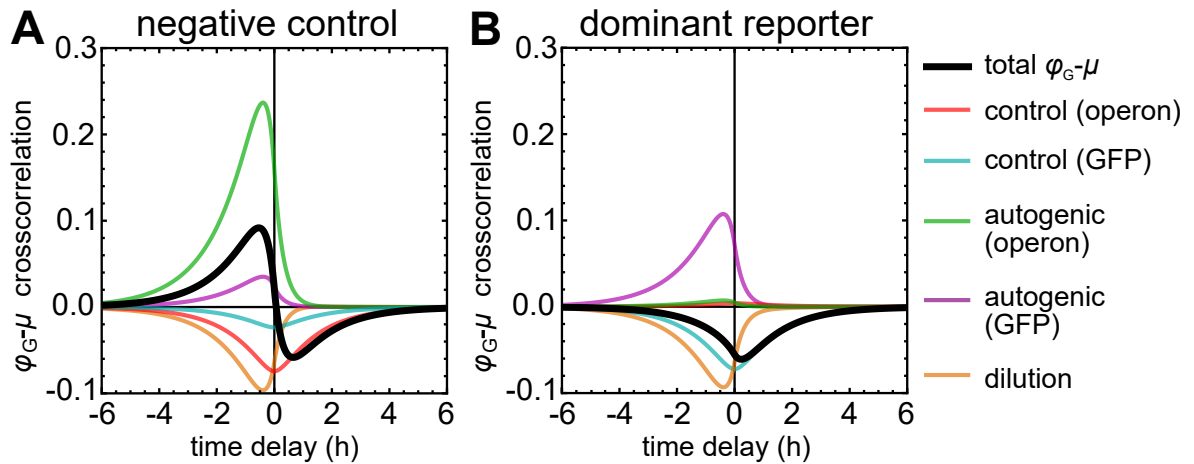
Figure 2.5: Deceptive concentration–growth cross-correlations. (A) Positive Pearson correlation despite a negative operon GCC, due to a dominant autogenic mode. Same parameters as Fig. 2.4F, but with $C_O^\mu = -0.035$. (B) Negative Pearson correlation despite a positive operon GCC, due to noisy GFP expression. Same parameters as Fig. 2.4F, but with operon noise much smaller than GFP noise (see Materials and Methods).

Control coefficients are routinely used in metabolic control analysis [45, 63, 77, 130] and have also been studied in the context of evolutionary optimization [10, 196]. In our linearized model, GCCs emerged as transfer coefficients, indicating that these quantities also affect the propagation of noise. Conversely, this suggests that GCCs could be inferred from noise-propagation measurements. For example, the Pearson correlation coefficient (cross-correlation at zero delay) between ϕ_i and μ might be used as an indication of control. However, we have seen in Fig. 2.3 that the φ–μ correlation involves several noise modes that are independent of the GCC. As a result, the signs of the Pearson correlation and the GCC do not necessarily agree (see Fig. 2.5A). In addition, the intrinsic noise and GCC of the reporter protein can result in a negative cross-correlation even if the operon's control is positive (Fig. 2.5B). Alternatively, the asymmetry of the control mode in the π–μ cross-

correlation could perhaps be exploited [86] (Fig. S2.4). Unfortunately, this asymmetry is also affected by other modes, such as the transmission mode, which can mask the effect (Fig. S4C). We conclude that, in any case, such results have to be interpreted with great caution, ideally guided by a quantitative model.

Future theoretical work should aim to relax assumptions and remove limitations. The assumed strict control of protein density can be relaxed by allowing density fluctuations. If these are long-lived, they will likely weaken the autogenic mode and introduce new modes of their own. Also, additional noise sources can be included that do not stem directly from protein synthesis. In particular, we ignored noise originating from cell division despite its importance [52, 149, 188]. In addition, gene regulation will affect some noise modes; this can be studied by allowing the f_i to depend on $\vec{\phi}$. It will also be interesting to include non-protein components of the cell, such as RNAs.

A further caveat is that the linear approximation used here is only reasonable if the noise is sufficiently weak. In fact, in the presence of strong non-linearities, the approach may even break down completely. For instance, it has been shown that cellular growth can be stochastically arrested when an enzyme whose product is toxic to the cell is expressed close to a threshold beyond which toxic metabolites build up to lethal doses [146]. In such circumstances, expression level noise in those enzymes can have a highly nonlinear effect on the cellular growth rate, resulting in subpopulations of growth-arrested cells [146]. That said, under more ordinary conditions linear models that describe noise in cellular networks have previously been used to great success [40, 86].

Throughout this document we have considered noise sources that act on each production rate independently. Alternatively, one could hypothesize that the observed fluctuations in protein concentrations might instead originate from noise in the allocation of the flux—that is, from fluctuations in the allocation coefficients f_i. This would be expected under the supposition that ribosomes are always fully occupied and translating at a constant, maximal rate, so that the relative rates of protein synthesis are determined solely by competition between different mRNAs based on their relative abundances and their translation initiation rates. Protein synthesis rates then become intrinsically correlated: an increase in the synthesis rate of one protein requires an simultaneous decrease in the synthesis rates of other proteins. In future work, such alternative models could be explored in detail. Preliminary simulations, however, show a striking symmetry in the ϕ_i–μ cross-correlation and a consistent asymmetry in the π_i–μ cross-correlation (for details see SI, section S2.6, and Fig. S2.5). This can be understood as follows. If an increase in a particular synthesis rate is always compensated by a decrease in other production rates, the noise does not affect the sum of all production rates nor the growth rate instantaneously. Therefore, no autogenic mode should be present. Notably, in our model it is the autogenic mode that explains the asymmetry in the measured ϕ_i–μ cross-correlations as well as the dominant symmetric mode in the π_i–μ cross-correlations under the fast growth condition. We conclude that noise on flux allocation alone cannot readily explain these experimental findings and additional noise sources would have to be included, such as the common noise as defined by Kiviet *et al*.

Lastly, we hope that this work will inspire new experiments that can confirm or falsify the assumptions and results presented above. In particular, single-cell measurements of mass-density of protein-density fluctuations [58, 119] could establish whether our assumption of density homeostasis is warranted. Also, additional single-cell measurements could determine whether expression noise indeed propagates between reporter proteins, adding

to their covariance, and whether the amplitude of the various noise modes scales with the GCCs and mass fractions as predicted.

Material and Methods

We here specify the parameters used for the many-protein model; see also SI section S2.5.2.

Growth rates and protein abundances. The Monod curve (Fig. 2.4B) is given by $\mu_0 = \mu_{max}\phi_{o,O}/(\phi_{half} + \phi_{o,O})$, with μ_0 the mean growth rate, $\phi_{o,O}$ the mass fraction of the *lac*-operon proteins, $\mu_{max} = 0.8\ \mathrm{h}^{-1}$, and $\phi_{half} = 0.005$. The three growth states correspond to three points on this curve, with $\phi_{o,O}/\phi_{half} = \{0.3, 1.3, 15\}$; this mass is shared equally among proteins Y, Z, and G. The mass fractions of the remaining proteins matched the proportions of the dataset [173].

Ornstein–Uhlenbeck noise sources. The amplitudes of all noise sources were uniquely fixed by the constraints that (i) the CV of each Lac protein was 0.15, (ii) the amplitude of N_O was 1.5 times that of N_G [42], and (iii) all other CVs agreed with the dataset [173]. All noise reversion rates were set to $\beta = 4\mu_{max}$.

GCCs. To select the GCCs, we first randomly assigned proteins ($\approx$ 25% of the total mass) to the non-metabolic sector H. After the *lac* reporter construct was added, the GCC of each protein $h \in H$ was set by equation 2.12. In each growth state, the GCC of the *lac* operon was calculated from the Monod curve, which yielded $C_O^\mu = \{0.77, 0.43, 0.063\}$. Assuming GFP is non-metabolic and the GCCs of Y and Z are equal, we set $C_G^\mu = -\phi_{o,G}$ and $C_Y^\mu = C_Z^\mu = \left(C_O^\mu - C_G^\mu\right)/2$. The GCCs of all other proteins were sampled from a probability distribution that respects equation 2.11 and assumes that proteins with a larger abundance tend to have a larger GCC (see SI section S2.5.2).

Author contributions. Conceptualization: ITK, LHJK, RH. Formal analysis: ITK, LHJK, RH. Supervision: RH. Visualization: ITK, LHJK, RH. Writing – original draft: ITK, RH. Writing – review & editing: ITK, LHJK, RH.

SUPPLEMENTARY INFORMATION CHAPTER 2

In this part, we provide detailed derivations and analyses pertaining to the results presented in the main text.

We aim to describe the stochastic behavior of a cell under conditions of steady-state exponential growth. A cell contains many protein species. Each of these may affect growth to some extent via an effect on metabolism. (Note that, here and below, we use the term metabolism in a broad sense; it is intended to encompass all catabolic and anabolic processes required for biomass production and cell growth, including protein synthesis.) As mentioned in the main text, our results are based on several crucial assumptions. In particular, we assume that

- the protein density of the cell is under tight homeostatic control,

- protein synthesis is inherently stochastic,

- the cellular growth rate is an *intensive*[1] quantity.

These assumptions are used repeatedly in the text below.

S2.1 DERIVATION OF THE LINEAR NOISE MODEL

We start with a detailed derivation of the linear noise model.

S2.1.1 *Abundances, volume, and growth rate*

The abundances of the protein species are given by the vector $\vec{n}(t) = \{n_1(t), n_2(t), \ldots\}$, where $n_i(t)$ is the copy number of protein species i at time t. We define the total protein mass of the cell by

$$M := \sum_i n_i, \tag{S2.1}$$

where we ignore that different proteins have different molecular weights. (To avoid clutter, we left out the explicit time dependence of the variables; below, we do so whenever this is unlikely to cause confusion.) Under the assumption that the protein density of a cell is tightly controlled, M is proportional to the volume of the cell. Therefore, we define the instantaneous growth rate μ as the rate at which M increases:

$$\mu := \frac{1}{M}\frac{dM}{dt}. \tag{S2.2}$$

---

[1] In this context, an intensive quantity is a quantity that does not change if the size of the cell is changed, provided the concentrations of all enzymes are kept constant. By contrast, an *extensive* quantity is a quantity that, under the same conditions, increases linearly with the size of the cell. It follows that any quantity that is a function of intensive quantities only, is itself an intensive quantity. Also, the ratio of two extensive quantities is an intensive quantity.

s2.1.2 *Proteome fractions and their dynamics*

The proteome fractions of the different protein species in the cell are given by the vector

$$\vec{\phi} := \frac{\vec{n}}{M}. \tag{S2.3}$$

Again, assuming that the mass M is proportional to the cellular volume, we can interpret the mass fractions as concentrations.

We denote the rate of protein synthesis by the vector $\vec{p}(t) = \{p_1(t), p_2(t), \ldots\}$ and assume that, under conditions of balanced exponential growth, degradation of proteins is negligible compared to protein dilution due to growth [122]. Then the time derivative of equation S2.3 yields the following system of dynamical equations:

$$\frac{d\vec{\phi}}{dt} = \vec{\pi} - \mu \vec{\phi}, \tag{S2.4}$$

where the vector $\vec{\pi} := \vec{p}/M$ is the protein synthesis rate per unit of protein mass.

By definition, the mass fractions add up to one:

$$\sum_i \phi_i = 1. \tag{S2.5}$$

Therefore, an increase in the concentration of one protein must go at the expense of the concentration of another. Applied to equation S2.4, this constraint also implies that

$$\mu = \sum_i \pi_i. \tag{S2.6}$$

That is, the growth rate is equal to the total rate of protein synthesis per unit of mass.

In reality, even if protein density is under strict control, it is likely to show fluctuations on short time scales. In mammalian cells, mass fluctuations have indeed been measured, but appeared to dissipate on time scales of less than $20\,\mathrm{s}$ [119]. By comparison, the auto-correlation times of the expression and growth-rate fluctuations observed in *E. coli* cells are very long—they are of the order of hours. On this time scale, we therefore expect equation S2.6 to be a reasonable approximation.

s2.1.3 *Stochastic protein-synthesis rate*

The protein synthesis rate of each protein species i is assumed to have the following form:

$$\pi_i = f_i J(\vec{n})/M + N_i. \tag{S2.7}$$

The first term on the right-hand side models the deterministic dependence of the protein synthesis rate on the protein content of the cell. The function $J(\vec{n})$ represents the global metabolic flux that fuels protein synthesis and growth. The vector $\vec{f} = \{f_1, f_2, \ldots\}$ indicates which fraction of this metabolic flux is allocated towards the synthesis of each of the protein species. As such, it obeys the constraint

$$\sum_i f_i = 1. \tag{S2.8}$$

In general, $\vec{f}$ is itself likely to be a function of intra- and extracellular variables, including the protein content of the cell, due to the dynamical regulation of gene expression. Further below we will reflect on this general case, but, unless indicated otherwise, we focus on the special case where $\vec{f}$ is constant.

The second term on the right-hand size of equation S2.7 models the stochasticity of protein synthesis. Each N_i is a colored Ornstein–Uhlenbeck noise source with zero mean, reversion rate β_i, and amplitude θ_i, obeying the following stochastic differential equation (Itō):

$$dN_i(t) = -\beta_i N_i \, dt + \theta_i \, dW_i(t), \tag{S2.9}$$

where each $W_i(t)$ is a Wiener process. Importantly, all noise sources are assumed to be mutually independent.

s2.1.4 *Stochastic growth*

Combined, equations S2.6 and S2.7 yield the following expression for the growth rate μ:

$$\mu = J(\vec{n})/M + \sum_i N_i. \tag{S2.10}$$

The first term, $J(\vec{n})/M$, is a function of the protein content of the cell; we call it $\mu_d(\vec{n})$. The assumption that the growth rate is an intensive quantity translates to the requirement that

$$\mu_d(\alpha\,\vec{n}) = \mu_d(\vec{n}), \tag{S2.11}$$

for any $\alpha > 0$. Because μ_d is intensive and M is extensive, the metabolic flux $J(\vec{n}) = M\mu_d(\vec{n})$ must be extensive, *i.e.*:

$$J(\alpha\,\vec{n}) = \alpha J(\vec{n}), \tag{S2.12}$$

for any $\alpha > 0$. This is indeed routinely assumed in the field of metabolic control analysis [77]. Using equation S2.11, we note that $\mu_d(\vec{n}) = \mu_d(\vec{n}/M) = \mu_d(\vec{\phi})$, so that equations S2.7 and S2.10 can be rewritten as

$$\pi_i = f_i\,\mu_d(\vec{\phi}) + N_i, \tag{S2.13}$$

$$\mu = \mu_d(\vec{\phi}) + \sum_i N_i. \tag{S2.14}$$

Equation S2.14 shows that all noise sources affect the growth rate in two ways: directly through the second term, and indirectly through the first, because of fluctuations in the composition $\vec{\phi}$.

Provided a function $\mu_d(\vec{\phi})$ is given, equations S2.4, S2.13 and S2.14 together fully specify the dynamics of $\vec{\pi}$, $\vec{\phi}$, and μ.

s2.1.5 *Linearization*

To obtain analytical results, we assumed that fluctuations in the concentrations $\vec{\phi}$ are small enough to warrant a linear approximation. Below, we describe the linearization of equations S2.4, S2.13, and S2.14 around the time average $\vec{\phi_0}$ of $\vec{\phi}$.

We start with equation S2.4. We introduce transformed variables:

$$\delta\vec{\phi} := \vec{\phi} - \vec{\phi_0}, \tag{S2.15}$$

$$\delta\vec{\pi} := \vec{\pi} - \vec{\pi_0}, \tag{S2.16}$$

$$\delta\mu := \mu - \mu_0, \tag{S2.17}$$

where $\mu_0 := \mu_d(\vec{\phi_0})$ and $\vec{\pi_0} := \vec{f}\mu_0$. In terms of these variables, equation S2.4 reads:

$$\frac{\delta\dot{\phi}_i}{\mu_0\phi_{0,i}} = \left(\frac{f_i}{\phi_{0,i}} - 1\right) + \left(\frac{f_i}{\phi_{0,i}}\right)\frac{\delta\pi_i}{\pi_{0,i}} - \frac{\delta\mu}{\mu_0} - \frac{\delta\phi_i}{\phi_{0,i}} - \frac{\delta\phi_i}{\phi_{0,i}}\frac{\delta\mu}{\mu_0}. \tag{S2.18}$$

To proceed, we exploit that $f_i = \phi_{0,i}$ to linear order. To see why, we first observe from equation S2.13 that

$$f_i = \frac{\mathbb{E}[\pi_i]}{\mathbb{E}\left[\mu_d(\vec{\phi})\right]}, \tag{S2.19}$$

where $\mathbb{E}[X]$ denotes the steady-state expectation value of variable X. Next, equation S2.14 implies that

$$\mathbb{E}\left[\mu_d(\vec{\phi})\right] = \mathbb{E}[\mu], \tag{S2.20}$$

and equation S2.4 that

$$\mathbb{E}[\pi_i] = \mathbb{E}[\mu\,\phi_i]. \tag{S2.21}$$

Therefore,

$$\frac{f_i}{\phi_{0,i}} = \frac{\mathbb{E}[\mu\,\phi_i]}{\phi_{0,i}\mathbb{E}[\mu]} = 1 + \mathrm{Cov}\left(\frac{\delta\phi_i}{\phi_{0,i}}, \frac{\delta\mu}{\mathbb{E}[\mu]}\right), \tag{S2.22}$$

which is 1 up to quadratic corrections. We can estimate the size of these corrections (the covariance term) as follows. We write:

$$\mathrm{Cov}\left(\frac{\delta\phi_i}{\phi_{0,i}}, \frac{\delta\mu}{\mathbb{E}[\mu]}\right) = R\,\eta_{\phi_i}\eta_\mu, \tag{S2.23}$$

where R is the correlation coefficient between ϕ_i and μ, and η_x is the coefficient of variation of variable x. Taking rather large estimates for each, we arrive at $R\,\eta_{\phi_i}\eta_\mu \lesssim 0.5 \times 0.4 \times 0.25 = 0.05$. This shows that quadratic corrections are small and we can indeed assume that $f_i = \phi_{0,i}$ to good approximation. (We also verified this using simulations; see below.) Based on this, equation S2.18 takes the following form after linearization:

$$\frac{\delta\dot{\phi}_i}{\mu_0\phi_{0,i}} = \frac{\delta\pi_i}{\pi_{0,i}} - \frac{\delta\mu}{\mu_0} - \frac{\delta\phi_i}{\phi_{0,i}}. \tag{S2.24}$$

We now move to equation S2.13. After linearization, it can be written as:

$$\frac{\delta\pi_i}{\pi_{0,i}} = \sum_j C_j^\mu \frac{\delta\phi_j}{\phi_{0,j}} + \nu_i, \tag{S2.25}$$

where we introduce growth-control coefficients (GCCs), defined as

$$C_j^\mu := \left[\frac{\phi_j}{\mu_d}\frac{\partial\mu_d}{\partial\phi_j}\right]_{\vec{\phi_0}}. \tag{S2.26}$$

(See section S2.3.1 for a further discussion of these quantities.) The Ornstein–Uhlenbeck noise terms $\nu_i := N_i/\pi_{0,i}$ are identical to the original noise terms N_i, except that their noise amplitudes ϑ_i are rescaled as $\vartheta_i = \theta_i/\pi_{0,i}$. Equation S2.25 shows that the synthesis of a protein species is affected by its own noise source, but also by fluctuations in the concentrations of all other proteins, in as far as these fluctuations affect the growth.

These same methods can be applied to equation S2.14, the linearized version of which is given by

$$\frac{\delta\mu}{\mu_0} = \sum_i C_i^\mu \frac{\delta\phi_i}{\phi_{0,i}} + \mathcal{N}. \tag{S2.27}$$

Here we introduced the shorthand

$$\mathcal{N} = \sum_i f_i \nu_i, \tag{S2.28}$$

which is a weighted sum of all noise sources.

We now use equations S2.25 and S2.27 to rewrite equation S2.24 as

$$\frac{\delta\dot{\phi}_i}{\mu_0\phi_{0,i}} = \nu_i - \mathcal{N} - \frac{\delta\phi_i}{\phi_{0,i}}. \tag{S2.29}$$

Equations S2.25, S2.27, and S2.29 together form the basis for the analyses in the main text.

To test the quality of the analytical results based on the linearization of the equations and the approximation that $f_i = \phi_{0,i}$, we simulated the nonlinear equations S2.4, S2.13 and S2.14 for a system containing just two proteins X and Y, using a convenient nonlinear function $\mu_d(\vec{\phi}) = (a_X\phi_X + a_Y\phi_Y)/(\phi_X + \phi_Y)$. (Note that μ_d is indeed an intensive function, as it should be. See p. 9 of this document for details on the simulation method.) We then numerically calculated the cross-correlations based on these simulations, and compared them to the corresponding analytical predictions. We chose the parameters of the simulation (including a_X and a_Y) such as to match the parameters used in Fig. 2.3A and B of the main text. In that figure, the simulation results based on the non-linear equations are plotted (gray diamonds) alongside the analytical predictions based on the linearized equations (solid gray line). They show excellent agreement, despite significant noise levels (coefficients of variation of ϕ_X, ϕ_Y, and μ were 0.09, 0.19, and 0.26, respectively). This shows that the approximations made in this section are excellent for sufficiently smooth functions $\mu_d(\vec{\phi})$.

s2.1.6 *Including dynamical gene regulation*

In the previous sections, we assumed that a fixed fraction f_i of the metabolic flux is allocated towards the synthesis of protein i. In this section we briefly describe the more general case where $\vec{f}$ depends on the composition of the cell, as expected due to transcriptional and translational regulation. For this case, analytical solutions are generally much harder to obtain. Nevertheless, we can gain some insight into the regulated system by studying the structure of the resulting equations.

We assume that $\vec{f}$ is an intensive quantity—that is, that it can be written as a function of the relative abundances $\vec{\phi}$. Then equation S2.13, for the production rate π_i, generalizes to:

$$\pi_i = f_i(\vec{\phi})\mu_d(\vec{\phi}) + N_i, \tag{S2.30}$$

with $0 \leqslant f_i(\vec{\phi}) \leqslant 1$ and $\sum_i f_i(\vec{\phi}) = 1$. This expression describes how the composition $\vec{\phi}$ of the cell differentially affects the production rate of each protein species. Linearizing equation S2.30 around $\vec{\phi_0}$, we find:

$$\delta\pi_i = \sum_j \frac{\partial}{\partial\phi_j} \left[f_i(\vec{\phi})\mu_d(\vec{\phi}) \right] \delta\phi_j + N_i$$

$$= f_i(\vec{\phi_0})\mu_0 \sum_j \left(\frac{\phi_{0,j}}{\mu_0} \frac{\partial\mu_d}{\partial\phi_j} + \frac{\phi_{0,j}}{f_i(\vec{\phi_0})} \frac{\partial f_i}{\partial\phi_j} \right) \frac{\delta\phi_j}{\phi_{0,j}} + N_i, \tag{S2.31}$$

which results in:

$$\frac{\delta\pi_i}{\pi_{0,i}} = \sum_j \left(C_j^\mu + C_j^{f_i} \right) \frac{\delta\phi_j}{\phi_{0,j}} + \nu_i. \tag{S2.32}$$

Here, new control coefficients emerge: regulational control coefficients (RCCs), defined as:

$$C_j^{f_i} := \left[\frac{\phi_j}{f_i} \frac{\partial f_i}{\partial\phi_j} \right]_{\vec{\phi_0}}. \tag{S2.33}$$

These RCCs quantify to what extent small changes in the cell's composition affect the allocation of the metabolic flux over the synthesis of various proteins species. As apparent from equation S2.32, the RCCs modulate the GCCs in an additive fashion.

Remarkably, the RCCs vanish from the expression for $\delta\mu$:

$$\delta\mu = \sum_i \pi_{0,i} \sum_j \left(C_j^\mu + C_j^{f_i} \right) \frac{\delta\phi_j}{\phi_{0,j}} + \sum_i \pi_{0,i}\nu_i$$

$$= \mu_0 \sum_j C_j^\mu \frac{\delta\phi_j}{\phi_{0,j}} + \mu_0 \sum_{i,j} f_i(\vec{\phi_0}) C_j^{f_i} \frac{\delta\phi_j}{\phi_{0,j}} + \mu_0 \mathcal{N}$$

$$= \mu_0 \sum_j C_j^\mu \frac{\delta\phi_j}{\phi_{0,j}} + \mu_0 \sum_j \left(\sum_i \frac{\partial f_i}{\partial\phi_j} \right) \delta\phi_j + \mu_0 \mathcal{N}$$

$$= \mu_0 \left(\sum_j C_j^\mu \frac{\delta\phi_j}{\phi_{0,j}} + \mathcal{N} \right). \tag{S2.34}$$

This derivation uses $\pi_{0,i} = \mu_0 f_i(\vec{\phi_0})$ and hinges on the fact that $\sum_i f_i(\vec{\phi}) = 1$, so that $\sum_i \partial f_i/\partial\phi_j = 0$.

The linearized dynamical equation for $\delta\phi_i$ in the presence of regulation becomes:

$$\frac{\delta\dot{\phi}_i}{\phi_{0,i}\mu_0} = \frac{\delta\pi_i}{\pi_{0,i}} - \frac{\delta\mu}{\mu_0} - \frac{\delta\phi_i}{\phi_{0,i}}$$

$$= \nu_i - \mathcal{N} - \frac{\delta\phi_i}{\phi_{0,i}} + \sum_j C_j^{f_i} \frac{\delta\phi_j}{\phi_{0,j}}. \tag{S2.35}$$

By comparing with equation S2.29 we concluded that gene regulation adds a single term that describes (to linear order) the combined effect on f_i of the fluctuations in the concentrations of all proteins species j.

S2.2 CALCULATING STATISTICAL PROPERTIES

S2.2.1 *Fourier transforms*

The linearized expressions (section S2.1.5) enable us to calculate various statistical properties. To do so, it is convenient to work with Fourier-transformed equations and variables. Based on equations S2.25, S2.27, and S2.29, we will now derive expressions for the Fourier transforms of the protein concentrations, the synthesis rates, and the growth rate. Further below, these expressions will be used to compute (co-)variances and correlations between the variables.

We denote the Fourier transform of variable y as $\widetilde{y}$. Taking the Fourier transform of both sides of equation S2.29 and using the transforms of equations S2.25 and S2.27, we find

$$\frac{\widetilde{\delta\phi_j}}{\phi_{0,j}} = \frac{\mu_0}{\mu_0 + \mathring{\imath}\omega}\left(\widetilde{\nu}_j - \widetilde{\mathcal{N}}\right),$$

(S2.36)

where ω is the Fourier frequency and $\mathring{\imath}$ denotes the imaginary unit (to distinguish it from indices i). To arrive at this result, we used that

$$\widetilde{\frac{\partial f}{\partial t}} = \mathring{\imath}\omega\widetilde{f}$$

(S2.37)

for any time-dependent function f.

By combining equation S2.36 with the Fourier transform of equation S2.27, we find a direct expression for the Fourier-transformed growth-rate deviations:

$$\frac{\widetilde{\delta\mu}}{\mu_0} = \sum_j C_j^\mu \frac{\mu_0}{\mu_0 + \mathring{\imath}\omega}\left(\widetilde{\nu}_j - \widetilde{\mathcal{N}}\right) + \widetilde{\mathcal{N}}$$

$$= \sum_j \left(\frac{\mu_0}{\mu_0 + \mathring{\imath}\omega}C_j^\mu + \phi_{0,j}\right)\widetilde{\nu}_j$$

(S2.38)

Here we used the definition of $\mathcal{N}$ (equation S2.28), the fact that $f_i = \phi_{0,i}$ to linear order, and the sum rule for GCCs presented in the main text:

$$\sum_i C_i^\mu = 0.$$

(S2.39)

(See section S2.3.2 for a derivation.) After analogous algebra, equation S2.25 yields

$$\frac{\widetilde{\delta\pi_i}}{\pi_{0,i}} = \sum_j \frac{\mu_0}{\mu_0 + \mathring{\imath}\omega}C_j^\mu\widetilde{\nu}_j + \widetilde{\nu}_i.$$

(S2.40)

S2.2.2 *Statistical quantities*

In this section, we first provide analytical expressions for cross-covariance functions between several variables [40]. Equations S2.36, S2.38, and S2.40 have the same mathematical form as analogous equations appearing in earlier models [40, 86]. Consequently, our derivations of the cross-covariance functions are directly parallel to their earlier work. The

resulting expressions are then used to calculate various stochastic quantities, including the coefficients of variation of the protein expression levels and the (cross-)correlations between growth rate and expression. Based on these, we provide details on several results presented in the main text, notably the complications in the distinction between intrinsic and extrinsic noise, and the mode decomposition of expression–growth cross-correlation functions.

Definitions and notation

The cross-covariance $\chi_{xy}(\tau)$ between variables $x(t)$ and $y(t)$ is defined as

$$\chi_{xy}(\tau) := \mathbb{E}\left[\delta x(t)\delta y(t+\tau)\right], \tag{S2.41}$$

where $\delta x(t) := x(t) - \mathbb{E}[x]$ and $\delta y(t) := y(t) - \mathbb{E}[y]$. In this notation, a positive value of $\chi_{xy}(\tau)$ means that the value of x co-varies with the value of y taken a time τ later; that is, y "follows" x. The convolution theorem states that

$$\chi_{xy}(\tau) = \int \frac{e^{i\omega\tau}\widetilde{\delta x}^*\widetilde{\delta y}}{2\pi}\,d\omega, \tag{S2.42}$$

where the star denotes complex conjugation. The Fourier transforms that were computed in the previous section can therefore be used verbatim to calculate (cross-co)variances for all variables μ, ϕ_i and π_i.

The functions $\chi_{xy}(\tau)$ directly yield various other important quantities. In particular, the variance of x is given by

$$\mathrm{Var}(x) = \chi_{xx}(0), \tag{S2.43}$$

and the coefficient of variation (CV) is defined as

$$\eta_x := \frac{\sqrt{\mathrm{Var}(x)}}{\mathbb{E}[x]} = \frac{\sigma_x}{x_0}, \tag{S2.44}$$

with σ_x the standard deviation of x. The cross-correlation function $R_{xy}(\tau)$ is given by

$$R_{xy}(\tau) = \frac{\chi_{xy}(\tau)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}}; \tag{S2.45}$$

at delay $\tau = 0$ it reduces to the Pearson correlation coefficient.

The cross-covariances of μ, ϕ_i and π_i are calculated by inserting S2.36, S2.38, and S2.40 into equation S2.42. In the process, we encounter products of noise sources $\widetilde{\nu_i}^*\widetilde{\nu_j}$ (Fourier-transformed cross-covariances of the noise sources), which, based on equation S2.9, can be expressed as:

$$\widetilde{\nu_i}^*\widetilde{\nu_j} = \delta_{ij}\frac{\vartheta_i^2}{\beta_i^2 + \omega^2}. \tag{S2.46}$$

Here, the Kronecker delta δ_{ij} reflects the mutual independence of the noise sources. After using this identity, all cross-covariances can be written as linear combinations of three families of functions, $B_i(\tau)$, $A_i(\tau)$ and $S_i(\tau)$, resulting from three Fourier integrals that can be solved explicitly with Cauchy's residue theorem.

The first function is

$$B_i(\tau) := \int \frac{e^{i\omega\tau}}{2\pi} |\widetilde{v}_i|^2 \, d\omega$$

$$= \frac{\vartheta_i^2}{2\beta_i} e^{-\beta|\tau|}. \tag{S2.47}$$

This function is symmetric; it is the auto-covariance of an Ornstein–Uhlenbeck process and therefore characterized by a single timescale β_i.

The second function is

$$A_i(\tau) := \int \frac{e^{i\omega\tau}}{2\pi} \frac{\mu_0}{\mu_0 - i\omega} |\widetilde{v}_i|^2 \, d\omega$$

$$= \vartheta_i^2 \mu_0 \begin{cases} \dfrac{e^{-\mu_0|\tau|}}{\beta_i^2 - \mu_0^2} - \dfrac{e^{-\beta_i|\tau|}}{2\beta_i(\beta_i - \mu_0)}, & \text{if } \tau \leqslant 0, \\[3mm] \dfrac{e^{-\beta_i|\tau|}}{2\beta_i(\beta_i + \mu_0)}, & \text{if } \tau \geqslant 0. \end{cases} \tag{S2.48}$$

This is an asymmetric function. It appears in the cross-correlation between variables X and Y when X is instantaneously affected by an Ornstein–Uhlenbeck noise source, while Y is affected by the same noise source with a delay characterize by rate μ_0. For example, the effect of noise in the synthesis of a protein on its concentration involves such a delay; as a result, the function $A_i(\tau)$ appears in the covariance between π_i and ϕ_i.

The third function is

$$S_i(\tau) := \int \frac{e^{i\omega\tau}}{2\pi} \frac{\mu_0^2}{\mu_0^2 + \omega^2} |\widetilde{v}_i|^2 \, d\omega$$

$$= \frac{\vartheta_i^2 \mu_0^2}{2(\beta_i^2 - \mu_0^2)} \left(\frac{e^{-\mu_0|\tau|}}{\mu_0} - \frac{e^{-\beta_i|\tau|}}{\beta_i} \right). \tag{S2.49}$$

This is again a symmetric function. It emerges whenever the covariance is taken between variables that are both affected by an Ornstein–Uhlenbeck process with the same delay characterized by rate μ_0. We note that

$$A_i(\tau) + A_i(-\tau) = 2S_i(\tau). \tag{S2.50}$$

Auto-covariances and coefficients of variation

In terms of the functions $B(t)$, $A(t)$, and $S(t)$, the auto-covariances for ϕ_i, π_i, and μ are given by

$$\frac{\chi_{\phi_i\phi_i}(\tau)}{\phi_{0,i}^2} = (1 - 2\phi_{0,i})S_i(\tau) + \sum_j \phi_{0,j}^2 S_j(\tau), \tag{S2.51}$$

$$\frac{\chi_{\pi_i\pi_i}(\tau)}{\pi_{0,i}^2} = B_i(\tau) + 2C_i^\mu S_i(\tau) + \sum_j \left(C_j^\mu\right)^2 S_j(\tau), \tag{S2.52}$$

$$\frac{\chi_{\mu\mu}(\tau)}{\mu_0^2} = \sum_j \left[C_j^\mu \left(C_j^\mu + 2\phi_{0,j}\right) S_j(\tau) + \phi_{0,j}^2 B_j(\tau) \right]. \tag{S2.53}$$

From these auto-covariances we obtain the CVs η_x:

$$\eta_{\phi_i}^2 = \left(\frac{\mu_0 (1 - 2\phi_{0,i})}{2\beta_i (\beta_i + \mu_0)} \right) \vartheta_i^2 + \sum_j \left(\frac{\mu_0 \phi_{0,j}^2}{2\beta_j (\beta_j + \mu_0)} \right) \vartheta_j^2 \tag{S2.54}$$

$$\eta_{\pi_i}^2 = \left(\frac{1}{2\beta_i} + \frac{C_i^\mu \mu_0}{\beta_i (\beta_i + \mu_0)} \right) \vartheta_i^2 + \sum_j \left(\frac{\mu_0 \left(C_j^\mu \right)^2}{2\beta_j (\beta_j + \mu_0)} \right) \vartheta_j^2 \tag{S2.55}$$

$$\eta_\mu^2 = \sum_j \left(\frac{C_j^\mu (C_j^\mu + 2\phi_{0,j}) \mu_0}{2\beta_j (\beta_j + \mu_0)} + \frac{\phi_{j,0}^2}{2\beta_j} \right) \vartheta_j^2 \tag{S2.56}$$

Here we used equation S2.43 and the special cases $B_i(0) = \vartheta_i^2/(2\beta_i)$ and $A_i(0) = S_i(0) = \vartheta_i^2 \mu_0 \left[2\beta_i (\beta_i + \mu_0) \right]^{-1}$ derived from equations S2.47, S2.48, and S2.49.

Intrinsic and extrinsic noise

Defining the shorthand

$$S_i := S_i(0) = \frac{\vartheta_i^2 \mu_0}{2\beta_i (\beta_i + \mu_0)} = \frac{\mathrm{Var}(N_i)}{\phi_{0,i}^2 \mu_0 (\mu_0 + \beta_i)} \tag{S2.57}$$

as a compound parameter associated to noise source i, where $\mathrm{Var}(N_i)$ is the variance of the Ornstein–Uhlenbeck process N_i, allows us to write equation S2.54 in the form of equation (14) of the main text, or as:

$$\eta_{\phi_i}^2 = \underbrace{(1 - \phi_{0,i})^2 S_i}_{\text{intrinsic}} + \underbrace{\sum_{j \neq i} \phi_{0,j}^2 S_j}_{\text{extrinsic}} . \tag{S2.58}$$

The two terms can be interpreted as intrinsic and extrinsic components of the noise. A discussion of this expression is given in the main text.

To see how the intrinsic and extrinsic components behave as a function of the protein's mean mass fraction $\phi_{0,i}$, we assume that the relative noise amplitude ϑ_i scales as $\phi_{0,i}^{-1/2}$, so that $S_i \sim \phi_{0,i}^{-1}$. Equation S2.58 then predicts two regimes: at low expression levels, the intrinsic noise dominates and the CV scales as $\phi_{0,i}^{-1/2}$. At high expression levels, the intrinsic component becomes negligible compared to the extrinsic component, and the CV approaches a plateau. This type of scaling has indeed been observed in experiments [173] and has been predicted by other theory as well [195].

Cross-covariances and cross-correlations

The second half of the main text focuses on the analysis of cross-correlations between gene expression and growth. Here we give explicit expressions for the cross-covariances between expression levels $\vec{\phi}$, protein production rates $\vec{\pi}$, and the growth rate μ; in conjunction with the CVs from equations S2.54, S2.55, and S2.56, these expressions can be used to to compute the cross-correlation functions.

The cross-covariances are:

$$\frac{\chi_{\phi_i \mu}(\tau)}{\phi_{0,i}\mu_0} = \underbrace{C_i^\mu S_i(\tau)}_{\text{Control}} + \underbrace{\phi_{0,i}A_i(\tau)}_{\text{Autogenic}} - \underbrace{\sum_j \phi_{0,j}\left[C_j^\mu S_j(\tau) + \phi_{0,j}A_j(\tau)\right]}_{\text{Dilution}}, \tag{S2.59}$$

$$\frac{\chi_{\pi_i \mu}(\tau)}{\pi_{0,i}\mu_0} = \underbrace{C_i^\mu A_i(-\tau)}_{\text{Control}} + \underbrace{\phi_{0,i}B_i(\tau)}_{\text{Autogenic}} + \underbrace{\sum_j C_j^\mu \left[C_j^\mu S_j(\tau) + \phi_{0,j}A_j(\tau)\right]}_{\text{Transmission}}, \tag{S2.60}$$

$$\frac{\chi_{\phi_i \phi_j}(\tau)}{\phi_{0,i}\phi_{0,j}} = -\phi_{0,i}S_i(\tau) - \phi_{0,j}S_j(\tau) + \sum_k \phi_k^2 S_k(\tau), \tag{S2.61}$$

$$\frac{\chi_{\pi_i \pi_j}(\tau)}{\pi_{0,i}\pi_{0,j}} = C_i^\mu A_i(-\tau) + C_j^\mu A_j(\tau) + \sum_k \left(C_k^\mu\right)^2 S_k(\tau). \tag{S2.62}$$

In these expressions we have indicated the four noise modes as defined and illustrated in the main text.

The cross-covariance between expression levels of different proteins $\chi_{\phi_i \phi_j}(\tau)$ is symmetrical. Any asymmetry that may be observed in this function will therefore remain unexplained by our model. Such an asymmetry could be explained by noise sources outside of protein production that act differently on the two proteins, or by gene regulation.

Whereas $\chi_{\phi_i \phi_j}(\tau)$ is fully determined by the expression levels $\vec{\phi}$ and the noise properties of the proteins, the cross-covariance between protein production rates $\chi_{\pi_i \pi_j}(\tau)$ contains information about the GCCs. Measurements on this function may therefore shed light on the growth-control properties of proteins.

S2.3   RESULTS ON GROWTH-CONTROL COEFFICIENTS

S2.3.1   *Relation between GCCs and FCCs*

As stated in the main text, the growth-control coefficient C_i^μ, flux control coefficient C_i^J, and mean mass fraction $\phi_{0,i}$ of a protein species i are tightly related through

$$C_i^J = C_i^\mu + \phi_{0,i}. \tag{S2.63}$$

Here, we provide a derivation. (See also [10, 196] for a similar derivation and application in an evolutionary context.)

By definition, $\mu_d(\vec{n}) := J(\vec{n})/M(\vec{n})$; equation S2.63 then follows from the product rule for differentiation:

$$
\begin{aligned}
C_i^J &= \left[\frac{\phi_i}{J(\vec{\phi})}\frac{\partial J(\vec{\phi})}{\partial \phi_i}\right]_{\vec{\phi_0}} \\
&= \frac{\phi_{0,i}}{\mu_d(\vec{\phi_0})M(\vec{\phi_0})}\left[\frac{\partial\left[\mu_d(\vec{\phi})M(\vec{\phi})\right]}{\partial \phi_i}\right]_{\vec{\phi_0}} \\
&= \frac{\phi_{0,i}}{\mu_d(\vec{\phi_0})}\left(\frac{1}{M(\vec{\phi_0})}\left[\frac{\partial\mu_d}{\partial \phi_i}\right]_{\vec{\phi_0}} + \mu_d(\vec{\phi_0})\left[\frac{\partial M}{\partial \phi_i}\right]_{\vec{\phi_0}}\right) \\
&= C_i^\mu + \phi_{0,i}.
\end{aligned}
\tag{S2.64}
$$

In this derivation, we used that $\partial M(\vec{\phi})/\partial \phi_i = 1$ and $M(\vec{\phi_0}) = 1$, which both follow from the definition $M(\vec{n}) \coloneqq \sum n_i$.

s2.3.2   *Sum rules*

In this section, we provide derivations of the sum rules for flux control coefficients (FCCs) and growth-control coefficients (GCCs)

In metabolic control analysis [63, 77] , the FCC of enzyme i with respect to a metabolic flux J has been defined as

$$C_i^J \coloneqq \left[\frac{\phi_i}{J} \frac{\partial J}{\partial \phi_i} \right]_{\vec{\phi_0}} \tag{S2.65}$$

where ϕ_i is the concentration of enzyme i. A classical result is the sum rule for FCCs:

$$\sum_i C_i^J = 1. \tag{S2.66}$$

This result is ultimately based on the common assumption that metabolic fluxes are *extensive* quantities; that is, for any $\alpha > 0$ and any composition $\vec{\phi}$ it is assumed that

$$J(\alpha\vec{\phi}) = \alpha J(\vec{\phi}). \tag{S2.67}$$

Differentiating both sides of this equation with respect to α and evaluating the result in $\alpha = 1$ yields:

$$\sum_i \phi_i \frac{\partial J}{\partial \phi_i} = J(\vec{\phi}), \tag{S2.68}$$

where the summation runs over all protein species. Dividing by J directly results in the sum rule of equation S2.65.

Our definition of growth-control coefficients (GCCs) proceeds analogously, as

$$C_i^\mu \coloneqq \left[\frac{\phi_i}{\mu_d} \frac{\partial \mu_d}{\partial \phi_i} \right]_{\vec{\phi_0}}. \tag{S2.69}$$

For GCCs, however, a different sum rule holds:

$$\sum_i C_i^\mu = 0. \tag{S2.70}$$

The derivation of this result is directly analogous to that for FCCs, except that μ_d is assumed to be *intensive* rather than extensive. That is, for any $\alpha > 0$ and any composition $\vec{\phi}$ it is assumed that

$$\mu_d\left(\alpha\vec{\phi}\right) = \mu_d\left(\vec{\phi}\right). \tag{S2.71}$$

Differentiating both sides with respect to α and evaluating the resulting expression in $\alpha = 1$ now gives

$$\sum_i \phi_i \frac{\partial \mu_d}{\partial \phi_i} = 0. \tag{S2.72}$$

Dividing by μ_d directly results in the sum rule of equation S2.70.

S2.4 APPLICATION TO THE TWO-PROTEIN TOY MODEL

In the main text, a toy model is presented in which the cell contains only two protein species, called X and Y. We here derive the mathematical expressions underlying these results.

S2.4.1 *Cross-correlations*

For the two-protein model, the general expressions of equations S2.45, S2.59 and S2.60 evaluate to the following cross-correlations $R_{\phi_Y \mu}(\tau)$ and $R_{\pi_Y \mu}(\tau)$:

$$R_{\phi_Y \mu}(\tau) = \frac{\overbrace{C_Y^\mu S_Y(\tau)}^{\text{Control}} + \overbrace{\phi_{0,Y} A_Y(\tau)}^{\text{Autogenic}} \overbrace{- \phi_{0,Y}\left[C_Y^\mu S_Y(\tau) + \phi_{0,Y} A_Y(\tau)\right] + \phi_{0,X}\left[C_X^\mu S_X(\tau) + \phi_{0,X} A_X(\tau)\right]}^{\text{Dilution}}}{\eta_{\phi_Y} \eta_\mu},$$

(S2.73)

$$R_{\pi_Y \mu}(\tau) = \frac{\overbrace{C_Y^\mu A_Y(-\tau)}^{\text{Control}} + \overbrace{\phi_{0,Y} B_Y(\tau)}^{\text{Autogenic}} + \overbrace{C_X^\mu \left[C_X^\mu S_X(\tau) + \phi_{0,X} A_X(\tau)\right] + C_Y^\mu \left[C_Y^\mu S_Y(\tau) + \phi_{0,Y} A_Y(\tau)\right]}^{\text{Transmission}}}{\eta_{\pi_Y} \eta_\mu},$$

(S2.74)

where the coefficients of variation in the denominators are given by

$$\eta_{\phi_Y}^2 = \left(\frac{(1 - \phi_{0,Y})^2 \mu_0 \vartheta_Y^2}{2\beta_Y(\beta_Y + \mu_0)} + \frac{\phi_{0,X}^2 \mu_0 \vartheta_X^2}{2\beta_X(\beta_X + \mu_0)}\right),$$

(S2.75)

$$\eta_{\pi_Y}^2 = \left(1 + \frac{\mu_0 C_Y^\mu \left(C_Y^\mu + 2\right)}{\beta_Y + \mu_0}\right)\frac{\vartheta_Y^2}{2\beta_Y} + \frac{\mu_0 \left(C_X^\mu\right)^2 \vartheta_X^2}{2\beta_X(\beta_X + \mu_0)},$$

(S2.76)

$$\eta_\mu^2 = \left(\frac{\mu_0 C_X^\mu \left(C_X^\mu + 2\phi_{0,X}\right)}{\beta_X + \mu_0} + \phi_{0,X}^2\right)\frac{\vartheta_X^2}{2\beta_X}$$
$$+ \left(\frac{\mu_0 C_Y^\mu \left(C_Y^\mu + 2\phi_{0,Y}\right)}{\beta_Y + \mu_0} + \phi_{0,Y}^2\right)\frac{\vartheta_Y^2}{2\beta_Y}.$$

(S2.77)

The expressions above can be further simplified by inserting the constraints

$$C_X^\mu + C_Y^\mu = 0,$$

(S2.78)

$$\phi_{0,X} + \phi_{0,Y} = 1.$$

(S2.79)

S2.4.2 *Regulation in the two-protein model*

In deriving the results above, we assumed that $\vec{f}$ is constant; that is, the synthesis rate of the proteins is not dynamically regulated in response to fluctuations in the protein mass fractions. Generally, if gene regulation is included, analytical results become hard to obtain. The simple case of the two-protein toy model is an exception, as we now show.

Assuming, as before, that $\vec{f}$ is an intensive quantity, we find the additional constraint (sum rule)

$$C_X^{f_Y} + C_Y^{f_Y} = 0.$$

(S2.80)

Furthermore, $\vec{f}$ denotes the fractional allocation of resources towards each protein species, so that

$$f_X + f_Y = 1. \tag{S2.81}$$

Therefore, the regulation has only one degree of freedom and is thus fully characterized by just one parameter:

$$C_X^{f_X} = C_Y^{f_Y} = -C_X^{f_Y} = -C_Y^{f_X}, \tag{S2.82}$$

which quantifies the auto-regulation of protein X, that of Y, the regulation of Y by X, and *vice versa*.

These constraints allow us to calculate the two-protein cross-correlations analytically. Equation S2.35 simplifies to

$$\begin{aligned}
\frac{\delta\dot\phi_Y}{\phi_{0,Y}\mu_0} &= C_X^{f_Y}\frac{\delta\phi_X}{\phi_{0,X}} + C_Y^{f_Y}\frac{\delta\phi_Y}{\phi_{0,Y}} + (1-\phi_{0,Y})(\nu_Y - \nu_X) - \frac{\delta\phi_Y}{\phi_{0,Y}}\\
&= -C_Y^{f_Y}\frac{-\delta\phi_Y}{1-\phi_{0,Y}} + C_Y^{f_Y}\frac{\delta\phi_Y}{\phi_{0,Y}} + (1-\phi_{0,Y})(\nu_Y - \nu_X) - \frac{\delta\phi_Y}{\phi_{0,Y}}\\
&= \left(\frac{C_Y^{f_Y}}{1-\phi_{0,Y}} - 1\right)\frac{\delta\phi_Y}{\phi_{0,Y}} + (1-\phi_{0,Y})(\nu_Y - \nu_X). \tag{S2.83}
\end{aligned}$$

The Fourier transform then reads

$$\frac{\widetilde{\delta\phi_Y}}{\phi_{0,Y}} = (1-\phi_{0,Y})\left(\frac{\mu_0}{\mu_Y + i\omega}\right)(\widetilde{\nu_Y} - \widetilde{\nu_X}). \tag{S2.84}$$

This equation differs only slightly from equation S2.36: apart from a constant prefactor $(1-\phi_{0,Y})$, a term μ_0 in the denominator is replaced by

$$\mu_Y = \mu_0\left(1 - \frac{C_Y^{f_Y}}{1-\phi_{0,Y}}\right). \tag{S2.85}$$

Because equations S2.36 and S2.84 have the same form, their solutions are analogous.

To illustrate the effect of regulation in this toy model, the expression–growth cross-correlation function is plotted in Fig. S2.2A, for different values of $C_Y^{f_Y}$. Positive auto-regulation widens the cross-correlation function, indicating increased auto-correlation timescales; negative auto-regulation narrows it, indicating decreased auto-correlation timescales. Without regulation, dilution by growth quenches fluctuations in the concentration ϕ_Y with the associated time scale μ_0^{-1}. In the presence of regulation this time scale is adjusted to μ_Y^{-1}. This directly determines the time scales of the cross-correlations.

Furthermore, regulation affects the CV of the expression of both X and Y, because the term μ_0 in the denominator of equation S2.75 is replaced by to μ_Y. This is shown in Fig. S2.2B. If $C_Y^{f_Y}$ is negative (negative auto-regulation), fluctuations are actively quenched; the variance of ϕ_Y is therefore reduced. Conversely, if $C_Y^{f_Y}$ is positive (positive auto-regulation) fluctuations are amplified; the variance in ϕ_Y is therefore increased.

s2.4.3 *Stochastic simulation of the two protein model*

To test the analytical results on the two-protein toy model, we used the Euler–Maruyama method to propagate the system of stochastic differential equations and thus calculated

the cross-correlations numerically. These simulations are based on the non-linear equations S2.4, S2.13 and S2.14, and a non-linear function $\mu_d(\vec{\phi})$ was used. As such, they demonstrate that, provided $\mu_d(\vec{\phi})$ is sufficiently smooth, the approximations used in the derivation of the linearized equations are appropriate. The results are included in Fig. 2.3A and B.

In this simulation we chose the parameters of the non-linear system such as to match the parameters used for the plots in Fig. 2.3A and B. Specifically, we specified the allocation coefficient as $f_Y = 0.33$, and the parameters of the noise sources as $\theta_Y = 0.5$, $\theta_X = 0.5$, $\beta_Y = 4/h$, and $\beta_X = 4/h$, as in Fig. 2.3A and B. For the the deterministic part of the growth rate we chose $\mu_d(\vec{\phi}) := (a_X\phi_X + a_Y\phi_Y)/(\phi_X + \phi_Y)$, with $a_X = 0.63/h$ and $a_Y = 1.37/h$, such that $\mu_0 = \mu_d(\vec{\phi_0}) \approx \mu_d(\vec{f}) = 1/h$ and the control coefficients would approximate the ones in Fig. 2.3 ($C_Y^\mu = -C_X^\mu = 0.25$).

We propagated the dynamical equations in discrete time steps Δt by repeating the following calculations. (Here, the index i is either X or Y.)

$$\begin{cases} \mu_d(t) = (a_X\phi_X(t) + a_Y\phi_Y(t))/(\phi_X(t) + \phi_Y(t)), \\ \pi_i(t) = f_i\,\mu_d(t) + N_i(t), \\ \mu(t) = \mu_d(t) + N_X(t) + N_Y(t), \\ \phi_i(t+\Delta t) = \phi_i(t)\,(1 - \Delta t\mu(t)) + \Delta t\,\pi_i(t), \\ N_i(t+\Delta t) = N(t)\,(1 - \beta_i\Delta t) + \sqrt{\Delta t}\,\theta_i dW(t). \end{cases} \tag{S2.86}$$

In the last line, $dW(t)$ is random variable drawn from a standard normal distribution each time step. We ran the simulation for roughly 30.000 doubling times, using time steps $\Delta t = 10^{-3}$.

After the simulation, we measured ϕ_Y to test whether $f_Y \approx \phi_Y$ as is assumed in section 2.5. Indeed, f_Y did not differ from ϕ_Y by more than 5%. We next verified that $\mu_0 := \mu_d(\vec{\phi_0}) \approx 0.993/h$ was close to $1/h$ and that $C_Y^\mu = \phi_{0,Y}(a_Y/\mu_0 - 1) \approx 0.249$ was close to 0.25, as intended.

s2.5   APPLICATION TO THE MANY-PROTEIN MODEL

As described in the main text, we applied the modeling framework to model a full cell, containing 1021 protein species in total. We here provide the mathematical details of this model. First we explain how the operon construct was represented. Then we describe our parameter choices and how they were informed by previously published experimental data.

s2.5.1   *GFP reporter fused to the lac operon*

As mentioned in the main text, Kiviet *et al.* performed experiments in which the expression levels of the genes *lacY* and *lacZ* were reported by a green fluorescent protein (GFP), whose gene was transcriptionally fused to the *lac* operon [86]. Thus, the GFP signal was used as a proxy for the concentration of the *lac* proteins LacY and LacZ. To include this construct in the model, we assumed that the synthesis noise of each protein, introduced by transcription of the operon and the translation of its genes, can be linearly decomposed into two components. The first component reflects noise originating from processes that similarly

affect all genes on the operon, adding to the correlation between their expression levels. In particular, this includes fluctuations in the copy number of the mRNA transcript due to transcriptional noise. The second component reflects noise originating from processes that affect each gene independently, including all post-transcriptional processes.

Formally, we included noise sources N_Y, N_Z, and N_G, respectively acting on LacY, LacZ, and GFP synthesis only, as well an additional noise source N_O that simultaneously affects the synthesis of all of these proteins. That is, equation S2.13 was modified to:

$$\pi_i = f_i \mu_d(\vec{\phi_0}) + N_i + I_i^O \frac{\pi_{0,i}}{\pi_{o,O}} N_O. \tag{S2.87}$$

Here, I_i^O is an indicator function that specifies whether protein i is encoded on the *lac* operon: $I_i^O = 1$ for $i \in O := \{Y, Z, G\}$ and $I_i^O = 0$ otherwise.

As before, all noise sources are mutually independent Ornstein–Uhlenbeck processes with zero mean. For simplicity, we incorporated the operon noise such that it affects the three genes in proportion with their average production rate. In the limit that $\nu_i = 0$ for $i \in O$ and $N_O \neq 0$, the three proteins behave as a single protein O, with the same scaling $N_j = \pi_{0,j} \nu_j$ as the other genes.

After linearization (cf. equation S2.25) we arrive at

$$\frac{\delta \pi_i}{\pi_{0,i}} = \sum_j C_j^\mu \frac{\delta \phi_j}{\phi_{0,j}} + \nu_i + I_i^O \nu_O. \tag{S2.88}$$

Here, the summation runs over all 1021 proteins, including Y, Z, and G. Moreover, we have defined $\nu_O := N_O/\pi_{o,O} = N_O/(\pi_{o,Y} + \pi_{o,Z} + \pi_{o,G})$. To obtain the cross-covariances, the analysis now proceeds exactly as before. Using the shorthands $\phi_{0,O} := \phi_{0,Y} + \phi_{0,Z} + \phi_{0,G}$ and $C_O^\mu := C_Y^\mu + C_Z^\mu + C_G^\mu$, the resulting (cross-)covariances are given by

$$\frac{\chi_{\phi_i \phi_i}(\tau)}{\phi_{0,i}^2} = (1 - 2\phi_{0,i})S_i(\tau) + I_i^O(1 - 2\phi_{o,O})S_O(\tau) + \sum_j \phi_{0,j}^2 S_j(\tau); \tag{S2.89}$$

$$\frac{\chi_{\pi_i \pi_i}(\tau)}{\pi_{0,i}^2} = B_i(\tau) + 2C_i^\mu S_i(\tau) + I_i^O \left[B_O(\tau) + 2C_O^\mu S_O(\tau) \right] + \sum_j \left(C_j^\mu \right)^2 S_j(\tau); \tag{S2.90}$$

$$\frac{\chi_{\mu\mu}(\tau)}{\mu_0^2} = \sum_j \left\{ \phi_j^2 B_j(\tau) + \left[\left(C_j^\mu \right)^2 + 2\phi_{0,j} C_j^\mu \right] S_j(\tau) \right\}; \tag{S2.91}$$

$$\eta_{\phi_i}^2 = (1 - 2\phi_{0,i}) \frac{\mu_0 \vartheta_i^2}{2\beta_i(\beta_i + \mu_0)} + I_i^O (1 - 2\phi_O) \frac{\mu_0 \vartheta_O^2}{2\beta_O(\beta_O + \mu_0)} + \sum_j \phi_{0,j}^2 \frac{\mu_0 \vartheta_j^2}{2\beta_j(\beta_j + \mu_0)}; \tag{S2.92}$$

$$\frac{\chi_{\phi_i \mu}(\tau)}{\phi_{0,i} \mu_0} = C_i^\mu S_i(\tau) + \phi_{0,i} A_i(\tau) + I_i^O \left[C_O^\mu S_O(\tau) + \phi_{0,O} A_O(\tau) \right] - \sum_j \phi_{0,j} \left[C_j^\mu S_j(\tau) + \phi_{0,j} A_j(\tau) \right]; \tag{S2.93}$$

$$\frac{\chi_{\pi_i \mu}(\tau)}{\pi_{0,i} \mu_0} = \phi_{0,i} B_i(\tau) + C_i^\mu A_i(-\tau) + I_i^O \left[\phi_{0,O} B_O(\tau) + C_O^\mu A_O(-\tau) \right] + \sum_j C_j^\mu \left[C_j^\mu S_j(\tau) + \phi_{0,j} A_j(\tau) \right]. \tag{S2.94}$$

The index j of the summations now runs over all proteins as well as the operon $j = O$, so that all noise sources are included.

Parametrization

In order to simulate a whole cell, we had to specify four parameters for each protein i, namely its abundance $\phi_{0,i}$, its control coefficient C_i^μ, and its noise properties ϑ_i and β_i. In addition, the mean growth rate μ_0 had to be set for each of the three growth states presented in the main text (low, intermediate, and high).

Monod curve

Under the experimental conditions, the mean growth rate μ_0 follows an empirical Monod curve $\mu_M(\phi_{o,O})$ as a function of the mean expression level $\phi_{o,O}$ of the *lac* operon [86]. To ensure that our parameters respect this relation, we constrain them with the Monod curve shown in Fig. 2.4B,

$$\mu_M(\phi_{o,O}) := \mu_{max} \frac{\phi_{o,O}}{\phi_{o,O} + \phi_{half}}, \tag{S2.95}$$

which is defined by a maximal growth rate of $\mu_{max} = 0.8~\text{h}^{-1}$ and $\phi_{half} = 0.005$. The parameter ϕ_{half} was taken clearly larger than the abundance of an average protein, since transporter proteins are highly expressed in growth media where they are utilized [155].

Choosing the average growth rate μ_0

The average growth rates in the three growth states were chosen in rough agreement with the experiments as $\mu_0 = 0.19, 0.45,$ and $0.75~\text{h}^{-1}$.

Choosing reversion rates $\vec{\beta}$

For simplicity, we assumed that the reversion rates β_i of all noise sources were identical and considerably larger than the average growth rate μ_0; that is, we set $\beta_i = \beta := 4\mu_{max}$ for all i.

Choosing $\vec{\phi_0}$ and $\vec{\vartheta}$ based on experimental data

Several experimental datasets are available that provide cell-wide estimates of *E. coli* protein abundances [1, 68, 155, 173]. In particular, Taniguchi *et al.* have measured the means and variances of 1018 protein abundances [173]. We used these measurements to fix the values of $\phi_{0,i}$ and ϑ_i for all proteins i in our model, as we now explain.

First we determined the average mass fractions $\vec{\phi_0}$. Given the mean growth rate μ_0 for a given growth condition, the average mass fraction of the operon, $\phi_{o,O}$, follows directly from the Monod curve described above. We then assumed equal properties for the proteins on the operon, i.e. $\phi_{o,G} = \phi_{o,Y} = \phi_{o,Z} = \phi_{o,O}/3$ and $\vartheta_{o,G} = \vartheta_{o,Y} = \vartheta_{o,Z}$. We denote the sum of all measured protein abundances as m, and the mass of the whole cell including the *lac* operon as M. Then $M = m/(1 - \phi_{o,O})$, and the average mass fraction of protein i can be calculated by dividing the measured abundance by M.

Next, we calculated the values of the noise amplitudes $\vec{\vartheta}$. Equation S2.92 shows that the squared coefficient of variation of any given protein species is a linear function of the squared noise amplitudes of all noise sources. Conversely, imposing the measured variances of all protein species results in a linear system of equations, the solution of which uniquely determines the ϑ_i^2 (for given μ_0, $\vec{\phi_0}$ and $\vec{\beta}$).

To do so explicitly, we assigned all proteins an index in the range 1 to $K + 3$, where K is the number of proteins that are not encoded on the *lac* operon (in this case $K = 1018$); proteins G, Y and Z were assigned indices 1 to 3, respectively. We set the amplitude of the operon noise source to $\vartheta_O = \alpha\vartheta_G = \alpha\vartheta_Y = \alpha\vartheta_Z$, where α defines the ratio of the amplitude of ν_O to the amplitudes of ν_G, ν_Y, and ν_Z. We used $\alpha = 1.5$ to obtain the results in the main text (Figure 4). This procedure allowed us to rewrite equation S2.92 in $((K + 3) \times (K + 3)$ dimensional) matrix form:

$$\vec{\eta_\phi}^2 = \vec{T}\vec{\vartheta^2}, \tag{S2.96}$$

where the vector $\vec{\vartheta^2}$ contains the squared noise amplitudes, $\vec{\eta}_\phi^2$ is the vector of squared CVs, and the elements of matrix $\vec{T}$ are given by

$$T_{ij} \quad = \quad \frac{\mu_0}{2\beta(\beta + \mu_0)} \quad \times \quad \begin{cases} (1 - \phi_{0,G})^2 + (1 - \phi_{0,O})^2\alpha^2 & \text{if } i = j = 1. \\ \phi_{0,G}^2 + (1 - \phi_{0,O})^2\alpha^2 & \text{if } j = 1 \neq i \leqslant 3, \\ \phi_{0,G}^2 + \phi_{0,O}^2\alpha^2 & \text{if } j = 1, i > 3, \\ \phi_{0,j}^2 & \text{if } i \neq j \neq 1, \\ (1 - \phi_{0,j})^2 & \text{if } i = j \neq 1, \end{cases} \tag{S2.97}$$

Lastly, $\vec{\vartheta^2}$ was obtained by numerically inverting equation S2.96.

Choosing the growth-control coefficients

Large-scale measurements of growth-control coefficients are not readily available. Therefore, we randomly sampled the GCCs using the following heuristic. First, we randomly assigned proteins from the database measured by Taniguchi *et al.* to the non-metabolic H sector, until the total H sector size was $\sim 25\%$ of the sum of measured protein abundances m. The size of this non-metabolic mass fraction has been directly estimated as 25% by O'Brien *et al.* [135]; it has been described earlier by Hui *et al.* as part of a growth-rate independent proteome fraction that amounts in total to about 60% of the total protein mass [68]. For these H proteins, we set $C_i^\mu = -\phi_i$ (note that ϕ_i partly depends on the expression level of the *lac* operon, as described earlier), in line with equation (11) from the main text.

Second, the GCCs of the *lac*-operon proteins Y, Z, and G were chosen in agreement with the Monod curve discussed above. We next used the derivative of the Monod curve $\mu_M(\phi_{o,O})$ to estimate C_O^μ:

$$C_O^\mu := \left[\frac{\phi_O}{\mu_d}\frac{\partial\mu_d}{\partial\phi_O}\right]_{\vec{\phi_o}} \approx \frac{\phi_{o,O}}{\mu_M}\frac{d\mu_M}{d\phi_{o,O}} = \frac{\phi_{half}}{\phi_{o,O} + \phi_{half}}. \tag{S2.98}$$

Here we implicitly assumed that the operon as a whole is never so strongly over-expressed as to have a negative control on the growth rate.

Lastly, we chose $C_Y^\mu = C_Z^\mu = (C_O^\mu - C_G^\mu)/2$ for simplicity.

Third, for the remaining proteins (which are neither in the H sector nor encoded on the *lac* operon), we assumed that the GCCs tend to scale with their abundance, and enforced the sum rule. In practice, we drew a uniformly distributed random number $x_i \in [0, 1]$ for each protein that was not in the H sector and set

$$C_i^\mu = \frac{x_i\phi_{0,i}}{\sum_{j\notin H} x_j\phi_{0,j}}\left(\phi_H - C_O^\mu\right). \tag{S2.99}$$

As in the main text, ϕ_H is the mass fraction of all H proteins combined. To determine the robustness of our results with respect to the sampling scheme, we also tried different GCC distributions, for example using exponentially distributed variables x_i. This yielded visually indistinguishable results.

s2.5.3 *Alternative choice for abundances and variances*

Although the data set obtained by Taniguchi *et al.* [173] is unique to report both abundances and variances for each measured protein, it suffers from a systematic underestimation of the protein masses [1, 155]. We therefore also estimated the protein abundances from another proteomics dataset, using a phenomenological model to calculate the variances in the absence of explicit measurements.

In this case, the abundances $\phi_{0,i}$ were taken from a dataset obtained from Arike *et al.* [1]. Given that the mean copy number of protein i is n_i, a theoretical minimum can be calculated for its variance; it is given by $n_i^2 \sigma_{ab} + n_i \beta_W$, with $\sigma_{ab} = 0.025$ and $\beta_W = 450$ [195]. The original paper reported variances that were typically higher than this lower limit [195]. To empirically match these results, we chose

$$\mathrm{Var}_i = n_i^2 \sigma_{ab} + n_i(\beta_W + x_i), \tag{S2.100}$$

with x_i an exponentially distributed random variable with mean 200.

The resulting cross-correlations are plotted in Fig. S2.3. They lead to the same qualitative conclusions as the analysis based on the data from Taniguchi *et al.*

s2.6 ALTERNATIVE MODEL: NOISE IN THE ALLOCATION OF THE FLUX

Throughout this document we have considered noise sources that act on each protein-synthesis rate independently. Alternatively, one could hypothesize that the observed fluctuations in protein concentrations instead originate from noise on the allocation of the available metabolic flux. Especially when translation is a highly rate-limiting step in protein production, so that mRNAs compete for a limited capacity of ribosomes, a stochastic increase in the synthesis of a given protein species may necessarily go at the expense of the synthesis rates of other protein species. In this section, we therefore explore a model in which such "allocation noise" is dominant instead of the pure "production noise" studied above.

In the modified model, each protein species is allotted a fluctuating fraction $f_i(t)$ of the total metabolic flux, with the additional constraint that $\sum_i f_i(t) = 1$ at all times. We end up with an adjusted version of equation S2.13:

$$\pi_i := f_i(t)\mu_d(\vec{\phi}), \tag{S2.101}$$

with

$$f_i(t) = \frac{f_{0,i} + N_i}{\sum_j f_{0,j} + N_j} = \frac{f_{0,i} + N_i}{1 + \sum_j N_j}. \tag{S2.102}$$

Here, the N_i are again independent Ornstein–Uhlenbeck processes. This also results in an adjusted version of equation S2.14:

$$\mu = \mu_d(\vec{\phi}). \tag{S2.103}$$

As before, in the small-noise limit μ_d can be linearized around the mean composition $\vec{\phi_0}$:

$$\mu_d(\vec{\phi}) := \mu_0 \left(1 + \sum_j C_j^\mu \left(\frac{\phi_j}{f_{0,j}} - 1 \right) \right), \qquad \text{(S2.104)}$$

where we assumed that $\phi_{0,i} \approx f_{0,i}$, as previously.

Unfortunately, production rates as defined in equation S2.101 are intrinsically correlated and therefore analytical expressions can not be obtained within our current framework. To gain insight in the cross-correlations generated by this system, we therefore numerically integrated equations S2.101 and S2.104 for a model cell containing 40 proteins with randomly sampled parameters (see next section for details about parameters and the simulation method). The cross-correlations obtained for one of the proteins (protein 1) are shown in Fig. S2.5A.

A striking observation is that the ϕ–μ cross correlations of the model with "allocation noise" (Fig. S2.5A and C) are always perfectly symmetrical, with their peak at zero delay ($\tau = 0$).

In itself, the presence of a symmetric mode is no surprise. Indeed, an increase in a protein species' synthesis rate, affects—with a delay—its concentration, and this in turn affects the growth rate, in proportion with the protein's GCC. Therefore, a symmetrical control mode is to be expected.

However, when noise acts dominantly on the allocation of the flux, the growth rate is not instantly affected by fluctuating production rates. Therefore, there is now pathway by which noise in a protein's production rate instantly affects the growth rate, and then—after a delay—the protein concentrations (either by dilution or by direct production). This removes the asymmetric autogenic mode that caused the asymmetry in the ϕ–μ cross-correlation as well as the symmetric mode in the ϕ–μ cross-correlation in the "production noise" model (see also Fig. 2.3C–F of the main text).

To check the the symmetric ϕ–μ cross-correlation is indeed a typical feature of the allocation noise model and not just a feature of the particular cell sampled, we analytically calculated cross-correlations for a similar cell (*i.e.* same protein variances, same mean protein abundances, same GGCs), but now according to the "production noise" model (*i.e.*, production rates according to equation S2.13). The resulting cross-correlations are shown in Fig. S2.5B and show clear asymmetry in the ϕ–μ cross-correlation (demonstrated in Fig. S2.5C).

Since the measured cross-correlations by Kiviet *et al.* showed clear asymmetrical features as well [86]—which in fact prompted them to propose a "common-noise" source—we conclude that noise on allocation alone cannot explain the observed cross-correlations.

s2.6.1 *Simulations of the "allocation-noise" model*

We here provide details on the simulations discussed in the previous section and presented in Fig. S2.5.

We simulated a cell with 40 proteins. To choose their mean protein fractions, we drew uniformly distributed numbers ($x_j \sim \mathcal{U}[0,1]$, where $\mathcal{U}[0,1]$ is the uniform distribution over the interval $[0,1]$) and normalized them to obtain protein fractions, *i.e.*, $\phi_{i,0} = f_{i,0} = x_i / \sum x_j$. Approximately 60% of the proteome was assumed to be metabolic. For the $\approx 40\%$ non-metabolic proteins, we set $C_h^\mu = -\phi_{h,0}$; for the remaining proteins we sampled GCCs

according to a uniform distribution, scaled with the protein's mean fraction, under the constraint of the sum rule. That is: for any metabolic protein i, draw an $x_i \sim \mathcal{U}[0, 1]$, and subsequently set $C_i^\mu = \left(\sum_{h \in H} \phi_j \right) (x_i \phi_i) / \left(\sum_{j \notin H} x_j \phi_j \right)$. We also sampled the noise amplitudes $\vec{\theta}_{AN}$ (here the label "AN" stands for "allocation noise") uniformly and then scaled them with $\sqrt{f_{i,0}}$, *i.e.*, $\theta_{AN,i} \sim \left(0.1 \sqrt{f_{i,0}} \right) \mathcal{U}[0, 1]$.

First we numerically integrated equations S2.101-S2.104 using the numerical integration scheme below to calculate the cross-correlation for the allocation-noise model.

$$\vec{\phi}(t + \Delta t) = \vec{\phi}(t) \left(1 - \Delta t \sum_i \pi_i(t) \right) + \Delta t \, \vec{\pi}(t), \tag{S2.105}$$

$$\vec{N}(t + \Delta t) = \vec{N}(t) \left(1 - \Delta t \, \beta \right) + \sqrt{\Delta t} \vec{\theta}_{AN} * dW(t), \tag{S2.106}$$

$$\vec{\pi}(t + \Delta t) = \mu_0 \left(\frac{\vec{f_0} + \vec{N}(t + \Delta t)}{1 + \sum_i N_i(t + \Delta t)} \right) \left(1 + \sum_i C_i^\mu \left(\frac{\phi_i(t + \Delta t)}{f_{0,i}} - 1 \right) \right). \tag{S2.107}$$

Here, $*$ denotes element-wise multiplication.

To define a similar cell in the "production-noise" model, we first measured $\vec{\eta}_{\phi,AN}$, in the "allocation noise" simulation and used a method similar to equation S2.96 to calculate which noise amplitudes $\vec{\theta}_{PN}$ (PN stands for "production noise") would create the same variances in the "production-noise" model:

$$\vec{\vartheta}_{PN}^2 = (T_{PN})^{-1} \, \vec{\eta}_{\phi,AN}^2.$$

With the matrix T_{PN} similar to equation S2.97:

$$(T_{PN})_{ij} = \frac{\mu_0}{2\beta (\beta + \mu 0)} \begin{cases} (1 - f_{i,0})^2, & \text{if } i = j, \\ f_{i,0}^2, & \text{else.} \end{cases}$$
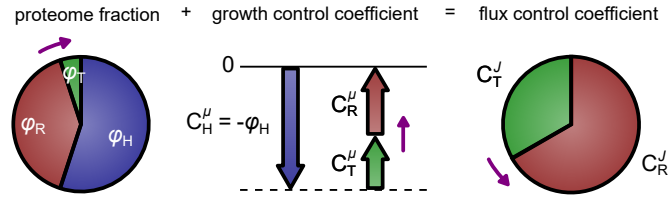
Figure S2.1: **Illustration of equation 2.9 of the main text.** Pictured is the relation between flux-control coefficients C_i^J, growth-control coefficients C_i^μ, and proteomic mass fractions ϕ_i for a cell containing just three protein species T, R, and H. Proteins H do not contribute to the global metabolic flux, so that $C_H^J = 0$. Purple arrows indicates the effect of a reduction in ϕ_T in favor of ϕ_R, which increases the growth control possessed by T.
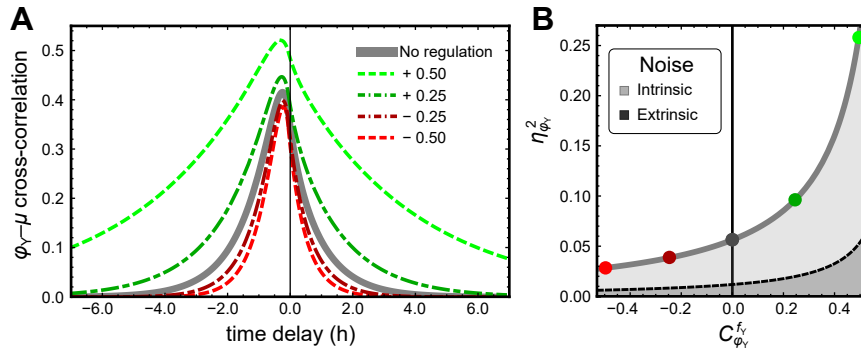
Figure S2.2: **The effects of regulation on noise parameters in the two-protein model.** *(A)* Cross-correlations between the expression level ϕ_Y of protein Y and growth rate μ in the two-protein model, for varying levels of positive (green) and negative (red) auto-regulation, as quantified by the regulatory control coefficient $C_Y^{f_Y}$. The curve plotted in gray is based on $C_Y^{f_Y} = 0$ (no auto-regulation), and all other parameters are chosen as in Fig. 2.3A; therefore the gray curve corresponds to the gray curve of Fig. 2.3A. *(B)* Analytical solution of the coefficient of variation of the concentration of protein Y in the two-protein model, under varying levels of auto-regulation. The intrinsic and extrinsic noise components are indicated by the two shades of gray. The colored circles indicate the parameter choices belonging to the corresponding curves of panel A.
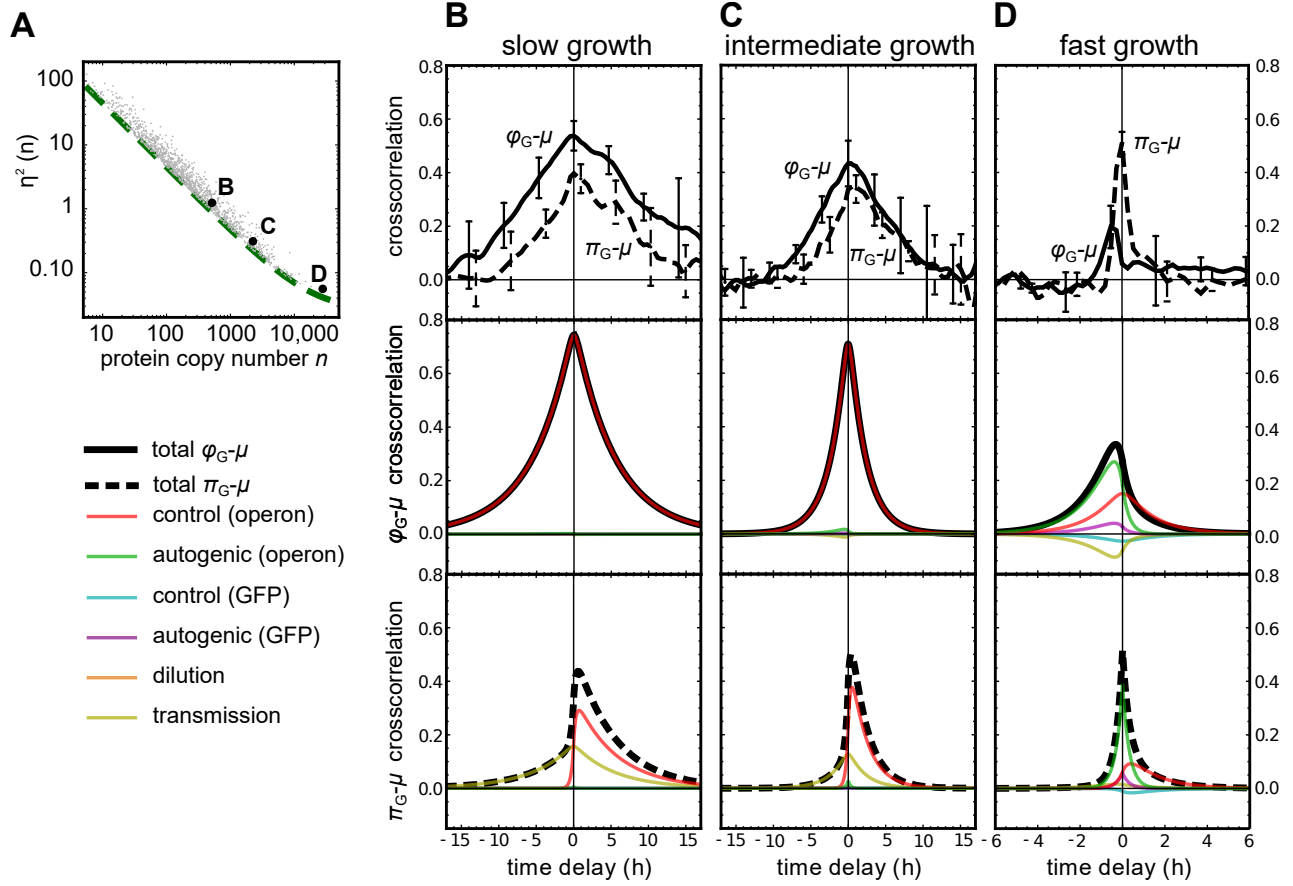
Figure S2.3: **Expression–growth cross-correlations in a many-protein model based on sampled variances.** Analysis of the model with protein abundances taken from Arike *et al.* [1], and variances sampled from a phenomenological noise model (see [195] and equation S2.100). This figure is equivalent to Fig. 2.4, except that it is based on different protein abundances and variances. *(A)* Distribution of protein abundances and variances. Each gray dot represents a protein; the black points indicate the abundance and variance of the GFP reporter under the three growth condition (equivalent to Fig. 2.4C). *(B)–(D)* Growth rate cross-correlations between GFP concentration and growth rate (top panels) and GFP synthesis rate and growth rate (bottom panels), for the three growth conditions (equivalent to Fig. 2.4D, E, and F).

Figure S2.4: **Production–growth cross-correlations corresponding to Fig. 2.5.** This figure shows the cross-correlations between GFP abundance and growth rate for the same parameters whose concentration–growth cross-correlations were studied in Fig. 2.5. *(A)* The operon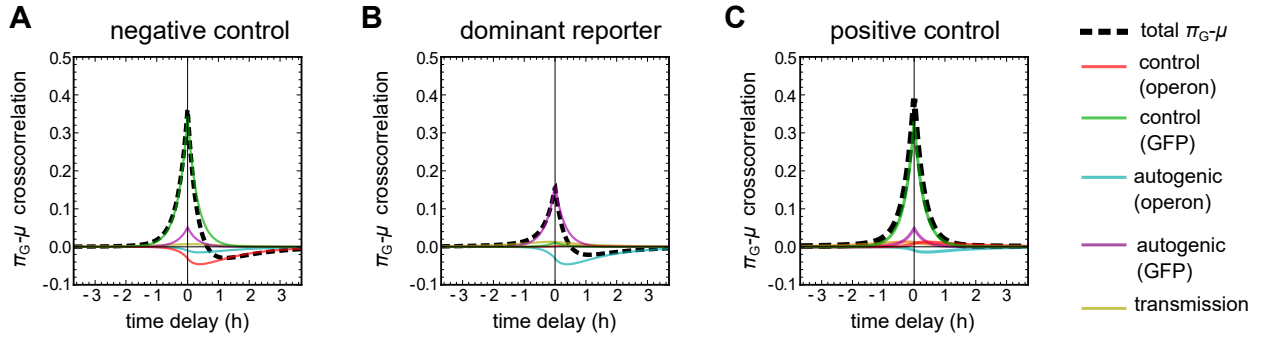 has a negative growth-control coefficient (cf. Fig. 2.5A). *(B)* Fluctuations in GFP are dominated by its private noise source N_G, and are therefore largely decoupled from the fluctuations in the rest of the *lac* operon (cf. Fig. 2.5B). *(C)* Highly symmetrical π_G–μ cross-correlation despite a (slightly) positive control of the operon, which is masked by the negative control carried by the reporter protein as well as by the asymmetrical transmission mode.
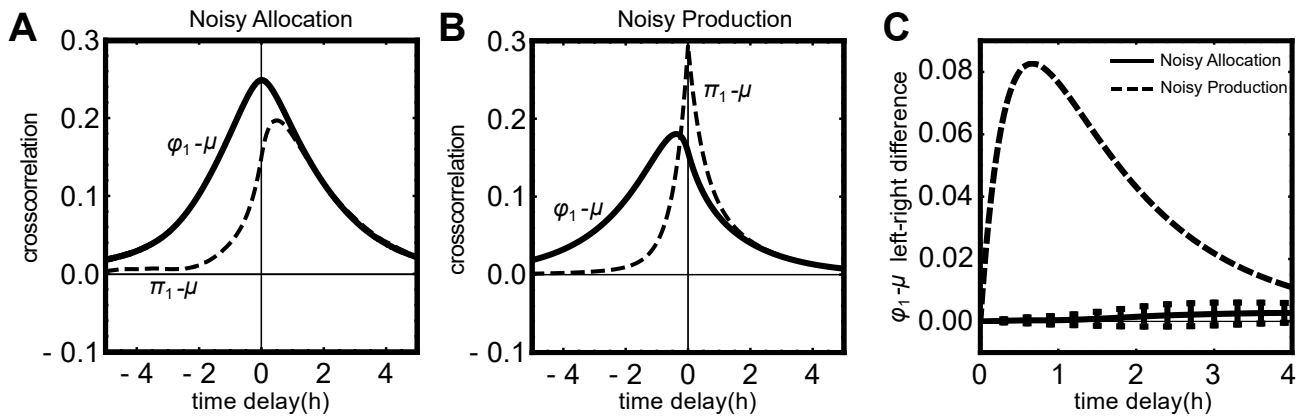

Figure S2.5: **Comparison of noisy-allocation and noisy-production models.** *(A)* Simulations of an alternative model in which the noise sources act on the allocation of the flux rather than on each protein synthesis rate independently. Shown are the ϕ_1–μ (solid line) and π_1–μ (dashed line) cross-correlations of protein 1 in a cell containing 40 protein species with arbitrary parameters (see section S2.6.1 for more details about the simulation). Here, $\phi_{0,1} = 0.027$ and $C_1^\mu = 0.022$. *(B)* Analytical results for the ϕ_1–μ (solid line) and π_1–μ (dashed line) cross-correlations of the same cell, but where noise again acts on each protein synthesis rate independently. The amplitudes of the noise sources were adjusted such that the variances of all protein species were identical to those in panel (A). *(C)* The asymmetry of the ϕ_1–μ cross-correlations $R(\phi_1, \mu)(\tau)$ shown in panels A (solid line) and B (dashed line), quantified as $R_{\phi_1\mu}(-\tau) - R_{\phi_1\mu}(\tau)$. In order to estimate error bars, we repeated the "noisy allocation" simulation of the exact same cell 10 times; error bars indicate the standard error of the mean. Note that, for the "noisy allocation" model, zero lies within the narrow error bars, showing that the cross-correlation is highly symmetric and therefore shows no evidence for an asymmetric mode akin to the autogenic or common-noise mode.

Laurens H.J. Krah, Rutger Hermsen

ABSTRACT

In bacterial cells, protein expression is a highly stochastic process. Gene expression noise moreover propagates through the cell and adds to fluctuations in the cellular growth rate. A common intuition is that, due to their relatively high noise amplitudes, proteins with a low mean expression level are the most important drivers of fluctuations in physiological variables. In this work, we challenge this intuition by considering the effect of natural selection on noise propagation. Mathematically, the contribution of each protein species to the noise in the growth rate depends on two factors: the noise amplitude of the protein's expression level, and the sensitivity of the growth rate to fluctuations in that protein's concentration. We argue that natural selection, while shaping mean abundances to increase the mean growth rate, also affects cellular sensitivities. In the limit in which cells grow optimally fast, the growth rate becomes most sensitive to fluctuations in highly abundant proteins. This causes abundant proteins to overall contribute strongly to the noise in the growth rate, despite their low noise levels. We further explore this result in an experimental data set of protein abundances, and test key assumptions in an evolving, stochastic toy model of cellular growth.

Stochasticity is inherent to gene expression [42, 144, 175]. Stochastic variation in the copy numbers of proteins is observed even under constant external conditions, and among individual cells in a population of isogenic bacteria. How cellular stochasticity, and noise in protein expression specifically, interferes with the functioning, survival and fitness of bacteria has been of great interest for many years [41, 100, 144, 160, 176, 185].

Noisy gene expression is indeed commonly accepted as the dominant mechanism behind the strong phenotypic variation that has been observed in populations of genetically identical cells [145]. In an exponentially growing population of cells, even the growth rate of individual cells is distributed surprisingly broadly [170]. Since cellular growth rate (and its population average) is often considered an important proxy for bacterial fitness, the growth rate —and how its variation is shaped by noisy gene expression— has received much attention [61, 71, 81, 156]. Notably, noise in the concentration of metabolic proteins is shown to propagate from the protein-level, via the metabolic network, to the instantaneous single-cell growth rate [86].

Commonly, noise is characterised in terms of the coefficient of variation (CV), defined as the standard deviation divided by the mean. In snapshots of bacterial populations, proteins with a higher mean expression level ($\mathbb{E}[X]$) generally have a lower coefficient of variation squared (CV^2) [4, 173]. For proteins with a low mean expression, noise is dominated by the intrinsic stochasticity of the chemical reactions involved [138, 168] and CV^2 scales as $1/\mathbb{E}[X]$ [25, 42, 47, 160, 173, 195]. For higher mean expression, noise levels decrease to eventually reach a plateau, where fluctuations in gene expression are dominated by extrinsic noise, such as noise resulting from cell division or environmental noise. Because of their larger noise levels, lowly expressed proteins are commonly assumed to be particularly important drivers of fluctuations in variables at the cellular level, such as the growth rate. At the same time the effects of the relatively small fluctuations of highly abundant proteins have largely been neglected.

However, while protein noise levels are mainly determined by mean abundance, the mean abundance itself is a product of evolution. Under many external conditions, bacteria indeed seem to tune their protein levels in order to grow, on average, at a near-optimal rate [17, 32, 51, 106, 165, 181]. Optimal gene expression for fast growth has also been an important and fruitful assumption in countless modelling studies and techniques concerning deterministic growth, including Flux Balance Analysis [56, 76, 131]. So far, however, the possible effects of natural selection on how noise in protein expression affects the noise in macroscopic variables such as the growth rate, have not been considered.

In this work, we therefore consider bacteria whose protein expression levels are shaped by natural selection acting on the population growth rate. For the extreme case of cells growing optimally fast, we obtain analytical predictions for the contribution of each protein to the noise in the growth rate as a function of its mean expression only. The main result, directly opposing common intuitions, is that proteins with a high mean expression are most important for the noise at the cellular level. The argument is, in short, that a protein's contribution to noise in the growth rate does not only depend on the protein's noise level, but also on the sensitivity of the growth rate to that protein's fluctuations. We show that when protein expression levels are optimised for fast growth, the growth rate becomes most sensitive to fluctuations of abundant protein species. This causes abundant proteins

to overall contribute strongly to the noise in the growth rate, despite their low noise levels. In a stochastic toy model of gene expression and growth, we verify and further investigate the role of natural selection on shaping noise propagation properties. Lastly, an analysis on experimental data of protein abundances and protein noise levels indicates that the common intuition –cells behave noisily because of the low copy number of certain molecular species– might be incorrect.

3.2 RESULTS

To grow, bacteria need to express a certain set of (metabolic) proteins. Together, these proteins create a metabolic flux used to build cellular components and new proteins. Since the cell's growth rate is limited by this metabolic flux, noise in the expression levels of the proteins involved propagates through the metabolic network to affect the growth rate [86]. For any fixed external environment, we therefore assume the existence of an unknown function $\mu(\vec{X})$ that describes the instantaneous rate of cellular growth as a function of the copy numbers of all proteins, $\vec{X}$. The growth rate is thus a deterministic function of stochastic variables. In a population snapshot, different individuals stochastically express different copy numbers of their proteins and hence the growth rates of individuals will differ.

To quantify how variation in the expression of protein i affects growth rate μ, we use previously defined Growth Control Coefficients, which measure the sensitivity of the growth rate to small changes in the copy number of protein i [87]:

$$C_i^\mu \equiv \left(\frac{X_i}{\mu} \frac{\partial \mu}{\partial X_i} \right) \Bigg|_{\mathbb{E}[\vec{X}]} . \tag{3.1}$$

Here, the expectation value is taken over the distribution of protein copy numbers across a population of cells. As we will show below, these GCCs offer a way to decompose and analyse the noise in the growth rate in terms of contributions by each of the noisy components, the proteins.

To arrive at a comprehensive and useful noise decomposition, we adhere to two simplifying assumptions. First, noise levels are assumed to be small, so that all protein abundances are close to their means. The growth rate can then be approximated as a linear function of the protein levels. Secondly, fluctuations in all protein species are assumed to be independent. In bacterial cells, this is certainly not the case. However in the case of correlated fluctuations, noise contributions can not be uniquely defined [18, 107]. Indeed, when two proteins correlate, and their joint fluctuations affect growth, the attribution of the noise contribution to either protein is arbitrary. Therefore, we here present the simplified case where all protein abundances are uncorrelated so that noise contributions can be uniquely defined and understood intuitively.

Under these assumptions noise in the growth rate, CV_μ^2, can be approximated as a sum of contributions from all protein species:

$$CV_\mu^2 \approx \sum_i CV_i^2 \left(C_i^\mu \right)^2 , \tag{3.2}$$

where CV_i is the coefficient of variation of the copy number of protein species i across a population of cells (see SI, section S3.1 for the derivation). In this equation, each protein's

contribution consists of two factors. The first factor is no surprise: the proteins' coefficient of variation which quantifies the fluctuations in the expression of that particular protein. The second factor is the protein's GCC, which quantifies how strongly these fluctuations actually affect the cellular growth rate.

3.2.1   *Distribution of Growth Control Coefficients*

To further quantify which proteins are important for noise in the growth rate, we need to gain more insight into how growth control is distributed among proteins. This distribution is not arbitrary due to three properties of the GCCs, which are discussed below.

Sum Rule

Firstly, the sum of the GCCs equals zero [87]:

$$\sum_i C_i^\mu = 0. \tag{3.3}$$

This sum rule originates from the so-called *intensivity* of the growth rate: if all protein copy numbers inside a cell are increased by the same factor, the cellular flux increases, but the mass increases as well, such that the cell's growth rate (mass increase per mass) stays the same. That the growth rate is to a good approximation an intensive variable has been shown in multiple experiments [86, 170] and is a common modelling assumption [76]. Moreover, it is analogous to the assumption that the metabolic flux and cellular mass are *extensive* variables, which has been used in Metabolic Control Analysis to derive a similar sum rules for fluxes [77].

H-proteins

Secondly, there is a set of proteins, here called H-proteins, that are crucial for the cell's survival, but do not contribute to metabolism or cellular growth. This set H includes 'housekeeping proteins' participating in, *e.g.*, stress-response, immunity, and DNA damage repair; in bio-engineering, H may also contain engineered pathways. In wild type *Escherichia coli*, the H-sector comprises an estimated $25 - 40\%$ of the total protein mass [135]. Even when H-sector proteins are not toxic or otherwise harmful to the cell, their control on the growth rate will still be negative. This is because their synthesis does take up resources that otherwise could go to growth-related enzymes. Previously, the GCC of such H-sector proteins has been calculated [87] to be equal to their mass fraction:

$$C_{i\in H}^\mu = -\phi_i. \tag{3.4}$$

Here we write $\phi_i := \mathbb{E}[X_i] / \sum_j \mathbb{E}[X_j]$ for the proteome mass fractions ϕ_i of each protein species i and ignore, for simplicity of notation, that different proteins have different masses. The mass fraction of the total H sector is denoted as ϕ_H.

Together with the sum rule, the presence of the H-sector has important consequences for the distribution of GCCs: because some proteins have negative GCCs, others must have a positive GCC.

Optimal Growth

Thirdly, natural selection tends to favour populations of cells that, on average, grow faster. This drives the (mean) expression levels of many proteins to be (near)-optimal for growth [32, 181]. We here show that this also affects their GCCs by considering the extreme case of a cell in which the expression levels of all proteins (except for those in the H sector) are fully optimised for growth.

To do so, evolution is treated mathematically as a constrained optimisation problem, where the mean growth rate, $\mathbb{E}[\mu]$, is optimised under two constraints. First, the cell's protein density is kept constant. Second, only a fixed fraction of the proteome $(1 - \phi_H)$ can be allocated towards proteins related to growth. To maximise the growth rate, only the (mean) protein abundances inside this fraction can be tuned by evolution while the total abundance must stay the same.

Formally, the optimisation can be done using Lagrange multipliers on a linearisation of μ (see SI, section S3.1). For all proteins that are not in the H sector, the result is the following important expression for the GCCs in the optimal state:

$$C^{\mu^*} = \left(\frac{\phi_H}{1 - \phi_H}\right)\phi_i^*. \tag{3.5}$$

Here, the asterisk indicates that the equation is only valid under optimality. Intuitively, the result can be understood as follows. In the optimal state, all partial derivatives of the growth rate must be equal: if the growth rate would increase more upon increasing $\mathbb{E}[X_i]$ than upon increasing $\mathbb{E}[X_j]$, increasing the expression of i at the expense of j would increase the growth rate, and hence the growth rate would not be optimal.

equation 3.5 reveals two important properties for cells optimised for growth. First, growth control is shared between all metabolic proteins: there is no single growth-limiting protein. Secondly, and most importantly, enzymes with a higher mean expression level have a proportionally larger control on the growth rate.

3.2.2   *Combining all factors*

The distribution of the GCCs in optimally growing cells (equation 3.5) can be combined with the experimentally observed scaling of coefficients of variations to predict the contribution of each protein species i to the noise in the growth rate. We write κ_i^P for this contribution, which is defined as the protein's relative contribution to the CV of μ as expressed in equation 3.2:

$$\kappa_i^P := \frac{\left(C_i^\mu\right)^2 CV_i^2}{CV_\mu^2} \approx \frac{\left(C_i^\mu\right)^2 CV_i^2}{\sum_j \left(C_j^\mu\right)^2 CV_j^2}. \tag{3.6}$$

Inspired by experimental data [173, 195], the intrinsic noise component is assumed to be inversely proportional to mean abundance:

$$CV_i^2 = F/\mathbb{E}[X_i], \tag{3.7}$$

with a fixed Fano Factor F. If we ignore the noise plateau caused by extrinsic noise sources, equation 3.7 sets the noise levels of all protein species. Note that for the highly expressed

proteins, noise levels are thus deliberately underestimated, resulting in a conservative estimate for their contribution to noise in the growth rate.

To now analyse noise propagation in optimally growing cells, equations 3.4, 3.5 and 3.7 are inserted in equation 3.6. This results in:

$$\kappa_i^{P*} = \begin{cases} \dfrac{\phi_H}{1 - \phi_H} \phi_i^* & \text{for } i \text{ not in } H, \\[2ex] \dfrac{1 - \phi_H}{\phi_H} \phi_i & \text{for } i \text{ in } H. \end{cases} \tag{3.8}$$

This equation is the pivotal finding of this study. It states that, in cells whose expression levels are optimised for growth, κ_i^P is proportional to ϕ_i, that is, proteins with a high mean expression contribute most strongly to fluctuations in the growth rate.

The implications of the above equations become clear when applied to a data set of measured protein abundances and protein noise levels in the model bacterium *E. coli* [155, 173]. Under the assumptions of equations 3.2 and 3.5, the top 5% most abundant protein species are estimated to contribute over 90% of the noise in the growth rate (Fig. 3.1, red dots). This contribution is significantly larger than might have been expected *a priori*: If all GCCs are assumed equal (Fig. 3.1, purple dots) these abundant proteins contribute only 40%, and if the GCCs of the optimal state are shuffled so that the correlation between a protein's abundance and its GCC is broken (Fig. 3.1, black dots) their contribution becomes negligible ($< 2\%$). On the other hand, the 50% least abundant proteins are estimated to contribute only 2% of the overall noise in the growth rate (Fig. 3.1, red curve), instead of $50 - 90\%$. For details of the analysis, see SI section S3.2 and Fig. S3.1.

3.2.3   *Stochastic Toy Model*

To examine if our results still hold for the large noise levels seen in living cells, we study noise propagation in a stochastic toy model of a growing, and evolving, cell. Specifically, we test if a positive scaling between ϕ and κ can still be observed for genotypes that have evolved by random mutations, under realistic noise levels.

To do so, we defined a highly simplified model of a growing cell with stochastic protein expression levels. To mimic the effects of evolution, we then employ random mutations to search for the mean protein expression levels that optimise the mean growth rate of such cells. Next, we characterised the noise propagation in such optimised cells to verify the predictions of equations 3.5 and 3.8.

The model cell consists of a linear metabolic pathway consisting of five reactions that import an external metabolite (m_1) and convert it to biomass (Fig. 3.2A). Each reaction is catalysed by a single enzyme species and inhibited by its own product. Additionally, a sixth protein species is expressed that is not metabolically active, representing the H-sector. Given the abundances $\vec{X}$ of all proteins, the instantaneous growth rate μ is defined as the steady state flux through the pathway divided by the total number of expressed proteins, including the H-sector. Note that the growth rate depends non-linearly on all protein abundances. (For more details, see SI, section S3.3.)

The abundances $\vec{X}$ themselves are stochastic: each X_i is distributed in the population according to a Gamma distribution [47, 173] characterised by mean $\mathbb{E}[X_i]$ and a Fano
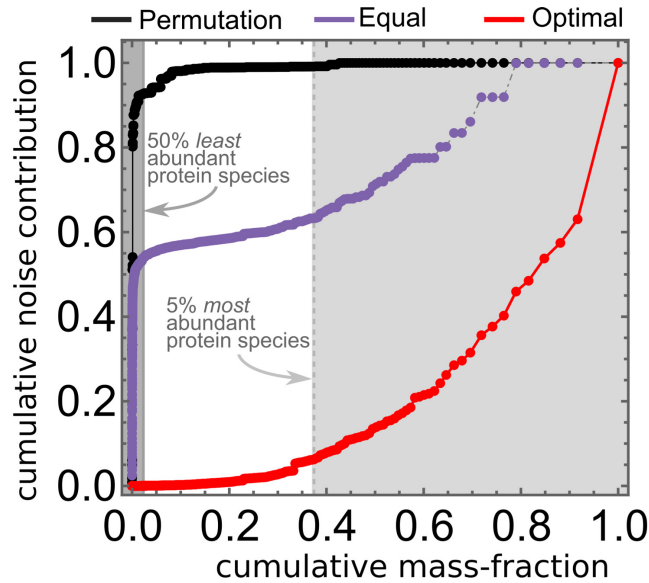
Figure 3.1: Cumulative noise contribution as a function of cumulative mass fraction, both calculated from protein abundances and noise levels measured in *E. coli* [173]. Protein species were ordered by their mass fraction ϕ_i, and the cumulative mass fractions and noise contributions subsequently calculated as $\sum_{j=1}^{i} \phi_j$ and $\sum_{j=1}^{i} \kappa_j$, respectively. GCCs are either set by equation 3.5 (red dots), all equal (purple dots), or a random permutation of the optimal GCCs (black dots).

factor F. The Fano factor is chosen the same for all proteins, consistent with equation 3.7, and sets the overall noise amplitude in the cell.

It is useful to distinguish three levels of description of the model cells: their kinotype, genotype, and phenotype. We introduce the 'kinotype' as the set of reaction parameters that fully characterise the enzymes in a cell's metabolic network: kinetic rates, Michaelis–Menten constants and inhibition parameters of the five reactions. We define a cell's 'genotype' as the mean abundances of all protein species. Lastly, a cell's 'phenotype' is given by the vector of the current protein abundances and the corresponding growth rate. The phenotype is therefore a multi-dimensional stochastic variable whose probability distribution depends on the genotype.

During an evolutionary trajectory, the genotype is repeatedly subjected to mutations that are subsequently either rejected or accepted. Mutations increase or decrease the mean abundance of one particular protein species, after which the mean expression of all other metabolic proteins is adjusted such that the total (mean) protein abundance remains fixed ($\sum_i \mathbb{E}[X_i] = \Omega = 10^4$ in all simulations, for more details see SI, section S3.3). Mutations thus affect the protein copy number distributions across the population and therewith also the probability distribution of the growth rate. A mutated genotype is accepted only if it increases the population mean growth rate, which is determined by sampling many phenotypes generated by that genotype. The evolutionary algorithm is halted when 100 consecutive mutations around a particular genotype are rejected. The resulting genotype is expected to be close to a local optimum, although it is not guaranteed to be exactly the mathematical optimum due to sampling error.
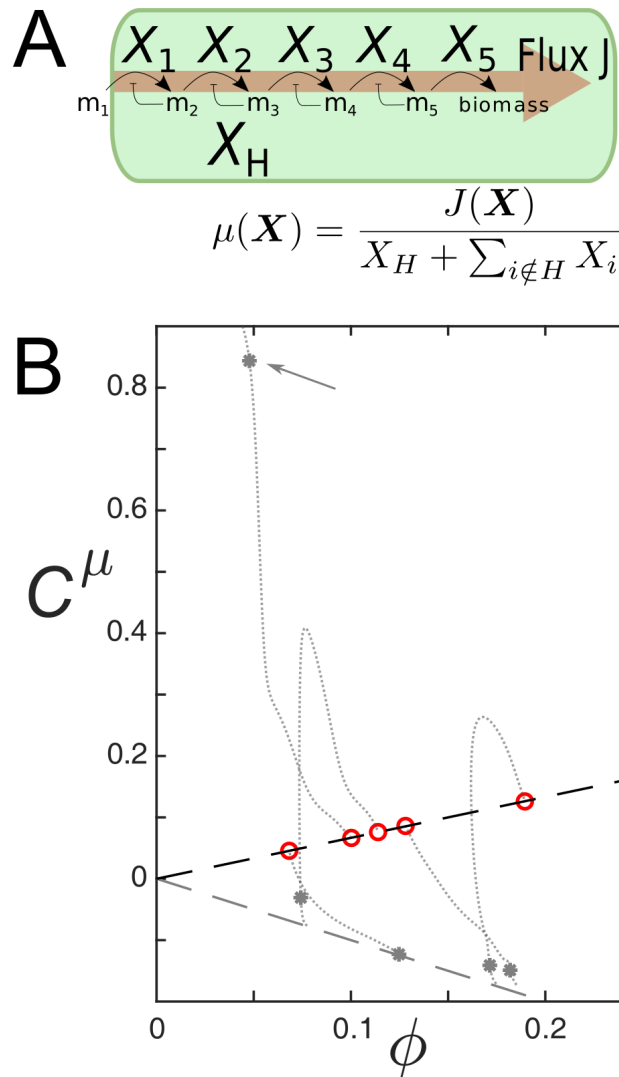
Figure 3.2: (A) Representation of the stochastic toy model. A cell expresses five metabolic protein species, each catalysing a single reaction in a linear reaction chain that imports and converts a fixed external metabolite, m_1, into biomass. The growth rate is defined as the steady state flux through the network, divided by the total number of expressed proteins, including the H-sector protein. (B) Example trajectory of the GCCs of metabolic proteins during the optimisation of a single kinotype (grey dotted lines). Grey dots are GCCs at an early stage of this process and the arrow indicates the rate-limiting protein species. Red point are the values of the GCCs in the optimal genotype, matching the predicted scaling for metabolic proteins (positive dashed line, negative dashed line is the prediction for H-proteins). See SI, table S3.1 for parameters.

Using a method adopted from [18] the noise contribution of each protein species is then measured by again sampling and analysing many phenotypes. These measured noise contributions were compared with our prediction, κ^P (equation 3.8). Lastly, the whole process above was repeated for many kinotypes (randomly generated; see SI, section S3.3), resulting in different optimal genotypes.

Low-Noise Regime

Before analysing the more realistic regime of high noise levels (large variance of the protein copy numbers), we first study the model in a low-noise regime (using $F = 1$, $\Omega = 10^4$, resulting in small copy-number variance), where our results (equation 3.2 and 3.8) are expected to hold well. When noise levels are this low, the mean growth rate $\mathbb{E}[\mu]$ is well approximated by the growth rate in the vector of mean abundances, $\mu(\mathbb{E}[\vec{X}])$. Instead of using the undirected, slow stochastic evolutionary algorithm described above, the growth rate was therefore optimised using a deterministic gradient-based hill climb algorithm (for details see SI, section S3.3).

During each step of the optimisation process, we measured the GCCs of the metabolic proteins to observe how they adjust during optimisation. A representative example of such a trajectory is shown in Fig. 3.2B (dotted grey lines). Early in the optimisation process, when genotypes are still far from optimal (grey dots), often one particular protein species (shown in the figure with an arrow) is strongly limiting growth ($C_i^\mu \approx 1$, indicating that the growth rate could be improved by increasing this protein's expression level). In contrast, the expression of other metabolic proteins is too high; those proteins have a negative GCC similar to H-sector proteins, indicating that almost all of their expression is a burden to the cell. Eventually, as fitter genotypes are found, growth control becomes shared among all proteins (Fig. 3.2B, dotted grey lines). When the optimisation algorithm simulation has found the optimal genotype, the predicted positive scaling between a protein's GCC and its mean abundance is obtained (Fig. 3.2B, red points, and equation 3.5).

Repeating the same process for multiple kinotypes ($n = 10$) confirms the generality of the positive scaling between C^μ and ϕ after optimisation (Fig. 3.3A, red points). Again, note that in an early stage of the optimisation process the distribution of the GCCs is markedly different: Although only metabolic proteins are shown in the figure, some of them have negative GCCs that resemble the GCCs of H-sector proteins (Fig. 3.3A, grey dots).

Next, we measured for each kinotype the noise contributions of each protein in the optimal genotype, and, for comparison, in a non-optimal genotype. For all kinotypes, the noise contributions in the optimal genotypes neatly follow our prediction (Fig. 3.3B, red points). In contrast, for non-optimal genotypes, noise contributions are dominated by only a few lowly expressed proteins (Fig. 3.3B, grey dots). The mean abundance of these proteins is below the optimal value, causing both their GCC and their CV to be large, resulting in a large noise contribution.

This analysis clearly highlights the fundamental role of evolution in shaping noise propagation properties: only in evolved cells that grow at an (almost) optimal rate a positive scaling exist between κ and ϕ.
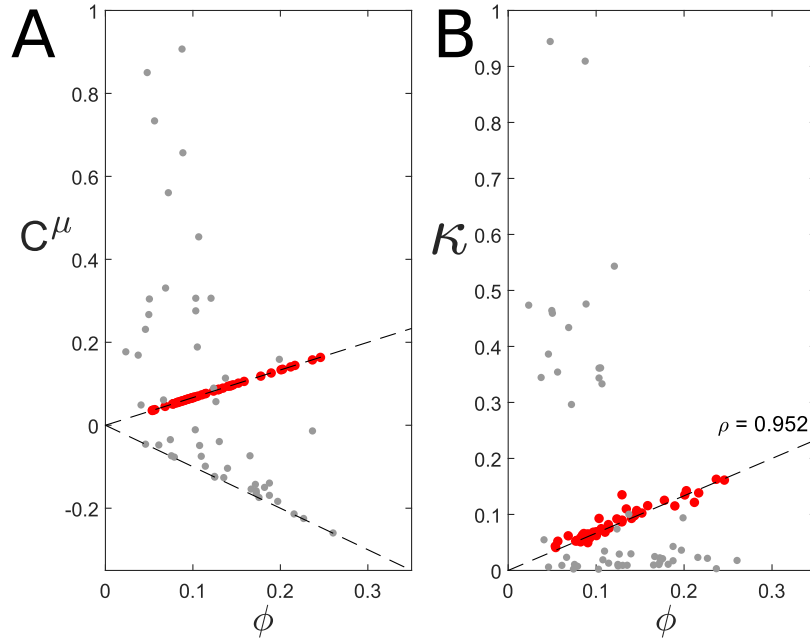
Figure 3.3: Perfect prediction in the case of linear noise for 10 different kinotypes. (A) Grey dots are GCCs of metabolic proteins in genotypes after 150 optimisation steps, which are not yet optimal. Red points are the GCCs in the optimal genotypes. Dashed lines are predictions for the values of GCCs for metabolic (positive) or H-proteins (negative) (B) Measured noise contributions compared to predicted noise contributions (dashed line). Red points are measured in the optimal genotype, grey points in the non-optimal genotypes.

High-Noise Regime

Next we study the toy model in a high-noise regime, where copy number variation matches observed variation in living bacteria more closely (F = 10, $\Omega = 10^4$, resulting in CVs up to 0.2, Fig. S3.5C). In this regime, the non-linear dependence of the growth rate on the protein abundances becomes important and might influence the mean growth rate. That is, a genotype that was optimal in the low-noise regime, does not necessarily yield the highest mean growth rate in the high-noise regime. Below, we therefore distinguish two genotypes for each kinotype: the 'Low-Noise' (LN) genotype, which is optimal in the low-noise regime, and the 'High-Noise' (HN) genotype', which evolved in the high-noise regime by random mutations. To efficiently find the HN genotype of each kinotype, we employed the evolutionary algorithm described above, starting with the LN genotypes.

For some kinotypes, the resulting HN genotypes indeed differed significantly from the LN ones. This can be understood as follows. In the LN genotype, when noise levels are low, control over the growth rate is shared between the proteins. In the same LN genotype, higher noise levels can increase the probability that in some sampled phenotypes a single protein species becomes the sole rate-limiting step. Indeed, some genotypes that were optimal in the low-noise regime generated many slow-growing phenotypes when noise levels were high (Fig. S3.2B). The low growth rates were often caused by a single protein species whose phenotypic expression was too low (Fig. S3.2A) and therefore became a bottle-neck.

From an allocation point of view, increasing the mean expression of lowly abundant, rate-limiting proteins is cheap: little additional resources are needed to cause a relatively
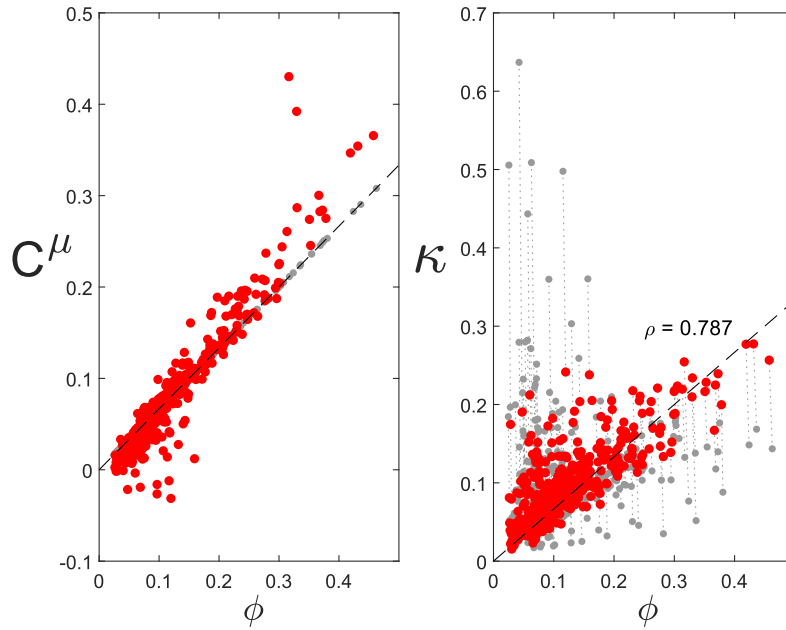
Figure 3.4: For 100 kinotypes, Low-noise genotypes (grey dots) compared with high-noise geno-types (red dots). Dashed lines are theoretical predictions for metabolic proteins. (A) Growth Control Coefficients. (B) Measured noise contributions.

large change in the protein's expression level. Indeed, the mean expression of potential bottlenecks increased during evolution, but mainly for lowly expressed protein species (Fig. S3.2B and Fig. S3.3).

The distribution of the GCCs is also different in the evolved HN genotypes compared to the LN genotypes (Fig. 3.4A, red points). GCCs are by definition linear measures and because in the high-noise regime the non-linearity of the growth rate become relevant, equation 3.5 is not expected to hold exactly anymore. Interestingly, however, the positive scaling between ϕ and C^μ remains, and becomes, if anything, even steeper. Again, this makes sense: when noise levels increase, lowly expressed proteins are, due to their larger CV, more likely to fluctuate down to levels that strongly limit growth (Fig. S3.3A). Increasing the mean expression of these proteins will reduce their GCC (Fig. S3.3B-C), while increasing the GCC of the other proteins via the sum rule (equation 3.3). The net result is an increase in the slope in Figure 3.4A.

Importantly, the positive scaling between κ and a protein's mean expression was again observed in the high-noise regime (Fig. 3.4B, red points). This is remarkable, since the mathematical prediction (equation 3.8) was derived under the assumption of small noise amplitudes. To understand why the positive scaling nonetheless persists in the high-noise regime, the same reasoning as for the GCCs can be followed. At the start of the evolution-ary trajectory, when the genotype is still the LN genotype, some lowly abundant protein species are dominant noise contributors (Fig 3.4B, grey points top-left corner). During evo-lution those proteins obtained a slightly higher expression, reducing both their GCC and their noise contribution κ (Fig. 3.4B, red points and Fig. S3.3B-C), and at the same time increasing the contributions of the other proteins species.

In this article we argued that highly expressed proteins play an important role in system-level noise properties despite their low noise levels. In summary, our argument is that, in cells optimised for growth, abundant protein species have large Growth Control Coefficients (GCCs). The product of a protein's CV and GCC –an indicator for the protein's noise contribution– is then predicted to be proportional to the protein's mean abundance. A crude estimate from an *E. coli* data set suggests that the 50% least abundant protein species contribute only ∼ 2% of the noise in the growth rate, instead of up to 90% as might have been expected *a priori* if the scaling of the GCCs is ignored. In contrast, the top 5% most abundant species are estimated to contribute well over 80% in optimised cells. In a simplistic toy model of cellular growth, we then showed that the positive scaling between a protein's mean abundance and its noise contributions persist even when noise levels are considerable and the growth rate is a non-linear function of protein abundances.

Of course, these results rely on the assumption of a perfect optimum. Experiments suggest, however, that bacteria are not always in such an optimum [181]. In certain fixed environments, adaptive mutants arise with higher growth rates, indicating that the wild-type growth rate is not optimal yet [69]. Assuming those adaptive mutations did not take place in the H-sector, it remains a question whether in wild-type cells abundant proteins also contribute most to noise in the growth rate. However, in our simulations the positive scaling between ϕ and κ persists even in the high noise regime, where genotypes were evolved by random mutations–and therefore are not necessarily perfectly optimal. Moreover, some mutants generated around the evolved genotype still displayed the positive scaling even when growing, on average, significantly slower (Fig. S3.6). Other generated mutants, however, did lose their positive scaling. It would therefore be interesting to further investigate why some non-optimal mutants still display the predicted scaling, but others do not.

While the toy model assumed particular enzyme dynamics, all mathematical predictions were derived without any assumptions concerning the underlying biochemistry. This implies that the strong contribution of highly expressed proteins to the noise in the growth rate is a general property of evolved biological systems as long as protein expression is in some way constrained.

The presence of such cellular constraints is crucial for our results. In this study, a tight constraint was imposed by assuming that the H-sector is completely static. However, a small relaxation of this constraint -*e.g.*, assuming that the allocation towards the H-sector has to be within a certain range- should yield similar results. Also other allocation constraints, *e.g.* a fixed total protein abundance allocated to a particular metabolic pathway, or a maximum density of membrane proteins [169], will result in a similar positive scaling between ϕ and κ in all constrained proteins.

We point out that, besides the growth rate, other cellular traits, such as stress response or antibiotic resistance, are important for bacterial fitness as well. Interestingly, noise in these traits can be analysed in the same way as noise in the cellular growth rate. Therefore, the argument suggests that noise in any intensive trait that has been optimised during the bacteria's evolutionary history should be dominated by highly expressed proteins.

Throughout this paper, we ignored all correlations between protein species [29, 127], because in the presence of such correlations, the contribution of noise in a particular protein to the noise in cellular growth rate becomes ill defined. One way to circumvent this

problem is to use a fine-grained model description, *e.g.* at the level of individual chemical reactions, in which the noise sources are inherently uncorrelated [97, 176], or to adopt a meta-modelling approach [107] where highly correlated protein species (*e.g.,* those coded in the same operon) are modelled as single, noise contributing units. That said, the method presented here allows for a more intuitive interpretation of noise contributions, because it directly relates noise contributions to observed protein abundances.

The results discussed above add to the realisation that global cellular constraints have intricate consequences for the overall physiology of evolved cells, from noisy gene expression [87], to metabolism [56, 169] and growth [198]. Our work highlights the holistic nature of noise propagation via the sum rule for the GCCs (equation 3.3). The sum rule specifically could have important consequences for biotechnology: tinkering with a specific part of the cell affects noise propagation properties of the entire system. For example, most synthetic proteins or pathways do not contribute to growth, but instead create by-products. Such pathways will thus have a negative GCC and hence increase the GCC -and therewith the propagation- of all metabolic proteins.

The stochastic toy model moreover revealed a trade-off between efficient resource allocation and the robustness of metabolism to expression noise [37]. Genotypes that are optimal in the low-noise regime allocate resources efficiently, but lack metabolic robustness at higher noise levels. Cells with different levels of expression noise therefore required different genotypes to grow, on average, the fastest (Fig. 3.4). Similar observations were also made in recent experiments in yeast [156]. Together, these observations can have consequences for Flux Balance Analysis-like techniques [131], where optimal growth states have so far been calculated mostly deterministically, *i.e.,* optimising the growth rate in the mean expression levels.

We conclude that noise in gene expression –and its propagation towards the growth rate– needs to be considered when discussing optimal growth, but also *vice versa*: when enzyme expression is optimised, this affects noise propagation in such a way that abundant protein species become most relevant for noise on a system level.

SUPPLEMENTARY INFORMATION CHAPTER 3

S3.1 DERIVATION OF EQUATIONS 3.2-3.8

Here we derive equations 3.2 and (therewith) 3.8.

S3.1.1 *Growth rate as an intensive function of protein copy numbers*

We assume that the growth rate μ is a deterministic, *intensive* function of the stochastic protein copy numbers $\vec{X}$. The intensivity of the growth rate implies that if all the copy numbers are multiplied by the same factor α, the growth rate does not change (*i.e.*, $\mu(\alpha\vec{X}) = \mu(\vec{X})$). Describing the growth rate as a function protein copy numbers is therefore equivalent to considering proteome mass fractions or concentrations. In our derivation, we start with a description based on copy numbers, but our final results will be in terms of proteome mass fractions.

We do not model the underlying causes of protein stochasticity, and instead assume the distribution of each protein species across a population is known. Protein copy numbers of individual cells thus differ, and because the growth rate of each individual cell is determined by protein copy numbers, also the growth rate differs between individuals cells. Writing the growth rate as a function of protein copy numbers moreover implicitly assumes that intrinsic fluctuations of all other cellular components (*e.g.*, nucleic acids, lipids, and metabolites) relax on timescales much faster than growth (seconds rather than minutes or hours), and their concentrations are thus determined by the protein copy numbers. Note that throughout the derivation, we do not consider any temporal dynamics of $\vec{X}$.

S3.1.2 *Linearisation of μ*

Assuming that $\vec{X}$ is always close to some linearisation point $\widetilde{\vec{X}}$, we can write to first order:

$$\mu(\vec{X}) \approx \mu(\widetilde{\vec{X}}) \left(1 + \sum_i \left(\frac{X_i}{\mu} \frac{\partial \mu}{\partial X_i} \right)\bigg|_{\widetilde{X}} \left(\frac{X_i - \widetilde{X_i}}{\widetilde{X_i}} \right) \right). \tag{S3.1}$$

In a snapshot of a population of cells, this linearisation then approximates the growth rate of individual cells given their stochastic vector $\vec{X}$.

S3.1.3 *Decomposition of CV*

When the noise in the copy numbers of different proteins is uncorrelated, it is straightforward to calculate CV_μ. First, using the definition of the Growth Control Coefficients

(equation 3.1) we can simplify equation S3.1 by setting $\widetilde{X}$ equal to the mean expression levels across the population of cells, $\mathbb{E}\left[\vec{X}\right]$:

$$\mu(\vec{X}) \approx \mu(\mathbb{E}\left[\vec{X}\right]) \left(1 + \sum_i C_i^\mu \left(\frac{X_i - \mathbb{E}\left[X_i\right]}{\mathbb{E}\left[X_i\right]}\right)\right). \tag{S3.2}$$

Then, we use the basic properties of the variance, *i.e.*, for any two uncorrelated stochastic variables Y_1 and Y_2, and scalars a, b, $\mathrm{Var}\{aY_1 + b + Y_2\} = a^2 \mathrm{Var}\{Y_1\} + \mathrm{Var}\{Y_2\}$, to write:

$$\mathrm{Var}\{\mu\} \approx \mu(\mathbb{E}\left[X\right])^2 \left(\sum_i \left(C_i^\mu\right)^2 \frac{\mathrm{Var}\{X_i\}}{\mathbb{E}\left[X_i\right]^2}\right). \tag{S3.3}$$

In the regime where equation S3.1 holds, we can also read from equation S3.2 that $\mathbb{E}\left[\mu\right] \approx \mu(\mathbb{E}\left[\vec{X}\right])$. This allows us to further simplify equation S3.3:

$$\frac{\mathrm{Var}\{\mu\}}{\mathbb{E}\left[\mu\right]^2} =: \mathrm{CV}_\mu^2 \approx \sum_i \left(C_i^\mu\right)^2 \mathrm{CV}_i^2, \tag{S3.4}$$

which is equation 3.2. This is also a special case of the variance decomposition in [176].

As mentioned, all expectation values are based on population distributions (as opposed to the distribution of a cell lineage over time). This means that the mean growth rate also equals the population growth rate.

s3.1.4 *Optimisation*

Next, we optimise the *mean* expression levels of the metabolic proteins to achieve the maximal *mean* growth rate. We assume that a fixed number of proteins has to be allocated for other things than growth (the H-sector proteins), and constrain the average total protein abundance $\sum_i \mathbb{E}\left[X_i\right]$ to a fixed value Ω. The optimisation can then be written as maximising $\mathbb{E}\left[\mu\right]$ over $\mathbb{E}\left[\vec{X}_{\notin H}\right]$, with the constraint $\sum_i \mathbb{E}\left[X_i\right] = \Omega$:

$$\underset{\mathbb{E}\left[\vec{X}_{\notin H}\right]}{\mathrm{Max}} \left[\; \mathbb{E}\left[\mu\right] \;\middle|\; \sum_{i \notin H} \mathbb{E}\left[X_i\right] = \Omega - \mathbb{E}\left[X_H\right] \;\right].$$

To find the optimal genotype, we use the Lagrange Multiplier method, which in this case takes the form

$$\mathcal{L} = \mathbb{E}\left[\mu\right] - \lambda \left(\sum_{i \notin H} \mathbb{E}\left[X_i\right] - (\Omega - \mathbb{E}\left[X_H\right])\right),$$

where we need to solve $\nabla \mathcal{L} = 0$. This gives:

$$\left.\frac{\partial \mathbb{E}\left[\mu\right]}{\partial \mathbb{E}\left[X_i\right]}\right|_{\mathbb{E}\left[\vec{X}\right]^*} = \lambda, \text{ for all } i \notin H, \tag{S3.5}$$

$$\sum_{i \notin H} \mathbb{E}\left[X_i\right]^* = \Omega - \mathbb{E}\left[X_H\right]. \tag{S3.6}$$

Here the asterisks indicate the optimal value of that variable. Using equation S3.1 we can calculate $\mathbb{E}\left[\mu\right]$ explicitly:

$$\mathbb{E}\left[\mu\right] \approx \mu(\widetilde{X}) \left(1 + \sum_i \left.\left(\frac{X_i}{\mu} \frac{\partial \mu}{\partial X_i}\right)\right|_{\widetilde{X}} \frac{\mathbb{E}\left[X_i\right] - \widetilde{X_i}}{\widetilde{X_i}}\right),$$

which allows us to calculate the partial derivatives of equation S3.5:

$$\frac{\partial \mathbb{E}\left[\mu\right]}{\partial \mathbb{E}\left[X_i\right]} = \mu(\widetilde{X}) \left.\left(\frac{X_i}{\mu} \frac{\partial \mu}{\partial X_i}\right)\right|_{\widetilde{X}} / \widetilde{X_i}, \text{ for all } i \notin H.$$

Next, we make a particular choice for the point of linearisation $\widetilde{X}$ by setting it equal to optimal genotype, $\mathbb{E}\left[\vec{X}\right]^*$. In this vector, the partial derivatives equal λ (equation S3.5):

$$\lambda = \left.\frac{\partial \mathbb{E}\left[\mu\right]}{\partial \mathbb{E}\left[X_i\right]}\right|_{\mathbb{E}\left[\vec{X}\right]^*} = \frac{\mu^* C_i^{\mu^*}}{\mathbb{E}\left[X_i\right]^*}. \tag{S3.7}$$

Here, the asterisk denotes that the equation is only valid at the vector with optimal mean expression levels. (Similar arguments were exploited in [10] and [38], although without the inclusion of the H-sector and stochastic gene expression.) Lastly, we can use equation S3.6, together with the sum rule for the GCCs (equation 3.3), to calculate λ explicitly:

$$0 = \sum_i C_i^{\mu^*} = \sum_{i \notin H} \frac{\lambda}{\mu^*} \mathbb{E}\left[X_i\right]^* + \sum_{i \in H} C_i^{\mu^*}$$

$$= \frac{\lambda}{\mu^*} \left(\Omega - \mathbb{E}\left[X_H\right]\right) - \frac{\mathbb{E}\left[X_H\right]}{\Omega},$$

and thus

$$\lambda = \frac{\mu^* \mathbb{E}\left[X_H\right]}{\Omega(\Omega - \mathbb{E}\left[X_H\right])} = \frac{\phi_H}{\Omega(1 - \phi_H)}, \tag{S3.8}$$

using to the definition of $\phi_i := \mathbb{E}\left[X_i\right]/\Omega$. Finally, by combining equation S3.7 and S3.8 we find the Growth Control Coefficient for all proteins that do not belong to the H-sector in the optimal growth state:

$$C_i^{\mu^*} = \frac{\phi_H}{(1 - \phi_H)} \phi_i^*. \tag{S3.9}$$

In chapter 3 we argue that, in optimised cells, proteins with a larger mass fraction obtain a larger GCC and noise contribution. To study what this entails in the context of realistic distributions of protein abundances, we applied this theory to two experimental data sets: Taniguchi *et al* (reference [173]) and Schmidt *et al* (reference [155]). The results are shown in Fig. 3.1 and S3.1. We here provide details of these calculations.

S3.2.1   *The data set of Taniguchi et al*

The data set from Taniguchi *et al* [173] contains a library essay of protein abundances (copy numbers) and copy number variances of over 1000 protein species of *E. coli*. We estimated protein mass fractions from this data set by assuming that they are similar to the relative copy numbers in the data set (*i.e.*, ϕ_i is estimated as the measured mean copy number of protein species i, divided by the sum of the mean copy numbers of all measured protein species, thus ignoring differences in molecular mass). Moreover, we assume that the measured expression noise in the different proteins is uncorrelated, even though extrinsic noise sources, which are responsible for the noise floor, are likely to add correlated noise to abundant protein species.

After estimating mass fractions, GCCs were set as follows. First, a fraction of ϕ_H of the proteome was randomly selected and assigned to the H-sector; their GCCs were set equal to minus their mass fractions (equation 3.4). The other protein species were not part of the H-sector and thus assumed to be "metabolic" proteins. For those we compared three cases. Firstly, the proteins were assigned the GCCs calculated for optimal cells ($C_i^\mu \propto \phi_i$, equation 3.5). Secondly, we considered the case in which all GCCs are equal. Thirdly, we took the GCCs of the optimal cells, but permuted them randomly, resulting in a cell in which the distribution of the GCCs is the same as in the optimal cell, but in which there is no relation between mass fraction and growth control. Next, for all three cases the noise contributions were calculated according to equation 3.6. Results of the analysis are displayed in Fig. S3.1.

We moreover analysed the effect of ignoring the noise floor by setting protein variances to $CV^2 = F/\mathbb{E}[X]$ ($F = 2$). As expected, a noise floor causes abundant protein species to relatively contribute more strongly (compare Fig. 3.1, with a noise floor, to Fig. S3.1A, without). Indeed, the noise floor mostly affects the CV of abundant protein species. However, the noise floor also strongly decreases the (relative) noise contribution of low-copy-number proteins (Fig. S3.1B).

S3.2.2   *The data set of Schmidt et al*

The extensive data set from Schmidt *et al* [155] contains measured protein abundances (copy numbers) of over 2000 *E. coli* protein species, under various growth conditions. Since also the molecular mass of each protein species was given, we included these in our calculation of mass fractions. We further analysed abundances measured during growth on glucose, a growth medium for which the overall expression level of catabolic proteins is experimentally shown to be optimal [181]. Since variance in expression levels was not measured, we set the CV for each protein species according to equation 3.7, but added

a noise floor to each protein species to mimic observed variances in Taniguchi *et al*, i.e., $CV_i^2 = F/\mathbb{E}[X_i] + n_f$, with $F = 2, n_f = 0.15$. Lastly, the same calculations were performed as described above for the data set of Taniguchi *et al*. The results are displayed in Fig. S3.1C.

Note that due to the many assumptions, the noise contributions are only rough estimates, and should not be considered quantitative predictions.

S3.3 DETAILS OF THE STOCHASTIC SIMULATION

S3.3.1 *Kinotype, genotype, phenotype and growth rate*

In our toy model, we simulated a linear chain of 5 proteins, where the flux through each of the first 4 steps is given by:

$$v_i = \frac{k_{cat,i} \, X_i \, m_i}{k_m + m_i + m_{i+1}/k_{i+1}}. \tag{S3.10}$$

Here, k_{cat} is the reaction rate, k_m is the Michaelis-Menten constant, and k is the inhibition constant, together called a kinotype. To define a kinotype, we sample k_{cat} and k_m uniformly from the interval $[0.1, 6.1]$ and k from $[1, 7]$. The external metabolite, m_1 is set to 10 and kept constant, simulating a fixed environment. The fifth protein in the reaction chain creates biomass and is not inhibited.

In a snapshot of a population of cells, the values of $\vec{X}$ are distributed according to some probability distribution $P(\vec{X})$, here assumed to be six independent Gamma distributions. Because the variance in the expression of each protein species is tied to its mean (equation 3.7), the Gamma distribution for each protein species is completely specified by its mean copy number. Together, the mean copy numbers are called a genotype and fully define the probability distribution $P(\vec{X})$. According to $P(\vec{X})$, we can sample for each individual in a population a vector $\vec{X}$. This vector is referred to as the individual's phenotype. Corresponding with the phenotype, an individual's growth rate can be calculated by integrating the system of ODEs for all metabolites (for $i \in \{2, 3, 4, 5\}$, $\dot{m}_i = v_{i-1} - v_i = 0$ with each v_i specified in equation S3.10) until steady state (10^6 time steps in Matlab 2019b). The dynamics of the metabolites are assumed fast, such that they are always in steady state relative to the sampled protein copy numbers. Moreover, stochasticity in the metabolic reactions is ignored for the same reason: metabolic fluctuations relax on timescales much faster than growth (seconds rather than minutes or hours).

The specific kinetics chosen (equation S3.10) then enforce that for all possible phenotypes a steady state flux exists. In the steady state, where all fluxes are equal, we define $v_i = v_{i-1} = J$, and:

$$\mu := J / \left(\sum_i X_i \right).$$

S3.3.2 *Evolution*

To simulate evolution, we start with an initial genotype and measure its mean growth rate (by sampling many phenotypes and calculating the corresponding growth rates), and subsequently mutate the genotype to search for genotypes that result in higher mean growth rates.

The initial genotype is constructed by sampling 5 uniform numbers, $\vec{x} \in [0,1]^5$ and setting $\mathbb{E}\left[\vec{X}\right]_{\text{initial}} = \vec{x}\phi_H / \sum_j x_j \Omega$, with $\phi_H = 0.4$ (the proteomic fraction allocated to the H-sector) and $\Omega = 10^4$, the (mean) total protein abundance, here assumed to be equal to cell size.

Genotypes were mutated in two different ways, depending on the noise amplitude chosen (value of F) in the simulation. In the low noise regime (F = 1), the next genotype was determined with a gradient-based hill climbing algorithm that uses the current GCCs:

$$\vec{X}_{t+1} = \vec{X}_t + \epsilon \, \delta\vec{X}_t / \left|\delta\vec{X}_t\right|^2 , \; \delta\vec{X}_t = \frac{\Omega}{5}\left(C^{\vec{\mu}}_t / \vec{\phi}_t - \sum_{j \notin H} C^{\mu}_{j,t} / \phi_{j,t}\right),$$

where ϵ is small (0.0002). This algorithm changes the genotype in the direction of the steepest growth rate increase.

In the high-noise regime (F = 10), changes in the genotype are due to random mutations, where one of the metabolic proteins is chosen at random and its mean abundance is changed according to a percentage drawn from a normal distribution (mean zero, variance 5%), after which the entire genotype is renormalised to enforce a fixed mean cell size of Ω. A mutant genotype is accepted only if it yields a higher mean growth rate (calculated over $2 \cdot 10^4$ sampled phenotypes). The evolutionary process is terminated when 100 mutants have been rejected. The full evolutionary process is repeated 15 times; from these 15 evolutionary trajectories, the genotype with the highest mean growth rate is chosen to be the HN genotype. (The number 15 is arbitrary, but deemed enough to ensure the simulation did not get stuck in a local optimum, while still being computationally feasible.)

s3.3.3   *Calculating noise contributions*

For a specific genotype, we measured noise contributions as follows. First, we sampled $2 \cdot 10^4$ phenotypes and calculated the corresponding growth rates. Then, we divided the distribution for each protein in 100 bins, and calculated the mean growth rate in each bin (sampling extra if less than 100 phenotypes fell in a particular bin). Afterwards, we calculated the weighted variance between these growth rates. This is an approximation of a conceptual decomposition method from Bowsher and Swain [18]:

$$\kappa_i = \frac{\text{Var}\left\{\mathbb{E}\left[\mu | X_i\right]\right\}}{\text{Var}\left\{\mu\right\}}. \tag{S3.11}$$

This method is a first order approximation of a full Global Sensitivity Analysis [107]. The approximation is only valid if the sum of all contributions is close to unity. This is indeed the case (see Fig. S3.5A). Moreover, sampling errors in κ are small (Fig. S3.5B). All Matlab and Mathematica codes are available upon request.
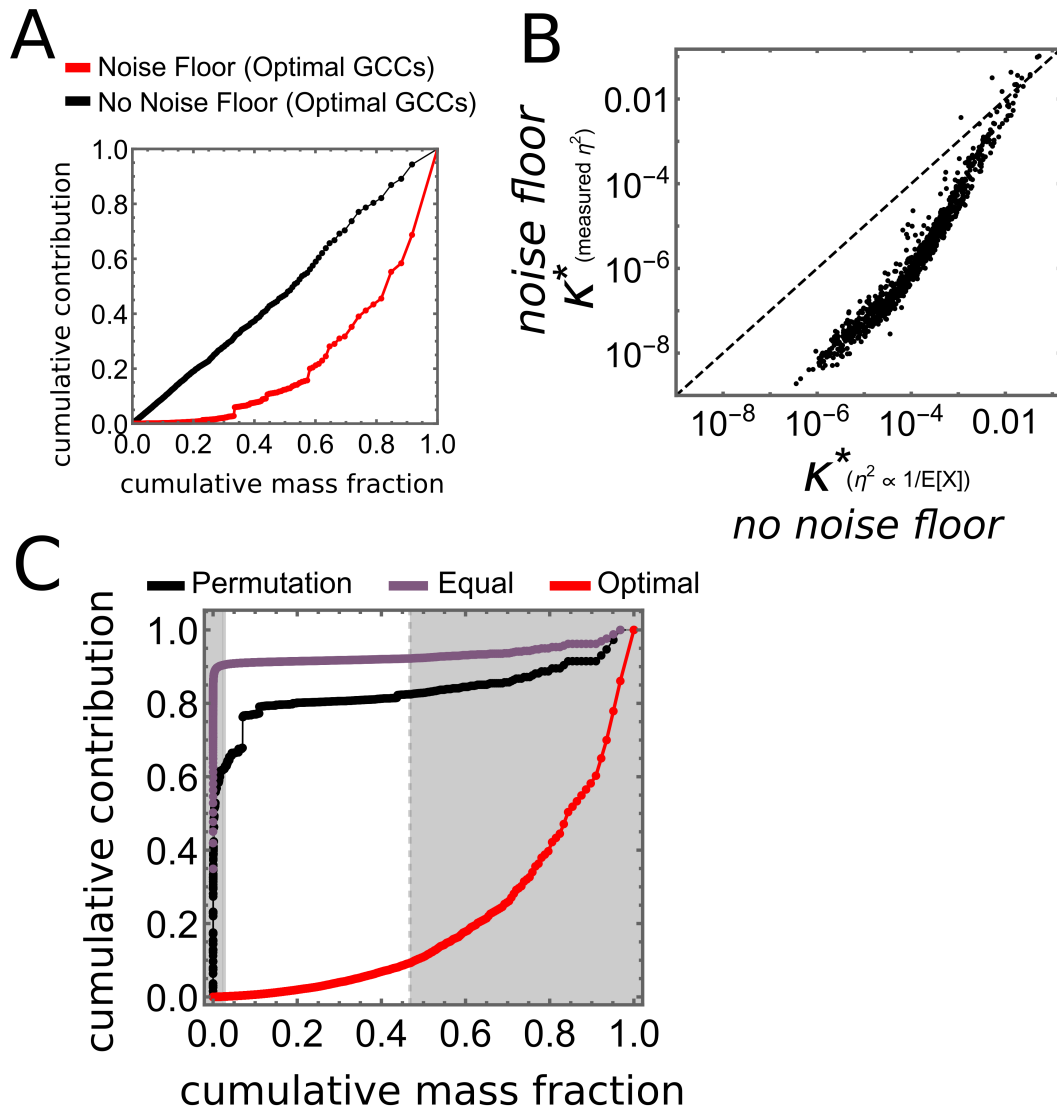
s3.4   SUPPLEMENTARY FIGURES

Figure S3.1: Examination of the effect of the noise floor and application of the theory of noise contribution to the data set from Schmidt *et al*. **(A)** Data from Taniguchi *et al*, but variances were now assumed to solely scale with mean protein abundance ($CV^2 = 1/\mathbb{E}[X]$), ignoring the noise floor ($n_f = 0$, see section S3.2, black dots in the figure). **(B)** Effect of adding the noise floor on predicted noise contributions in an optimally growing *E. coli* cell. Adding a noise floor (in this case using measured variances) increases the noise contribution of a few very abundant protein species (their noise levels increase), but also causes many low copy number protein species to contribute relatively less. **(C)** Cumulative noise contributions ($\sum \kappa$) against cumulative mass fractions ($\sum \phi$) estimated from the Schmidt *et al* data set. GCCs were again set according to optimal growth (red dots), equal (purple dots), or shuffled (black dots). Shaded areas indicate the 50% least abundant protein species (left) and the top 5% most abundant species (right).
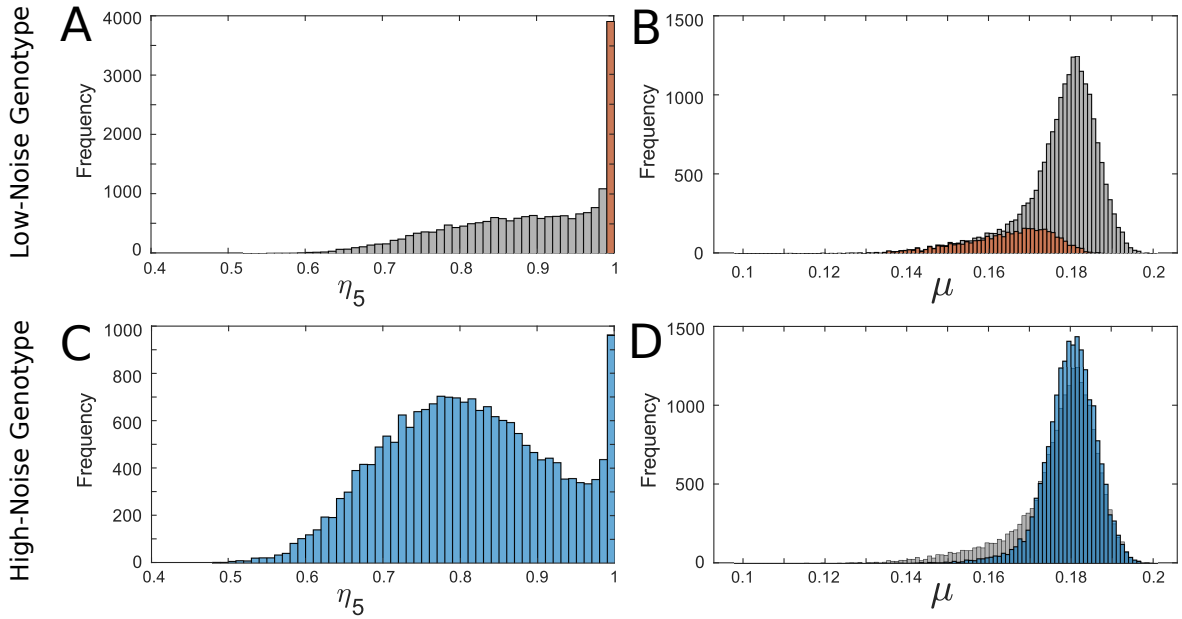
Figure S3.2: Example of highly skewed distributions in an optimised kinotype. (We picked this kinotype because it most clearly showed the effect of evolution on the distributions.) **(A)** Distribution of the efficiency of the fifth protein ($\eta_i := \frac{k_{cat,i}}{\mu} \frac{X_i}{\sum_j X_j}$). **(B)** Distribution of growth rates for the optimal Low-Noise genotype in the high noise regime. Red areas in (A) and (B) correspond to the same phenotypes. **(C)** The same as (A), but for the evolved High-Noise genotype. **(D)** the same as (B), but for the evolved, High-Noise genotype. Grey distribution is the distribution in the LN-genotype for comparison.
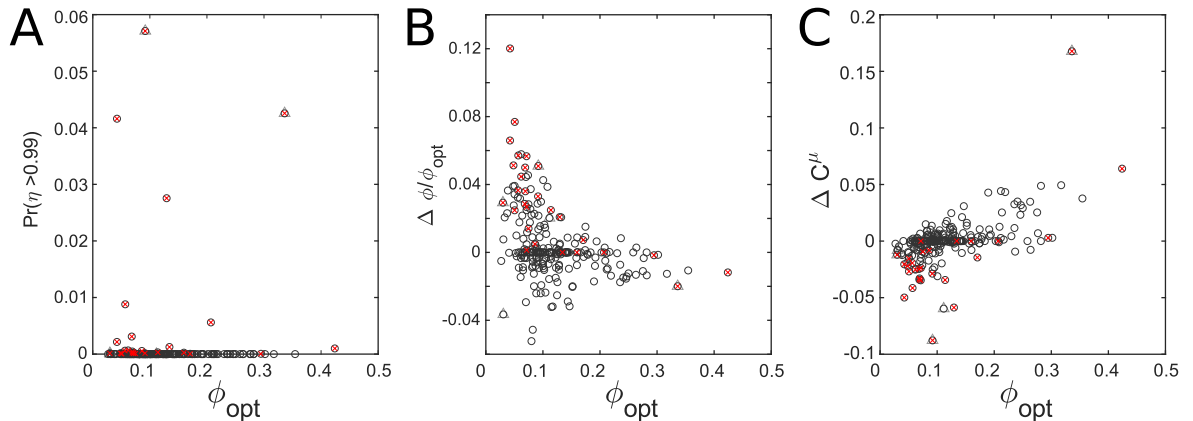

Figure S3.3: **(A)** Probability a protein's efficiency is very close to unity, calculated over $2 \cdot 10^4$ phenotypes. Although a higher efficiency on first glance seems good, an efficiency close to unity indicates that this protein might be limiting growth. All proteins for with $p(\eta > 0.99) > 0$ are marked with a red cross. Two outliers (triangle) protein species from the same kinotype. **(B)** Changes in genotype when noise levels increase, as a function of the mean ϕ in the optimal genotype. Red crosses are those proteins that had a non-zero probability of getting an efficiency close to one in the optimal genotype with high noise. Note that the highly expressed proteins with $p(\eta > 0.99) > 0$ are not increased, probably because this required the allocation of too much additional resource. **(C)** Changes in mean genotype coincide with a change in C^μ.
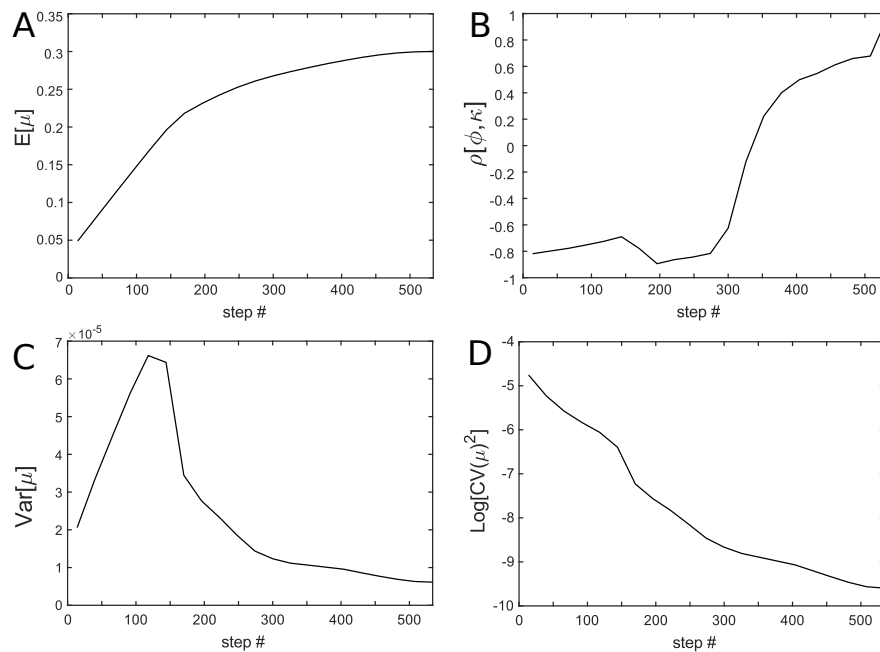
Figure S3.4: Example of the optimisation algorithm (gradient-based hill climb algorithm). **(A)** Growth rate increases each step. **(B)** Correlation coefficient increases and switches sign during optimisations. **(C)** Variance decreases during most parts of the optimisation process. **(D)** CV_μ^2 decreases. Parameters of this example kinotype are equal to those in Fig. 3.1B.
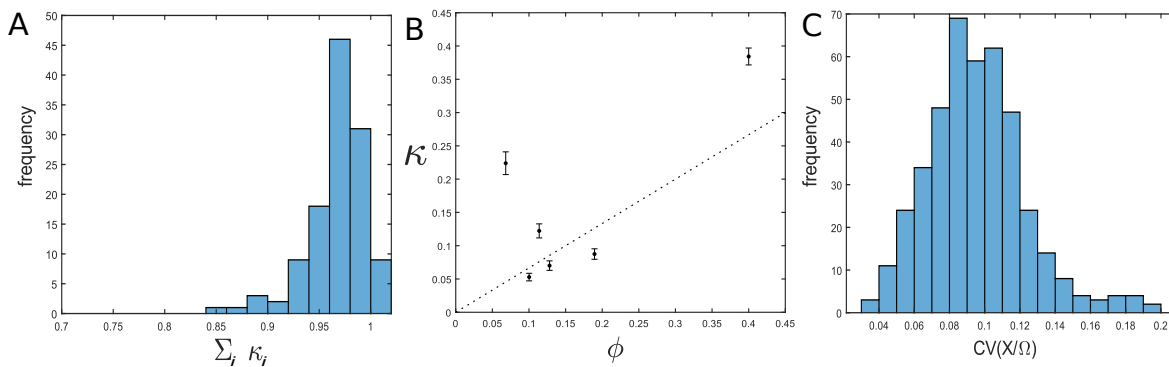

Figure S3.5: Examination of the noise decomposition method and noise levels. **(A)** The sum of all first order noise contributions is close to unity, indicating that a first order Global Sensitivity Analysis captures the variance contributions well. **(B)** Noise contributions for the example kinotype in the LN genotype in the high noise regime. Error bars indicate 2 sd over 125 repeated sampling of $2 \cdot 10^4$ phenotypes. Sampling errors in κ are within reasonable bounds. **(C)** Distribution of CVs as encountered in HN genotypes.
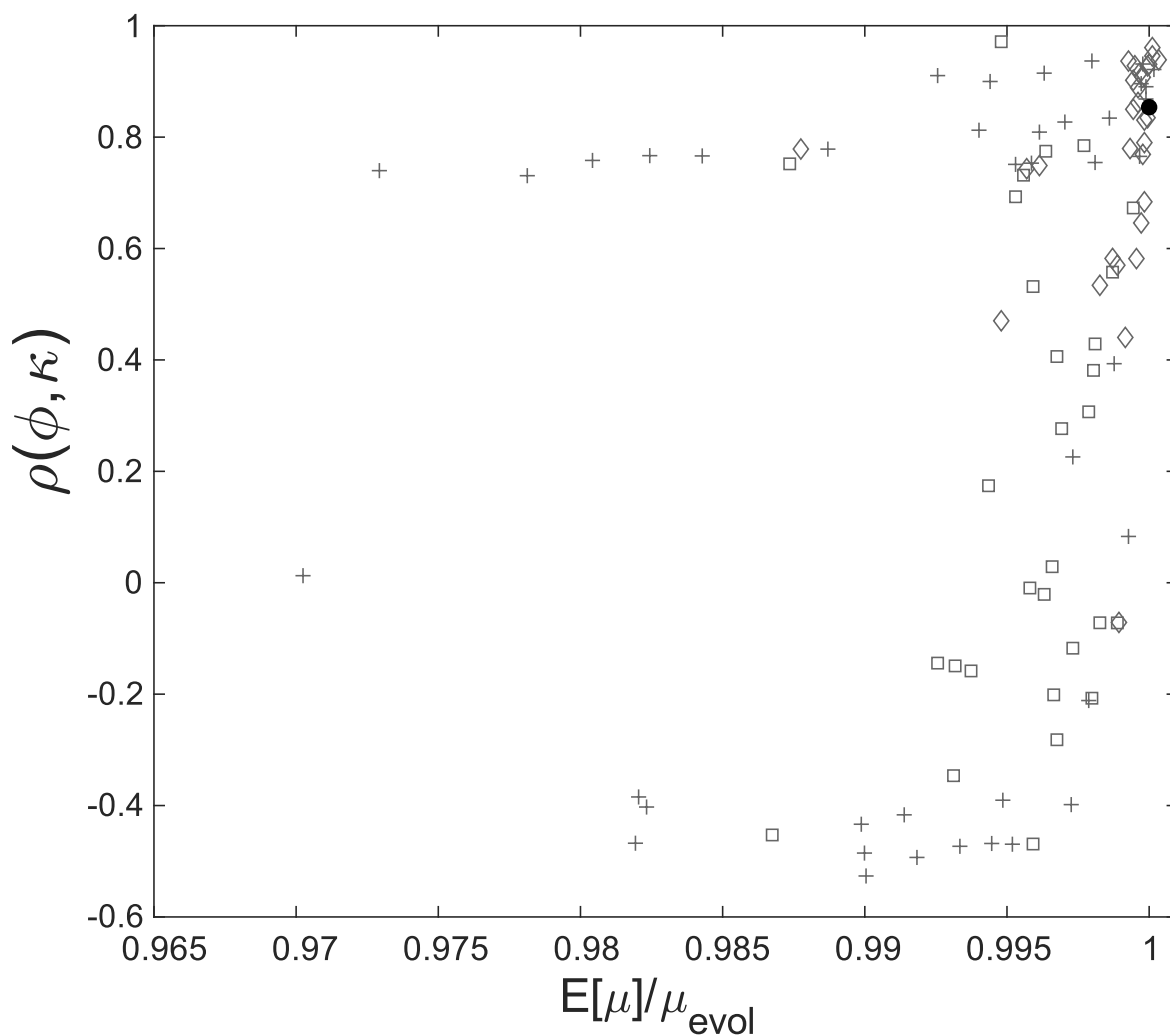
Figure S3.6: Examination of the sensitivity of the positive scaling between φ and κ for kinotype 2 around the evolved genotype (black circle) in the high noise regime (α = 10). Grey dots denote mutant genotypes, for which their mean growth rate (relative to the mean growth rate of the evolved genotype) is plotted against the correlation coefficients of φ and κ for that particular genotype.

Mutants are created by three different methods: (1) mutation in a single protein with a normally distributed step size with standard deviation 15% (40 genotypes, plus-sign markers), (2) with 5% (30 genotypes, diamond-sign markers), and (3) mutations in all protein species with standard deviation 5% (30 genotypes, square-sign markers). After the mutations, genotypes were re-normalized to make sure total cell size, and ϕ_H remained constant.

Table S3.1: Kinetic parameters, k_{cat} and inhibition constant (k_{inhi}) for the five metabolic protein species in the kinotype used to create figure Figs. 3.2 and S3.4. Additionally, the initially sampled (relative) protein abundances are given.

Protein	k_{cat}	k_{inhi}	$\phi_{initial}$
1	3.4048	0.2153	0.19177
2	4.3489	4.3831	0.0080544
3	1.8454	3.6435	0.18137
4	3.1650	6.5335	0.084103
5	5.4577	1.8425	0.1347

DYNAMICAL REGULATION IN BACTERIAL CELLS DURING STEADY STATE GROWTH

4

Martijn Wehrens*, Laurens H.J.Krah*, Benjamin D. Towbin, Rutger Hermsen, Sander J. Tans

*These authors contributed equally

ABSTRACT

Bacteria sense changes in their external environment and regulate genes accordingly. But even in a constant environment the concentrations of proteins and environmental sensors such as metabolites, fluctuate, because gene expression is intrinsically stochastic. It is known that stochastic fluctuations can transmit between nodes in a (regulatory) network. However, to what extent regulatory networks respond to and in turn shape these fluctuations is unknown. To address this question, we studied the cAMP-CRP regulatory network during steady state growth. We studied the interplay between gene expression and growth in wild type *Escherichia coli* and compared it to a mutant whose cAMP concentration is experimentally held fixed and therefore does not react to intracellular fluctuations. Specifically, we simultaneously measured the singe-cell growth rate and the expression of two fluorescent reporters: one under cAMP-CRP control and one constitutive. These observations show that cAMP-CRP regulation affects the dynamics of gene expression and growth even at sub-cell-cycle timescales and in a fixed environment. The measured (cross-)correlations between reporter concentrations and the growth rate were analysed with a mathematical model of stochastic gene expression, metabolism and growth. The model explains the main differences between the wild type and the non-regulating mutant, and between the two reporters, by changing only a single parameter that represents cAMP-CRP regulation. To benchmark the model, we designed additional experiments, the results of which were qualitatively predicted by adjusting specific model parameters.

The contents of this chapter will be submitted to Nature Communications.

4.1 INTRODUCTION

Bacteria display striking regulatory abilities to adjust to different environments. When supplied with different carbon sources, bacteria make vast changes to their proteome composition to adjust metabolic resource allocation and growth rate [68, 89, 155, 158]. Often, this regulation is tuned such that the amount of catabolic enzymes expressed is neither wasteful nor limiting [17, 128, 157, 181]. Such an apparently fine-tuned relation between gene expression and the population growth rate stands in sharp contrast to the heterogeneity observed at the single cell level. Under constant conditions, isogenic populations of bacteria show large heterogeneity in protein expression and even growth rate [41, 42, 144, 173]. Between one individual and the next, growth rates can readily differ by more than a factor of two [86, 170]. With these observations in mind, we speculated that the regulatory mechanisms that adjust metabolism upon changes in external conditions might also respond to internal fluctuations that arise stochastically.

The origins of noise in gene expression -and its interaction with regulatory networks-have been investigated extensively. Heterogeneity can originate from the inherent chemical stochasticity of gene transcription [52] and bursty translation [25, 47]. Expression noise can subsequently propagate through gene regulatory networks [40, 64, 139] and affect the expression of downstream genes. In turn, specific network motifs influence noise amplitudes and timescales [39, 148]. Correlations between fluctuations of constitutively expressed genes have been observed, which is classically defined as extrinsic noise [42, 168]. Suggested contributions to extrinsic noise include fluctuations in the abundances of components involved in central cellular functions, such as ribosomes and transcription factors [42] and processes such as the cell cycle [188] and metabolism [86].

It was found that noise transfers not only through gene regulatory networks but also through the metabolic network [86]. Temporal fluctuations in limiting metabolic enzymes were shown to produce correlated fluctuations in the growth rate. Metabolic fluctuations may, however, propagate even more widely since metabolite pools and fluxes influence the expression of many other genes through regulatory interactions. This way, metabolic fluctuations might confer stochastic fluctuations to regulatory proteins, which in turn contribute to extrinsic and global cellular noise dynamics. In this paper, we set out to investigate whether the interaction between metabolism and regulation contributes to the global stochastic dynamics of the cell.

To address this question experimentally, we focused on the transcriptional regulation by the cAMP receptor protein (CRP) in *Escherichia coli* (*E. coli*). CRP regulates over 180 genes, and is thought to be a main regulator of genes related to catabolism (the C-sector) [49, 95, 155, 198]. CRP is activated by the second messenger cyclic AMP (cAMP). In turn, the synthesis of cAMP is inhibited by catabolic metabolites, such that C-sector expression decreases when the concentration of intermediate metabolites increases, for example due to changes in the external nutrients. The result is a regulatory feedback loop. At the population level, this feedback underlies a negative relationship between the proteome mass fraction involved in catabolism (ϕ_C) and the growth rate (λ) under variation of the carbon source; called the 'C-line' [68, 198]. Moreover, for many carbon sources the cAMP-CRP regulatory feedback is shown to result in optimal C-sector expression [181].

To investigate the dynamics of stochastic fluctuations, we used time lapse microscopy to obtain time series data of C-sector expression and instantaneous growth rate at the single

cell level. Central to our work is the comparison of wild-type *E. coli* cells to mutated *E. coli* cells that show the same population-averaged growth rate and C-sector expression but whose cAMP-CRP regulatory network cannot respond to internal metabolite fluctuations. Disrupting the propagation of internal fluctuations through a regulatory connection while maintaining a proper response to the fixed external conditions is non-trivial. We achieved this by using an *E. coli cyaA cpdA* null mutant [181]. This strain is unable to degrade or synthesise cAMP, so that the cAMP-CRP regulatory feedback is broken. Importantly, cAMP was provided externally at a constant concentration, resulting in what we call cAMP-fixed cells (Fig. 4.1A). Although these cells can no longer respond to internal metabolite concentrations, we can artificially set C-sector expression to match external nutrient conditions by providing a well-chosen cAMP concentration (cAMP-fixed*). Comparison of cAMP-fixed* cells with their wild type counterparts allowed us to study the interaction between stochastic metabolite fluctuations, the cAMP-CRP regulatory network, and the growth rate.

We found that the relationship between fluctuations in ϕ_C and λ differs between cAMP-fixed* cells and the wild type. This indicated that fluctuations in gene expression and growth rate are influenced by the stochastic, metabolic input received by regulatory networks, even in a fixed environment.

To better understand the observed dynamics, we created a coarse-grained, mathematical model that captures the stochastic dynamics of gene expression, metabolism and growth. This model was able to reproduce temporal relationships between gene expression and growth from a mechanistic perspective and could explain the main difference between the wild type and the cAMP-fixed* cells as a change in a single parameter related to cAMP-CRP regulation. We subsequently validated this model against a second set of experiments; model predictions qualitatively matched experimentally observed dynamics. Together, the experiments and the model show that *E. coli* reacts to internal stochastic metabolic fluctuations in a fixed environment as if these fluctuations were caused by a changing environment.

4.2 RESULTS

4.2.1 *Establishing cAMP-fixed* cells*

To obtain cAMP-fixed* *E. coli* cells, we first determined the cAMP concentration required to match WT population-averaged C-sector expression and growth rate. We measured growth rate at bulk level for various cAMP concentrations (Fig. S4.3). This showed that cAMP-fixed cells supplied with 800 µM cAMP displayed WT growth rates (Figs. S4.3 and S4.5A). We later confirmed in single cell experiments that at this cAMP concentration C-sector expression also matches that of the WT (Fig. S4.5C). We thus refer to cAMP-fixed cells supplied with 800 µM cAMP as cAMP-fixed* cells.

4.2.2 *Probing single-cell stochastic dynamics under constant conditions*

To probe stochastic fluctuations at sub-cell-cycle resolution, we used time-lapse microscopy at high image acquisition rate to measure the instantaneous growth rate λ and C-sector expression of individual cells growing on minimal medium within micro-colonies. To determine C-sector expression we introduced two previously designed promoters [181] fused to
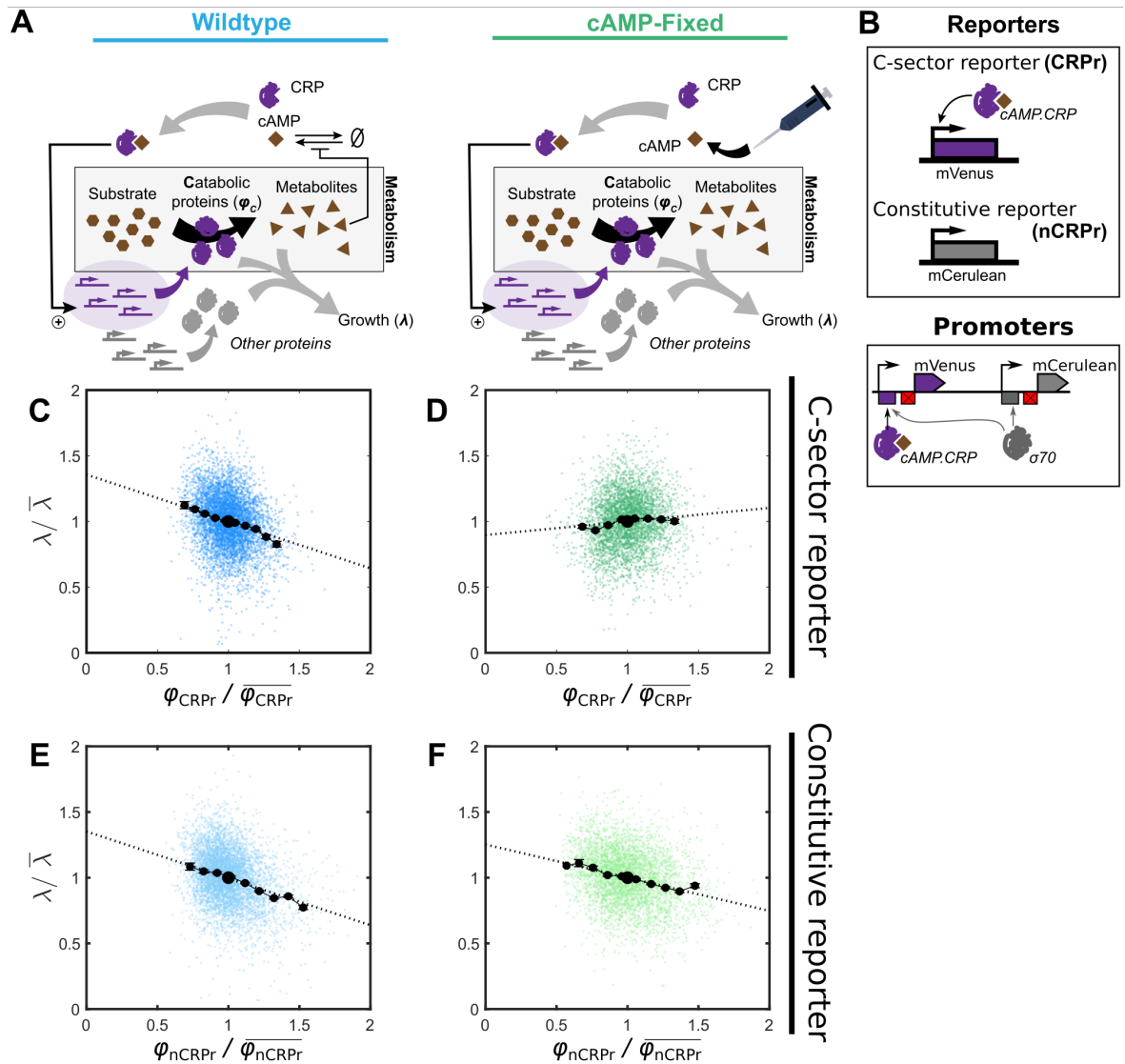
Figure 4.1: **Removing cAMP feedback alters dynamics of a regulated reporter only.** (**A**) Cartoon of a bacterial cell and the difference between wild type and mutant. Shown are the processes of metabolism, protein expression, cAMP-CRP regulation, and growth (λ). Here ϕ_C represents the expression level of the C-sector; the total concentration of all catabolic proteins that are regulated by cAMP-CRP and that import nutrients and convert them into internal metabolites (including cAMP itself). In the cAMP-fixed strain, cAMP is neither synthesised nor degraded, and instead supplied externally to experimentally tune ϕ_C. (**B**) The two reporters and their promoters that were used in this study: a C-sector reporter whose transcription was regulated by cAMP-CRP (CRPr), and a constitutive reporter, nCRPr. Crossed red block is the scrambled *lacI* site. (**C-F**) Scatter plot of instantaneous growth rate (λ) against single-cell relative expression of the reporters (ϕ_{CRPr} or ϕ_{nCRPr}). Dashed lines are linear regressions, black dots indicate binned averages. Data points are from a single growing colony (other colonies showed the same trends, Fig. S4.5).

fluorescent reporter genes into the bacterial genome of both WT and cAMP-fixed cells. In previous studies, *lac* expression has been used as a readout for the C-sector [198]. Here we use a *lac* promoter with a scrambled *lacI* site, such that it is always sensitive to cAMP-CRP, and fused it to mVenus (here called the CRP-regulated reporter: CRPr). Additionally, to control for changes in dynamics between WT and cAMP-fixed cells not related to CRP, we fused a constitutive promoter to mCerulean (the non-CRP regulated reporter: nCRPr). To create this constitutive promoter, we further modified the CRPr promoter by replacing the CRP-binding site by a σ^{70} consensus site, so that translation initiation occurs constitutively instead of requiring CRP to recruit σ^{70}. See also Figs. 4.1B and S4.11, and Table S4.2.

4.2.3   *Behaviour of cAMP-fixed* cells suggests CRP responds to stochastic metabolite fluctuations*

Removal of the cAMP-CRP metabolic feedback had a clear influence on the joint fluctuations of C-sector expression and the growth rate in single cells (Fig. 4.1C-D). In WT cells, fluctuations in the CRPr reporter concentration (ϕ_{CRPr}) are negatively correlated with fluctuations in λ, as apparent from the negative regression slope in Fig. 4.1C. In cAMP-fixed* cells, a positive correlation is found instead (Figs. 4.1D and S4.5A). The difference is statistically significant (p = 0.0031, two sample t-test, S4.5). The negative $\phi_{CRPr} - \lambda$ slope observed in WT cells might reflect the population-level negative relationship between C-sector expression and growth rate due to cAMP-CRP regulation under carbon source variation (the aforementioned C-line). In line with this hypothesis, the negative slope seems to be removed when the cAMP-CRP feedback is removed (Fig. 4.1C to D). To verify that changes in the slope between ϕ_{CRPr} and λ were due to the disruption of CRP regulation specifically, rather than due to global changes of the dynamics in cAMP-fixed cells, we looked at the relationship between the concentration of the constitutive reporter (ϕ_{nCRPr}) and λ. In WT and cAMP-fixed* cells, the $\phi_{nCRPr} - \lambda$ relationship was similar, and negative (Figs. 4.1E-F and S4.5A). This similarity (p = 0.93, Welch's t-test) was expected, because for the nCRPr reporter there is no difference between WT and cAMP-fixed* cells: the nCRPr reporter is not regulated by cAMP-CRP in either strain. In conclusion, because the stochastic dynamics of the CRPr reporter, but not the nCRPr reporter, differ between WT and cAMP-fixed* cells, it seems that the cAMP-CRP regulatory network reacts to internal fluctuations, even in a fixed environment.

4.2.4   *A mathematical model can be used to mechanistically understand temporal dynamics*

Though these observations on WT and cAMP-fixed* cells are indicative of interesting underlying CRP dynamics, a mechanistic understanding of the $\phi_{CRPr} - \lambda$ relationships was lacking. Why do WT cells show a negative $\phi_{CRPr} - \lambda$ relation and cAMP-fixed* cells a positive $\phi_{CRPr} - \lambda$ relation? And what is the origin of the negative relations of the nCRPr reporter? The dynamics of these reporters could be shaped by multiple cellular processes [16, 40, 86, 87], such that multiple (extrinsic) noise sources contribute to each slope while only some of these sources might be affected by removing the cell's ability to respond to internal metabolite concentrations. Therefore, we next sought to dissect the multiple cellular processes that shape the dynamics of the two reporters and the growth rate in both strains. To this end, we compared temporal data on gene expression and growth with a stochastic mathematical model.

We extended the stochastic cell model presented by Kiviet *et al* [86]. This model does not attempt to capture every biological detail, but instead describes the dynamics of coarse-grained biological variables phenomenologically. Our extended model uses linear stochastic differential equations (SDEs) to describe the temporal dynamics of protein production rates (π), protein concentrations (ϕ), metabolism (M), and growth rate (λ). We explicitly modelled π and ϕ of the C-sector (π_C and ϕ_C), the C-sector reporter (π_{CRPr} and ϕ_{CRPr}), and the constitutive reporter (π_{nCRPr} and ϕ_{nCRPr}). Concentrations ϕ depend on the growth rate λ (since volume growth involves the dilution of cellular components) and production π. To simulate the stochasticity of bacteria cells, we include independent phenomenological noise sources (Ornstein-Uhlenbeck processes) directly affecting λ, π, and M, and coefficients that describe transfer of noise between variables. We model cAMP-CRP regulation specifically as a negative feedback from metabolism towards π_C and π_{CRPr} (see Fig. 4.2A and SI section S4.1.1 for a full description). Transfer coefficients, time scales, and noise amplitudes can be fit to experimental data, allowing us to mechanistically describe the dynamics of cAMP-CRP regulation, metabolism and growth.

4.2.5 *Cross-correlations reveal temporal dynamics of gene expression and growth*

To obtain a more complete understanding of the temporal dynamics of π, ϕ and λ, and to compare our experiments to our mathematical model, we calculated cross-correlation functions. Cross-correlation functions quantify the correlation between two time series after one is shifted with respect to the other by a delay time τ. This can help characterise temporal dynamics and give insights into how noise is transmitted between biological quantities [40, 86, 88, 176]. For example, if noise that originates in signal A has a downstream and delayed effect on signal B, the cross-correlation between A and B will peak at a positive delay time. The cross-correlation at zero delay is the instantaneous correlation between A and B, which can for instance arise when both are affected by a signal C.

We re-analysed our time-series data of WT and cAMP-fixed* cells (Fig. 4.1C-F) to calculate cross-correlations between reporter concentrations (ϕ_{CRPr} and ϕ_{nCRPr}) and λ and between reporter production rates (π_{CRPr} and π_{nCRPr}) and λ (Fig. 4.2B, material and methods). We observed that the WT $\phi_{CRPr} - \lambda$ correlation ($R_{(\phi_{CRPr}-\lambda)}$) is negative at $\tau = 0$ (Fig. 4.2B), reflecting the negative slope between ϕ_{CRPr} and λ also observed in Fig. 4.1C. For cAMP-fixed* cells, $R_{(\phi_{CRPr}-\lambda)}$ was positive at $\tau = 0$ (Fig. 4.2B), similar to the positive slope observed in Fig. 4.1D. In addition, more features of the stochastic dynamics became clear. For example, both $R_{(\phi_{CRPr}-\lambda)}$ and $R_{(\phi_{nCRPr}-\lambda)}$ had a negative peak at negative delay ($\tau < 0$) in WT cells. And, not only $R_{(\phi_{CRPr}-\lambda)}$ but also $R_{(\pi_{CRPr}-\lambda)}$ was significantly higher in cAMP-fixed* cells compared to WT cells (Fig. 4.2B, top-right). Lastly, the constitutive reporter nCRPr again showed similar behaviour between WT and cAMP-fixed* cells (Fig. 4.2B).

4.2.6 *Mathematical model explains dynamics and confirms role of cAMP-CRP regulation*

To understand these cross-correlation functions further, we used our mathematical model to predict expressions for them. These expressions contained free parameters for noise transfer and noise amplitudes, which allowed us to fit the model's analytical cross-correlations to our experimental data.
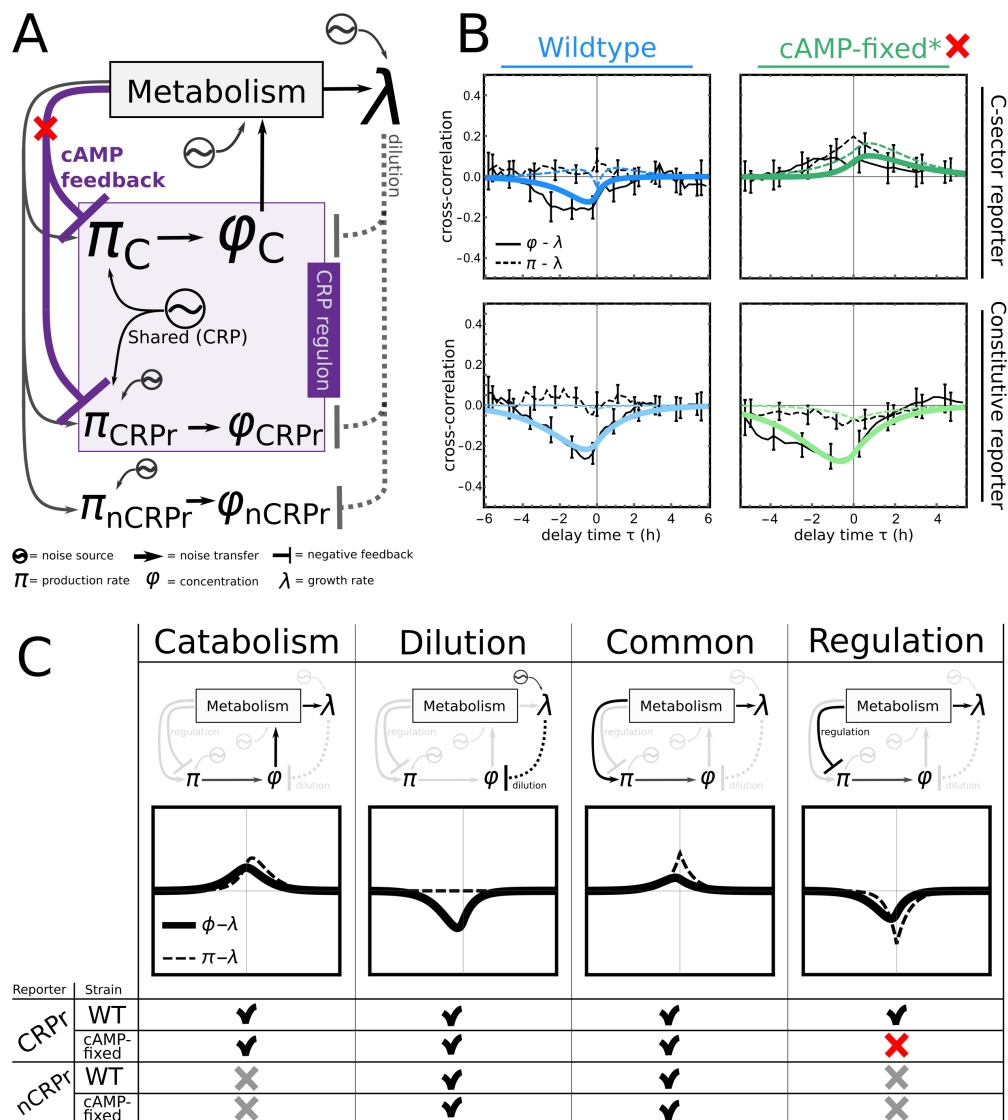
Figure 4.2: **Mathematical model pinpoints dynamical role of regulation.** (**A**) Cartoon of the mathematical model, which considers fluctuations in the growth rate (λ) and the production rates (π) and concentrations (ϕ) of the C-sector (π_C and ϕ_C), the C-sector reporter CRPr, and the constitutive reporter nCRPr. Black arrows indicate noise transfer; only fluctuations in ϕ_C affect metabolism. Metabolism affects growth and protein productions rates. Regulation reacts to metabolic fluctuations and transfers to π_C and π_{CRPr}. In the mutant, regulation is removed (red cross). (**B**) Cross-correlation functions between the protein production rate $\pi(t)$ and $\lambda(t)$ (dashed lines) and between concentrations $\phi(t)$ and $\lambda(t)$, in the wild type and the mutant. Coloured lines are model fits, black lines are cross-correlations calculated from data and error bars indicate standard error (SI section S4.4), shown for only some data points. (**C**) Interpretation, shape, and absence (crosses)/presence (checks) pattern of the four underlying noise modes that together shape the cross-correlations in (B). Only the regulation mode is completely removed due to the mutation (red cross). Cartoons indicate the direction and route of noise transfer for each specific mode.

We fitted analytical cross-correlations to data from both reporters, both for the WT and cAMP-fixed* cells (Fig. 4.2B). To confirm that the different dynamics in the WT and cAMP-fixed* cells could be explained by the removal of cAMP feedback, we allowed only the transfer parameter T_R associated with cAMP-CRP regulation to differ between the wild type and the cAMP-fixed* cells: it was constrained to negative values in WT to represent negative feedback, but set to zero in cAMP-fixed* cells to represent the lack of feedback; see crossed-out feedback arrow in Fig. 4.2A. All other parameters, including noise transfer parameters, noise amplitudes and timescales, were kept fixed (for a more detailed explanation of the model fit and fitted parameters see SI, section S4.5). Despite these strict constraints, the model quantitatively fits the data for both reporters and both strains; in particular, it reproduces the drastic change between WT and cAMP-fixed* cells for the CRPr reporter and the absence of change between WT and cAMP-fixed* in the nCRPr reporter (Fig. 4.2B). In conclusion, the model confirms that changes in the temporal dynamics of gene expression and growth, as visualised by their cross-correlations, can be explained by changes in cAMP-CRP regulation.

4.2.7   *Model cross-correlations are composed of terms reflecting mechanistic phenomena*

Next, we aimed to further understand how the shape of each cross-correlation is determined by underlying noise propagation mechanisms. We found that the cross-correlation functions of the model can be written as a sum of four noise modes, which we termed the catabolism, dilution, common, and regulation mode (Fig. 4.2C; see SI section S4.2 for mathematical decomposition). Each of the four modes has a different shape and amplitude, and hence together they determine the shape of the cross-correlations. Therewith, the noise modes link the abstract shapes of the cross-correlations to biologically interpretable phenomena.

First, the *catabolism* mode originates from fluctuations the production rate of catabolic proteins that are to some extent rate limiting. Such fluctuations will result in a higher concentration, and likewise a higher growth rate. Therefore, this mode contributes a positive peak at positive delay time to the $\pi - \lambda$ cross-correlation, and a symmetrical peak to the $\phi - \lambda$ cross-correlation. Second, the *dilution* mode results from fluctuations in the growth rate that dilute all protein concentrations, resulting in a negative correlation with negative delay. Third, the *common* mode is the result of fluctuations that affect all protein production rates simultaneously, as well as the growth rate. In result, this mode is symmetrical for the $\pi - \lambda$ cross-correlation, and has a negative delay for the $\phi - \lambda$ cross-correlation. Lastly, the *regulation* mode represents all noise that transfers via the cAMP-CRP regulatory network, pinpointing the exact contribution of the cAMP-CRP regulatory network to the cross-correlations. In cells with faster metabolism and growth, cAMP-CRP regulation suppresses the production of C-sector proteins, resulting in a symmetrical but negative $\pi - \lambda$ cross-correlation and, with a delay, also a negative $\phi - \lambda$ cross-correlation.

Importantly, due to the biological wiring of the system not all modes are present in all cross-correlations. The absence and presence of the noise modes can already help to qualitatively understand the shape of the cross-correlation functions for a specific reporter and strain (Fig. 4.2C, lower table). First, the CRPr reporter cross-correlation contains a catabolism mode, whereas the nCRPr reporter does not. Note that, although neither reporter directly influences the growth rate, the C-sector reporter CRPr can be seen as a

proxy for expression of the C-sector, which does influence the growth rate. Therefore, a part of the catabolism mode of the C-sector can be observed in the CRPr cross-correlations, but not in the cross-correlations for the constitutive reporter. Second, the regulation mode is only present in the cross-correlations for the C-sector reporter CRPr in the wild type, because only this reporter is regulated via the cAMP-CRP regulatory network.

We noted that the (cross-)correlations between ϕ_{CRPr} and λ (Fig. 4.1C and Fig. 4.2B) and between ϕ_{nCRPr} and λ (Figs. 4.1E and 4.2B) looked similar in WT cells. The model fit, however, indicates that they are composed of different modes. For the constitutive reporter nCRPr, in both WT and cAMP-fixed* cells, catabolism and regulation modes are absent and the main contribution comes from the dilution mode. Cross-correlations of the C-sector reporter CRPr, on the other hand, additionally contain the catabolism and regulation modes, which largely cancel out, resulting in WT correlations with a shape similar to those of the nCRPr reporter. In the cAMP-fixed* cells, the negative regulation mode is absent in CRPr correlations and the catabolism mode becomes visible, resulting in a positive (cross-)correlation. Taken together, these observations show that temporal dynamics can be modelled as a linear combination of modes, consistent with the idea that multiple cellular processes, including metabolism and regulation, shape cellular heterogeneity.

4.2.8 *Cross-correlations under non-optimal C-sector expression corroborate multi-modal noise model*

To further test whether the mathematical model and the noise modes properly describe the stochastic dynamics of cAMP-CRP signalling, we sought ways to experimentally change the amplitudes of certain modes. In the cAMP-fixed cells, we therefore examined cells for which we fixed the cAMP concentration to levels such that the average C-sector expression was below ($cAMP-fixed^{low}$) or above ($cAMP-fixed^{high}$) wild type expression. As expected, $cAMP-fixed^{low}$ showed lower, and $cAMP-fixed^{high}$ showed higher mean C-sector reporter expression (Fig. 4.3A, black dots). Strikingly, the constitutive nCRPr reporter showed the opposite: it showed higher expression in $cAMP-fixed^{low}$ cells, and lower expression in $cAMP-fixed^{high}$ (Fig. 4.3B, positions of the orange and red clouds). These observations may be understood in terms of resource limitations: when the C-sector takes up a larger fraction of the proteome, other protein expression must necessarily decrease [158] (Figs. S4.10 and 4.3C). Changes in C-sector expression moreover greatly influenced the average growth rate. The average growth rate was highest for cAMP-fixed* cells (Fig. 4.3B, middle black dot, and Fig. S4.10C). An under-expressed C-sector (Fig. 4.3A, left black dot) resulted in lower mean growth rate, consistent with a growth-limiting catabolism -the cell does not express enough catabolic enzymes to support a higher growth rate [32, 86, 181]. On the other hand, C-sector over-expression also resulted in slow population-averaged growth rate (Fig. 4.3A, right black dot), possibly because expressing catabolic proteins superfluously is costly and consumes resources better spent on other cellular components, such as ribosomes.

Interestingly, the stochastic trend at the single cell level for a given external cAMP level typically differed from the mean trend, *i.e.*, the regression lines through each of the data clouds (dashed lines in Figs. 4.3A, B) were not tangent to the curve that best fits the means (solid lines in Fig. 4.3A and B, and Fig. S4.10). For the constitutive reporter, the stochastic trend and the mean trend even had a different sign in the high CAMP condition (Fig. 4.3B,
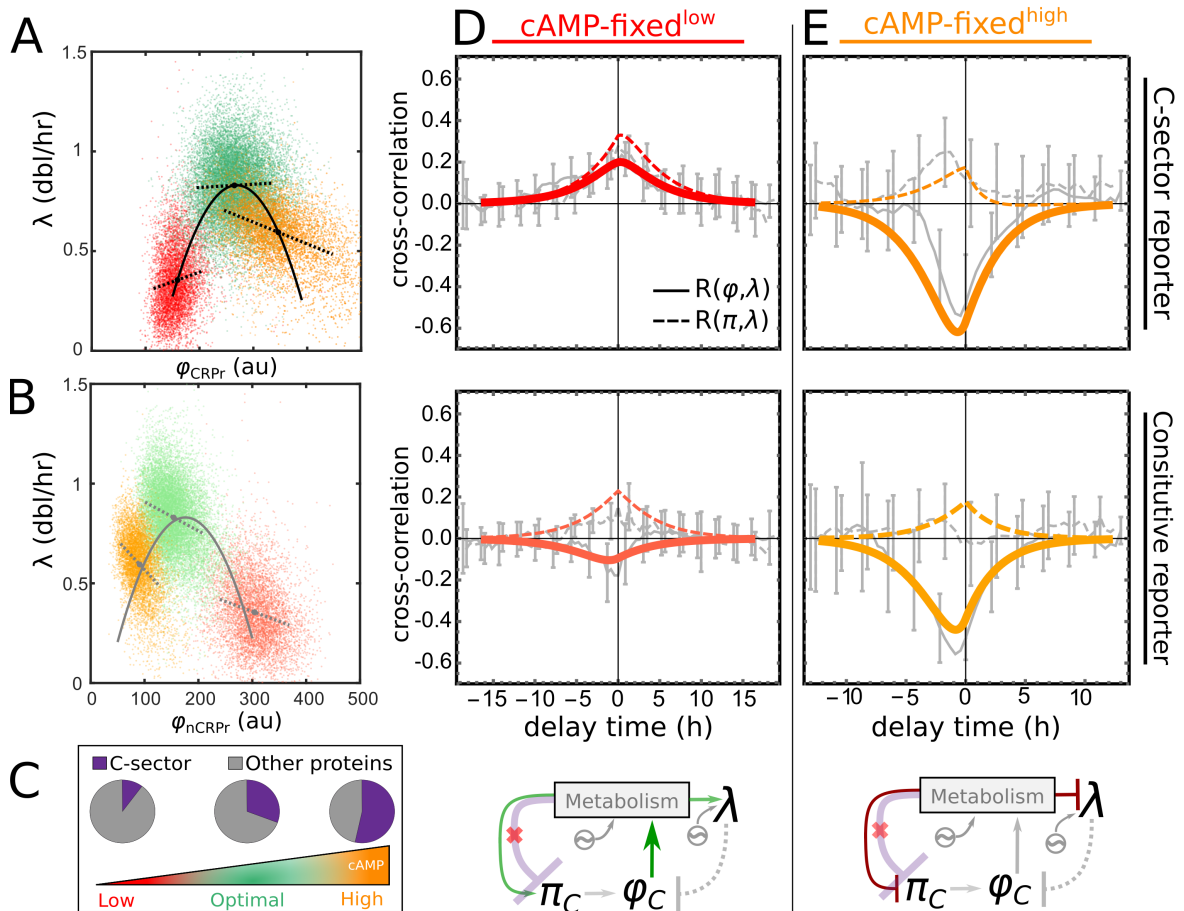
Figure 4.3: **Too low or high cAMP dampens growth and changes noise-mode amplitudes.** (**A**) Scatter plot of the C-sector reporter against the growth rate, under three conditions: low external cAMP (red cloud), optimal cAMP (green cloud, same condition as in Figs. 4.1D, 4.2B), high external cAMP (orange cloud). Black dots indicate averages, dashed lines are linear regressions (extending 2 std to each side), black curve is a 2nd order polynomial fit to the means. (**B**) Same as in B, but then for the constitutive reporter. Grey parabola is calculated from a sum constraint of both reporters (SI section S4.6). (**C**) Cartoon showing how increasing the external cAMP concentration increases the size of the C-sector in the mutant strain, but represses other proteins. (**D-E**) Measured cross-correlations (grey lines with error bars indicate se) for both reporters for low cAMP (80 μM) and high cAMP (5000 μM), together with model predictions (coloured lines) resulting from minimal parametric changes compared to the wild type fit. Model cartoons (bottom panels) indicate changes in transfer parameters (green: increase, red: decrease) with respect to cAMP-fixed* cells (SI section S4.5).

orange cloud). This highlights a fundamental difference between cells that transiently have a larger C-sector due to stochasticity, and cells that have a larger C-sector in response to external cues.

Next, we again used cross-correlation analysis and our mathematical model to study noise propagation in the $cAMP - fixed^{low}$ and $cAMP - fixed^{high}$ cells. Notably, we found that the most dramatic change in the cross-correlation function occurs when over-expressing the C-sector (high cAMP), for the C-sector reporter. At optimal cAMP, it showed positive correlations with no delay (Fig. 4.2B, top-right), while at high cAMP, the C-sector reporter showed a strongly negative correlation with a slightly negative delay (Fig. 4.3E, top panel). The other condition, low cAMP, also showed changes to the cross-correlation functions compared to the optimal cAMP concentration (Fig. 4.2B, right), but mostly in terms of amplitudes (Fig. 4.3D, top and middle panel).

These changes can again be explained by the noise propagation modes, only three of which are now active because the regulation mode is disabled in cAMP-fixed cells. Specifically, the catabolism mode was expected to increase in amplitude in $cAMP - fixed^{low}$ cells because in these cells the C-sector metabolic enzymes are rate limiting. This should increase the coupling from ϕ_C to metabolism, but also from metabolism to the growth rate and protein production rates (Fig. 4.3D, model cartoon). For instance, temporary upward fluctuations in C-sector proteins might alleviate the metabolic bottleneck caused by their insufficient mean expression, thus yielding larger increases in the growth rate than at optimal cAMP. As a result, positive correlations increase in magnitude and become symmetrical (Fig. 4.3D, top panel) or negative correlations become less negative (Fig. 4.3D, middle panel) compared to the optimal cAMP condition (Fig. 4.2D, left panels). Increasing the corresponding parameters in the model (Fig. 4.3D, model cartoon, SI section S4.5) readily reproduced the experimentally observed cross-correlations qualitatively (Fig. S4.6A), although the absolute value of cross-correlation was over-estimated. A more quantitative fit (Fig. 4.3D) can be obtained by increasing the noise amplitudes of the reporters to decorrelate the signals (see SI section S4.5). Possibly, such differences in noise amplitudes are caused by differences in average expression levels and in the mean growth rate, as noise amplitudes are generally proportional to the mean [4, 173, 195].

Conversely, in $cAMP - fixed^{high}$ cells the C-sector is over-expressed and takes up valuable cellular resources superfluously. We hypothesised that stochastic fluctuations that result in higher C-sector expression would take up even more superfluous resources, and further dampen the growth rate and protein production rates. Such negative control over the growth rate -a higher concentration leading to slower growth- is often seen in bulk experiments [22, 181] and has been suggested to play a role in noise propagation as well [87, 99]. The strongly negative $\phi_{CRPr} - \lambda$ cross-correlation observed (Fig. 4.3E, top panel) is indeed consistent with a negative catabolism mode, and setting the model's corresponding transfer parameters (Fig. 4.3E, bottom panel, SI section S4.5) to a negative value again readily reproduced important features observed in cross-correlations of the $cAMP - fixed^{high}$ (Fig. S4.6B). However, the observed high-cAMP cross-correlations seem to have shorter timescales relative to the mean growth rate, compared to the other conditions (Fig. S4.6B). A quantitative fit was possible (Fig. 4.3E), but again required additional tuning of noise amplitudes and timescales (SI section S4.5). Taken together, these experiments show that our coarse-grained mechanistical model, together with the noise propagation modes, is able

to capture and explain many important features of the stochastic dynamics of metabolism and growth.

4.3   DISCUSSION

Many metabolites are known to play a role in the cell's regulatory response to external variation. Recently, it has become increasingly clear that metabolic networks are noisy [48, 86, 134, 184], raising the question of whether metabolic regulatory networks also react to stochastic internal variation. By experimentally disrupting the ability of the cAMP-CRP regulation system to respond to internal metabolite fluctuations, whilst maintaining proper CRP activity with regard to external conditions, we were able to probe this question. Experimental single cell measurements and mathematical modelling allowed us to observe and interpret the temporal dynamics of cAMP-CRP regulation responding to fluctuating metabolite concentrations, C-sector expression, and cell growth. We found that these dynamics can be understood in terms of several noise propagation modes, and are consistent with noise propagation from internal metabolite concentrations to cAMP-CRP activity, and from thereon to the C-sector and to the whole of the cell.

Whilst CRP is an important master regulator, other regulatory and metabolic proteins also respond to concentrations of cellular metabolites and small other molecules. For instance, (p)ppGpp is a global modulator of proteins related to translation and metabolism including ribosomal expression, which also directly influences the cellular growth rate [166, 204]. Our results may also apply to such other regulators, as these could also respond to internal fluctuations of small molecules and propagate them through the cell. Since stochastic fluctuations could occur in all metabolites that exert allosteric control, noise likely propagates through the cell via many complex routes, potentially involving metabolites, metabolic pathways, expression of groups of functionally related genes, and growth.

To understand such stochastic dynamics and pinpoint the effect of various noise propagation mechanisms, we constructed a mathematical model that was able to dissect observed noisy dynamics of the cAMP-CRP system in terms of four underlying noise modes. These noise modes were then further validated in a second set of experiments that under-, or over-expressed the C-sector, which changed the dynamics in ways that could be explained by the model and are consistent with biological expectations. Noise in metabolism has a central role in this model, suggesting that noise in metabolic pathways and proteins could strongly affect cellular heterogeneity. Our phenomenological model treats the underlying intricate biological processes as a black box, but it is interesting to speculate about the mechanisms producing metabolic noise. Despite regulation of the C-sector, each C-sector protein species is subject to intrinsic noise. This likely contributes to heterogeneity in metabolite concentrations and metabolic fluxes [48, 179, 192]. The metabolic noise in our model therefore could be viewed as the compound result of the fluctuations of multiple catabolic proteins and their effect on metabolic fluxes. In our model, the cAMP-CRP feedback receives direct input from this internal metabolic noise source and transmits this noise to the entire C-sector. In that sense, noise reverberates through the cell; even if the provenance of noise may be in transcriptional and translational processes, other cellular processes can transmit, modify and amplify this noise, and start to act as noise sources themselves, such that it becomes difficult to disentangle source from intermediary. Thus, fluctuations in metabolite concentrations may emerge as a result of noise in catabolic protein expression, and

in turn contribute to fluctuations in the cAMP-CRP sensory-regulatory network and its downstream regulatory effects.

We estimated noise transfer parameters in our model by fitting them to single-cell experimental data obtained in steady state. However, this work suggests that regulatory interactions respond both to changes in external conditions and to stochastic metabolite fluctuations within the cell. Given this dual role of regulatory interactions, future models that describe stochastic dynamics could be further sophisticated by taking into account information determined from bulk experiments that probe the proteome's dynamical response to a change in the environment. After all, the behaviour of the regulatory network is shaped by its physical parameters that should remain identical independent of whether it is responding to stochastic fluctuations or changes in external conditions. Bulk experiments could provide robustly parameterized non-linear functions that describe how regulatory networks respond to changes in their molecular input -which might for example be metabolite concentrations- or how they respond to changes in other cellular parameters such as growth or C-sector expression [43, 98, 181]. Linearization of such functions should yield parameters that can describe the amount of noise that would be transferred by these interactions. These transfer parameters can then be used to construct single cell models as presented here. Such an approach would validate the link between stochastic responses and responses to external cues of regulatory networks, and further strengthen the validity of stochastic models, as their parameters are not fit to the data they describe, but inferred from independent experiments.

This work adds to the growing literature that studies noise in gene expression in the context of metabolism and growth [31, 86, 87, 108, 118, 160, 162, 167]. This contrasts with earlier, theoretical work that mainly focused on particular regulatory networks [64, 85, 164] and their influence on expression noise. There, a common finding is that negative feedback reduces noise amplitudes [23, 66, 115]. Consistently, our model hints at a reduction of noise levels due to the negative feedback by cAMP-CRP (see SI section S4.5). However, in our experiments coefficients of variation of both reporters, as well as the growth rate, were slightly smaller in wild type cells compared to cAMP-fixed* cells, but the differences were too small to be conclusive (Fig. S4.4).

Our work is a new step towards understanding the counter-intuitive differences between the population-averaged bacterium, and the noisy, individual bacterium. By studying not only different steady states, but also the dynamics of individuals around these steady states, we shed light on how the behaviour of a population of cells is shaped by the regulatory wiring and sensory metabolites inside each individual cell that continuously reacts to intricate internal stochastic processes.

Material and Methods

Strains. All strains used were based on wild type strain MG1655 (CGSC 8003, bBT12). The CRPr and nCRPr promoters were based on the lac operon promoter, with respectively the lacI binding site or both lacI and CRP binding site scrambled [181]. To obtain the C-sector reporter (CRPr) and constitutive reporter (nCRPr), we fused these promoters to mCerulean and mVenus sequences respectively. The reporters where then inserted into the chromosomes of the bBT12 strain and a cyaA cpda null mutant strain constructed earlier (bBT80), using a lambda red protocol [181]. See SI, Table S4.2 for strain details and Fig. S4.11 for promoter sequences.

Single cell experiments. Micro-colonies of cells were grown on gel pads, imaged under a microscope, and analyzed by computer as described earlier[86, 192]. Briefly, polyacrylamide gel pads (approx. 5 mm x 5 mm x 1 mm in size) were pre-soaked in M9 minimal medium supplemented with lactose (0.01 % g/mL), uracil (0.2 mM), Tween20 (0.001%) and the desired concentration of cAMP (Sigma Aldrich) if applicable. Pads were placed in a sealed glass chamber created by a microscope slide and a 2nd glass cavity slide, covered by a glass cover slip. Cells were pre-grown overnight in the same medium, and 1 μl of exponentially growing culture (OD 0.005) was then inoculated on the gel pad at the start of the experiment. Everything was done at 37 °C, and the glass chamber with pad and cells was then placed in a customised scaffold, and imaged under a microscope with a customised incubation chamber at 37 °C. For the WT, $cAMP-fixed^{low}$, cAMP-fixed* and $cAMP-fixed^{high}$ conditions, we respectively processed time series data from 6, 2, 4 and 2 micro-colonies.

Microscopy. We used a Nikon, TE2000 microscope, equipped with 100X oil immersion objective (Nikon, Plan Fluor NA 1.3), cooled CMOS camera (Hamamatsu, Orca Flash4.0), xenon lamp with liquid light guide (Sutter, Lambda LS), GFP, mCherry, CFP and YFP filter set (Chroma, 41017, 49008, 49001 and 49003), computer controlled shutters (Sutter, Lambda 10-3 with SmartShutter), automated stage (Märzhäuser, SCAN IM 120 x 100) and an incubation chamber (Solent) allowing precise 37 °C temperature control. An additional 1.5X lens was used, resulting in images with pixel size of 0.0438μm. The microscope was controlled by MetaMorph software, which allowed us to automatically take pictures at set intervals. Image acquisition intervals were adjusted to doubling times to obtain multiple fluorescent images per cell cycle; phase contrast images were taken every 60-90 seconds, CFP and YFP fluorescent images (150-200 ms exposure time) were taken at intervals ranging from 13.5-26 minutes.

Image analysis. Series of phase contrast images were analysed with a custom Matlab (Mathworks) program originally derived from Schnitzcells software [199]. Cells were segmented and tracked to follow cells and lineages through time. For each frame, cell lengths were determined by fitting a 3rd order polynomial to the curved segmentation regions. Cells were assumed to have a constant width. Growth rates (dbl/hr) were determined by fitting an exponential function to the cell lengths over multiple frames (5 to 9). To determine the production rate per volume, first the sum of the fluorescence signal (a.u.) over all pixels that make up a cell was calculated. If on frame n also a fluorescence image was taken, we then calculated the slope of a linear fit through three points $n - l$, n, and $n + l$ (where l is the frame interval at which fluorescence pictures are taken), the resulting number is divided by the total number of pixels of the cell in frame n to obtain the production rate. Concentrations were determined by dividing the sum of the fluorescence signal by the total number of pixels in a cell. To determine scatter plots and correlations, only frames where fluorescence images were taken are considered.

Cross-correlation analysis. We define the cross-correlation between signals f and g as $R_{f,g}(\tau) = S_{f,g}(\tau)/\sqrt{\sigma_f^2 \sigma_g^2}$, with $S_{f,g}(\tau) = 1/(N - |\tau|) \sum_{n=0}^{N-|\tau|-1} f(n)g(n + \tau)$, where n equals discrete units of time as frame numbers, τ is a delay in number of frames, N the total number of time points in the data series and σ_f^2 the variance of f (equal to $S_{f,f}(0)$). See SI section S4.4 for more details about data weights and statistics.

Mathematical model As mentioned, our model consists of stochastic differential equations, and includes parameters for the protein production rates (π), protein concentrations (ϕ), metabolism (M), and growth rate (λ), Ornstein-Uhlenbeck noise sources N and noise transfer coefficients T that couple equations for $\partial\pi/\partial t$, $\partial M/\partial t$, and $\partial\lambda/\partial t$; concentrations are determined by $\partial\phi/\partial t = \pi - \phi\lambda$. This model is solved analytically to predict cross-correlations between the quantities. See the supplement for an extensive description of the model, procedures to fit the model to experimental data, a statistical null model for the cross-correlations, and a toy model that describes the mean behaviour of π, ϕ, and λ for CRPr and nCRPr in different conditions as observed in 4.3.

Script and data availability. Matlab scripts and mathematica notebooks used to create all the figures can be requested by the authors, and can be found at:

https://github.com/Jintram/DynamicalRegulationBacterialCells.

SUPPLEMENTARY INFORMATION CHAPTER 4

S4.1 MATHEMATICAL ANALYSIS OF THE NOISE MODEL

Note on notation: Where the main text talks about the C-sector reporter (CRPr) and the not CRP-regulated constitutive reporter (nCRPr), this document uses the notation 'Y' for the C-sector reporter and 'H' for the constitutive reporter.

S4.1.1 *Model definition and parameter interpretation*

Similar to Kiviet *et al* [86] a linear noise propagation model can be constructed using the diagram of Fig. S4.1. The model captures the interplay between the production rate (π) of certain proteins, the concentrations (ϕ) of these proteins, metabolism (M), and the growth rate (λ).

We model the total C-sector in a coarse-grained manner as a single protein, C, whose fluctuations directly influence (the rate of) metabolism. Additionally, we model the dynamics of the two fluorescent reporters (for ease of notation here called Y and H) that are not metabolically active and hence do not control metabolism.

We assume that the variables fluctuate around a particular average state $\{\lambda_0, M_0, \vec{\pi_0}, \vec{\phi_0}\}$; the model then describes the dynamics of small, relative deviations of each variable from its respective average, $\frac{\delta X(t)}{X_0} := \frac{X(t) - X_0}{X_0}$. (Below, the explicit time dependence will often be omitted.) These deviations, referred to as "noise", propagate through the system according to the arrows in Fig. S4.1. For example, fluctuations in the production rate π_C affect ϕ_C, whose fluctuations affect M. Via M the noise then further transfers to the production rates and to the growth rate. Noise in the growth rate in turn affects the dilution of all proteins (dashed line). The logarithmic gains/transfer coefficients T_{AB} describe the strength of noise propagation from A to B.

The role for metabolism is crucial in our model because many noise routes pass through metabolism, but also because the cAMP-CRP regulatory network reacts to metabolic fluctuations. The interpretation of M is therefore not straightforward. M could be interpret as the rate of metabolism. Additionally, we assume that with a higher rate of metabolism, the concentrations of certain internal metabolites (in particular keto-acids such as OAA) also increase. The metabolite OAA is known to inhibit the production of cAMP in wild type cells [198] (see Fig. 4.1A), therewith triggering the cAMP-CRP regulatory feedback when M increases (represented by the regulatory parameter T_R in the model, Fig. S4.1). Although we model metabolism with a single variable M, it thus represents both the rate of metabolism and the saturation level of metabolic precursors such as OAA.

Certain variables are directly influenced by phenomenological, independent coloured noise sources, notated as N_i: Ornstein–Uhlenbeck Noise sources that each have a specific timescale. These noise sources offer a way to model the combined effect of many stochastic processes that together influence the stochastic dynamics of certain variables. For example, N_{π_Y} summarises processes that intrinsically influence the production rate of the Y reporter,
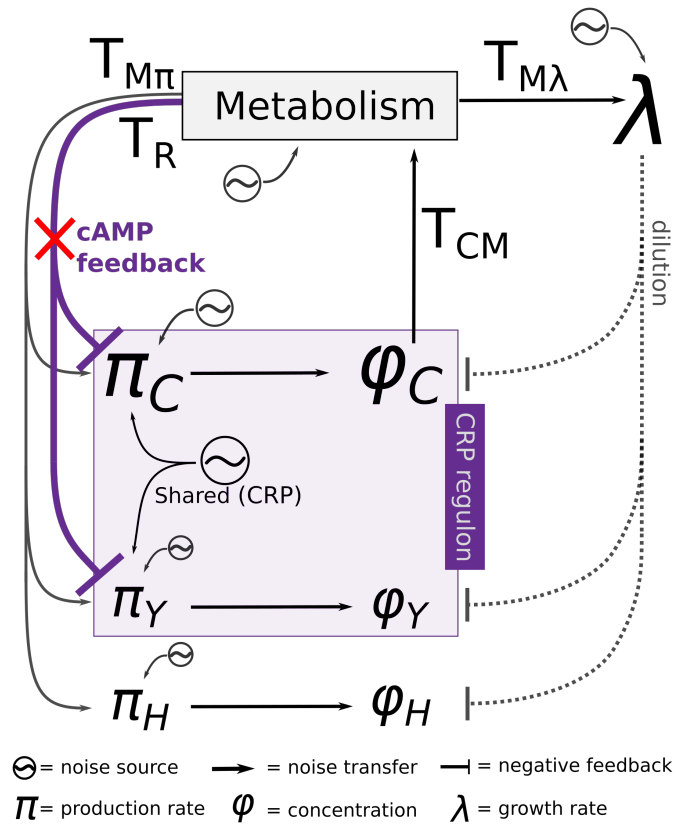
Figure S4.1: Diagram of the Noise Routes, including C reporter (Y) and constitutive reporter (H). An additional shared noise source is added between π_C and π_Y, here called N_s, that represents noise in the sensory mechanism (for example CRP-concentration fluctuations). The transfer of noise from Metabolism (M) to production exist of general transfer ($T_{M\pi}$) and, for C and Y, regulatory transfer (T_R).

such as transcription and translation. N_M, the noise source acting on metabolism, is primarily the result of fluctuations in the concentration of individual protein species that are part of the C-sector. The fluctuations of all these proteins together are expected to cause the metabolic flux to fluctuate, even if the total size of the C-sector would remain constant. Fluctuations in the concentrations of individual proteins are diluted via growth, so that the N_M noise source has a timescale close to λ_0. N_λ has a similar interpretation. N_s is a shared noise source between the C-sector and the C-sector reporter Y, mainly interpreted as fluctuations in the production of CRP-regulated proteins due to fluctuations in the concentration of CRP itself. Even when the cAMP concentration is experimentally kept fixed in the mutant strain, fluctuations in the CRP concentration are expected to influence the cAMP-CRP concentration and therewith the production rate of both the C-sector and its reporter Y.

Based on the considerations above, we arrive at the following set of equations:

$$\frac{\delta\lambda}{\lambda_0} = N_\lambda + T_{M\lambda}\frac{\delta M}{M_0}, \tag{S4.1}$$

$$\frac{\delta M}{M_0} = N_M + T_{CM}\frac{\delta\phi_C}{\phi_{C,0}}, \tag{S4.2}$$

$$\frac{\delta\pi_C}{\pi_{0,C}} = N_{\pi_C} + N_s + (T_{M\pi} + T_R)\frac{\delta M}{M_0}, \tag{S4.3}$$

$$\frac{\delta\pi_Y}{\pi_{0,Y}} = N_{\pi_Y} + N_s + (T_{M\pi} + T_R)\frac{\delta M}{M_0}, \tag{S4.4}$$

$$\frac{\delta\pi_H}{\pi_{0,H}} = N_{\pi_H} + T_{M\pi}\frac{\delta M}{M_0}, \tag{S4.5}$$

$$\left(\frac{\delta\dot{\phi}_C}{\lambda_0\phi_{0,C}}\right) = \frac{\delta\pi_C}{\pi_{0,C}} - \frac{\delta\phi_C}{\phi_{0,C}} - \frac{\delta\lambda}{\lambda_0}, \tag{S4.6}$$

$$\left(\frac{\delta\dot{\phi}_Y}{\lambda_0\phi_{0,Y}}\right) = \frac{\delta\pi_Y}{\pi_{0,Y}} - \frac{\delta\phi_Y}{\phi_{0,Y}} - \frac{\delta\lambda}{\lambda_0}, \tag{S4.7}$$

$$\left(\frac{\delta\dot{\phi}_H}{\lambda_0\phi_{0,H}}\right) = \frac{\delta\pi_H}{\pi_{0,H}} - \frac{\delta\phi_H}{\phi_{0,H}} - \frac{\delta\lambda}{\lambda_0}. \tag{S4.8}$$

The last three equations are linearisations of the time derivative of the protein concentrations ($\dot{\phi}_i = \pi_i - \lambda\phi_i$), see also [86]. The dynamics of each noise source N_x is given by its stochastic differential equation,

$$dN_x = -\beta_x N_x dt + \theta_x dW_t, \tag{S4.9}$$

where β_x is the noise source's timescale, θ_x its amplitude, and $W(t)$ a Wiener Process.

s4.1.2 *Metabolic timescale, solution in Fourier space*

The system of equations S4.1-S4.8 above can be solved in Fourier space. For each variable $A(t)$ we denote the Fourier Transform of $\delta A(t)$ as $\widetilde{A}(\omega)$. Next, we introduce a metabolic timescale that scales with the mean growth rate: $\lambda_E = \lambda_0(1 - T_{CM}T_{MC})$, where $T_{MC} = T_R + T_{M\pi} - T_{M\lambda}$. This timescale can be interpret as the time it takes before the protein concentration equilibrates with its protein production rate (note that producing proteins takes time, such that a higher production rate now, only results in a higher concentration some

time later). During exponential growth, proteins are generally produced at a (mean) rate proportional to λ_0, such that production and dilution are balanced. In effect, fluctuations in protein concentrations generally decay on a timescale of λ_0. However, in the case of C-sector proteins, synthesised proteins directly influence the rate of metabolism ($T_{CM} \neq 0$), and (depending on the parameters $T_R, T_{M\pi}$ and $T_{M\lambda}$) metabolism in turn influences the production rate and protein concentration. This changes the timescale of C-sector fluctuations from λ_0 to λ_E. Using this notation, we solve the linear system on the previous page in Fourier Space and arrive at:

$$\begin{cases} \frac{\widetilde{\delta\lambda}}{\lambda_0} &= N_\lambda + T_{M\lambda}N_M + \frac{\lambda_0}{\lambda_E+i\omega}(T_{CM}T_{M\lambda})\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right], \\ \frac{\widetilde{\delta\phi_C}}{\phi_{C0}} &= \frac{\lambda_0}{\lambda_E+i\omega}\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right], \\ \frac{\widetilde{\delta\pi_C}}{\pi_{C0}} &= N_{\pi_C}+N_s+(T_R+T_{M\pi})N_M+\frac{\lambda_0}{\lambda_E+i\omega}T_{CM}(T_R+T_{M\pi})\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right]. \end{cases}$$

$$(S4.10)$$

$$\begin{cases} \frac{\widetilde{\delta\phi_Y}}{\phi_{y,0}} &= \frac{\lambda_0}{\lambda_0+i\omega}\left[N_{\pi_Y}+N_s+T_{MC}N_M-N_\lambda\right]+\frac{\lambda_0^2 T_{CM}T_{MC}}{(\lambda_E+i\omega)(\lambda_0+i\omega)}\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right], \\ \frac{\widetilde{\delta\pi_Y}}{\pi_{y,0}} &= N_{\pi_Y}+N_s+(T_R+T_{M\pi})N_M+\frac{\lambda_0}{\lambda_E+i\omega}T_{CM}(T_R+T_{M\pi})\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right]. \end{cases}$$

$$(S4.11)$$

$$\begin{cases} \frac{\widetilde{\delta\phi_H}}{\phi_{H,0}} &= \frac{\lambda_0}{\lambda_0+i\omega}\left[N_{\pi_H}+(T_{M\pi}-T_{M\lambda})N_M-N_\lambda\right]+\frac{\lambda_0^2 T_{CM}(T_{M\pi}-T_{M\lambda})}{(\lambda_E+i\omega)(\lambda_0+i\omega)}\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right], \\ \frac{\widetilde{\delta\pi_H}}{\pi_{H,0}} &= N_{\pi_H}+T_{M\pi}N_M+\frac{\lambda_0}{\lambda_E+i\omega}T_{CM}T_{M\pi}\left[N_{\pi_C}+N_s+T_{MC}N_M-N_\lambda\right]. \end{cases}$$

$$(S4.12)$$

s4.1.3   *Calculated variances and cross-covariances*

Definition of cross-covariance

The solution in Fourier Space (equations S4.10 - S4.12) can be used to calculate variances and cross-covariances of the variables. Recall that the cross-covariance of variables A/A_0 and B/B_0 is given by:

$$\chi_{(A,B)}(\tau) = \frac{1}{2\pi}\mathcal{F}^{-1}\left(\frac{\widetilde{A(t-\tau)}^*}{A_0}\frac{\widetilde{B(t)}}{B_0}\right), \tag{S4.13}$$

where $\mathcal{F}^{-1}(.)$ denotes the Inverse Fourier Transform. In the product $\frac{\widetilde{A}^*}{A_0}\frac{\widetilde{B}}{B_0}$ we can safely ignore terms of $N_i N_j^*$ when $i \neq j$, for they will not contribute to the cross-correlation due to independence of the noise sources (* here denotes complex conjugation). Consequently, the expansion of the product is always linear in the absolute values of the noise sources $|N_i|^2$. This feature is important, because it assures that the cross-covariance can be written as $\chi_{(A,B)}(\tau) = \frac{1}{2\pi}\mathcal{F}^{-1}\left(\sum_i f_i(\omega)|N_i|^2\right)$, where $f_i(\omega)$ are complex functions to be determined and the summation runs over all noise sources. Because the Inverse Fourier Transform is a linear operator, we only need to calculate the Inverse Fourier Transform of each term, $f_i(\omega)|N_i|^2$, separately. The Inverse Fourier Transform of each term is given in section S4.1.3.

Analytical expressions for the coefficients of variation

From the above equations (S4.10 - S4.13) one can also derive the coefficient of variation of all variables. Concretely, the coefficient of variation of variable A is related to its auto-covariance at zero delay:

$$\eta_A^2 := \sigma_A^2/A_0^2 = \mathcal{X}_{(A,A)}(0). \tag{S4.14}$$

For example, the coefficients of variation for ϕ_C and λ can be calculated using the Inverse Fourier transform as follows:

$$\eta_{\phi_C}^2 := \mathcal{F}_{(\tau=0)}^{-1}\left[\frac{\widetilde{\delta\phi_C}^*}{\phi_{0,C}}\frac{\widetilde{\delta\phi_C}}{\phi_{0,C}}\right] = \mathcal{F}_{(0)}^{-1}\left[\frac{\lambda_0^2}{\lambda_E^2+\omega^2}\left(|N_{\pi_C}|^2+|N_s|^2+T_{MC}^2|N_M|^2+|N_\lambda|^2\right)\right]. \tag{S4.15}$$

$$\eta_\lambda^2 = \mathcal{F}_{(\tau=0)}^{-1}\left[|N_\lambda|^2+T_{M\lambda}^2|N_M|^2+\frac{2\lambda_0\lambda_E}{\lambda_E^2+\omega^2}T_{CM}T_{M\lambda}\left(T_{M\lambda}T_{MC}|N_M|^2-|N_\lambda|^2\right)\right] \tag{S4.16}$$

$$+ (T_{CM}T_{M\lambda})^2\eta_{\phi_C}^2. \tag{S4.17}$$

(Here we have used the identity $\frac{\lambda_0}{\lambda_E+i\omega}+\frac{\lambda_0}{\lambda_E-i\omega}=\frac{2\lambda_0\lambda_E}{\lambda_E^2+\omega^2}$.) To calculate these (Inverse) Fourier Transforms we use a set of identities that can be found later in this document (section S4.1.3, equations S4.28-S4.38). Here we present the results (using the notation $\beta_X^i := \beta_X+\lambda_i$, with $i=0,E$):

$$\begin{cases}
\eta_{\phi_C}^2 &= \frac{\lambda_0^2}{2\lambda_E}\left(\frac{\theta_{\pi_C}^2}{\beta_{\pi_C}\beta_{\pi_C}^E}+\frac{\theta_s^2}{\beta_s\beta_s^E}+T_{MC}^2\frac{\theta_M^2}{\beta_M\beta_M^E}+\frac{\theta_\lambda^2}{\beta_\lambda\beta_\lambda^E}\right), \\[2mm]
\eta_\lambda^2 &= \frac{\theta_\lambda^2}{2\beta_\lambda}\left(1-\frac{2\lambda_0}{\beta_\lambda^E}T_{CM}T_{M\lambda}\right)+\frac{\theta_M^2}{2\beta_M}T_{M\lambda}^2\left(1+\frac{2\lambda_0}{\beta_M^E}T_{CM}T_{MC}\right)+(T_{CM}T_{M\lambda})^2\eta_{\phi_C}^2, \\[2mm]
\eta_{\pi_C}^2 &= \frac{\theta_{\pi_C}^2}{2\beta_{\pi_C}}\left(1+\frac{2\lambda_0}{\beta_{\pi_C}^E}T_{CM}(T_R+T_{M\pi})\right)+\frac{\theta_s^2}{2\beta_s}\left(1+\frac{2\lambda_0}{\beta_s^E}T_{CM}(T_R+T_{M\pi})\right) \\[2mm]
&\quad +\frac{\theta_M^2}{2\beta_M}(T_R+T_{M\pi})^2\left(1+\frac{2\lambda_0}{\beta_M^E}T_{MC}T_{CM}\right)+T_{CM}^2(T_R+T_{M\pi})^2\eta_{\phi_C}^2.
\end{cases} \tag{S4.18}$$

$$\begin{cases}
\eta_{\phi_Y}^2 &= \frac{\theta_{\pi_Y}^2\lambda_0}{2\beta_{\pi_Y}\beta_{\pi_Y}^0}+\frac{\theta_s^2\lambda_0}{2\beta_s\beta_s^0}\left(1+T_{CM}T_{MC}\frac{\lambda_0(\beta_s+\lambda_0+\lambda_E)}{\lambda_E\beta_s^E}\right)+\frac{\theta_M^2\lambda_0}{2\beta_M\beta_M^0}T_{MC}^2\left(1+T_{CM}T_{MC}\frac{\lambda_0(\beta_M+\lambda_0+\lambda_E)}{\lambda_E\beta_M^E}\right) \\[2mm]
&\quad +\frac{\theta_\lambda^2\lambda_0}{2\beta_\lambda\beta_\lambda^0}\left(1+T_{CM}T_{MC}\frac{\lambda_0(\beta_\lambda+\lambda_0+\lambda_E)}{\lambda_E\beta_\lambda^E}\right)+\frac{\theta_{\pi_C}^2\lambda_0^2}{2\beta_{\pi_C}\beta_{\pi_C}^0}T_{CM}^2T_{MC}^2\frac{\lambda_0(\beta_{\pi_C}+\lambda_0+\lambda_E)}{\beta_{\pi_C}^E\lambda_E(\lambda_0+\lambda_E)}, \\[2mm]
\eta_{\pi_Y}^2 &= \frac{\theta_{\pi_Y}^2}{2\beta_{\pi_Y}}+\frac{\theta_s^2}{2\beta_s}\left(1+2\frac{\lambda_0}{\beta_s^E}T_{CM}(T_R+T_{M\pi})\right)+\frac{\theta_M^2}{2\beta_M}(T_R+T_{M\pi})^2\left(1+2\frac{\lambda_0}{\beta_M^E}T_{CM}T_{MC}\right) \\[2mm]
&\quad +T_{CM}^2(T_R+T_{M\pi})^2\eta_{\phi_C}^2.
\end{cases} \tag{S4.19}$$

$$\begin{cases}
\eta_{\phi_H}^2 &= \frac{\theta_{\pi_H}^2\lambda_0}{2\beta_{\pi_H}\beta_{\pi_H}^0}+\frac{\lambda_0^3T_{CM}^2(T_{M\pi}-T_{M\lambda})^2}{2\lambda_E(\lambda_0+\lambda_E)}\left(\frac{\theta_{\pi_C}^2(\beta_{\pi_C}+\lambda_0+\lambda_E)}{\beta_{\pi_C}\beta_{\pi_C}^0\beta_{\pi_C}^E}+\frac{\theta_s^2(\beta_s+\lambda_0+\lambda_E)}{\beta_s\beta_s^0\beta_s^E}\right) \\[2mm]
&\quad +\frac{\theta_M^2\lambda_0(T_{M\pi}-T_{M\lambda})^2}{2\beta_M\beta_M^0}\left(1+T_{CM}T_{MC}\frac{\lambda_0(\beta_M+\lambda_0+\lambda_E)}{\lambda_E\beta_M^E}\right) \\[2mm]
&\quad +\frac{\theta_\lambda^2\lambda_0}{2\beta_\lambda\beta_\lambda^0}\left(1+T_{CM}(T_{M\pi}-T_{M\lambda})\frac{\lambda_0(\beta_\lambda+\lambda_0+\lambda_E)}{\beta_\lambda^E(\lambda_0+\lambda_E)}\left(T_{CM}(T_{M\pi}-T_{M\lambda})\frac{\lambda_0}{\lambda_E}+2\right)\right) \\[2mm]
\eta_{\pi_H}^2 &= \frac{\theta_{\pi_H}^2}{2\beta_{\pi_H}}+\frac{\theta_M^2}{2\beta_M}T_{M\pi}^2\left(1+2\frac{\lambda_0}{\beta_M^E}T_{CM}T_{MC}\right)+T_{CM}^2T_{M\pi}^2\eta_{\phi_C}^2.
\end{cases} \tag{S4.20}$$

Analytical expressions for the cross-covariances

Now we are in the position to present explicit expressions for the cross-covariances, which, normalised by the above variances, will result in cross-correlations: $R_{(A,B)}(\tau) = \chi_{(A,B)}(\tau)/\sqrt{\sigma_A^2 \sigma_B^2}$. The cross-covariances use functional forms A, B, S, D^1, D^2, D^3, which are described on the next page, and the relation $A(t) + A(-t) = 2\frac{\lambda_E}{\lambda_0}S(t)$):

$$\chi_{(\phi_C,\lambda)}(t) = T_{CM}T_{M\lambda}\left[S_{\pi_C}(t) + S_s(t) + T_{MC}^2 S_M(t) + S_\lambda(t)\right] + T_{MC}T_{M\lambda}A_M(t) - A_\lambda(t).$$
(S4.21)

$$\chi_{(\pi_C,\lambda)}(t) = T_{CM}^2 T_{M\lambda}(T_R + T_{M\pi})\left[S_{\pi_c}(t) + S_s(t) + T_{MC}^2 S_M(t) + S_\lambda(t)\right] + (T_R + T_{M\pi})T_{M\lambda}B_M(t)$$
$$+ T_{CM}T_{M\lambda}\left[A_{\pi_C}(-t) + A_s(-t)\right] + \frac{2\lambda_E}{\lambda_0}T_{CM}T_{M\lambda}T_{MC}(T_R + T_{M\pi})S_M(t)$$
$$- T_{CM}(T_R + T_{M\pi})A_\lambda(t).$$
(S4.22)

$$\chi_{(\phi_Y,\lambda)}(t) = T_{MC}T_{M\lambda}A_M^0(t) - A_\lambda^0(t)$$
$$+ T_{CM}T_{M\lambda}\left[D_s^1(t) + T_{MC}^2 D_M^1(t) + D_\lambda^1(t)\right] + T_{CM}T_{MC}\left[T_{MC}T_{M\lambda}D_M^2(t) - D_\lambda^2(t)\right]$$
$$+ T_{CM}^2 T_{MC}T_{M\lambda}\left[D_{\pi_C}^3(t) + D_s^3(t) + T_{MC}^2 D_M^3(t) + D_\lambda^3(t)\right].$$
(S4.23)

$$\chi_{(\pi_Y,\lambda)}(t) = T_{M\lambda}(T_R + T_{M\pi})B_M(t) + \frac{2\lambda_E}{\lambda_0}T_{CM}T_{MC}T_{M\lambda}(T_R + T_{M\pi})S_M(t)$$
$$+ \quad T_{CM}\left[T_{M\lambda}A_s^E(-t) - (T_R + T_{M\pi})A_\lambda^E(t)\right]$$
$$+ \quad T_{CM}^2(T_R + T_{M\pi})T_{M\lambda}\left[S_{\pi_C}(t) + S_s(t) + T_{MC}^2 S_M(t) + S_\lambda(t)\right].$$
(S4.24)

$$\chi_{(\phi_H,\lambda)} = (T_{M\pi} - T_{M\lambda})T_{M\lambda}A_M^0(t) - A_\lambda^0(t) + T_{CM}T_{M\lambda}\left[(T_{M\pi} - T_{M\lambda})T_{MC}D_M^1(t) + D_\lambda^1(t)\right]$$
$$+ T_{CM}(T_{M\pi} - T_{M\lambda})\left[T_{MC}T_{M\lambda}D_M^2(t) - D_\lambda^2(t)\right]$$
$$+ T_{CM}^2(T_{M\pi} - T_{M\lambda})T_{M\lambda}\left[D_{\pi_C}^3(t) + D_s^3(t) + T_{MC}^2 D_M^3(t) + D_\lambda^3(t)\right].$$
(S4.25)

$$\chi_{(\pi_H,\lambda)} = T_{M\pi}T_{M\lambda}B_M(t) + \frac{2\lambda_E}{\lambda_0}T_{M\pi}(T_{MC}T_{CM}T_{M\lambda})S_M(t) - T_{CM}T_{M\pi}A_\lambda(-t)$$
$$+ T_{CM}^2 T_{M\pi}T_{M\lambda}\left[S_{\pi_C}(t) + S_s(t) + T_{MC}^2 S_M(t) + S_\lambda\right].$$
(S4.26)

The variances and cross-covariances presented here can be further studied and checked with the Mathematica notebook/supplementary file 'VarianceChecker.nb', which is available online: https://github.com/Jintram/DynamicalRegulationBacterialCells.

Functional forms of the building blocks

In the above equations, A, S, B, and D^1, D^2 and D^3 are functional forms that will be described below. The functional forms are found by complex integration using Cauchy's Residue theorem and are given by:

$$A^j(t) = \int \frac{e^{i\omega t}}{2\pi} \frac{\lambda_0}{\lambda_j - i\omega} \frac{\theta^2}{\beta^2 + \omega^2} d\omega, \quad \text{for } j \in \{0, E\}, \text{but superscript "E" is implicit} \quad (S4.27)$$

$$= \frac{\theta^2 \lambda_0}{2\beta} \begin{cases} \dfrac{2\beta e^{\lambda_j t}}{\beta^2 - \lambda_j^2} - d\dfrac{e^{\beta t}}{\beta(\beta - \lambda_j)}, & \text{for } t < 0, \\[3mm] \dfrac{e^{-\beta t}}{\beta_x + \lambda_j}, & \text{for } t \geqslant 0. \end{cases} \quad (S4.28)$$

$$S(t) = \int \frac{e^{i\omega t}}{2\pi} \frac{\lambda_0^2}{\lambda_E^2 + \omega^2} \frac{\theta^2}{\beta^2 + \omega^2} d\omega \quad (S4.29)$$

$$= \frac{\theta^2 \lambda_0^2}{2(\beta^2 - \lambda_E^2)} \left(\frac{e^{-\lambda_E |t|}}{\lambda_E} - \frac{e^{-\beta |t|}}{\beta} \right) \quad (S4.30)$$

$$D^1(t) = \int \frac{e^{i\omega t}}{2\pi} \frac{\lambda_0^2}{(\lambda_E + i\omega)(\lambda_0 - i\omega)} \frac{\theta^2}{\beta^2 + \omega^2} d\omega \quad (S4.31)$$

$$= \theta^2 \lambda_0^2 \begin{cases} \dfrac{e^{\lambda_0 t}}{(\beta^2 - \lambda_0^2)(\lambda_0 + \lambda_E)} - \dfrac{e^{\beta t}}{2\beta\beta_x^E(\beta - \lambda_0)}, & \text{for } t < 0, \\[3mm] \dfrac{e^{-\lambda_E t}}{(\beta^2 - \lambda_E^2)(\lambda_0 + \lambda_E)} - \dfrac{e^{-\beta t}}{2\beta\beta^0(\beta - \lambda_E)}, & \text{for } t \geqslant 0 \end{cases} \quad (S4.32)$$

$$D^2(t) = \int \frac{e^{i\omega t}}{2\pi} \frac{\lambda_0^2}{(\lambda_E - i\omega)(\lambda_0 - i\omega)} \frac{\theta^2}{\beta^2 + \omega^2} d\omega \quad (S4.33)$$

$$= \theta^2 \lambda_0^2 \begin{cases} \dfrac{e^{\beta t}}{2\beta(\beta - \lambda_0)(\beta - \lambda_E)} + \dfrac{1}{\lambda_0 - \lambda_E} \left(\dfrac{e^{\lambda_E t}}{\beta^2 - \lambda_E^2} - \dfrac{e^{\lambda_0 t}}{\beta^2 - \lambda_0^2} \right), & \text{for } t < 0, \\[3mm] \dfrac{e^{-\beta t}}{2\beta\beta^E\beta^0}, & \text{for } t \geqslant 0 \end{cases} \quad (S4.34)$$

$$D^3(t) = \int \frac{e^{i\omega t}}{2\pi} \frac{\lambda_0^3}{(\lambda_E^2 + \omega^2)(\lambda_0 - i\omega)} \frac{\theta^2}{\beta^2 + \omega^2} d\omega \quad (S4.35)$$

$$= \frac{\theta^2 \lambda_0^3}{2} \begin{cases} \left(\dfrac{e^{\beta t}}{\beta(\beta - \lambda_0)(\beta^2 - \lambda_E^2)} - \dfrac{2e^{\lambda_0 t}}{(\beta^2 - \lambda_0^2)(\lambda_0^2 - \lambda_E^2)} + \dfrac{e^{-\lambda_E t}}{\lambda_E(\beta^2 - \lambda_E^2)(\lambda_0 - \lambda_E)} \right), & \text{if } t < 0, \\[3mm] \dfrac{e^{-\lambda_E t}}{\lambda_E(\lambda_0 + \lambda_E)(\beta^2 - \lambda_E^2)} - \dfrac{e^{-\beta t}}{\beta\beta^0(\beta^2 - \lambda_E^2)}, & \text{if } t \geqslant 0 \end{cases} \quad (S4.36)$$

$$B(t) = \int \frac{e^{i\omega t}}{2\pi} \frac{\theta^2}{\beta^2 + \omega^2} d\omega \quad (S4.37)$$

$$= \frac{\theta^2}{2\beta} e^{-\beta |t|}. \quad (S4.38)$$

Note that the functions D^1 and S are very similar: the only difference is a single timescale change from λ_0 to λ_E. Functions D^2 and D^3 appear only in terms representing contributions of noise that propagates a full circle, reverberating through the entire cell, starting in ϕ_C, passing through metabolism and growth, and eventually reaching the reporters.

S4.2   ANALYSING THE CROSS-CORRELATIONS

The cross-correlation between two signals A and B (covariances of A and B divided by their standard deviations) contains temporal information about the dynamical interplay of the two signals. For example, if noise first reaches signal A and only later arrives at B, their cross-correlation will peak at a positive delay time. Generally, the complicated expressions for the cross-correlations between A and B indeed consists of several terms that each correspond to a different mechanism by which noise can travel from some source to A and B. Insight into these different mechanisms can be gained by splitting the mathematical description of the cross-correlation into different 'modes' that can each be interpreted as a different way in which noise can propagate through the cell. However, such a decomposition is not unique and hence subject to convention. Therefore, we first discuss a decomposition from Kiviet *et al* [86] using the ϕ_C-λ cross-covariance, $\chi_{\phi_C,\lambda}$ as an example. Next, we propose a new decomposition that is better suited to the system investigated in this work.

s4.2.1   *The ϕ_C-λ cross-correlation: Old Decomposition*

First, we split up the ϕ_C-λ cross-covariance (see equation S4.21) in the same way as in Kiviet *et al*, using three modes:

- **Catabolic mode**. This mode contains the contributions of all paths that transfer via catabolism (*i.e.* from ϕ_C to M and from M to λ), where one route ends at the growth rate, and the other at protein concentration:

$$T_{CM} T_{M\lambda} \left(S_{\pi_C} + S_s + T_{MC}^2 S_M + (-1)^2 S_\lambda \right).$$

- **Common mode**. This mode originates from phenomenological noise source that directly influences the growth rate, and the production rates (which in turn influences the concentrations):
$$T_{MC} T_{M\lambda} A_M.$$

- **Dilution mode**. This mode represents direct transfer from growth to protein concentration:
$$-A_\lambda.$$

s4.2.2   *The C-λ cross-correlations: New decomposition*

In the current work, we argue to define the modes slightly different. The major difference is in the interpretation of the *common* mode, which is now not defined as contributions to the cross-correlations originating from a particular noise source, but rather as all contributions that are sensed commonly by all proteins in the cell. Specifically, noise in the production of particular proteins can contribute to fluctuations of M, from where it will transfer to all proteins commonly. In other words, noise that originates in a particular part of the cell can propagate through the cell and partly *become* common noise, transferring further to all proteins equally. Additionally we introduce the regulation mode, which contains all

contributions to the cross-correlations that rely on the regulation and hence scale linearly with T_R.

With these conventions, the decomposition of the cross-correlation between ϕ_C and λ becomes as follows. For some terms we have added red brackets that divide the transfer parameters into specific groups of transfer parameters that are associated with a particular noise route through the cell (Figs. S4.1 and S4.2). From every noise source one can indeed follow two distinct paths, one to each of the variables for which the cross-correlation is calculated. However, sometimes noise routes are merged for clarity of reading. Then, the red brackets are omitted.

- $\chi_{(\phi_C,\lambda)}$:
    1. **Control**: $(T_{CM}T_{M\lambda})(S_{\pi_C} + S_s)$.
    2. **Dilution**: $-A_\lambda$.
    3. **Common**: $(T_{M\pi} - T_{M\lambda})T_{M\lambda}A_M + (T_{CM}T_{M\lambda})\left[T_{MC}^2 S_M + (-1)^2 S_\lambda\right]$.
    4. **Regulation**: $T_R T_{M\lambda} A_M$.

- $\chi_{(\pi_C,\lambda)}$:
    1. **Control**: $(T_{CM}T_{M\lambda})(A_{\pi_C}(-t) + A_s(-t))$.
    2. **Dilution**: 0.
    3. **Common**: $T_{M\pi}T_{M\lambda}B_M + (T_{CM}T_{M\pi})(T_{CM}T_{M\lambda})\left[(S_{\pi_C} + S_s + T_{MC}^2 S_M + (-1)^2 S_\lambda\right] + 2\frac{\lambda_E}{\lambda_0}T_{MC}T_{CM}T_{M\pi}T_{M\lambda}S_M - (T_{CM}T_{M\pi})A_\lambda$.
    4. **Regulation**: $T_R T_{M\lambda} B_M + (T_{CM}T_R)(T_{CM}T_{M\lambda})\left[S_{\pi_C} + S_s + T_{MC}^2 S_M + S_\lambda\right] + 2\frac{\lambda_E}{\lambda_0}T_{MC}T_{CM}T_R T_{M\lambda}S_M - (T_{CM}T_R)A_\lambda$.

The above equations for $\chi_{(\phi_C,\lambda)}$ and $\chi_{(\pi_C,\lambda)}$ apply to both wild type *E. coli* cells and the cAMP-fixed *cyaA cpdA* null mutant studied in the main text. This allows us to pinpoint the role of regulation in shaping the cross-correlation. The parameter T_R that represents the cAMP-CRP regulation feedback is expected to be negative in the wild type ($T_R < 0$ due to cAMP-CRP regulation). In the cAMP-fixed strain however, the negative feedback is abolished and $T_R = 0$. Such a parametric switch will qualitatively change the cross-correlations $R_{(\phi_C,\lambda)}$ and $R_{(\pi_C,\lambda)}$. For example, a negative contribution with functional form of A_M (asymmetrically, left-skewed function with timescale λ_0) is effectively removed from $R_{(\phi_C,\lambda)}$.

Note that setting $T_R = 0$ not only influences the presence of the regulation mode. Additionally, T_R also influences the coefficients of variation (equations S4.18 - S4.20), and amplitudes and timescales of the other modes via $\lambda_E := \lambda_0(1 - (T_R + T_{M\pi} - T_{M\lambda}))$. Most modes are therefore expected to show slight, quantitative differences between wild type and cAMP-fixed cells, but only the regulation mode will show a strong qualitative difference: it is completely absent in cAMP-fixed cells.

s4.2.3   *Summary of the new decomposition for the reporters*

C-sector reporter concentration: $\chi(\phi_Y, \lambda)$

1. **Control**: $(T_{CM} T_{M\lambda}) D_s^1$.

2. **Dilution**: $-A_\lambda^0$.

3. **Common**: $(T_{M\pi} - T_{M\lambda}) T_{M\lambda} A_M^0 + T_{CM} T_{MC} (T_{MC} T_{M\lambda} D_M^2 - D_\lambda^2)$
   $+ (T_{CM}(T_{M\pi} - T_{M\lambda}))(T_{CM} T_{M\lambda}) \left[D_{\pi_C}^3 + D_s^3 + T_{MC}^2 D_M^3 + D_\lambda^3 \right]$
   $+ (T_{MC} T_{CM}(T_{M\pi} - T_{M\lambda})) T_{M\lambda} D_M^1 + (T_{CM} T_{M\lambda}) D_\lambda^1$.

4. **Regulation**: $T_R T_{M\lambda} A_M^0 + (T_{CM} T_R)(T_{CM} T_{M\lambda}) \left[D_{\pi_C}^3 + D_s^3 + T_{MC}^2 D_M^3 + D_\lambda^3 \right]$
   $+ (T_{MC} T_{CM} T_R) T_{M\lambda} D_M^1$.

C-sector reporter production rate: $\chi(\pi_Y, \lambda)$

1. **Control**: $(T_{CM} T_{M\lambda}) A_s(-t)$.

2. **Dilution**: 0.

3. **Common**: $T_{M\pi} T_{M\lambda} B_M + (T_{CM} T_{M\pi})(T_{CM} T_{M\lambda}) \left[S_{\pi_C} + S_s + T_{MC}^2 S_M + S_\lambda \right]$
   $+ 2 \frac{\lambda_E}{\lambda_0} T_{M\pi} (T_{MC} T_{CM} T_{M\lambda}) S_M - (T_{CM} T_{M\pi}) A_\lambda$.

4. **Regulation**: $T_R T_{M\lambda} B_M + (T_{CM} T_R)(T_{CM} T_{M\lambda}) \left[S_{\pi_C} + S_s + T_{MC}^2 S_M + S_\lambda \right]$
   $+ 2 \frac{\lambda_E}{\lambda_0} T_R (T_{MC} T_{CM} T_{M\lambda}) S_M - (T_{CM} T_R) A_\lambda$.

Constitutive reporter concentration: $\chi(\phi_H, \lambda)$

1. **Control**: 0.

2. **Dilution**: $-A_\lambda^0$.

3. **Common**: $(T_{M\pi} - T_{M\lambda}) T_{M\lambda} A_M^0 + T_{CM} T_{MC} (T_{MC} T_{M\lambda} D_M^2 - D_\lambda^2)$
   $+ (T_{CM}(T_{M\pi} - T_{M\lambda}))(T_{CM} T_{M\lambda}) \left[D_{\pi_C}^3 + D_s^3 \right.$
   $\left. + T_{MC}^2 D_M^3 + D_\lambda^3 \right] + (T_{MC} T_{CM}(T_{M\pi} - T_{M\lambda})) T_{M\lambda} D_M^1 + (T_{CM} T_{M\lambda}) D_\lambda^1$.

4. **Regulation**: 0.

Constitutive reporter production rate: $\chi(\pi_H, \lambda)$

1. **Control**: 0.

2. **Dilution**: 0.

3. **Common**: $T_{M\pi} T_{M\lambda} B_M + (T_{CM} T_{M\pi})(T_{CM} T_{M\lambda}) \left[S_{\pi_C} + S_s + T_{MC}^2 S_M + S_\lambda \right]$
   $+ 2 \frac{\lambda_E}{\lambda_0} T_{M\pi} (T_{MC} T_{CM} T_{M\lambda}) S_M - (T_{CM} T_{M\pi}) A_\lambda$.

4. **Regulation**: 0.

s4.2.4   *Example: interpretation of the decomposition of the $\phi_Y - \lambda$ cross-correlations*

To clarify the interpretation of the expressions above, we now discuss the decomposition of the Y–λ cross-correlation function in detail.

1. **Catabolism mode**. This mode is a consequence of fluctuations in the expression level of the catabolic sector (C-sector) that directly influence the rate of metabolism and the growth rate. Even though the reporter Y is itself not metabolically active, it does share a noise source (N_s) with the C-sector. As a result, the (cross-)correlation $R_{(\phi_Y, \lambda)}$ also contains the catabolism mode, reflecting that ϕ_Y correlates with ϕ_C (due to the shared noise source N_s), and ϕ_C with λ (due to the effect that fluctuations in ϕ_C have on the growth rate).

$$T_{CM}T_{M\lambda}D_s^1.$$

Note that indeed only the shared noise source N_s transfers directly to growth from π_Y, instead of reverberating through metabolism (and therewith becoming 'common' noise). The function D^1 appears (instead of the function S as found in [86]) because noise transferring from metabolism, via production to the reporters (Y or H) has a delay timescale of λ_0, whereas noise transferring to ϕ_C has timescale λ_E. The logic behind the functions D^1, D^2, and D^3 is that: (D^1) noise transfers once via ϕ_C and once directly to ϕ_Y, obtaining timescales λ_E and λ_0 respectively; (D^2) one route is instantaneous, but the other makes a 'double' loop, affecting first ϕ_C and picking up timescale λ_E and then transferring to ϕ_Y, picking up timescale λ_0; (D^3) one route has timescale λ_E and the other route is 'double', picking up both λ_E and λ_0 timescales. For example, compare the two terms that originate at noise source N_M in the second line of equation S4.21. The first one can be written as $(T_{MC})(T_{MC}T_{CM}T_{M\lambda})D_M^1$ and has time scales λ_E for the route that transfers via metabolism and λ_0 for the route that goes from metabolism to ϕ_Y with parameter T_{MC} (Fig. S4.2A). The other term can be written as $(T_{MC}T_{CM}T_{MC})(T_{M\lambda})D_M^2$, and here one path is instantaneous ($T_{M\lambda}$) and the other transfers first from M to ϕ_C with timescale λ_E and afterwards to ϕ_Y with timescale λ_0 (Fig. S4.2B).

2. **Dilution mode**. This represents direct transfer from growth to protein concentration.

$$-A_\lambda^0.$$

3. **Common mode**.

$$(T_{M\pi} - T_{M\lambda})T_{M\lambda}A_M^0 + T_{CM}T_{MC}\left(T_{MC}T_{M\lambda}D_M^2 - D_\lambda^2\right)$$
$$+ (T_{CM}(T_{M\pi} - T_{M\lambda}))(T_{CM}T_{M\lambda})\left[D_{\pi_C}^3 + D_s^3 + T_{MC}^2 D_M^3 + D_\lambda^3\right]$$
$$+ (T_{MC}T_{CM}(T_{M\pi} - T_{M\lambda}))T_{M\lambda}D_M^1 + (T_{CM}T_{M\lambda})D_\lambda^1.$$

Common noise now also includes noise in the C-sector, and particularly its effect on the reporters after it reverberated through the cell. Here, one can for example examine the term $(-T_{CM}T_{M\lambda})(-T_{CM}(T_{M\pi} - T_{M\lambda}))D_\lambda^3$, where both the path that
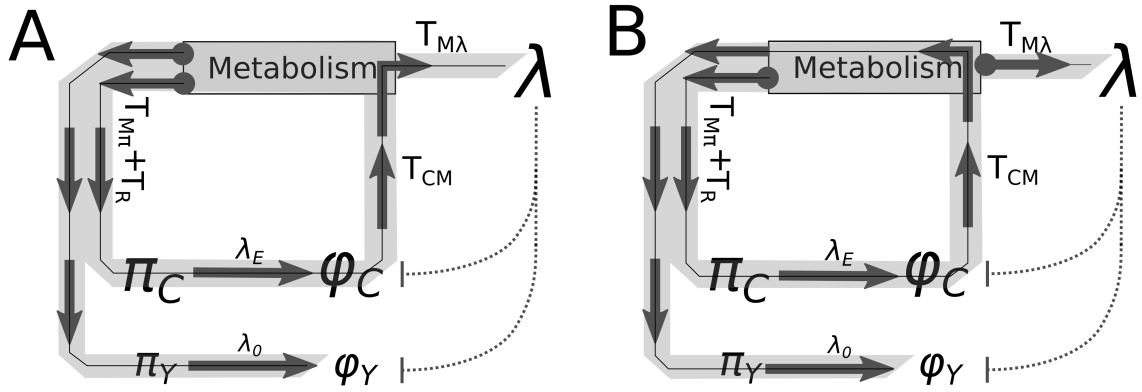
Figure S4.2: Example analysis of the noise routes (shown here is a part of the network.) (A) Example of a M-noise route that appears in the $\phi_Y - \lambda$ cross-correlation. Transfer coefficients can be split up to represent the paths to ϕ_C and to λ separately as: $(T_{MC})(T_{MC}T_{CM}T_{M\lambda})$. (B) Another example of a M-noise route: $(T_{MC}T_{CM}T_{MC})(T_{M\lambda})$.

ends up in λ, $(-T_{CM}T_{M\lambda})$, and the path that ends up in ϕ_Y, $(-T_{CM}(T_{M\pi} - T_{M\lambda}))$, make a route passing ϕ_C, picking up a timescale of λ_E and after that one route transfers to ϕ_Y, picking up a timescale λ_0. Another interesting example is the second term: $-T_{CM}(T_{M\pi} - T_{M\lambda})D_\lambda^2$. This represents noise in the growth rate that diluted ϕ_C, which has affected metabolism, but from there transfers equally to all other proteins (with coefficient $T_{M\pi} - T_{M\lambda}$). Therefore, this term is now also regarded as common noise.

4. **Regulation**

$$T_R T_{M\lambda} A_M^0 + (T_{CM} T_R)(T_{CM} T_{M\lambda})\left[D_{\pi_C}^3 + D_s^3 + T_{MC}^2 D_M^3 + D_\lambda^3\right] + (T_{MC} T_{CM} T_R)T_{M\lambda} D_M^1.$$

In the decompositions we often have the term $T_{M\pi} - T_{M\lambda}$. This represents the net transfer from M to ϕ via π and from M to ϕ via λ. Red brackets can help to identify the noise route.

In this section, we show that two parameters can be effectively scaled away from the system, resulting in fewer parameters to be fitted. First, since cross-correlations are dimensionless measures, we can set one noise amplitude to 1. We pick θ_s, since this noise source has to be present to find any difference between the two reporters in cAMP-fixed cells. Second, note that we do not measure M, so that we can scale away T_{CM}, and consider fluctuations in $\delta M/(T_{CM}M_0)$ with scaled noise source $N'_M = N_M/T_{CM}$.

Now, to do the parameter reductions formally, we start with the original system (equations S4.1-S4.8):

$$\frac{\delta\lambda}{\lambda_0} = N_\lambda + T_{M\lambda}\frac{\delta M}{M_0}, \tag{S4.39}$$

$$\frac{\delta M}{M_0} = N_M + T_{CM}\frac{\delta\phi_C}{\phi_{C,0}} = T_{CM}\left(\frac{N_M}{T_{CM}} + \frac{\delta\phi_C}{\phi_{C,0}}\right), \tag{S4.40}$$

$$\frac{\delta\pi_C}{\pi_{0,C}} = N_{\pi_C} + N_s + (T_{M\pi} + T_R)\frac{\delta M}{M_0}, \tag{S4.41}$$

$$\frac{\delta\pi_y}{\pi_{0,y}} = N_{\pi_Y} + N_s + (T_{M\pi} + T_R)\frac{\delta M}{M_0}, \tag{S4.42}$$

$$\frac{\delta\pi_H}{\pi_{0,H}} = N_{\pi_H} + T_{M\pi}\frac{\delta M}{M_0}. \tag{S4.43}$$

Defining $\delta\hat{M} = \frac{\delta M}{T_{CM}M_0}$, $\hat{N_M} = N_M/T_{CM}$, $\hat{T_{M\lambda}} = T_{CM}T_{M\lambda}$, $\hat{T_{M\pi}} = T_{CM}T_{M\pi}$ and $\hat{T_R} = T_{CM}T_R$. This system can then be transformed to:

$$\frac{\delta\lambda}{\lambda_0} = N_\lambda + T_{M\lambda}\left(T_{CM}\delta\hat{M}\right) = N_\lambda + \hat{T_{M\lambda}}\delta\hat{M}, \tag{S4.44}$$

$$\delta\hat{M} = \hat{N}_M + \frac{\delta\phi_C}{\phi_{C,0}}, \tag{S4.45}$$

$$\frac{\delta\pi_C}{\pi_{0,C}} = N_{\pi_C} + N_s + (T_{M\pi} + T_R)\left(T_{CM}\delta\hat{M}\right) = N_{\pi_C} + N_s + (\hat{T_{M\pi}} + \hat{T_R})\delta\hat{M}, \tag{S4.46}$$

$$\frac{\delta\pi_y}{\pi_{0,y}} = N_{\pi_Y} + N_s + (T_{M\pi} + T_R)\left(T_{CM}\delta\hat{M}\right) = N_{\pi_Y} + N_s + (\hat{T_{M\pi}} + \hat{T_R})\delta\hat{M}, \tag{S4.47}$$

$$\frac{\delta\pi_H}{\pi_{0,H}} = N_{\pi_H} + T_{M\pi}\left(T_{CM}\delta\hat{M}\right) = N_{\pi_H} + \hat{T_{M\pi}}\delta\hat{M}. \tag{S4.48}$$

Next, we divide all noise sources by θ_s, since all cross-covariances and all variances are linear in terms of θ_i^2 (the noise amplitude of each independent noise source i). Thus, if we scale all noise sources with θ_s, the resulting cross-correlations (cross-covariances divided by variance) do not change. How the parameter reduction, and the scaling by θ_s affects the functional form of the cross-correlation between signals A and B is shown below, where we split the cross-covariance and the variances in contributions of each noise source. We explicitly write the contribution of noise source N_M, associated with noise amplitude θ_M and only later replace its parameters with $\hat{\theta_M} := \theta_M/T_{CM}$. The contribution of each noise source i to the cross-covariance is written as f_i (time dependence is omitted for readability)

and we write g_A and h_B for the noise source's contribution to the variance of A and B respectively. Then:

$$
\begin{aligned}
R_{(A,B)}(\tau) &= \frac{\left(\frac{\theta_M}{T_{CM}}\right)^2 f_M + \sum_{i \neq M} \theta_i^2 f_i(\tau)}{\sqrt{\left(\frac{\theta_M}{T_{CM}}\right)^2 g_{A,M}(\tau) + \sum_{i \neq M} \theta_i^2 g_{A,i}}\sqrt{\left(\frac{\theta_M}{T_{CM}}\right)^2 h_{B,M} + \sum_{i \neq M} \theta_i^2 h_{B,i}}} \\[2ex]
&= \frac{\left(\frac{\theta_M/\theta_s}{T_{CM}}\right)^2 f_M + f_s + \sum_{i \neq s,M} (\theta_i/\theta_s)^2 f_i}{\sqrt{\left(\frac{\theta_M/\theta_s}{T_{CM}}\right)^2 g_{A,M} + g_{A,s} + \sum_{i \neq s,M} (\theta_i/\theta_s)^2 g_{A,i}}\sqrt{\quad \cdots}} \\[2ex]
&= \frac{\left(\hat{\theta}_M\right)^2 f_M + f_s + \sum_{i \neq s,M} \hat{\theta}_i^{\,2} f_i}{\sqrt{\left(\hat{\theta}_M\right)^2 g_{A,M} + g_{A,s} + \sum_{i \neq s,M} \hat{\theta}_i^{\,2} g_{A,i}}\sqrt{\left(\hat{\theta}_M\right)^2 h_{B,M} + h_{B,s} + \sum_{i \neq s,M} \hat{\theta}_i^{\,2} h_{B,i}}}.
\end{aligned}
$$
(S4.49)

Here, $\hat{\theta}_i = \theta_i/\theta_s$, and $\hat{\theta}_M = \frac{\theta_M/\theta_s}{T_{CM}}$. By considering only $\hat{\cdot}$ - parameters, we effectively removed two parameters from the model.

s4.4   DATA ANALYSIS

s4.4.1   *Calculating cross-correlations from data*

Segmentation was done using the software Schnitzcells, developed by the Elowitz lab [199], with custom scripts written by Daan Kiviet, Philippe Nghe and Noreen Walker. Tracking was done in line with Kiviet *et al* [86]. Cell length is used by fitting a 3rd (or, in some cases 5th) order polynomial through the cell area. Cross-covariance S and cross-correlation, R, between two signals in discrete time is then defined as:

$$
S_{f,g}(\tau) = \frac{1}{N-1} \sum_{n=0}^{N-|\tau|-1} \hat{f}(n)\hat{g}(n+\tau),
$$
(S4.50)

$$
R_{f,g}(\tau) = \frac{S_{f,g}(\tau)}{\sqrt{S_{f,f}(0)S_{g,g}(0)}}.
$$
(S4.51)

Here, hats indicate mean-subtracted signals.

Cells that are born earlier in the experiments appear in more lineages. When calculating cross-correlations along lineages, we must thus be careful to not count such cells repeatedly. Therefore, we introduce for each data point a weight, representing in how many branches the point occurs. The resulting cross-correlation is a composite cross-correlation with contributions of points from multiple branches i. Lastly, we also introduce time-

average-subtracted variables, since averages can change slightly during experiments. We therefore define a composite cross-covariance and cross-correlation:

$$\mathcal{S}_{f,g}(\tau) := \frac{1}{W_{total,\tau}} \sum_i \frac{1}{N_i - |\tau|} \sum_{n=0}^{N_i - |\tau|} w_{n,i,\tau} \hat{f}_i(n) \hat{g}_i(n + \tau), \tag{S4.52}$$

$$\mathcal{R}_{f,g}(\tau) = \frac{\mathcal{S}_{f,g}(\tau)}{\sqrt{\mathcal{S}_{f,f}(0) \mathcal{S}_{g,g}(0)}}, \tag{S4.53}$$

$$\hat{f}_i(n) = f_i(n) - \langle f_i \rangle_n, \tag{S4.54}$$

$$w_{n,i,\tau} = 1/K_{n,i,\tau}, \tag{S4.55}$$

$$W_{total,\tau} = \sum_{n,i} w_{n,i,\tau}. \tag{S4.56}$$

Here, the summations run are over all branches i and time points n. Weights are indicated with w, where $K_{n,i,\tau}$ is the total number a specific point pair $\hat{f}_i(n) \hat{g}_i(n + \tau)$ was used. Throughout the manuscript we refer to the composite cross-correlation $\mathcal{R}$ as the cross-correlation R. The mean-subtracted signal $\hat{f}_i(n)$ is now recalculated in each branch, for each time point to compensate for a changing overall average during the experiment.

s4.4.2 *Averaging multiple experiments and estimating error bars*

The cross-correlations calculated per microcolony are averaged to create Figs. 4.2B and 4.3D-E, see also Fig. S4.7. Here, we explain how we averaged the multiple (independent) experiments (each microcolony being an experiment) and how we calculated error bars. We use a Mixed Model Estimate, assuming that each individual measurement $y_{i,j}$ from experiment i is determined by the average of interest, $\langle y \rangle$, plus two noise sources: within experimental noise $\xi_j(i)$, and between-experiments noise ξ_i:

$$y_{i,j} = \langle y \rangle + \xi_i + \xi_j(i).$$

Here, ξ_i is the noise that determines the off-set of the mean from experiment i, and $\xi_j(i)$ the noise on individual measurement j in group i. Assume: $\mathbb{E}[\xi_i] = \mathbb{E}[\xi_j] = 0$ and $\mathrm{Var}\{\xi_j\} = s_i^2$, the variance within experiment i, which might differ between experiments, and $\mathrm{Var}\{\xi_i\} = s_\mu^2$, the variance between the means of each experiment. The within experiment variance is estimated by dividing each microcolony into four lineages (from the moment there were four cells in the microcolony, we followed each of their lineages separately), and calculate and compare statistics along each lineage. With these notions for ξ_i

and $\xi_j(i)$, we write n for the number of experiments and n_i for the number of subgroups within experiment i. Then:

$$\mathbb{E}\left[y_{i,j}\right] = \langle y \rangle, \tag{S4.57}$$

$$\text{Var}\left\{\mathbb{E}\left[y_{i,j}\right]\right\} = \frac{\text{Var}\left\{\sum_i \sum_j \langle y \rangle + \xi_i + \xi_j(i)\right\}}{\left(\sum_i n_i\right)^2} \tag{S4.58}$$

$$= \frac{1}{\left(\sum_i n_i\right)^2}\text{Var}\left\{\sum_i n_i \xi_i + \sum_i \sum_j \xi_j\right\} \tag{S4.59}$$

$$= \frac{1}{\left(\sum_i n_i\right)^2}\left(\sum_i n_i^2 \text{Var}\{\xi_i\} + \sum_i n_i \text{Var}\{\xi_j\}\right) \tag{S4.60}$$

$$= \frac{s_\mu^2 \sum_i n_i^2 + \sum_i n_i s_i^2}{\left(\sum_i n_i\right)^2}. \tag{S4.61}$$

To calculate $\langle y \rangle$, we again use knowledge of within-experiment variances to calculate the weight factor of each microcolony: let $y_i(t)$ be the measured mean value in experiment i at time t, with within-experiments error $s_i(t)$. Then:

$$\langle y(t) \rangle = \frac{\sum_i^n s_i^{-2}(t) y_i(t)}{\sum_j^n s_j^{-2}(t)} =: \frac{\sum_i w_i(t) y_i(t)}{W(t)}.$$

Here, $w_i(t) = 1/s_i^2(t)$ and $W(t) = \sum_{j=1}^n w_i(t)$. That is, more precise measurements (those with smaller within-experiment error s_i) obtain a higher weight.

s4.4.3 *Null-expectation for the cross-correlations*

To confirm that the measured cross-correlations correspond to real biological signals inside the cells, we performed a permutation analysis on the time-series data. We kept the temporal information of the data, but randomised at each time point the growth rate and expression data for all the cells in the colony. Any biological correlations between variables should therewith be removed. Repeating this randomisation 50 times, and each time re-calculating cross-correlations, indeed gives a band of cross-correlations around zero, allowing us to infer what kind of signals could still be explained purely by technical noise (see for example Fig. S4.9). Any part of the originally measured cross-correlations that fall outside this band can then be concluded to stem from a real biological signal.

s4.5 FITTING PROCEDURE

s4.5.1 *Wild type and cAMP-fixed* cells*

The full mathematical model, with the reduced number of parameters, was fitted to the cross-correlations with their error bars using a weighted least-square fitting procure in Mathematica. We fitted $R_{(\phi,\lambda)}$ and $R_{(\pi,\lambda)}$ for both reporters and for wild type and cAMP-fixed* cells simultaneously.

In all fits, we set $\theta_{\pi_C} = 0$, for three reasons. First, the variable ϕ_C represents the entire C-sector, so that random, intrinsic fluctuations in the total size of this sector are expected

Parameter	Confined to	Best Fit	95% Confidence Interval
θ_{π_C}	0	-	-
θ_s	1	-	-
T_{CM}	1	-	-
T_R	$[-2,2]$; 0	**-1.67**	[-2.28,-1.25]
$T_{\hat{M}\lambda}$	$[-1,1]$	**0.05**	[0.045, 0.54]
$T_{\hat{M}\pi}$	$[-1,1]$	**-0.179**	[-0.22, -0.14]
$\hat{\theta}_{\pi_Y}$	$[0,5]$	**0.0**	[0,0.74]
$\hat{\theta}_{\pi_h}$	$[0,5]$	**1.14**	[1.06,1.24]
$\hat{\hat{\theta}}_M$	$[0,10]$	**0.90**	[0.83,0.97]
$\hat{\theta}_\lambda$	$[0,5]$	**0.13**	[0.125,0.143]
β_π	>2	**2**	[1.8, 2.16]
β_μ	>0.5	**0.78**	[0.75,0.82]

Table S4.1: Best fit parameters of the wild type and optimal mutant, including predefined parameter constraints and best-fit confidence ranges based on statistical analysis. Parameter T_R was set to 0 for *cAMP-fixed* (MUT) cells. All other parameters were fitted using a minimisation of the squared distance of 8 analytical curves (cross-correlations $\phi - \lambda$ and $\pi - \lambda$ for both reporters. and for *WT* and *cAMP-fixed** cells) to the cross-correlation data (Fig. 4.2B from the main text). For the precise interpretation of $\hat{\cdot}$-parameters, see section 'Parameter Reduction'.

to be small. Second, any noise source that directly influences the total size of the C-sector should also affect the CRP-reporter Y, because C and Y are regulated and expressed similarly. Such noise sources are therefore also captured by the phenomenological noise source N_s that summarises shared noise sources between the C-sector and Y. Third, fluctuations in the concentrations of each of the individual proteins of the C-sector likely transfer differently to metabolism than joined fluctuations of the entire C-sector. The effect of the intrinsic fluctuations of individual C-sector protein species on the metabolic flux is instead captured by the noise source N_M. Fluctuation in any individual C-sector protein could indeed potentially influence the flux catalysed by the entire C-sector, causing fluctuations in metabolism.

For cAMP-fixed* cells we predefined that $T_R = 0$, and for wild type cells $-2 < T_R < 2$. All other parameters were not allowed to differ between WT and cAMP-fixed* cells. We thus fitted 9 free parameters: $\{\hat{T}_R$ (only for WT cells), $T_{\hat{M}\lambda}, T_{\hat{M}\pi}, \hat{\theta}_{\pi_Y}, \hat{\theta}_{\pi_h}, \hat{\hat{\theta}}_M, \hat{\theta}_\lambda, \frac{\beta_\pi}{\lambda_0}, \frac{\beta_\mu}{\lambda_0}\}$. (The hat-parameters are defined in the section 'Parameter Reduction', section S4.3; T_{CM} and θ_s, are set to unity and removed from the model by scaling.) The timescale β_π is the time-scale of the production noise rates, and β_μ is the timescale of growth/dilution-related noise sources (θ_s, θ_M, θ_λ), their value given is relative to the mean growth rate, λ_0.

In table S4.1 we present best-fit parameters, with their 95% confidence interval. The 95% confidence ranges of the parameters were estimated by changing that parameter until the increase in (weighted) sum of squared residuals was statistically significant (as determined by using F-statistics).

Although the model is able to reproduce the experimentally observed cross-correlations, we are hesitant to over-interpret the exact numerical values of the fitted parameters. For example, the wild type and optimal mutant's best-fit-parameter for the transfer from M to π, $\hat{T}_{M\pi}$, is small, but negative (Table S4.1). Such a negative parameter is counter-intuitive, for it results in a negative common mode (fluctuations in metabolism increase the growth rate, but lower the production rates). Presence of this negative common mode is only reflected in the mutant's negative π_H-λ cross-correlation, and, importantly, it is also this negative cross-correlation that causes the best-fit parameter to become negative; *i.e.* not including $R_{(\pi_H, \lambda)}$ in the fitting procedure causes the best-fit value for $\hat{T}_{M\pi}$ to become slightly positive (data not shown). Extensions of the model with alternative mechanistic explanations that could explain the negative correlation between π_H and λ, such as competition for ribosomes at the single-cell level, or a negative control from the reporters, did also not fit the negative cross-correlation well (data not shown; Mathematica notebooks available on: https://github.com/Jintram/DynamicalRegulationBacterialCells). Possibly, this specific negative cross-correlation could be caused by an experimental artefact that heavily influences the fitted parameters. Indeed, similar constitutive reporters measured in earlier work [86, 167] show a positive correlation between π_H and λ.

The model can predict preliminary coefficients of variation for the wild type and mutant cells, gaining some preliminary insight in whether cAMP-CRP feedback potentially plays the role of quenching noise. Filling the parameters from Table S4.1 into Equations S4.15 and S4.16, we find:

$$\eta^2_{\phi_C, WT}/\eta^2_{\phi_C, cAMP-fixed*} = 0.85, \tag{S4.62}$$

$$\eta^2_{\lambda, WT}/\eta^2_{\lambda, cAMP-fixed*} = 0.92. \tag{S4.63}$$

However, for the reported noise amplitudes in table S4.1 the absolute values for the CVs do not both seem realistic, *i.e.*, $\eta^2_{\phi_C, WT} \approx 0.3$ and $\eta^2_{\lambda, WT} = 0.014$. Still, this analysis might suggest a role for the cAMP-CRP mechanism to quench noise.

From the best-fit parameters above (Table S4.1), we qualitatively reproduced the Low cAMP cross-correlations by changing the transfer parameters to $\{T_R = 0, \hat{T}_{M\lambda} = 0.6, \hat{T}_{M\pi} = 0.45\}$. The timescales of each noise source relative to the growth rate (which is lower under this condition than under optimal cAMP levels) were kept constant. Although the resulting cross-correlations reproduce many qualitative features of the measurements, the mathematical model over-estimates the overall amplitude of the cross-correlation (Fig. S4.6A). This could perhaps be explained by experimental error: Independent measurement noise in any two variables reduces their correlation. In the model, such decorrelation can be mimicked by increasing the noise levels of θ_{π_Y} and θ_{π_H} to $\{5, 3.5\}$ respectively. An increase in those noise parameters does not affect the shape of the cross-correlation, but only decreases the overall amplitude of the cross-correlation (decorrelation).

The cross-correlation between ϕ_C and λ in the high-cAMP regime is hypothesised in the main text to posses a negative catabolism mode. When ϕ_C is over-expressed, the metabolic flux might still increase very slightly with an upward fluctuation of ϕ_C. However, the flux *per invested protein* will decrease. Since π and λ are relative quantities (both scaled with total protein mass), the negative control of ϕ_C effectively results in negative transfer from M to λ and from M to π.

Setting $\{T_R = 0, \hat{T}_{M\lambda} = -0.8, \hat{T}_{M\pi} = -0.47\}$ while keeping all other parameters as in Table S4.1) indeed qualitatively explained many featured of the measured cross-correlations

(Fig. S4.6B). However, relative to the mean growth rate, the experimentally measured cross-correlations decay faster than in the other conditions. Additionally, the high-cAMP cross-correlations are again over-estimated by the model. Another clear mismatch is that, for these parameters, the model predicts a strong common mode (which can be seen most clearly in the π_H-λ cross-correlation) that does not seem to be present in the data (although one microcolony did show a clear positive π_H-λ cross-correlation, see bottom right panel in Fig. S4.7). Interestingly, just as in the wild type, the cross-correlations of both reporters look similar in cAMP-fixed cells under the high-cAMP condition. Possibly, the noise amplitude of the CRP protein (N_s) is lower under this condition, such that the C-sector reporter is less correlated with the C-sector and behaves more similar to the constitutive reporter. A semi-quantitative fit can made (see also Fig. 4.3) by tuning the parameters further: $\left\{ \hat{T_{M\lambda}} = -0.6, \hat{T_{M\pi}} = -0.25, \hat{\theta}_{\pi_Y} = 2.5, \hat{\theta}_{\pi_h} = 4, \hat{\theta}_M = 1.8, \hat{\theta}_\lambda = 0.8, \beta_\pi = 2\lambda_0, \beta_\mu = 1.2\lambda_0 \right\}$. This corresponds with a decrease in θ_s of roughly a factor 2, and with a slight increase in θ_λ, also roughly a factor 2.

S4.6   TOY MODEL OF THE MEANS OF THE TWO REPORTERS

From the observation that the sum of the concentration of both reporters is, on average (Fig. S4.10A), approximately constant, we were inspired to also write equations for the population-level average behaviour. We here derive a phenomenological toy model that describes how the average concentrations, production rates and growth rates of cAMP-fixed cells change under changing external cAMP concentration.

Generally, we can write for the (average) total concentration of the catabolic sector ϕ_C and for the constitutive reporter H:

$$\frac{\partial \phi_C}{\partial t} = \pi_C - \phi_C \lambda = (f_C(x) - \phi_C)\lambda, \tag{S4.64}$$

$$\frac{\partial \phi_H}{\partial t} = \pi_H - \phi_H \lambda = (f_H - \phi_H)\lambda. \tag{S4.65}$$

Here $f_C(x)$ is a regulation function that determines the fraction of all metabolic flux allocated towards production of ϕ_C as a function of some internal metabolite (in this case, x reflect the internal cAMP concentration). However, f_H is also not necessarily a constant and can depend on resource allocation. We will show below that f_H is indeed not constant.

From the experiments we observe that $\phi_Y + \phi_H$ is approximately a constant. Assuming that Y is a good reporter of the C sector, this is equivalent to:

$$\phi_C + \phi_H = T.$$

(Note that this ignores proteomic shifts that result from a changes in the ribosomal sector, or any other sector that is not modelled here.) In the steady state, $\phi_H = f_H$, and therefore also $\pi_H = \phi_H \lambda = (T - f_C(x))\lambda = f_H \lambda$. We can thus rewrite the differential equation for H as:

$$\frac{\partial \phi_H}{\partial t} = (T - f_C(x))\lambda - \phi_H \lambda.$$

In steady state, this suggest that the production rates (as directly quantified from the experiments), divided by the growth rates should be equal to the concentration for both the reporters (Fig. S4.10A).

In the mutant, the growth rate declines when ϕ_C is either over, or under-expressed, such that the mean growth rates fit a 2nd order polynomial nicely:

$$\lambda(\phi_C) := g_\lambda(\phi_Y) = a\phi_Y^2 + b\phi_Y + c,$$

for best-fit parameters $\{a = -3.94 \cdot 10^{-5}, b = 2.12 \cdot 10^{-2}, c = -2.02\}$. Here we consider ϕ_C, the metabolic sector, well-reported by the metabolic reporter ϕ_Y. Using this polynomial, the relationships between mean protein production rates/mean protein concentrations, and the mean growth rate can be related (Figs. S4.3 and S4.10):

$$\{\phi_Y, \lambda\} = \{\phi_Y, g_\lambda(\phi_Y)\}, \tag{S4.66}$$

$$\{\pi_Y, \lambda\} = \{\phi_Y g_\lambda(\phi_Y), g_\lambda(\phi_Y)\}, \tag{S4.67}$$

$$\{\phi_H, \lambda\} = \{\phi_H, g_\lambda(T - \phi_H)\}, \tag{S4.68}$$

$$\{\pi_H, \lambda\} = \{\phi_H g_\lambda(T - \phi_H), g_\lambda(1 - \phi_H)\}, \tag{S4.69}$$

$$\{\pi_Y, \pi_H\} = \{\phi_Y g_\lambda(T - \phi_H), \phi_H g_\lambda(T - \phi_H)\}. \tag{S4.70}$$

These relationships match the experiments strikingly well (Fig. S4.10).
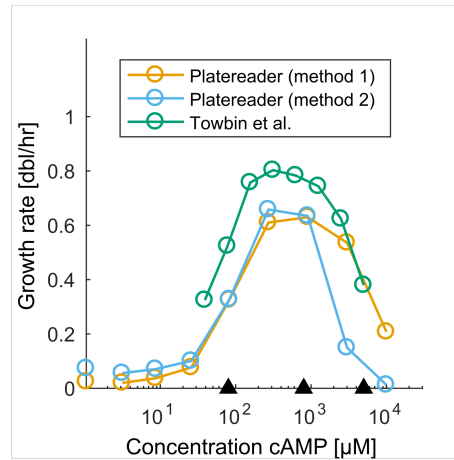
Figure S4.3: Growth rates in minimal medium supplemented with lactose and various cAMP concentrations. Measured exponential phase growth rates of the *cyaA cpdA* null mutant (cAMP-fixed cells) at different concentrations of cAMP as measured in a plate reader. Black triangles refer to the low, optimal and high cAMP concentrations respectively. Method 1 refers to a procedure where the growth rate was determined as an exponential fit over a manually selected part of the bacterial density curve, whereas in method 2 we fitted an exponential to the part of the bacterial density curve that fell between two threshold values. Additionally, this figure shows data from a similar experiments performed by Towbin *et al* [181].
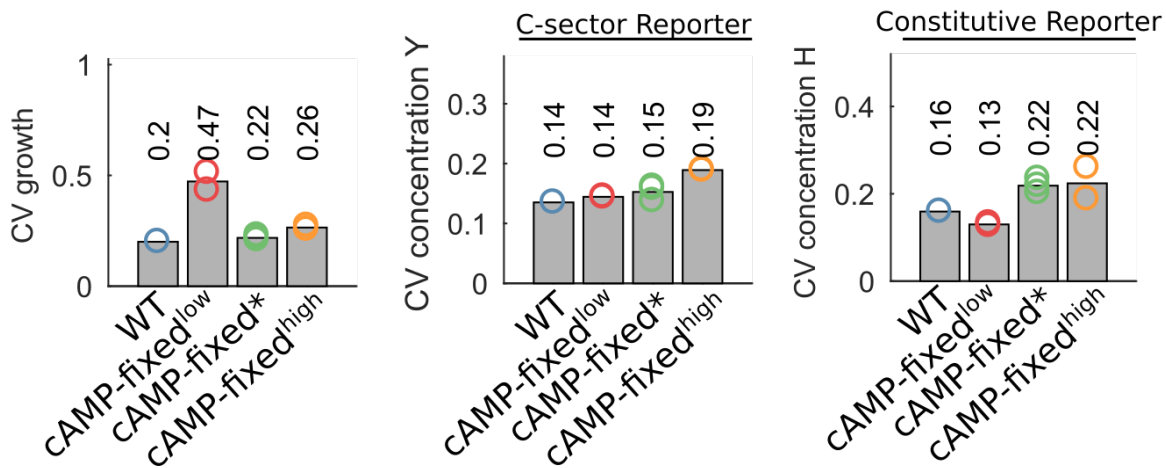
Figure S4.4: Coefficients of variation of the growth rate (left panel) and the concentrations of the C-sector reporter (middle panel) and the constitutive reporter (right panel) for the different conditions. Shown in the figure are only the experiments performed with a similar microscope such that their absolute values were comparable.
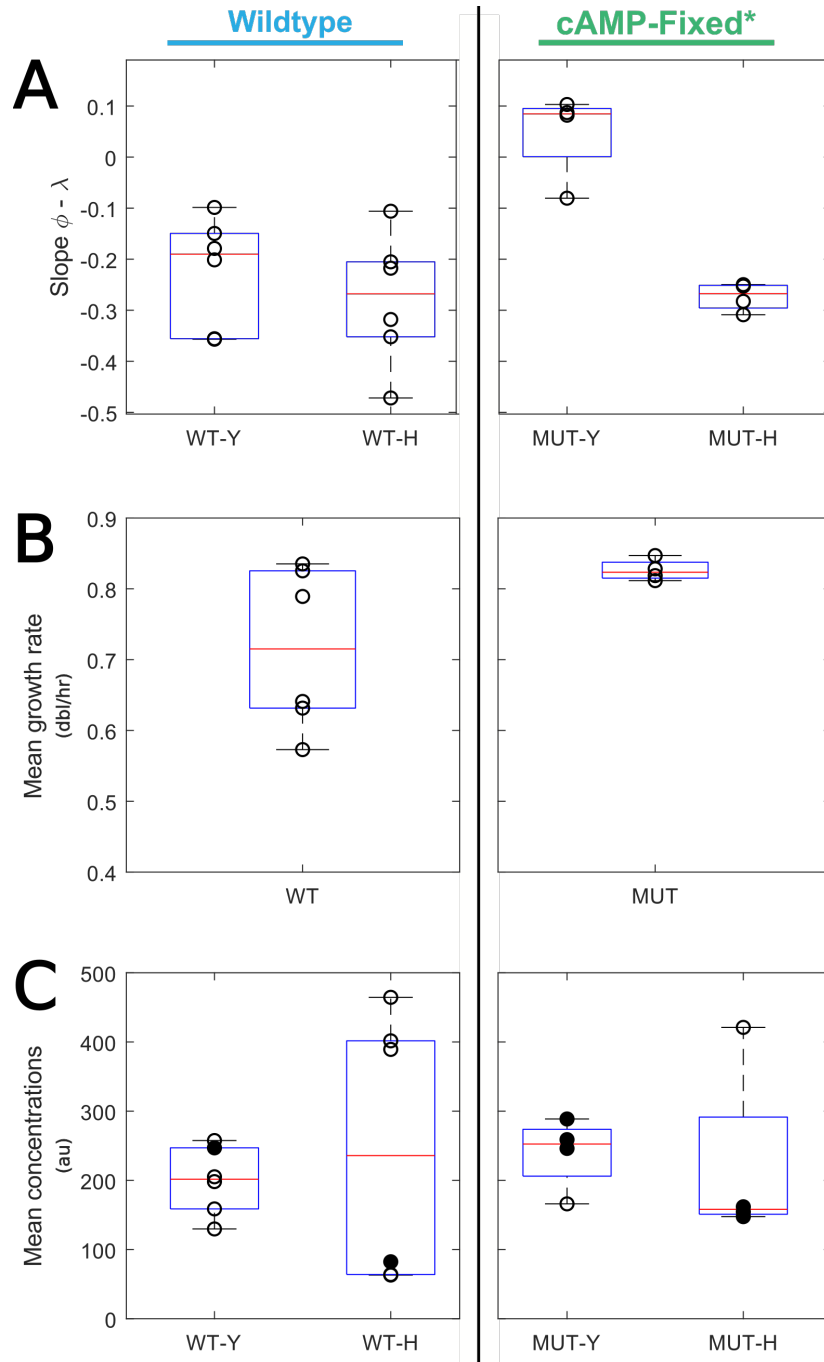
Figure S4.5: Average experimental values measured in different colonies. (**A**) Regression slope between φ and λ for the wild type and cAMP-fixed* cells (MUT on the figure axis). Slopes for the Y reporter differ significantly between wild type and mutant cAMP-fixed* cells (p = 0.0031, two-sample t-test), but not for the H reporter (p = 0.93, Welch's t-test). Note that to calculate these regression slopes only relative fluctuations are relevant, so that only relative fluorescence signals are relevant. (**B**) Average growth rate per colony, showing a large variance in growth-rate measurements. Difference between the mean growth rates of WT and cAMP-fixed* cells is not significant (p = 0.063, Welch's t-test). (**C**) Average fluorescence per colony. Filled circles are values measured with a standardised microscope setting (and thus only those absolute values can be compared). Y: C-sector reporter, H: constitutive reporter.
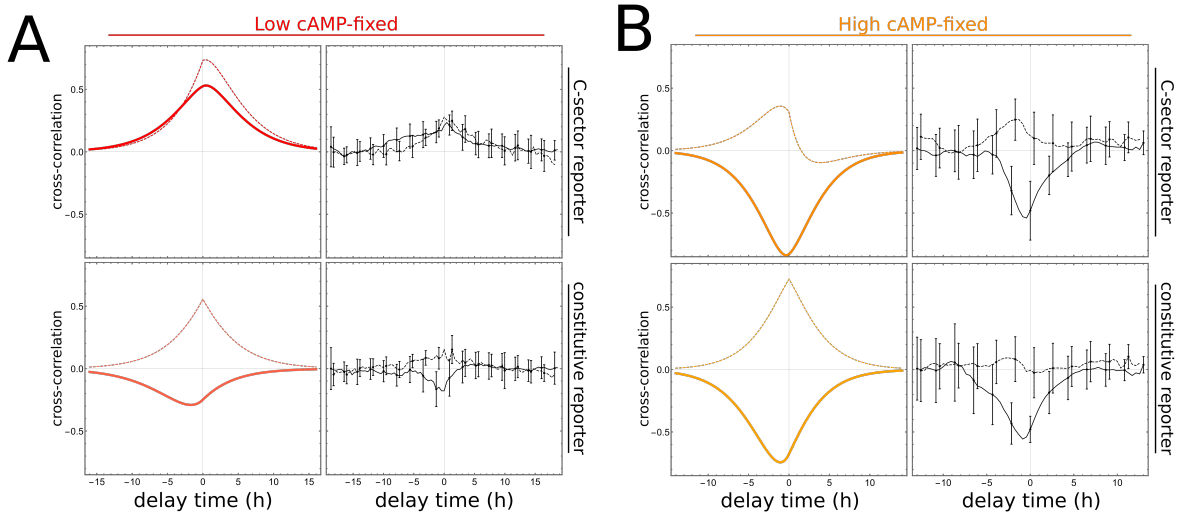
Figure S4.6: Qualitative model prediction for the cross-correlations as measured in the cAMP-fixed cells using low-cAMP (80 μM, panel A) and high-cAMP (5000 μM, panel B) conditions, together with the measured cross-correlations. (A) Only the transfer parameters from M to λ ($\hat{T_{M\lambda}}$ = 0.6) and from M to π ($\hat{T_{M\pi}}$ = 0.45) differ compared to the best-fit values from the wild type/cAMP-fixed* condition. (B) Same, but now $\hat{T_{M\lambda}}$ = −0.83, and $\hat{T_{M\pi}}$ = −0.47.

Identifier	Manuscript shorthand	Description
ASC838		Wild type MG1655, also known as strain bBT12 and CGSC number 8003. Known mutations: λ−, Δfnr-267, rph-1. (No resistance modules.)
ASC839		*cyaA*, *cpda* null mutant. Also known as strain bBT80. Based on ASC838. (No resistance modules.)
ASC990	wild type (WT)	Wild type strain, except for Δ(galk)::nCRPr-mCerulean-kanR and Δ(intc)::CRPr-mVenus-cmR. (Kanamycin and chloramphenicol resistant.)
ASC1004	cAMP-fixed	Strain based on ASC839 (ΔcyaA Δcpda), introduced Δ(galk)::s70-mCerulean-kanR and Δ(intc)::rcrp-Venus-cmR. (Kanamycin and chloramphenicol resistant.)

Table S4.2: Additional details on the strains that were used; ASC990 and ASC1004 were used in the manuscript, and are based on ASC838 and ASC839 respectively.
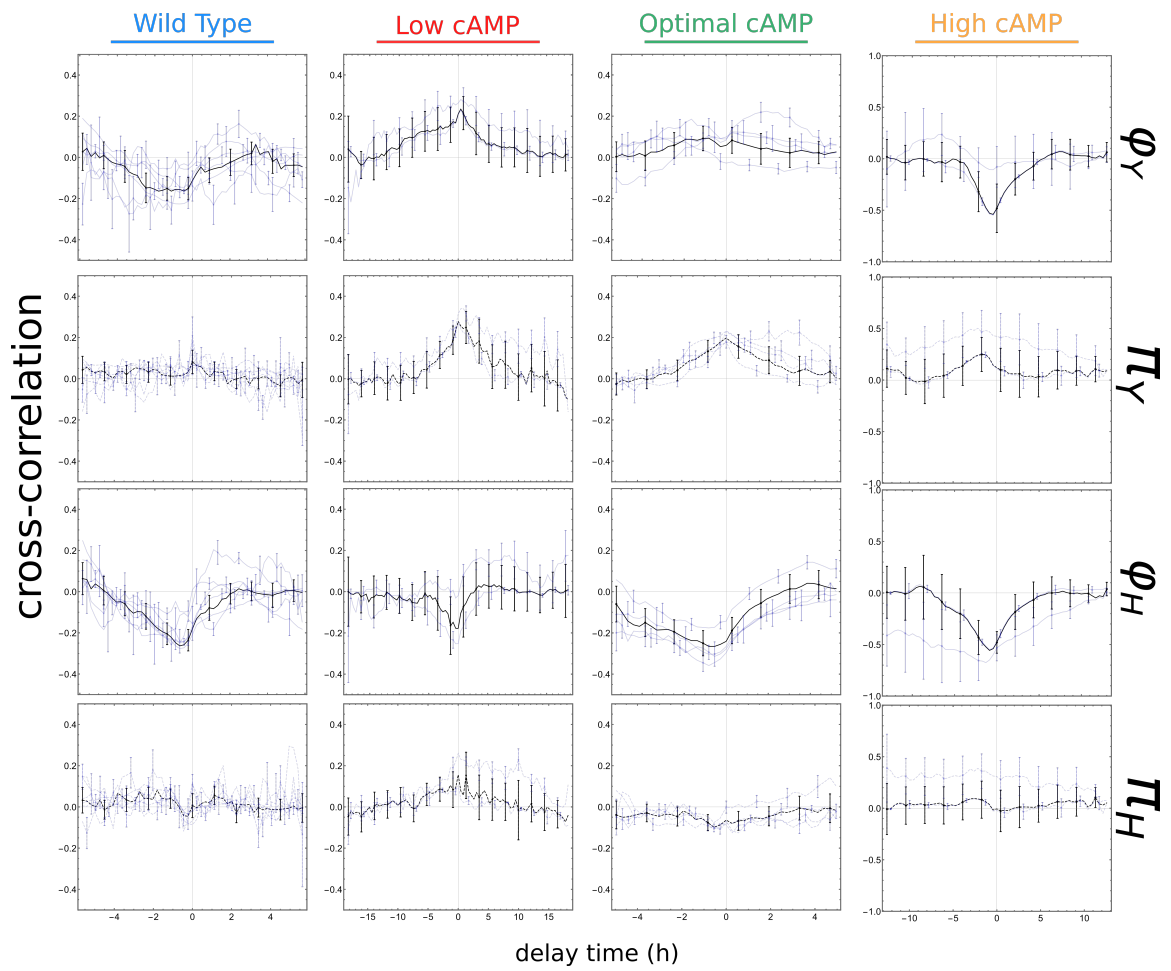
Figure S4.7: All measured cross-correlations (thin blue lines) in independent replicates (independent colonies), together with their weighted averages (thick black lines), for all conditions. Y: C-sector reporter, H: constitutive reporter.
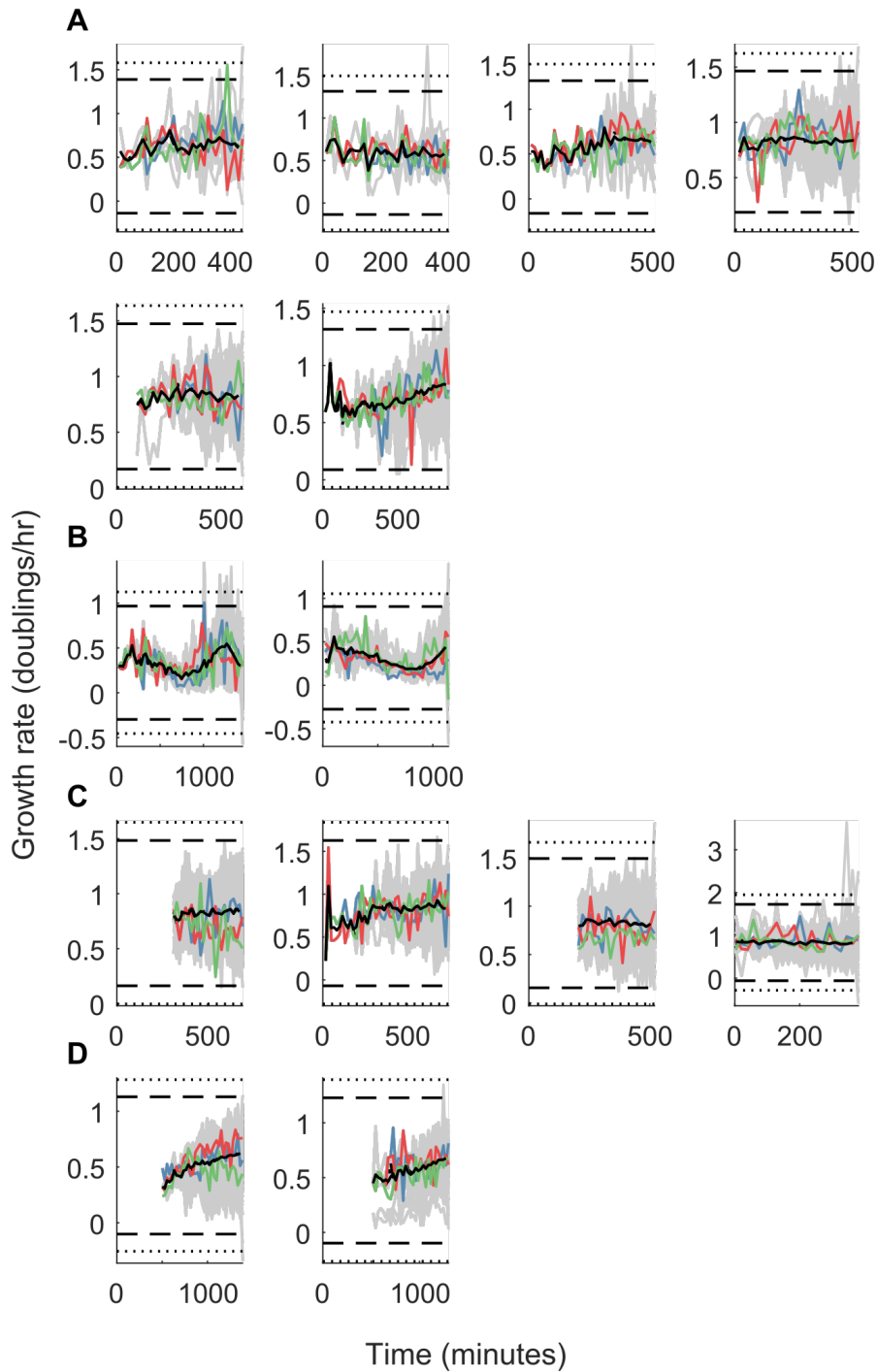
Figure S4.8: Growth rates during the experiments. Each panel plots growth-rate data for a single colony; panels are grouped by growth condition. The grey lines show single lineage traces, the black lines the population average. Coloured lines highlight example single lineage traces to illustrate single cell behaviour. Dashed and dotted lines indicate 4σ and 5σ boundaries from the overall mean respectively. As before, the displayed conditions are (A) wild type cells, (B) cAMP-fixed* cells growing on 80 μM cAMP, (C) cAMP-fixed cells growing on 800 μM cAMP and (D) mutant cells growing on 5000 μM cAMP.
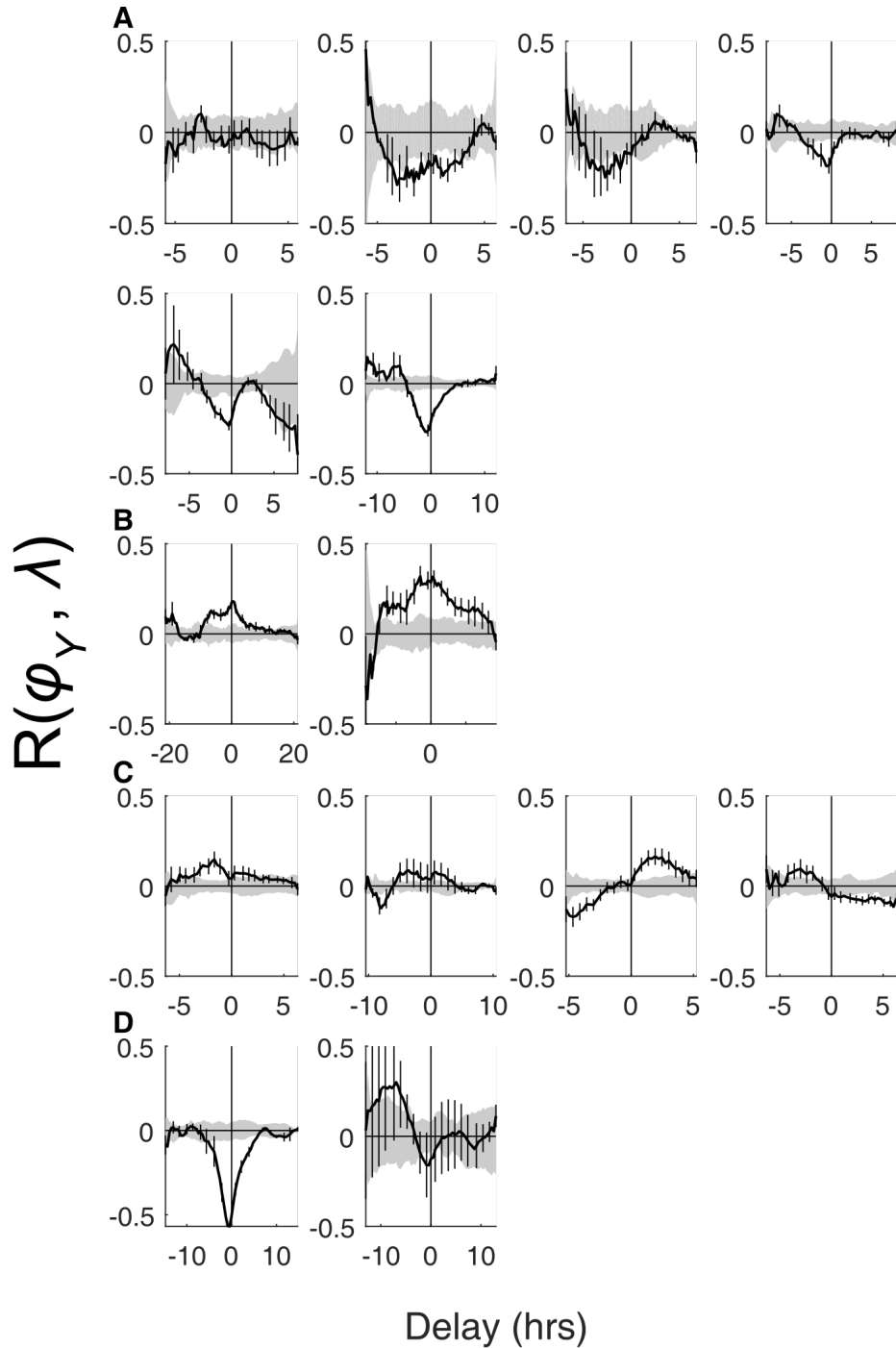
Figure S4.9: Cross-correlations between the C-sector reporter, ϕ_Y, and λ, together with their null expectation (gray areas around 0, see section S4.4.3 for details of the calculation). The black lines in this figures are the light-blue lines in top panels of Figure S4.7. Error bars are calculated by dividing each microcolony into four parts and comparing statistics in each part. As before, the displayed plots are from independent microcolonies growing under the following conditions: (A) wild type cells, (B) cAMP-fixed cells growing on 80 μM cAMP, (C) cAMP-fixed cells growing on 800 μM cAMP and (D) cAMP-fixed cells growing on 5000 μM cAMP.
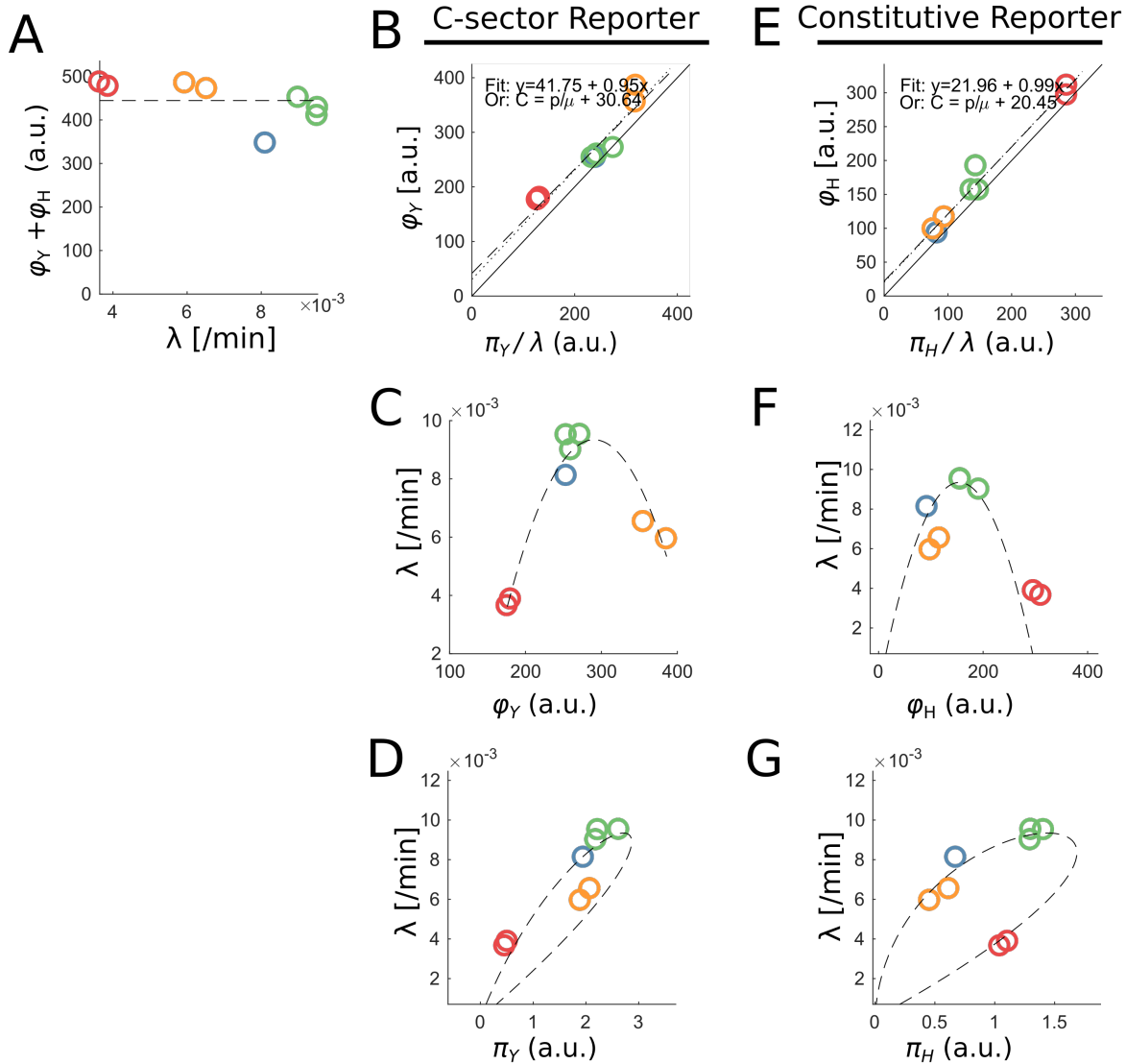
Figure S4.10: Toy model fits the mean behaviour of the reporters and the growth rate (see section S4.6 for details). (A) The sum of reporter concentrations is approximately constant in all conditions. (B) Steady state relationship $\phi_Y = \pi_Y/\lambda$ (black straight line) holds closely for all conditions. The best fit (dashed line), however, has a slight offset. (C) Fitted parabolic relationship of ϕ_C between the growth conditions. See also S4.3). (D) Relationship between π_Y and λ as calculated from the toy model (S4.6). (E) Steady state relationship for the C-sector reporter. (F-G) The C-sector reporter concentrations and production rates for each condition fall on the curve calculated from the toy model (not fitted). Colour coding is as in other figures (blue: wild type, red: low cAMP mutant, green: medium cAMP mutant, orange: high cAMP mutant). Y: C-sector reporter, H: constitutive reporter.

```
PLAC.   GGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTC

CRPr.   -----CGTCAGGAGGAGAGGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTC

nCRPr.  -----CGTCAGGAGGAGAGGGGCAGTGAGCGCAACGCAATCAGATCAAATGTGTCGTTTC


pLAC.   ATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGA

CRPr.   ATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGCATGGATAA

nCRPr.  CATAGGCACCCCAGGCTTGACACTTTATGCTTCCGGCTCGTATAATGTGTGCATGGATAA


pLAC.   GCGGATAACAATTTCACACAGGAAACAGCTATG

CRPr.   GTAGCTAGGAATTTCACACTGCAAACAGCTATG

nCRPr.  GTAGCTAGGAATTTCACACTGCAAACAGCTATG
```

Figure S4.11: Overview of promoter sequences used in this manuscript. The top row indicates the original LacZ upstream region including start codon ATG (NCBI; gene ID 945006, NC_000913.3), whilst the 2nd and 3rd row give the sequence of the engineered CRPr and nCRPr promoters. Colour indicates CRP binding sites according to [67] (yellow), [102] (green) or both (purple), and the LacI binding site (blue) according to [67]. In grey, changes in the engineered promoters are indicated.

<div style="text-align: right; font-size: 4em;">5</div>

BRIDGING POPULATION DYNAMICS AND SINGLE CELL STOCHASTICITY

Laurens H.J. Krah, Rutger Hermsen

ABSTRACT

With a surprisingly small number of regulatory molecules, the bacterium *Escherichia coli* regulates large groups of proteins in cohort (so-called 'protein sectors') to adjust to changing external carbon conditions. Numerous studies have illuminated how *E. coli* populations distribute their resources between the various protein sectors, and what the resultant growth rates are. However, the regulation of the protein sectors at the population level does not provide the full picture. At the single-cell level, stochasticity is present in the expression levels of all protein species, even in a fixed environment. Such noise in gene expression propagates, via metabolism, through the cell and to the instantaneous cellular growth rate. Noise is therewith also likely to arrive at those molecules that regulate the expression of protein sectors. This reveals a fundamental link between regulation and noise propagation: the cell's (internal regulatory) networks that govern the dynamical changes in response to a changing external condition, also shape noise propagation properties. In this chapter, we derive a course-grained stochastic model of a growing cell that uses a single internal variable to regulate its catabolic (C) and ribosomal (R) protein sectors. The model can simulate the cell's dynamic response to both external and internal fluctuations. Under the assumption that regulation optimises protein allocation for fast growth, all regulatory network parameters are fixed by fitting the model's deterministic behaviour to population-level data. Then, we show that many features of the resultant stochastic dynamics qualitatively match observation in single-cell data.

The contents of this chapter reflect the current state of this research.

5.1 INTRODUCTION

In the previous chapter, we established that the stochastic dynamics of metabolism, gene expression and growth are affected by regulation. Noise in metabolite concentrations was shown to propagate, via the cAMP-CRP regulatory network, to the expression level of catabolic proteins and therewith eventually affect the growth rate. Using a stochastic mathematical model, we studied the dynamical interplay between regulation, gene expression and growth under four conditions: wild type *Escherichia coli* cells, and *E. coli cyaA cpdA* null mutant cells ("cAMP-fixed") with, on average, a too low, a too high, or an optimal C-sector expression. For each condition, the linear model required different noise transfer parameters to explain the observed dynamics.

However, the noise transfer parameters in each condition are themselves a consequence of the (regulatory and metabolic) networks embedded in the cells. The same networks do not only shape the stochastic dynamics of regulation and growth that we aimed to describe, but also ensure a proper steady state (average expression and mean growth rate) is obtained to match the external nutrient conditions. In other words, regulatory properties, steady-state properties, and noise propagation properties are all different aspects of the same dynamical system. In this chapter, we therefore set out to derive a single simple model that can explain these three aspects simultaneously.

Luckily, gene regulation and growth rates of *E. coli* are well studied at the population level. In reaction to the external nutrients encountered, bacteria regulate large parts of their proteome in order to promote growth. As discussed in the Introduction, *E. coli* adjust many protein species with a similar function in cohort. These cohorts are often referred to as proteome sectors. For steady state growth, an incredible wealth of data is available concerning the expression levels of catabolic proteins (C-sector, ϕ_C, Fig. 5.1A) as well as ribosome-associated proteins and RNA (R-sector, ϕ_R, Fig. 5.1B). Additionally, the regulatory mechanisms that control C- and R-sector expression are well studied. Expression of catabolic proteins is governed by the allosteric binding of the metabolite cAMP to the transcription factor CRP [198] (see Introduction and Chapter 4). Ribosomal expression is inhibited by the secondary messenger (p)ppGpp, which in turn is synthesised under amino-acid starvation [204]. Vice versa, high amino-acid concentrations send a positive expression signal to ribosomal expression. Therewith, regulation of the ribosomes via (p)ppGpp is the likely mechanistic origin of the R-line; the observed positive relationship between ribosomal proteins and growth rate under variation of the nutrient condition (Fig.5.1B).

In summary, both the C- and the R-sector receive input from internal metabolite concentrations that are in turn affected by the 'metabolic state'. Although the sectors are controlled by distinct molecules (cAMP or (p)ppGpp), both sectors correlate strongly with a single variable, the growth rate. In fact, both the C- and R-sector change whenever the growth rate changes, either by varying the nutrient supplied in the medium, or for example by inhibiting the translation rate through antibiotics [158]. Yet, the growth rate itself is not the regulatory agent, itself being a result of the internal wiring on the cell. Taken together, a likely hypothesis is that the internal signals which determine C-sector expression, R-sector expression, and the growth rate, are strongly correlated. Possibly, the regulation of both sectors can thus be modelled with a single, abstract variable that signals the metabolic state. Using experiments and models, two recent papers (Towbin *et al*, 2017 [181], and Erickson *et al*, 2017 [43]) underlined this idea.
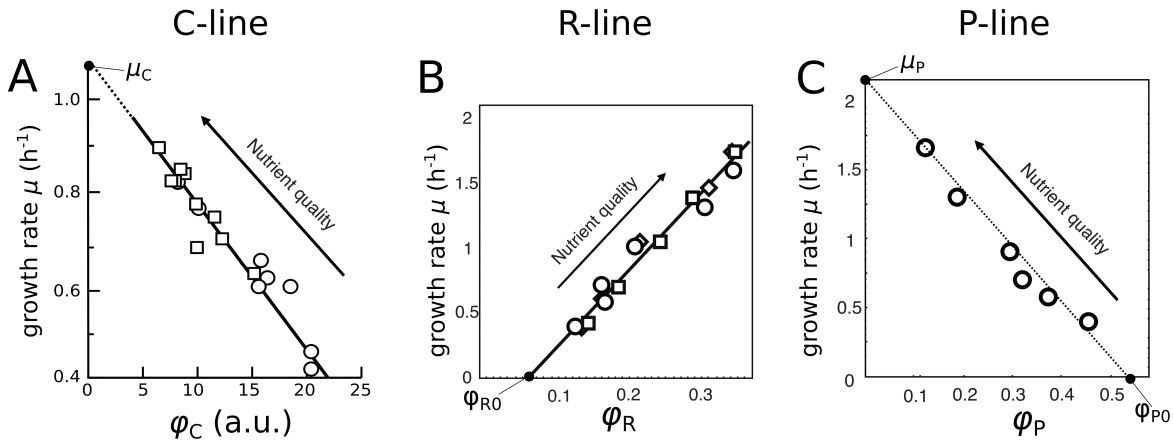
Figure 5.1: Growth laws as experimentally observed. (**A**) Concentration of a sector protein (LacZ) declines with the growth rate. Thick line is a linear fit. Extrapolation of the fitted line (dashed line) reaches a growth rate of $\mu_C \approx 1.1$ when $\phi_C = 0$. Data from Hermsen *et al*, 2015 [65]. (**B**) Ribosome-associated proteins and RNA increase with nutrient conditions. Data from Scott *et al*, 2010 [158]. (**C**) The expression level of useless proteins (P) decreases with nutrient condition. Dashed line is a fit by eye, extrapolating to $\mu_P \approx 2.1$. Data from Scott *et al*, 2010 [158].

Towbin *et al* used the same *E. coli* cAMP-fixed mutant strain as discussed in the previous chapter [181]. They measured, for both the wild type and the cAMP-fixed mutant strain, steady-state exponential growth rates on various nutrient sources. Additionally, on each carbon source they measured the steady-state growth rates of the mutant strain for different C-sector expression levels (experimentally controlled via the concentration of externally applied cyclic AMP). The authors show that for many carbon sources growth rates decline if the wild type ratio between C- and R-sector expression is disrupted, indicating that for many environments *E. coli*'s proteome allocation is optimal for growth [181]. Their experimental results were very neatly matched with their mathematical model which predicts steady state C-sector expression and the corresponding growth rate. In their model, a single variable gave regulatory input to determine the expression level of both the C- and the R-sector. This variable was, on an abstract level, associated with biomass precursors, and hence positively influenced by C-sector proteins (carbon import and catabolism), but consumed by the ribosomes to create biomass and cell growth.

Erickson *et al* instead focused on the temporal dynamics of gene expression and growth of *E. coli* cultures after an environmental change. In "upshift" experiments, cells were first grown on a single carbon substrate to which after some time a second, richer carbon source was added. Alternatively, in "downshift" experiments cells grew on two carbon sources, one of which was depleted during the exponential growth phase, resulting in slower growth on only a single carbon source. During these nutrient transitions, optical density (OD) and the expression levels of the C- and R-sector were measured. In the upshift experiments, the population's growth rate increased sharply when a richer carbon source was added, and afterwards slowly converged to a new steady growth rate [43, 98]. At the same time, cells shifted their proteome allocation towards more investment in ribosomes, and less in catabolic proteins. Again, the authors present a mathematical model that neatly describes the measured dynamics. In their model, a single regulatory variable,
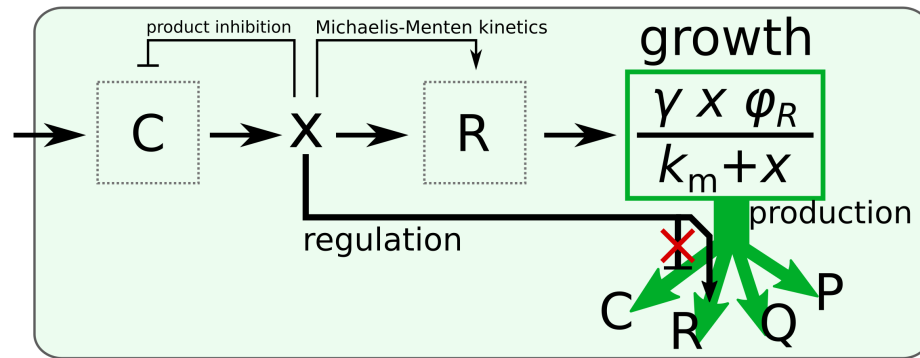
Figure 5.2: Schematic representation of the Erickson-Towbin dynamical regulation model. The red cross denotes the effect of the mutation as discussed in Chapter 4.

here representing the saturation level of the ribosomes, regulates both C- and R-sector expression. The functional form of these regulatory functions was directly derived from steady-state relations between growth rate and expression (*i.e.*, the growth laws, Fig. 5.1), using the critical assumption that at any moment the cell responds to the internal variable and distributes its resources as if it is currently in steady state. This regulatory strategy has recently independently been shown to be an optimal strategy, called 'qORAC' (for 'Specific Flux (q) Optimisation by Robust Adaptive Control' [141]). The model by Erickson *et al* reproduced, without requiring any fitting, the observed temporal dynamics during nutrient upshifts and downshifts for co-utilised substrates.

In this chapter, we first aim to derive an elegant unifying model that explains the data from Towbin *et al*, and, without additional fitting, also the data from Erickson *et al*. Next, we will investigate whether the same model can also explain the noise-transfer properties and the stochastic dynamics in each condition as discussed in the previous chapter. To do so, we must first combine all previous ideas: consider a single internal variable that regulates the C- and R-sector according to the qORAC principle, taking the proteome allocation constraints into account. However, our model should include considerably more. To explain the dynamics of both the C-sector reporter and the constitutive reporter (see previous chapter), we need to include the expression of a constitutively expressed protein. Furthermore, to model the dynamics of the mutant strain (cAMP-fixed cells, see previous chapter), we need to derive additional regulation functions for the mutant. Lastly, we need to include noise sources in line with the noise models seen in earlier chapters.

In the following, we will first formulate a deterministic model that merges the ideas discussed above. We will show that this model can indeed explain the steady-state growth rate and expression data for both the wild type and mutant data as obtained by Towbin *et al*, and also the dynamical data of the upshift experiments performed by Erickson *et al* on the wild type. Next, we extend the model such that it can also describe the dynamics of the cAMP-fixed mutant (which does not actively regulate its C-sector but can still regulate its R-sector). Lastly, we further extend the model to include stochasticity, which allows us to study the system's stochastic dynamics. We compare the resulting cross-correlations with the data presented in the previous chapter, and additionally compare simulated time traces of (stochastic) dynamics of the cAMP-fixed mutant strain to experimental data.

5.2 RESULTS

5.2.1 *A minimal three-sector model of gene-regulation and growth*

For simplicity, we start with a cell with only three proteome sectors: ribosomes (ϕ_R), catabolic proteins (ϕ_C) and a constant fraction ϕ_Q that consists out of maintenance proteins. Together, these three fractions make up an entire cell and we assume that the proteome sectors are subject to a strong allocation constraint, such that at any moment

$$1 = \phi_C + \phi_R + \phi_Q. \tag{5.1}$$

Like Towbin *et al* and Erickson *et al*, we consider a single internal variable x that represents biomass precursors. The C-sector consists of metabolic proteins that import nutrients and convert these to this biomass precursor. The R-sector contains ribosomes that further convert x into biomass and cell volume. We assume that ribosomal activity saturates as a function of x and that, for high concentrations of x, import and conversion of external nutrients slows down due to product inhibition. Then, the time evolution of x can be written as:

$$\frac{\partial x}{\partial t} = \frac{k k_i \phi_C}{k_i + x} - \frac{\gamma x \phi_R}{k_m + x}. \tag{5.2}$$

Here, the parameter k tunes the external nutrient quality, γ is the maximal translation rate of the ribosomes, k_m is the saturation constant of the ribosomes and k_i is the inhibition constant of the C-sector. The parameters k_i, k_m and γ are assumed to be invariant under variation of the environment, allowing us to adopt units such that $k_i = 1$. In equation 5.2 we identify the population's growth rate, μ, as the total (mass-normalised) rate of biomass formation:

$$\mu := \frac{\gamma x \phi_R}{k_m + x}. \tag{5.3}$$

The internal variable x is additionally assumed to give direct regulatory input to the production rates of ϕ_C and ϕ_R (Fig. 5.2). We introduce regulation functions $\chi_C(x)$ and $\chi_R(x)$ that take on positive values between 0 and $1 - \phi_Q$ and determine the fraction of the flux (μ) allocated towards the production of C and R sector proteins respectively. Additionally, their sum is constrained to $1 - \phi_Q$ at any time. Then, the expression levels of the C and R-sectors are a balance between protein production and dilution due to growth:

$$\frac{\partial \phi_C}{\partial t} = (\chi_C(x) - \phi_C)\,\mu, \tag{5.4}$$

$$\frac{\partial \phi_R}{\partial t} = (\chi_R(x) - \phi_R)\,\mu. \tag{5.5}$$

Similar to Towbin *et al*, we assume that the regulation functions for the C- and R-sectors optimise the steady-state exponential growth rate for different carbon sources, *i.e.*, maximise μ for all values of the parameter k. Towbin *et al* choose an *ad hoc*, fixed functional form of the regulatory functions. Instead, we here derive the optimal regulatory functions analytically (see SI, section S5.1). The result is surprisingly elegant:

$$\chi_C(x) = (1 - \phi_Q)\frac{k_m(1 + x)}{k_m + 2k_m x + x^2}, \tag{5.6}$$

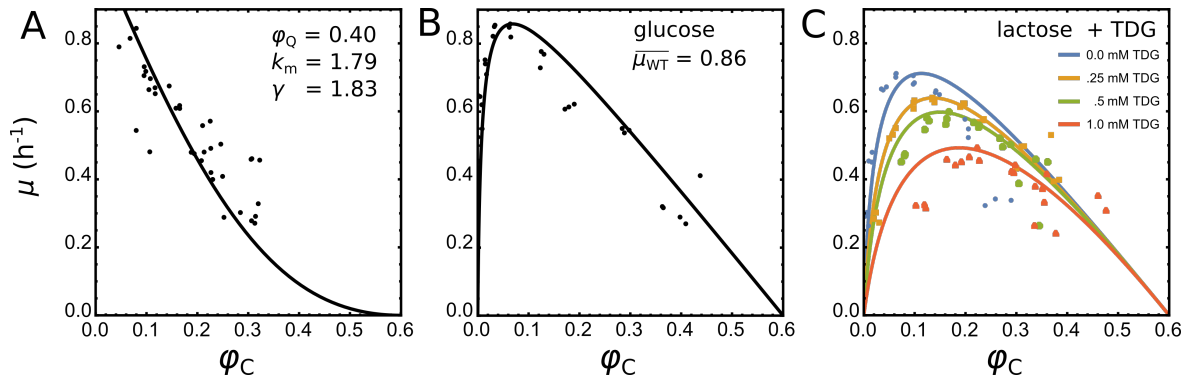$$\chi_R(x) = (1 - \phi_Q)\frac{x(k_m + x)}{k_m + 2k_m x + x^2}. \tag{5.7}$$

Figure 5.3: Fit to the data from Towbin *et al*. All three panels use the same parameters values of k_m, γ and ϕ_Q. (**A**) The relationship between wild type ϕ_C and μ for different nutrient conditions. Solid line is the model fit, constrained to have a y-axis intercept at $\mu_C \approx$ 1.1 h^{-1}. Fitted parameter values are shown in the panel. (**B**) For growth on glucose, the relationship between ϕ_C and μ in the mutant strain, under variation of external cAMP. (**C**) Same as in (B), but now for growth on lactose, and for growth on lactose plus various concentrations TDG (a lactose import inhibiter). For each condition, the wild type's growth rate is used to determine the parameter k. See section S5.2 for more details of the fitting procedure.

Given these regulatory functions, the system is fully determined by only four independent parameters: k, k_m, γ and ϕ_Q. Only the first, k, depends on the nutrient condition.

5.2.2 *The minimal model can fit both steady-state and dynamical data*

Expression and growth rate of WT and cAMP-fixed cells

The model can be fit to the data by Towbin *et al* (see Figs. 5.3 and S5.1). In a population of wild-type *E. coli* cells, Towbin *et al* measured C-sector expression (using a CRP-regulated reporter protein) and growth rate on various carbon sources (Fig. 5.3A). For the mutant strain, they measured the effect of C-sector expression on the growth rate, again for growth on various substrates (Fig. 5.3B-C) and Fig. S5.1).

The three parameters of the model, ϕ_Q, k_m and γ were simultaneously fited to all these data. Additionally, we incorporated constraints to the fit values based on previous knowledge. First, Q-sector expression level has been estimated to be between 0.3 and 0.5 [68, 135]. Second, for a specific value of ϕ_Q, the parameter γ is constrained by the highest carbon-sustained growth rate ($\mu_C \approx 1.1 h^{-1}$, the intercept of the C-line, see Fig. 5.1A). Namely, in our model this highest possible growth rate would be achieved in a hypothetical situation where the ribosomes make up a fraction of $1 - \phi_Q$ of the cell and are translating at their maximum rate, γ. We therefore constrain γ using the relationship: $\gamma(1 - \phi_Q) = \mu_C \approx 1.1 h^{-1}$. Third, given values for ϕ_Q, γ, and k_m the value of the parameter k for a particular nutrient condition follows directly from the average wild-type growth rate on that nutrient.

The resulting fits are shown as the solid lines in Figs. 5.3 and S5.1 and show an excellent agreement with the data.

Nutrient upshift and downshift experiments

Next, we tested whether the regulation functions derived earlier in this document can also predict expression levels and growth rates for cells that are not currently in a steady state exponential growth state. Erickson *et al* measured the optical density of a population of wild-type *E. coli* cells during nutrient up and downshifts. Additionally, they measured C- and R-sector expression levels over time. When during growth on a certain carbon source a second, richer, nutrient is suddenly added to the medium, growth rates first increase sharply, and, with a slower timescale, gradually increase towards a final, steady state, growth rate (Fig. 5.4B, see also [98]). During the process, the cells produce more ribosomal proteins, and less catabolic proteins (Fig. 5.4A) until eventually a steady state is reached in accordance with the growth laws (Fig. 5.1A and B). An example of this behaviour is shown in Fig. 5.4, where the cells initially grow on succinate, until gluconate is added to the medium.

As an example of the dynamics of our model, we show that it can reproduce the succinate-gluconate nutrient upshift, Fig. 5.4. After fitting the model's parameters to the data from Towbin *et al* (Fig. 5.3), all parameters are fixed except for k. The value of k before the nutrient upshift can be derived from the steady-state growth rate on succinate only, $\mu_{\text{succinate}} = 0.45$ h^{-1}. Likewise, the k value for gluconate can be derived from the steady state growth rate on gluconate alone ($\mu_{\text{gluconate}} = 0.91$ h^{-1}). Then, the system (equations 5.3 - 5.5) can be numerically integrated using a quasi-steady state assumption on x. All other timescales, *i.e* the timescales of regulation and adjustment of the growth rate, are fixed by the assumption that at any time, the cell regulates its proteome as if x is currently in an optimal steady state (*i.e.* the qORAC assumption). See section S5.2 for more details.

The resulting curves are shown as dashed curves in Fig. 5.4, and are in perfect agreement with the data from Erickson *et al*, increasing our trust in the validity of the model.

5.2.3   *Model extensions: dynamics of cAMP-fixed cells and stochasticity*

Dynamics in cAMP-fixed cells: including constitutive expression

In Chapter 4 we discussed the *E. coli cyaA* and *cpdA* null mutant that is unable to synthesise or degrade cAMP. Instead, its C-sector expression can be controlled experimentally by applying cAMP externally. In this mutant strain, which was also studied by Towbin *et al*, it is clear that C-sector regulation no longer directly responds to x. However, given the constraint $\chi_C + \chi_R = 1 - \phi_Q$ it seems unavoidable that, by disrupting C-sector regulation, R-sector regulation is affected as well. Moreover, the effect of x on χ_R likely still indirectly affects χ_C. As it is, the model cannot make a prediction on the dynamics of the cAMP-fixed cells.

We therefore extended the model in order to understand how the expression of the C- and R-sector responds to the disruption of C-sector regulation. We assume that the functions $\chi_C(x)$ and $\chi_R(x)$ are the result of a competition between C-sector mRNA and R-sector mRNA for translation by ribosomes. (Another option would be to assume that competition exists for binding to transcriptases [95], in which case similar mathematics can
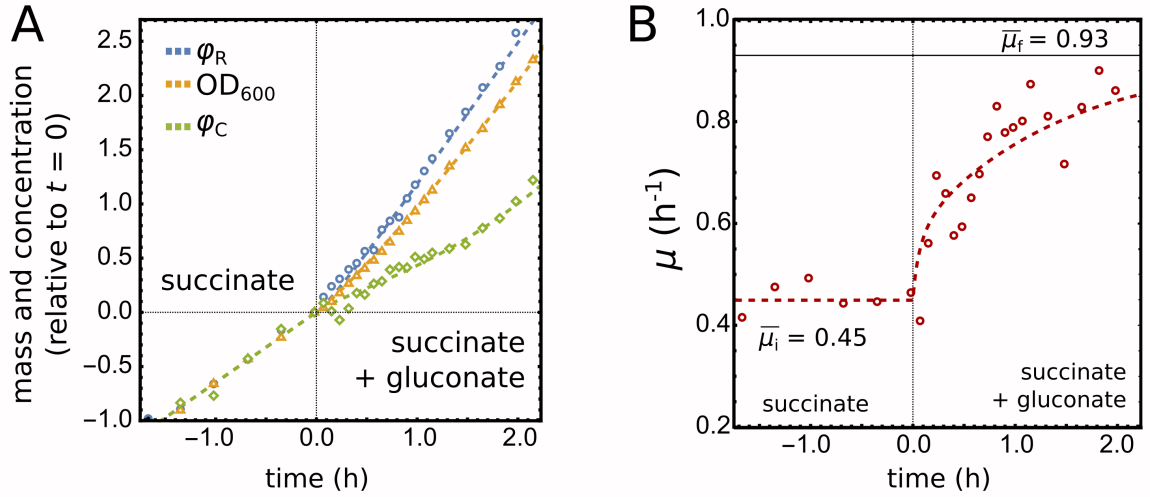
Figure 5.4: Without any free parameters, the model reproduces data from Erickson *et al.* (**A**) Optical cell density (OD600), ϕ_R and ϕ_C during an nutrient up-shift experiment. Values are relative to time $t = 0$, when gluconate was added to the medium. (**B**) Growth rate increases after gluconate is added. Horizontal solid line is the assumed final, steady state growth rate (slightly lower than reported by Erickson *et al*, who reported $\mu_f = 0.98$). The growth rates on each of the two substrates were used to find the parameter k for that particular carbon source. The final growth rate μ_f was then predicted according to the growth rate formula for co-utilised carbon sources [65], see section S5.2 for more details of the matching procedure.

be derived.) That is, we assume that the regulation functions $\chi_i(x)$ ($i = C$, R) can be written as:

$$\chi_i = (1 - \phi_Q) \frac{m_i(x)}{T(x)}, \tag{5.8}$$

where $m_i(x)$ is the mRNA level of sector i and $T(x)$ is the total size of the (C- and R-sector) mRNA pool. The mutations discussed in Chapter 4, *i.e.*, deletion of *cyaA* and *cpdA*, coding for the proteins that synthesise and degrade cAMP, disrupt how the level of C-sector mRNA reacts to x. We therefore assume that the C-sector mRNA level, $m_C(x)$, obtains a constant value e_C in cAMP-fixed cells.

Unfortunately, the wild-type functions $m_i(x)$ are underdetermined given the functions $\chi_i(x)$: if all $m_i(x)$ are multiplied by some function $f(x)$, all $\chi_i(x)$ functions remain unaltered. However, we next show that this issue can be solved by extending our model with a constitutively expressed protein P. Inclusion of such a protein will allow us to integrate more experimental data and resolve the function $T(x)$. (And, later on we will conveniently identify this protein with the constitutive reporter construct used in the experiments of Chapter 4.)

We assume that the mRNA level of protein P, m_P is constant and competes with C- and R-sector mRNA for ribosomal binding. The dynamics of ϕ_P are then analogous to those of ϕ_C and ϕ_R:

$$\frac{\partial \phi_P}{\partial t} = \left((1 - \phi_Q) \frac{m_P}{m_P + T(x)} - \phi_P \right) \mu. \tag{5.9}$$

The steady state protein expression level of P, ϕ_P, becomes:

$$\overline{\phi_P} = \chi_P(\overline{x}) = (1 - \phi_Q)\frac{m_P}{m_P + T(\overline{x})} = (1 - \phi_Q)\frac{1}{1 + T(\overline{x})/m_P},$$ (5.10)

where bars indicate quantities that are in steady state. At the same time, expression of P has its effect on the expression levels of ϕ_C and ϕ_R, due to the increased competition for translation. Importantly, the expression of ϕ_R is slightly decreased, which has an effect on the steady state growth rate in a particular environment k:

$$\overline{\phi_R} = (1 - \phi_Q)\frac{m_R(\overline{x})}{m_P + T(\overline{x})} \equiv \frac{\chi_R(\overline{x})}{1 + m_P/T(\overline{x}|k)},$$ (5.11)

$$\mu(\overline{x}) = \frac{\gamma\overline{x}}{k_m + \overline{x}}\overline{\phi_R} = \frac{\gamma\overline{x}}{k_m + \overline{x}}\left(\frac{\chi_R(\overline{x})}{1 + m_P/T(\overline{x})}\right).$$ (5.12)

In wild-type cells, the concentration of a constitutive protein ϕ_P is known to decrease with the growth rate under variation of the nutrient condition k; a linear empirical relationship known as the P-line [158] (Fig. 5.1C):

$$\overline{\phi_P} = \overline{\phi_{P_0}}\left(1 - \frac{\mu(\overline{x})}{\mu_P}\right).$$ (5.13)

Here, $\overline{\phi_{P_0}}$ is the P protein's mass fraction in the limit of zero growth (k = 0). Even though the expression of constitutively expressed genes is by definition not regulated directly, the P-line demonstrates that it is nevertheless reduced upon improvement of the nutrient conditions, possibly due to changes in the total amount of C- and R-sector mRNA.

Using equation 5.12, we can equate equations 5.10 and 5.13, and solve for $T(\overline{x})$, the total mRNA of the C and R-sector in an optimally regulated cell that possesses a small constitutive mRNA expression m_P. Again, in line with Erickson *et al* [43], we assume that the cell reacts to any value of x as if the system is currently fully in steady state. Together, this results in:

$$T(x) = \frac{\mu_P(k_m + 2k_m x + x^2)}{k_m\mu_P(1 + 2x) + x^2(\mu_P - \gamma(1 - \phi_Q))}.$$ (5.14)

Here, we arbitrarily scaled $T(x)$ (and therewith all mRNA levels) so that $T(0) = 1$. The value of the parameter μ_P (intercept of the P-line, equation 5.13, Fig. 5.1C) can be directly taken from literature: $\mu_P \approx 2.1\ h^{-1}$.

Now that we have discovered the functional form of $T(x)$, the mutant's regulation functions follow straightforwardly:

$$m_R(x) = \chi_R(x)T(x)/(1 - \phi_Q),$$

$$\chi_C^M(x|e_C) = (1 - \phi_Q)\frac{e_C}{e_C + m_P + m_R(x)},$$ (5.15)

$$\chi_R^M(x|e_C) = (1 - \phi_Q)\frac{m_R(x)}{e_C + m_P + m_R(x)},$$ (5.16)

$$\chi_P^M(x|e_C) = (1 - \phi_Q)\frac{m_P}{e_C + m_P + m_R(x)}.$$ (5.17)

The parameter e_C determines the expression level of C-sector mRNA in the mutant strain and reflects the experimentally controllable external cAMP concentration (see Chapter 4). The higher the external cAMP concentration, the higher also e_C and ϕ_C. In the following, we assume a small value of $m_P = 0.0101$ as to not disrupt the rest of the cell significantly.

Including stochasticity by adding noise sources

In line with the previous chapters we assume that noise originates from the stochastic production of proteins. However, we here make the explicit distinction between noise that originates from transcription and translation. Noise on the mRNA level only disturbs the allocation of resources, where translational noise also affects the total rate of protein production (and hence, the growth rate). We write the following stochastic equation for the production rate π_i of protein sector $i \neq Q$:

$$\pi_i = (1 - \phi_Q) \left(\frac{m_i(x) + N_{A_i}}{\sum_k m_k(x) + N_{A_k}} \right) \mu_d(x, \phi_R) + N_{\pi_i}. \tag{5.18}$$

In this equation, N_{A_j} and N_{π_j} are Ornstein–Uhlenbeck noise sources, each obeying its own independent stochastic differential equation:

$$dN_t = -\beta N_t dt + \theta dW_t, \tag{5.19}$$

with β setting the timescale and θ the noise amplitude. The noise sources N_{A_i} represent noise in the production rate of mRNA for protein i and result in noise in the allocation of flux; N_{π_j} represent additional noise in the production of protein j. The variable μ_d is defined below.

The growth rate, μ, is a crucial variable that appears at multiple positions in the equations, with subtly different interpretations. For example, μ can represent the average total metabolic flux, the total rate of biomass formation, and also the rate of the system's dilution. At the population level, during exponential growth, these quantities must be equal. However, when considering the stochastic dynamics inside a single cell, equivalence is not guaranteed. For example, the rate of volume increase (dilution) and the rate of metabolism might differ due to stochasticity. Therefore, we introduce more explicit variables:

$$\mu_d := \frac{\gamma x}{k_m + x} \phi_R, \tag{5.20}$$

$$\mu_M := \sum_i \pi_i = (\phi_Q \mu_d + N_{\pi_Q}) + \sum_{i \neq Q} \left[(1 - \phi_Q) \left(\frac{m_i(x) + N_{A_i}}{\sum_k m_k(x) + N_{A_k}} \right) \mu_d + N_{\pi_i} \right]$$

$$= \mu_d + \sum_i N_{\pi_i}, \tag{5.21}$$

$$\mu_V := \mu_M + N_v. \tag{5.22}$$

We thus distinguish (i) the deterministic flux, μ_d, determined by the internal metabolite x and the (current) ribosomal fraction, ϕ_R, (ii) the mass-growth rate, μ_M, defined as the total rate of mass increase per mass, and (iii) μ_V the rate of volume increase that sets the rate of dilution. For simplicity, we assume a constant cellular density, *i.e* $N_V = 0$ such that any moment, the rate of biomass increase equals the rate of volume increase. However, the total translation rate of the ribosomes, μ_M, is only equal to the deterministic flux μ_d on average. Additionally, μ_M includes noise, both in the allocation of flux (transcriptional noise), as well as in the total protein production rate (translational noise).

Lastly we introduce a noise source N_x that acts on the internal metabolite x. A possible origin of this noise is variation in the flux through the C- and R-sector for given values of ϕ_C and ϕ_R. Since ϕ_C and ϕ_R represent coarse-grained proteome sectors composed of many proteins, for particular values of ϕ_C and ϕ_R the composition of each sector can still

vary. Such fluctuations in the composition of the proteome sectors are likely to affect the carried flux and hence the concentration of the internal metabolite x.

Constraining the parameters of noise sources

By adding noise we inevitably introduce many new parameters: each Ornstein–Uhlenbeck process is characterised by a noise amplitude and a timescale. To constrain the number of parameters, we assume the following. First, we assume that the variances of transcription and translation scale with their respective means –a typical scaling for Poisson processes. Such a scaling is achieved by scaling the noise amplitudes of the corresponding noise sources, *i.e.* $\theta_{\pi_i}^2 \propto \overline{\phi_i}$ and $\theta_{A_i}^2 \propto m_i(\overline{x})$. (See Appendix B for a proof that a scaling of $\theta_{\pi_i}^2$ with $\overline{\phi_i}$ indeed results in a variance of ϕ_i in accordance with a Poisson process.) Although x is not the result of a single Poisson process, we still assume a similar scaling for the noise source N_x, namely $\theta_x^2 \propto \overline{x}$. Second, we assume that the timescales of N_{π_i} and N_{A_i} are fast for all i ($\beta_{\pi_i} = \beta_{A_i} = 3\,\overline{\mu_d}$). Noise on x, however, has an assumed timescale close to the (mean) growth rate ($\beta_x = 0.8\,\overline{\mu_d}$). In the following, we keep these timescales fixed.

Even if the scaling of the noise sources is fixed, we still need to set their absolute levels. For this, we introduce the parameters α_x, α_A, and α_π, that set the absolute noise levels of x, allocation, and production rates respectively. For example, the noise amplitude for the production rate of C-sector proteins then becomes: $\theta_{\pi_C}^2 = \alpha_\pi^2 \overline{\phi_C}$.

5.2.4 *The extended model largely reproduces cross-correlation data*

We are now finally in the position to test our model by comparing its stochastic dynamics to the measured cross-correlations presented in Chapter 4.

Since the model's dynamics are determined from the cell's population-level behaviour, we have only three parameters left that influence the cell's stochastic properties: α_x, α_A, and α_π. When calculating cross-correlations in a small noise regime, only the relative amplitudes of the noise sources matter. Therefore, we set $\alpha_x = 1\text{x}10^{-2}$, leaving only α_A and α_π as free parameters.

We tried to recreate all the cross-correlations seen in the previous chapter for particular values for α_A and α_π. As before, the model parameter k could be inferred from the mean growth rate of the wild type cells ($\overline{\mu}_{WT} \approx 0.75$ dbl/hr). Then, values for the parameter e_C (representing externally applied cAMP) for the low, optimal and high cAMP concentrations were calculated from the mean growth rates in these conditions ($\overline{\mu}_{low} \approx 0.35$ (dbl/hr), $\overline{\mu}_{high} \approx 0.45$ (dbl/hr, and $\overline{\mu}_{optimal} = \overline{\mu}_{WT}$). For each condition, cross-correlations were calculated over a long simulated time trace of a single cell.

The parameters and cross-correlations resulting from a fit-by-eye are shown in Fig. 5.5 and reproduce many important features. First, the cross-correlations of the C-sector reporter (both ϕ-μ and π-μ) increase in value from wild type to cAMP-fixed* cells. Importantly, the ϕ-μ cross-correlation changes sign in accordance with the data. Second, cross-correlations of the C-sector reporter further increase in magnitude when the C-sector becomes rate limiting at low cAMP, as is also observed in the data. Third, the ϕ-μ cross-correlation of the C-sector reporter again changes sign and becomes negative (yet more symmetrical) when going from optimal cAMP levels to high cAMP levels, while the π-μ cross-correlation stays positive. Again, the same qualitative differences be seen in the data.
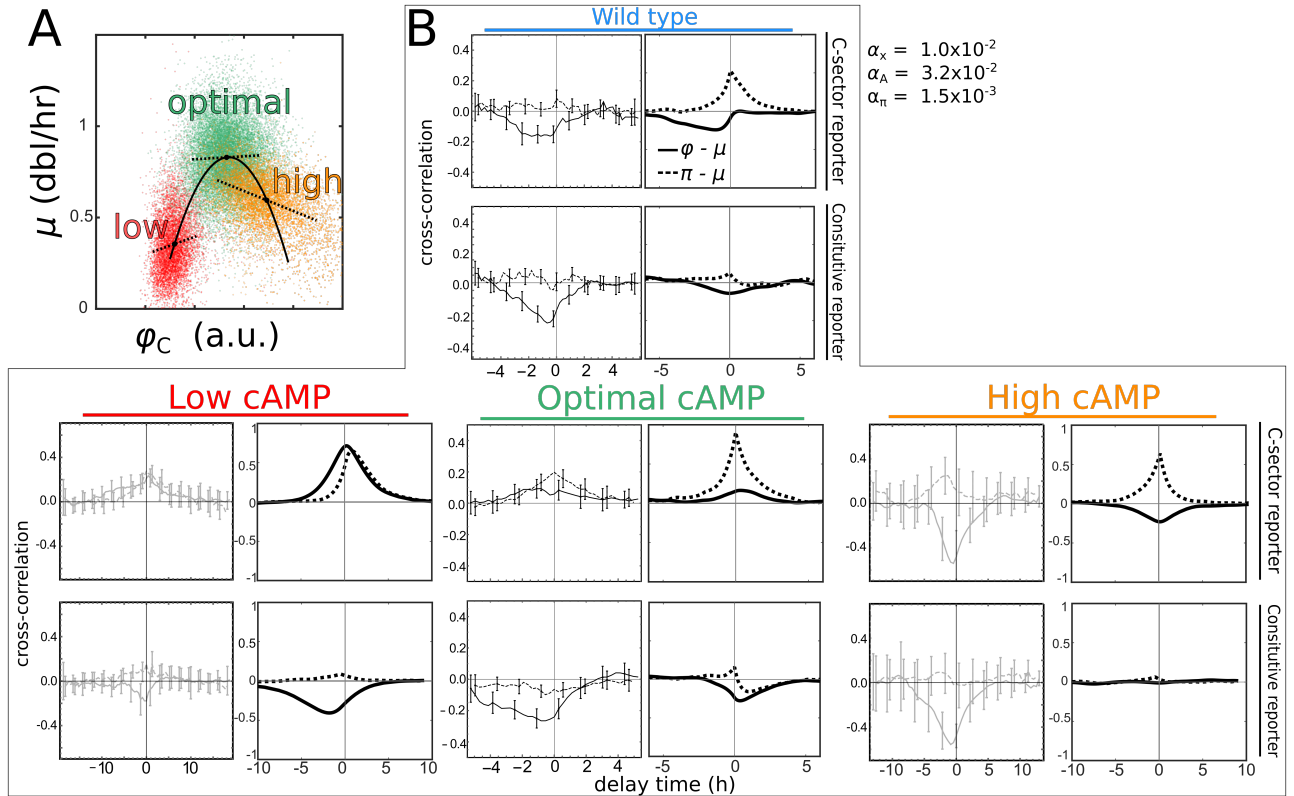
Figure 5.5: **(A)** Overview of the three mutant conditions with low, optimal and high cAMP concentrations. **(B)** Overview of all measured cross-correlations and simulated cross-correlations from the model. 'Fitted' values for α_x, α_A, and α_π are shown in the figure.

Fourth, cross-correlations for the constitutive reporter stay negative when going WT to cAMP-fixed* and cAMP-fixed cells supplied with low cAMP.

However, the model also produces some obvious mismatches with the data. Most striking is the lack of any signal for the constitutive reporter at high cAMP, where the data shows a clearly negative cross-correlation between ϕ and μ. Possibly, the lack of signal in our model is due to the low abundance of the constitutive reporter, and, consecutively, a relatively high intrinsic noise level under this condition. Additionally, our model does not correctly match the amplitudes of the cross-correlations, sometimes over- and sometimes under-estimating them compared to the data. Still, the fact that the model reproduces so many qualitative features with only two parameters for all the 16 curves is impressive.

5.2.5 *Reproducing cAMP pulse-experiments*

Lastly, we further tested the model against preliminary time series experiments performed by M. Wehrens at AMOLF [191]. Here, mutant cells were grown in micro-colonies on minimal medium with lactose, but with a pulsating external cAMP concentration that alternated between a too low (45 μM) or too high (2100 μM) cAMP concentration compared to the optimal (800 μM) (for more details see SI). The first five cAMP cycles were fast, pulsating with 1 hour intervals. Afterwards, three cycles with 5-hour intervals followed (Fig. 5.6). Cells were thus continuously pushed out of steady state and had to rearrange

their proteome to follow the external cue given by cAMP. The expression of the C-sector reporter, as well as the instantaneous growth rates, were measured over time (Fig. 5.6).

Such an experiment can be simulated by our model via numerical integration. The resulting model simulations are shown in Fig. 5.6. Here, the parameters α_x, α_A, and α_π were scaled such that the resulting $CV_\mu = 0.2$, closely resembling the measured CV of optimal WT cells, see Chapter 4, Fig. S4.4).

Again, many qualitative features are readily explained. First, the C-sector concentration has a smaller coefficient of variation than the growth rate (Fig. 5.6A,B), which seems to become smaller for smaller C-sector expression and increase with increasing C-sector expression. Second, during the fast cycling, shortly after cAMP switches from low to high the growth rate generally decreases; conversely, the growth rate increases when cAMP concentration is switched from high to low (Fig. 5.6, $0 < t < 6$). Our model shows that this is because, during the 1 hour pulses, cells consistently sit on the higher half of the optimum curve (high ϕ_C and low μ, Fig. 5.5A). Third, after a long period of low cAMP, an upshift in the cAMP concentration quickly results in much faster growth rates (Fig. 5.6, $t = 11$), after which the growth slowly decreases again. Fourth, the rise in growth rate is much slower when cAMP is lowered again (Fig. 5.6, $t > 16$).

The extremely fast rise of the growth rate, and seemingly much slower rise of ϕ_C, when the cells change from low to high cAMP has long been a puzzle (Fig. 5.7A). However, our model is now able to give an hypothesis for this observation. Namely, when cAMP is low and C-sector expression is low, cells are extremely sensitive to changes in C-sector expression (C-sector proteins have a high Growth Control Coefficient, Chapter 2). A very short delay between C-sector expression and the fluorescence signal of the C-sector reporter will then already allow cells to seemingly grow faster than possible with too little C-sector expression. Such a small delay could be due to the finite maturation time of fluorescent proteins, on a timescale of 5 to 18 minutes for mVenus (5 to 15 minutes [187]). Our model suggests that a delay time of 10 minutes could already explain the observed behaviour of a rapid increase in the growth rate at constant ϕ_C (Fig. 5.7B, C), giving a possible answer to the puzzle.

5.3 DISCUSSION

In this chapter, we derived a deterministic model that explained the population-level behaviour (steady states as well as dynamics) of *E. coli* cells under the assumption of optimal regulation. With a minimal number of parameters, a stochastic expansion of this model qualitatively matched multiple features of single cell stochastic dynamics. Therewith, the model suggests that the cell's steady state, dynamical response, and noise propagation properties are three sides of the same coin. In other words, the networks that shape the steady state response to changes in the external environment also determine the dynamical response, *and* the system's noise propagation properties.

The model is firmly based on experimental observations, integrating experimental data from four studies from three different labs [43, 158, 181, 191]. Importantly, the model therewith shows that the ideas upon which the models of these studies are based can be unified within a single framework of a growing stochastic cell.

At the same time, data integration poses challenges and problems. For example, in the different studies different *E. coli* strains were used, and growth rates and the expression
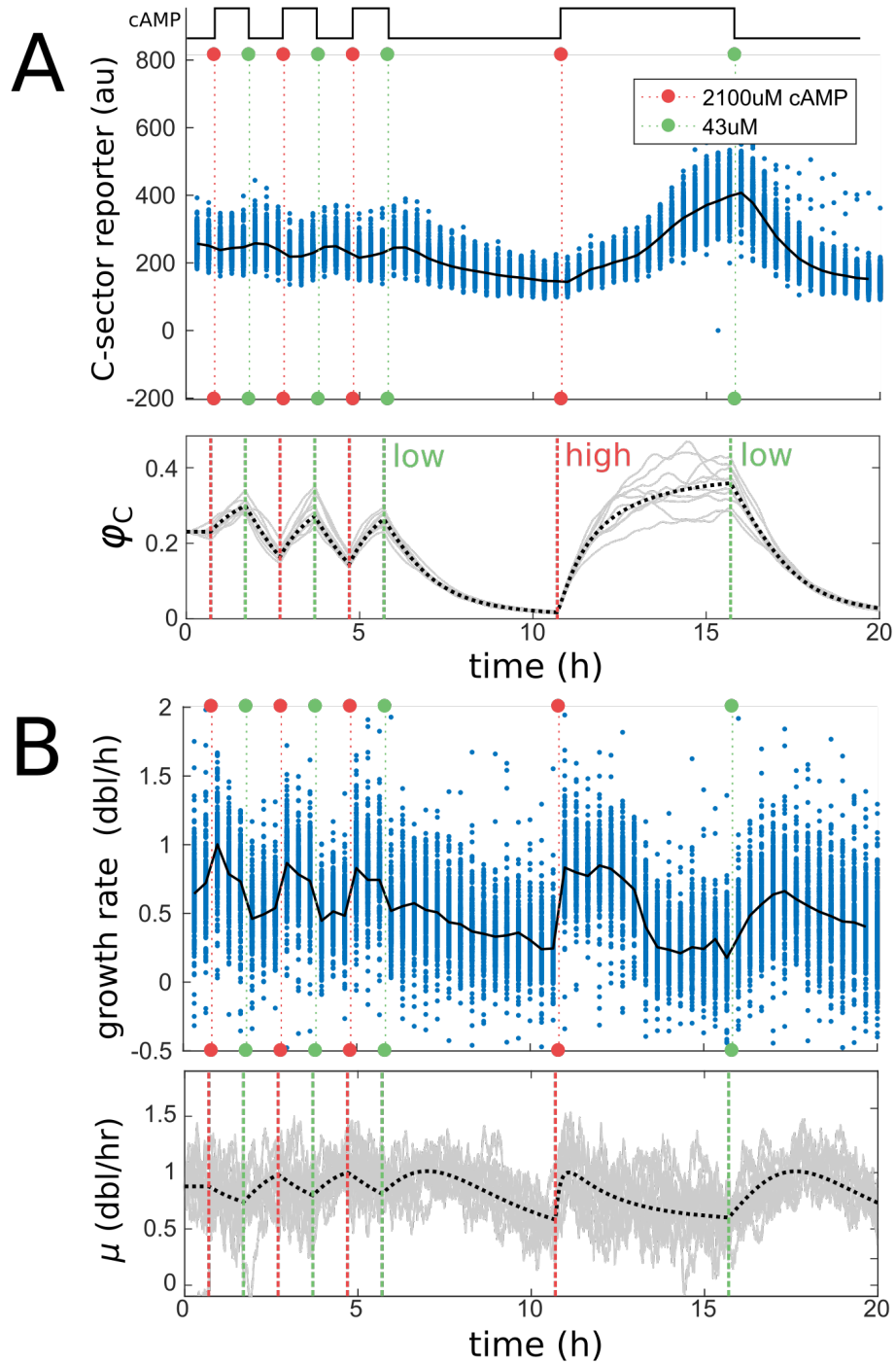
Figure 5.6: Data from the cAMP-pulse experiments and the corresponding model simulation. C-sector reporter (**A**) and growth rate (**B**) over time. Blue dots are single cell observations, and the black line is the population average. Time between data points is 20 minutes. Data courtesy of M. Wehrens and S. Tans (AMOLF), data-figures reproduced with permission from [191]. Below, the model simulations are shown: black dashed lines are deterministic curves (all noise amplitudes zero), grey lines are 11 individual stochastic cell traces. Here, $k = 50$, $e_{Clow} = 0.2$, $e_{Chigh} = 159$, $\alpha_x = 1 \times 0.02$, $\alpha_A = 3.2 \times 0.02$, $\alpha_\pi = 0.15 \times 0.02$ and $m_P = 0.01$.
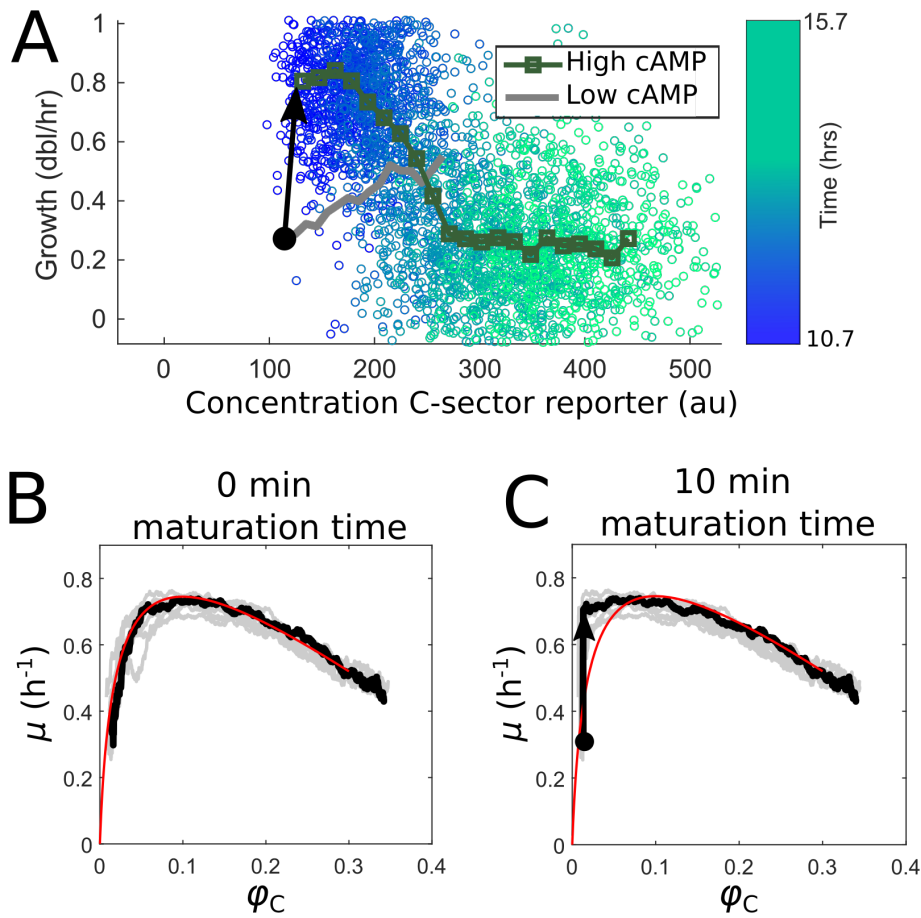
Figure 5.7: (**A**) Growth rate of individual cells against the concentration of the C-sector reporter during high cAMP (same data as in Fig. 5.6 for $10.7 < t < 15.7$). Colour indicates time. Thick, blocked curve is the average over time during the high cAMP cycle, thick grey line is the average during the preceding low cAMP cycle. The black arrow indicates where cells seemingly instantaneously jump from low growth rate and low C-sector expression, to high growth rates and low C-sector expression. Data courtesy of M. Wehrens and S. Tans (AMOLF), data panels reproduced with permission from [191]. (**B**) Model simulation of the same up shift from low to high cAMP, where cell traces (grey lines) follow the deterministic trend (red line). Thick black line is an example single cell trace (**C**) same as in (B), but C-sector signal is plotted with a 10 minutes delay. Now, a similar jump from low μ and low ϕ_C, to high μ and low ϕ_C can be observed.

levels of proteome sectors were measured differently in each study. We choose to scale experimental data such that measured values under the same growth conditions overlapped as much as possible. Alternatively, a different set of parameters could be fitted for each *E. coli* strain and condition. This may improve the quality of the fits, but also increases the number of parameters.

Generally, we have tried to keep the number of parameters as low as possible. For example, we assumed that the parameters k_i (inhibition constant of C-sector flux) and k_m (saturation constant of the ribosomes) have a fixed value in all nutrient conditions. Since different carbon sources enter the cell via different pathways, this assumption is not necessarily valid. Additionally, product inhibition modelled with the parameter k_i sometimes causes a mismatch when comparing the model's predicted growth rate on co-utilised carbon sources with the 'growth rate composition formula' [65] (data not shown), since during the derivation of that formula production inhibition is assumed to be negligible. A better match between our model and the composition formula might be obtained by tuning k_i in each environment independently, again at the cost of an increased number of parameters.

The stochastic extension of the model leaves a lot of room for exploration. Inspired by previous chapters, we considered noise sources acting on the production rates (similar to chapter 2) and on the internal biomass precursor x (similar to the metabolic noise source in Chapter 4). In chapter 4 we additionally assumed independent noise in the growth rate, resulting in an independent dilution mode (chapter 4, Fig. 4.2C). Especially for the constitutive reporter, the dilution mode had an important contribution to the cross-correlations. In our current model, the amplitude of the dilution mode is fixed because we model the growth/dilution rate explicitly instead of phenomenologically. Noise in the rate of dilution originates from noise in the production rate of other proteins (same as in chapter 2). Possibly this underestimates the dilution mode when only a small number of different protein species is modelled. We tried to add a specific noise source to μ_V (equation 5.22) to simulate noise in the rate of dilution. However, this did not yield realistic cross-correlations (see SI sec S5.3 and Fig. S5.3). Still, the effect of other noise sources could be studied within the context of this model. For example, it would be interesting to study the effect of partitioning noise during division on the cross-correlations [176], or, more generally, the effect of noise that is shared between x and μ, similar to common noise (chapter 4, [86]).

The potential of this model is not yet fully realised. The model should be able to reproduce more population-level observations, such as the R-line, the P-line (Fig. 5.1B, C) and the P-line under increased expression of ϕ_P (reference [158], figure 4C). Additionally, the model should reproduce the perturbed C- and R-lines when translational inhibitors are added to the growth medium (reference [159], figure 2A). The framework could furthermore be extended with more protein species, or with more protein sectors such as the anabolic sector 'A' [68, 95, 198]. In a specific environment, the model could be linearised in order to calculate noise transfer parameters and compare these values to those found in chapter 4 or reference [86].

Taken together, the model presented here is a useful tool that summarises our current understanding of the stochastic dynamics and regulatory wiring of *E. coli*. The model can be used to test different hypotheses, study the interplay between regulation and stochasticity, and to gain a more intuitive understanding of the complex behaviour of bacterial cells.

SUPPLEMENTARY INFORMATION CHAPTER 5

S5.1 DERIVATION OF OPTIMAL REGULATION FUNCTIONS

To derive the functional forms shown in equation 5.6, we first analyse the steady state of the system. In steady state, x is a function of ϕ_C and ϕ_R (see equation 5.2), but ϕ_C and ϕ_R are related due to the proteome constraint (equation 5.1), thus (bars indicate steady state):

$$\overline{x}(\phi_C, \phi_R) = \overline{x}(\phi_C, 1 - \phi_Q - \phi_C) = \overline{x}(\phi_C). \tag{S5.1}$$

We can therefore further write the steady-state growth rate as a function of ϕ_C alone:

$$\overline{\mu}(\phi_C) = \frac{\gamma \overline{x}(\phi_C)(1 - \phi_Q - \phi_C)}{k_m + \overline{x}(\phi_C)}. \tag{S5.2}$$

Next, we ask which proteome fraction ϕ_C (and consequently ϕ_R) gives the optimal growth rate, μ^*. We thus solve for ϕ_C^*:

$$\left.\frac{\partial \overline{\mu}(\phi_C)}{\partial \phi_C}\right|_{\phi_C^*} = 0. \tag{S5.3}$$

Because the solution depends on the nutrient quality k, we can express the optimal proteome allocation as a function of k, $\phi_C^*(k)$:

$$\phi_C{}^*(k) = (1 - \phi_Q)\frac{\gamma(k - 2kk_m + \gamma) + \sqrt{kk_m\gamma}(k - \gamma)}{k^2 + 2k(1 - 2k_m)\gamma + \gamma^2}. \tag{S5.4}$$

Filling in equation S5.4 into equation S5.1 results in the steady state value of x, given optimal proteome allocation, as a function of the external environment:

$$\overline{x}^*(k) := \overline{x}(\phi_C^*(k)) \tag{S5.5}$$

$$= \frac{\sqrt{k^5 k_m \gamma} + \sqrt{k^3 k_m \gamma^3} - 2\sqrt{k^3 k_m^3 \gamma^4} - kk_m\gamma(k - \gamma)}{\gamma\left(\sqrt{kk_m\gamma^3} + k(k + \gamma - 2k_m\gamma - \sqrt{kk_m\gamma})\right)}. \tag{S5.6}$$

Inverting this relation, we can find the 'observed' environment by the cell, $k(\overline{x}^*)$, *i.e.*, the inferred nutrient quality that should be present if x is in steady state and the proteome is optimally allocated for this environment:

$$k(\overline{x}^*) = [\overline{x}^*(k)]^{-1} = \frac{\overline{x}^{*2}\gamma}{k_m}. \tag{S5.7}$$

In our model, the expression levels of proteome sectors are determined via x. The internal variable x is not only a sensor of the external environment k, but also sets the expression level of the C sector, such that the cell can adequately adjust to its surroundings. We are therefore interested in finding, given a particular environment k, the relationship between

the steady state optimal proteome fraction of the C-sector (ϕ_C^*) and the corresponding steady state of x, *i.e.* inserting equation S5.7 in equation S5.4:

$$\phi_C^*(k(\overline{x}^*)) = (1 - \phi_Q)\frac{k_m(1 + \overline{x}^*)}{k_m + 2k_m\overline{x}^* + \overline{x}^{*2}}. \tag{S5.8}$$

The next step is to find the regulatory function $\chi_C(x)$ that decides, given the current value of x not necessarily in steady state, what part of the cell's resources should go to the production of C-sector proteins. Like Erickson *et al*, we here employ the qORAC principle [141], which states that resources should always be distributed as if the current value of x is in its optimal steady state. For any value of $x > 0$ this result in the following regulatory function:

$$\chi_C(x) = \phi_C^*(k(x)) = (1 - \phi_Q)\frac{k_m(1 + x)}{k_m + 2k_m x + x^2}. \tag{S5.9}$$

s5.2 FITTING PROCEDURE

s5.2.1 *Towbin data*

Before we can fit our model to the data from Towbin *et al*, these data need to be transformed in two ways.

First, the expression level of the C-sector is measured via a fluorescent reporter under CRP control. We assume a constant scaling factor between fluorescence and the total size of the C-sector, ϕ_C. This scaling factor, α_ϕ adds one more fit parameter.

Second, growth rates in the study from Towbin *et al* [181] are consistently lower than those reported by the Hwa lab [43, 65]. A reasonable explanation for this discrepancy is the use of different measurement techniques and experimental conditions. Towbin *et al* measure the growth rate in small wells (using a 96-well plate reader), while Erickson *et al* measure growth rates in batch cultures with a much larger volume. Since our model is supposed to reproduce both the growth curves from Erickson *et al* and the steady-state growth rates of Towbin *et al*, we scaled all growth rates measured by Towbin *et al* to match those of Erickson *et al*. Luckily there are two substrates, glucose and maltose, for which both labs have measured growth rates. For a single scaling factor, $\alpha_\mu = 1.27$, the growth rates in these two environments as measured by Towbin *et al* match the growth rates measured in the Hwa lab (see Table 1 in reference [65]). Therefore, we scale all the growth rates measured by Towbin *et al* with this factor before fitting the remaining parameters ϕ_Q, γ, k_m, and α_ϕ.

s5.2.2 *Erickson data*

The data from Erickson *et al* mainly consists of temporal data of nutrient up and downshifts and requires numerical integration of the system's ODEs. Additionally, the model must be altered to allow for multiple carbon sources. Analogous to Erickson *et al*, we write for the flux through the C-sector, relative to the total cell mass:

$$j_c(x) = \frac{k_1\phi_{C,1} + k_2\phi_{C,2}}{1 + x}. \tag{S5.10}$$
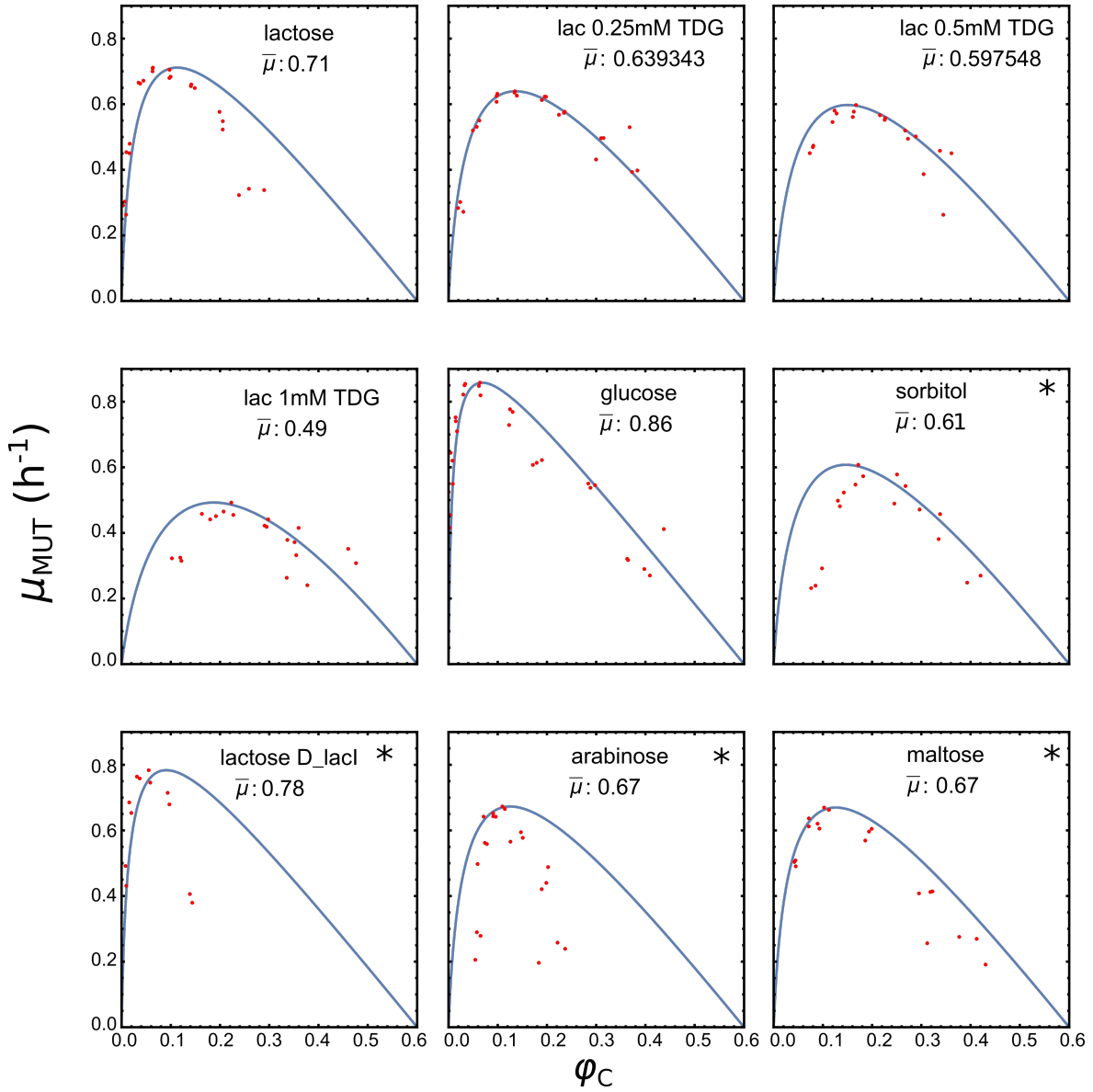
Figure S5.1: All data from Towbin *et al*, for which the wild-type growth rate, μ_0, equals the optimum of the mutant growth rates. In each panel is denoted: the nutrient condition, and the wild-type growth rate (scaled by $\alpha_\mu = 1.27$ to quantitatively match growth rates from the Hwa lab). Panels labelled with an asterisk were not included in the fit, but instead followed from the fitted parameters and the wild type growth rate. The parameter α_ϕ, that scales between CRP-reporter fluorescence and ϕ_C, was fitted to be $\alpha_\phi = 0.502$.

The proteome fractions $\phi_{C,1}$ and $\phi_{C,2}$ are only produced when their respective carbon sources are present. The particular carbon sources discussed (succinate and gluconate) are co-utilised [65]. We therefore introduce the following binary functions, h_1 and h_2 that track which nutrient is available:

$$\chi_{C,1}(x) = h_1\chi_C(x), \tag{S5.11}$$

$$\chi_{C,2}(x) = h_2\chi_C(x), \tag{S5.12}$$

with

$$h_1 = 1, \tag{S5.13}$$

$$h_2 = \begin{cases} 0, \text{ for } t < 0, \\ 1, \text{ for } t \geqslant 0. \end{cases} \tag{S5.14}$$

At $t = 0$, the cell suddenly starts to create more ϕ_C than possible (total production rate of C-sector proteins is now $2\chi_C(x)\mu(x)$, such that the total production rate of proteins exceeds $\mu(x)$). We therefore assume that $\phi_{C,2}$ does not add to the cellular mass, such that the total rate of cell-mass increase is still confined to $\mu(x)$ at any time point. Instead, $\phi_{C,2}$ is interpret as an importer protein that carries catabolic flux, but whose expression level is negligibly low compared to the whole C-sector.

Below, we list the growth rates according to Erickson *et al* on the two substrates shown in Fig. 5.4, succinate and gluconate, separately. The final growth rate, μ_f, achieved during steady state growth when both carbon sources are co-utilised, can then be calculated using the growth rate composition formula derived in [65].

$$\mu_1 = \mu_{\text{succinate}} = 0.45 \text{ h}^{-1}, \tag{S5.15}$$

$$\mu_2 = \mu_{\text{gluconate}} = 0.91 \text{ h}^{-1}, \tag{S5.16}$$

$$\mu_C = 1.1 \text{ h}^{-1}, \tag{S5.17}$$

$$\mu_f = \frac{\mu_1 + \mu_2 - 2\mu_1\mu_2/\mu_C}{1 - \mu_1\mu_2/\mu_C^2} = 0.93 \text{ h}^{-1}. \tag{S5.18}$$

The resulting composite growth rate is slightly lower than the final growth rate reported in Erickson *et al*, who found $\mu_{12} = 0.98 \text{ h}^{-1}$ (although they state $\mu_{12} = 0.95 \text{ h}^{-1}$ in their supplementary text). However, the value of $\mu_f = 0.93 \text{ h}^{-1}$ found here seems to be in better correspondence with their data (see horizontal solid line in Fig. 5.4B).

s5.2.3   *Stochastic simulations*

Stochastic simulations were done in Matlab by numerical integration. At each time point, OU-values were calculated according to the Euler-Maruyama method (first order). Given the current value for these stochastic noise sources, the next value of ϕ can be deterministically calculated using the Euler integration method: $\phi_i(t + dt) = \phi_i(t)(1 - dt\,\mu_M(t)) + dt\,\pi_i(t)$, with $dt = 10^{-3}$. The variable $x(t)$ was assumed to be in quasi steady state and hence a direct function of $\phi_C(t)$, $\phi_R(t)$, albeit with addition of the noise source $N_x(t)$. The growth rate and production rates then followed according to equations 5.20 and 5.18 respectively.

For a particular condition (low, optimal, and high cAMP for the cAMP-fixed cells, and the wild type) we calculated cross-correlations (Fig. 5.5) as follows. We first let the system
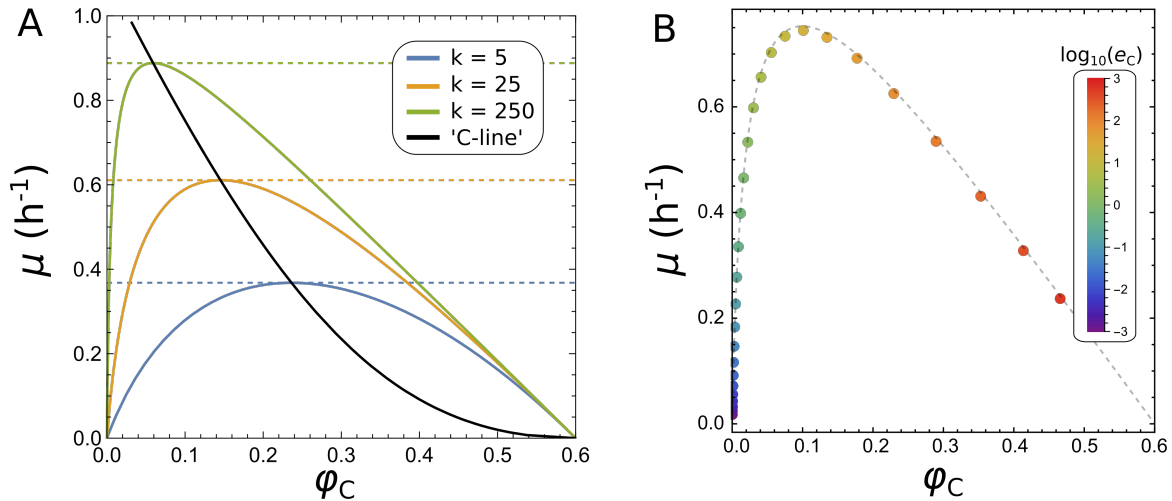
Figure S5.2: **(A)** Example of the mutant growth rate as a function ϕ_C for different values of k. Dashed lines are the wild type growth rates for those values of k. The black line represents the optimal growth rate under variation of k (*i.e.*, the C-line). **(B)** The relationship between the model parameter e_C (representing the external cAMP concentration) and ϕ_C and μ (equation 5.15). Here, k = 70, resulting in $\mu_{WT} = 0.753$ h^{-1}, a value that closely resembles the wild type growth rate in chapter 4. The grey dashed curve is the mutant's growth rate without any P protein. The small mismatch between the points and the curve is due to the very low expression of P sector proteins ($\phi_{P,0} = 0.01$).

run without noise until the steady state associated with the current condition was reached. The mean values in this steady state were then used to calculate the scaling of the noise amplitudes. Next, we turned on the noise, simulated a single long time trace (10000 hours) and calculated cross-correlations along this single lineage using Fast Fourier Transforms. We made sure that absolute noise levels were small such that the system at any times remained within physically reasonable limits.

For the cAMP pulse traces we first let the system run until steady state in the low and high e_C values without noise, to infer the noise amplitudes in those two regions. Then, we start with a medium value for e_C and pulsate the variable e_C from low to high at the same time switching times as in the experiments (t $\in \{0.7, 1.7, 2.7, 3.7, 4.7, 5.7, 10.7, 15.7\}$, see Fig. 5.6).

s5.2.4 *Experimental methods*

Experimental methods for the stochastic experiments discussed in this chapter can be found either in the previous chapter (for the cross-correlations), or in reference [191] (for the pulsed cAMP data).

s5.3 AN INDEPENDENT DILUTION MODE

In the current version of our model, we cannot tune the dilution mode freely. We here show why including the noise source N_V (that injects independent noise in the rate of volume
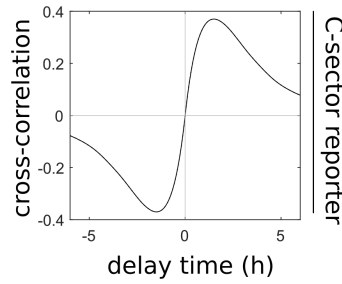
Figure S5.3: Cross-correlations for the wild type C-sector reporter with only noise source N_V non-zero ($\theta_V = 0.001$). Cross-correlations for all conditions and for both reporters were similar to this one.

increase, and therewith also in the rate of dilution, see equation 5.22) does not yield a dilution mode.

When only the noise source N_V is present the resulting cross-correlations do not reflect the Dilution mode as found in Chapter 2 [87]. Instead, the observed cross-correlations are perfectly anti-symmetrical, *i.e.*, $R(t) = -R(-t)$ (Fig. S5.3). Here, we explain why.

- Cross-correlations between the concentration of any protein species i and the growth rate μ as a result of N_V consist out of two functional forms that we do have seen before:

  1. **Dilution:** Fluctuations in the rate of dilution causes the growth rate to increase instantaneously, but protein concentrations decline after a certain delay. The functional form of this mode is $-A(t)$ (see Chapter 2, equation S2.48).

  2. **Control/Catabolic:** Fluctuations in the concentrations of proteins (due to noise in N_V earlier) have an effect on the growth rate. In the current framework, the growth rate is a function of $\bar{x}$ and ϕ_R. Since $\bar{x}$ is only determined by the ratio between ϕ_C and ϕ_R, the noise term N_V does not influence $\bar{x}$. However, ϕ_R directly and linearly influences the growth rate, resulting in the mode $S(t)$ (Chapter 2, equation S2.49). In other protein species this mode is present because of their perfect correlation with ϕ_R as N_V affects all proteins species equally.

- The amplitudes of both modes are equal, because both effects have a linear influence on the system. Both modes also have equal timescales, because they originate from the same noise source and reverberate through the same system.

- In Chapter 2 the relation between A and S is calculated: $A(t) + A(-t) = 2S(t)$; see equation S2.50.

- The cross-correlation for any protein species and the growth rate thus is: $R_{\phi_i - \mu}(t) = S(t) - A(t) = \frac{A(t) + A(-t)}{2} - A(t) = \frac{A(-t) - A(t)}{2}$, which is clearly an anti-symmetrical function.

Since traces of such an anti-symmetrical mode have not been seen in experiments, we set $\mu_V = 0$ in all simulations. This is identical to the assumption that the total rate of mass increase μ_M equals the rate of volume increase, *i.e.* a strong density constraint.

6

DISCUSSION AND OUTLOOK

6.1 SUMMARY OF THE CHAPTERS

This thesis focuses on the interplay between noise in the expression of proteins, and noise in the cellular growth rate. A biological cell is a highly constrained growing and evolving system with countless feedbacks. To further sharpen our intuition about how biochemical noise could reverberate through such a system, we have extensively used mathematical modelling, backed up, whenever possible, with experiments.

Chapter 2 describes a conceptual framework of gene expression, metabolism and growth. Using this framework, we were able to analyse how the noise in all protein species together influences the noise in the growth rate. We investigated a model cell where stochasticity originates in the gene expression of every protein species, but nowhere else. We calculated how the noise in all these protein species could then propagate through the cell to not only impact the growth rate, but also the temporal dynamics of a particular fluorescent reporter protein which is measured experimentally. Useful for our calculations were the *Growth Control Coefficients*; parameters that capture the extent to which the growth rate is sensitive to changes in a particular protein species' concentration. Therewith, GCCs quantify to which extent a protein species is growth-limiting and also summarise how strongly noise in that protein species propagates to the growth rate. From our calculations it became clear that the stochastic dynamics of *any* protein species are, mainly via the GCCs, influenced by the fluctuations of *all* protein species.

If all proteins together cause fluctuations in the cell's growth rate, a natural question is: how much is a protein species *a priori* expected to contribute to the noise in the growth rate? In **Chapter 3** we investigated this question. We showed that the GCCs are not only useful to describe noise propagation, but are moreover an indication of how optimal gene expression is. A protein with a (too) high GCC indicates that the growth rate would improve if the protein's (mean) expression increases. Whenever, in the organism's evolutionary history, natural selection favoured a high growth rate and average gene expression was adjusted to accommodate this, the GCCs were coincidentally altered as well. If gene expression is optimal for growth, the GCCs are distributed non-random, resulting in a scaling of the GCCS with the protein's mean abundance. Counter-intuitively, this results in a strong contribution of abundant protein species, despite their often low noise levels.

Chapter 4 added gene regulation to the discussion of noise propagation. If the cell has evolved regulatory networks that adjust the (mean) expression of proteins in reaction to a change in the external environment, those regulatory networks might also react to internal, stochastic fluctuations in a fixed environment. In this chapter we combined experiments and theory to confirm that the cAMP-CRP regulatory network indeed reacts to noise in metabolism. Again, the mathematical model pointed towards the importance of the noise originating from 'hidden variables': noise in the expression of all those proteins whose concentrations are not explicitly measured. The stochastic expression of the many (metabolic)

proteins that together generate a fluctuating metabolic flux was key to explaining the observed stochastic dynamics.

Chapter 5 integrates data from population-level experiments on proteome allocation to derive the dynamical regulatory wiring of the catabolic and the ribosomal proteome sectors. Since the regulatory networks that govern a population's response to a changing environment are embedded in every single cell, we hypothesised that also the single-cell stochastic dynamics are shaped by these networks. Therefore, we tried to explain all data from Chapter 4 with this regulatory model, with only a single free parameter. Qualitatively, many observed features were readily reproduced.

6.2   GENERAL REFLECTIONS

The use of mathematics in biology is tricky. When after extensive analysis, extreme simplifications, and bold assumptions, an analytically traceable formula is found, the mathematician's pitfall is to treat that equation as the truth. Although a quantitative fit is rewarding and insightful (for example Chapter 4), the search for quantitative fits might not be the most useful quest yet. Instead, we would here like to focus on qualitative ideas and concepts that reappear repeatedly in multiple models. We believe that together, Chapters 2–5 offer conceptual tools that go beyond the exact mathematical equations.

For example, the repeated appearance of noise modes with very comparable mathematical form, shape, and interpretation (Chapters 2, 3, and 4 of this thesis, but also Kiviet *et al* [86], and references [40] and [16]) fortifies our trust that such modes are independent of the exact biological wiring or fitted parameters. The noise modes moreover offer a way to reason about noise propagation without the trouble of deriving exact mathematical equations. The findings in Chapter 2, where the model framework was used that ignored regulation, may seem in sharp contrast with the findings in Chapters 4 and 5, where we make explicit that regulation does influence the cell's stochastic dynamics. However, the noise modes offer a way to explain these differences. In their original paper, Kiviet *et al* [86] include a common noise term which directly influenced protein production and the growth rate. This common noise resulted in a noise mode similar in shape to the autogenic mode in Chapter 2, and, notably, also similar to (minus) the regulation mode in Chapter 4. The fitted value for the amplitude of the common noise may thus include a negative part associated with regulation, which went unnoticed in the original study from Kiviet *et al*. In the framework of Chapter 2, negative regulation would decrease the protein's Growth Control Coefficient (see Chapter S2), making it again impossible to disentangle the influence of regulation on the stochastic dynamics.

Another returning concept that is not fundamentally embedded in the exact mathematical equations, but was merely pointed out by them, is the net effect of noise in the expression of 'background' proteins that together influence metabolism, the growth rate, and the concentration of a specific protein of interest. These concepts are hard to pinpoint in experiments, although similar ideas were investigated in experiments (and models) discussed in references [31, 118], where solely the entropy of all metabolic reactions resulted in observed distributions of the growth rate and metabolic fluxes. Additionally, recent work argues that noise in the concentrations of metabolic proteins might be selected for throughout evolution, as such expression noise increases metabolic noise, and leads to diverse phenotypes that are Pareto optimal [48].

Lastly, the GCCs, derived in Chapter 2 and further analysed in Chapter 3, are useful concepts even if their exact numerical value cannot easily be determined. The coefficients summarise and combine intuitions about rate-limitation, noise propagation, and evolutionary optimisation. Possible relations between those three concepts are non-trivial and were thus far often overlooked. An understanding of GCCs and their sum rule can moreover be useful indirectly. For example, the realisations that (*i*) proteins can have a non-zero GCC even in evolved, optimal systems and (*ii*) proteins can also have a negative control over the growth rate, were fundamental in the parametrisation of the models discussed in Chapter 4.

6.3 HOLISTIC CELL MODELS: ROLES OF GLOBAL CELLULAR CONSTRAINTS

A recurring theme in almost all chapters has been the importance of (global) cellular constraints. On the population-level, the necessity to view any part of the cell in the context of the whole cell has become increasingly clear. Such a view of a highly connected cell, where due to the many constraints all parts are related, is sometimes referred to as a 'holistic cell principle' [44, 120, 153]. A prime example is the tight connection between growth rate and the expression of almost all genes according to the growth laws [68, 158, 198]. In the interpretation of a classical knock-out experiment the distinction should be made between a direct effect of the knocked-out gene, and any effects that co-occur with a change in growth rate.

At the population level, the cellular constraints that underlie the cell's connectedness are well studied and adequately experimentally tested. However, our models suggest that, if present at the single-cell level, the global constraints also shape single-cell noise propagation properties. In that sense, understanding cellular noise again asks for a *holistic* approach: noise in the expression of all protein species, noise in any part of the cell, leaves traces in the stochasticity of any other aspect of the cell.

6.3.1 *Density constraints*

At the population-level, cellular density is remarkably constraint. Bacterial cells strive for homeostasis, since (the speed of) intracellular reactions are highly dependent on the cellular density [9, 90, 126, 142]. Again, homeostatic mechanisms that ensure a fixed, population-level density are expected to be present at the single cell level. In mammalian cells, tiny density fluctuations on very short timescales have been observed [119], indicating that in principle, cellular density could be regulated on short timescales, at the single-cell level.

In our model of Chapter 2, the density constraint caused the *autogenic* mode: the production of proteins leads to mass growth, and via the density constraint also to volume growth. The question remains if, in bacterial cells, the density is constrained to the level that volume increases any time a protein is produced, and, if so, on which timescale the volume lags behind. In our model we assumed an extreme limit: the volume instantaneously follows the production of every single protein that is translated. Still, we know that the cellular volume does react to mass-growth, such that some trace of the autogenic mode is expected to be present in real cells, even if in reality the density constraint is not as strict as in our model.

6.3.2   *Resource allocation constraints*

Mounting evidence suggests that resource allocation plays a fundamental role in many observed phenomena, such as the appearance of the growth laws [128, 157, 198] and overflow metabolism [5, 27, 57, 117]. Recently, resource allocation was even shown to play a regulatory role by passively suppressing anabolic protein expression upon carbon limitation [95]. At the population level, it is moreover known that steady state exponential growth rates decline when the expression of burdensome, useless proteins is increased.

However, to our knowledge, such resource allocation constraints have not been shown to be relevant for the dynamics of a single protein species. Nor is the presence of the resource allocation constraint experimentally observed at the single cell level. In collaboration with Andreas Milias-Argeitis and Floor Schukking from the Rijks Universiteit Groningen we therefore examined whether fluctuations in the expression of a fluorescent reporter could transfer to the growth rate, not due to toxicity effects, but due to the reporter's increased claim on limited resources. In the experiments, *E. coli* was grown in batch culture, and gene expression noise was induced via an opto-genetic construct, meaning that the reporter's expression was controlled by the strength of an external light source (Fig. 6.1A, [6, 125, 151]). Preliminary results (Fig. 6.1B and C) indicated that a relation indeed exists between the light flux, reporter concentration and the growth rate. Moreover, when the light source is turned off again, cells seem to return to their normal growth rate. These experiments might indicate that resource allocation can also dynamically affect the growth rate of population of cells. So far, similar experiments at the single-cell level did not yield conclusive results. Whether the resource-allocation constraint is present at the single cell level, and whether this constraint is also felt by the stochastic fluctuations of individual protein species, is therefore not yet established. Still, in the models of Chapters 2 and 3 it is exactly this constraint that results in the negative GCCs (and therewith noise transfer!) of non-growth-related proteins.

6.3.3   *Growth Control Coefficients and Sum Rules*

The GCCs from Chapters 2 and 3 bring additional constraints to light. Since the sum of all the GCCs is shown to be 0, and the GCC of non-growth-related proteins (H-sector) is calculated to be negative, the GCC of at least one metabolic protein species is constrained to be positive. This has massive, general implications for noise propagation from the level of gene expression to growth. Fundamentally, noise propagation seems to be unavoidable; even in evolutionarily optimised cells fluctuations in protein expression influences the instantaneous growth rate. Expression titration experiments, however, sometimes show a clear optimal expression level, where growth rates are maximal for the wild-type expression level and either an increased or decreased expression level seems to negatively influence the growth rate (similar to Chapters 3 and 4, see also [22, 141, 181]). Naively, the GCCs of protein species that lie at the top of such an optimum curve are expected to be zero. Still, such an optimum does not necessarily imply that the protein under consideration does not propagate noise to the growth rate: it is possible that in the titration-experiments, which only measure steady-state growth rates, the cell keeps the H-sector, via (auto)regulation, constant. Stochastic fluctuations around each steady state might still propagate to the growth
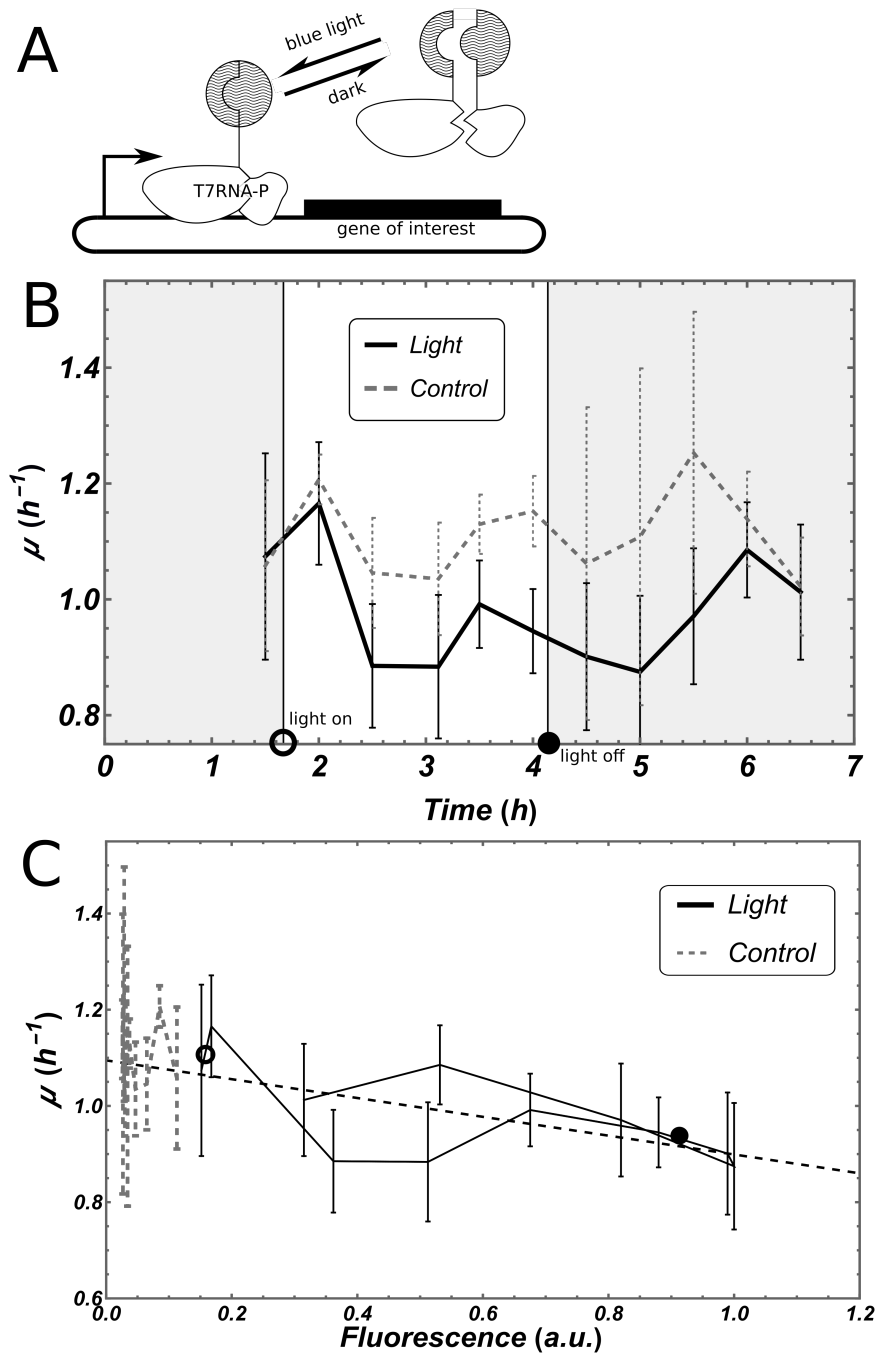
Figure 6.1: (A) Cartoon of the opto-genetic construct used, where two sides of a spliced T7 RNA Polymerase (foreign for *E. coli*) were fused to light-inducible dimerization domains via linkers at the split site [6, 7]. (B) Preliminary data of a population of cells with an optogenetic construct growing fully in the dark (control) or growing under a switching light source (open circle: light on, black dot: light off). A causal correlation between light and fluorescence (B), as well as between fluorescence and growth rate (C) seems to be observed. Growth rate is determined according to OD measurements, while at each time point samples were taken whose fluorescence was analysed using a flow cytometer. Data courtesy of Floor Schukking.

rate, because then the fluctuations in expression level would be partly at the expense of the H sector.

The sum rule constraint is particularly interesting in the field of synthetic biology. Expressing synthetic pathways that are, from the perspective of the cell, undesirable, increases the size of the H-sector. In turn, this increases the amplitude of all GCCs via the sum rule, resulting in a stronger noise propagation from the expression level to the growth rate. In recent articles, interest has risen for potential (functional) roles of stochasticity in multiple biotechnological applications [2, 3, 62, 79, 143].

Lastly, it is worth mentioning that Chapter 3 hinted at an extra trade-off between evolutionary optimisation and the GCCs: the cell cannot escape a certain distribution of the GCCs, if it wants to grow optimally fast.

6.3.4 *Other constraints*

Once started with the principle of a holistic cell, it is hard to imagine which constraints would not affect the cell's physiology and stochastic dynamics. Although we included numerous constraints in this thesis, many more constraints were not considered. A constraint discussed in recent literature is that of a maximum number of transporters that fit on the cell membrane, without hindering its structural stability [169, 205]. This constraint has a large potential impact on metabolism and the metabolic fluxes that proteins carry [56, 57]. To an extent, such a constraint might also influence noise propagation properties.

Throughout this work we ignored any toxic effects of high internal (metabolite) concentrations. Toxicity effects are widespread and constrain the total desirable concentration of multiple metabolites. Moreover, toxicity possibly influences the GCCs of particular proteins, having a direct effect on noise propagation properties.

Countless studies show (other) trade-offs or Pareto-optimality fronts [10, 13, 22, 48, 84, 96, 110, 118, 184, 193], which are all, in one way or another, effects from constraints. Therefore, the holistic realisation that cellular constraints also shape the stochastic properties of cells, is likely to become increasingly important.

6.4   EVOLUTION OF CELL MODELS

While the work on this thesis was in progress, quantitative models concerning balanced population growth, as well as on the single-cell level took a huge leap forward. First, the growth laws offered tools to understand the behaviour of a population of cells during steady state exponential growth [128, 158, 181, 198]. Soon after, focus was shifted to how a population's growth rate and expression patterns change during nutrient up or downshifts, when the cells do not necessarily grow in steady state. During such a transition *E. coli* cells actively (and passively) rearrange their proteome, which takes a considerable amount of time. Interestingly, the coarse-grained view of the proteome established by the growth laws again offered interesting insights. Examples of coarse-grained models that shed light on the cell's internal dynamics, as well as on how the proteome sectors are regulated are [43, 98].

At the same time, models have been developed to explain the stochasticity at the single cell. First, research mainly focused on neatly describing all the stochastic processes involved in gene expression, such transcription and translation, protein degradation, transcription factor binding, etc [47, 138, 168]. Today, this is still an active (and heavily mathe-

matical) research area [26, 60, 72, 73, 111, 190]. However, other models appeared that considered noise phenomenologically, *i.e.* they coarse-grained all processes underlying stochastic gene expression into a single abstract noise source. Instead, these models focused on including growth-related feedbacks. Examples are Chapters 2 and 4 in this thesis, but also references [16, 40, 86, 163]. So far, these models describe stochastic fluctuations during steady state exponential growth in a particular, pre-defined environment. The model's noise propagation parameters, which are likely to depend on the nutrient conditions, are fitted in each environment independently, giving the modeller unwanted degrees of freedom.

A reasonable and feasible next step would be to combine the coarse-grained view of the proteome sectors with the coarse-grained view of gene expression noise. Here, a fundamental realisation is that the same networks that manage the population's response to changes in the external environment, are embedded inside each individual cell. At the single-cell level, those regulatory networks are therefore expected to also shape the cell's response to stochastic fluctuations in its *internal* environment. Preliminary steps in creating such a model were made in Chapter 5. We created a model of a growing cells that matches recent dynamical population-level data concerning nutrient up and downshifts [43]. Additionally, the model reproduces not only the growth laws, but also population-level behaviour when the growth laws are experimentally perturbed [181]. For any environment, noise transfer parameters can then possibly be derived by making a local linearisation. Such an approach thus fixes all the noise transfer parameters, leaving only the noise amplitudes and timescales as free parameters. We see that after adding noise, the model indeed reproduces most observations from Chapter 4, with only two parameters.

In contrast to coarse-grained models, full-cell models have also taken large steps forward in the last years. Together with the frameworks described in this thesis, full-scale cell models that include stochasticity might become feasible in the future.

6.4.1 *Concluding remarks*

The unavoidable stochastic nature of microbes makes it hard to fully understand the functioning of a single cell. Since all creatures, from funghi to humans are build up from cells, an increased understanding of these fundamental building blocks could have its consequences for all life. Quantitative biology, together with applied mathematics, tries, with small steps, to increase intuition and understanding from the bottom up, starting with small molecules, protein building blocks, genes, etc.

However, by studying the cell as a highly stochastic, yet surprisingly robust, self-regulating factory we can learn more than just the fundamentals of life. While trying to understand those tiny bacterial cells, we are bound to learn ingenious general design principles, robust ways of regulating (even in noisy environments), and new ways to deal with, or even utilise, errors.

APPENDIX: COMPARISON BETWEEN OU AND ME

TWO DESCRIPTIONS OF GENE EXPRESSION

As discussed in the introduction, we make a distinction between two ways to model stochastic processes. One option is to describe a stochastic process with rates that are fully determined by the current state of the system (*i*). The other option is to model the process according to a deterministic differential equation, but with a rate on which a stochastic noise sources acts (*ii*).

Here, we compare the means and variances of both options for an example system: the time evolution of the concentration of a protein ϕ that does not in any way influence the growth rate itself. Instead, the growth rate μ is assumed constant. Assuming simple protein production dynamics, where the protein's production rate is either fixed, or has a fixed mean, and the concentration dilutes with the cell's growth rate μ. Such a system might look like, with both the descriptions:

$$(i) \quad \begin{cases} \text{DNA} \xrightarrow{k_1} \text{mRNA} \xrightarrow{k_2} \text{protein } (\phi), \quad \text{mRNA} \xrightarrow{\gamma_1} \emptyset, \quad \text{protein} \xrightarrow{\gamma_2} \emptyset, \\ \frac{\partial p(\phi)}{\partial t} = \frac{\partial [\gamma_2 \phi p(\phi)]}{\partial x} + k_1 \int_0^\phi w(\phi, \phi') p(\phi') d\phi'. \end{cases}$$

(A2.1)

$$(ii) \quad \begin{cases} d\pi_t = -\beta \left(\pi_t - \mathbb{E}\left[\pi \right] \right) dt + \theta \, dW_t, \\ \frac{\partial \phi}{\partial t} = \pi(t) - \mu\phi. \end{cases}$$

(A2.2)

The first set of equations (*i*) form a continuous master equation that describes how the probability of finding a certain protein concentration $\phi = n/V$ changes of time. These equations are derived by Friedman *et al* [47] and further discussed in the Introduction. The second set of equations, (*ii*), describe an Ornstein-Uhlenbeck (OU) process that acts on an ODE. Here W_t is a Wiener process, with the dW_t being independent Gaussian increments with variance t, *i.e.*, $dW_t = \mathcal{N}\left(0, \sqrt{t} \right)$. The parameter β can be interpret as the driving force back towards $\mathbb{E}\left[\pi \right]$, the mean value of the production rate, either by regulation, or simply by relaxation of the fluctuations. Note that the second equation of (*ii*) describes the change in ϕ deterministcally given the stochastic protein production rate $\pi(t)$.

Method (*ii*) will thus yield sample paths of $\phi(t)$ for a particular stochastic path $\pi(t)$, whereas method (*i*) will describe how the probability density function $p(\phi)$ changes of time. Here, we will test to what extent the two methods are comparable by calculating their averages, variances and timescales of fluctuations.

ANALYSING METHOD (*ii*): INTEGRATING THE ORNSTEIN-UHLENBECK PROCESS

To calculate the mean and variance over time from method (*ii*), we must first get an expression for $\phi(t)$ by integrating the second equation and inserting $\pi(t)$, the integral of the OU process:

$$\phi(t) = e^{-\mu t}\left(\phi(0) + \int_0^t e^{\mu\tau}\pi(\tau)d\tau\right) \tag{A2.3}$$

$$= e^{-\mu t}\left(\phi(0) + \int_0^t e^{\mu\tau}\left[\mathbb{E}\left[\pi\right] + \theta\int_0^\tau e^{-\beta(\tau-s)}dW_s\right]d\tau\right)$$

$$= e^{-\mu t}\phi(0) + e^{-\mu t}\left(\int_0^t e^{\mu\tau}\mathbb{E}\left[\pi\right]d\tau + \theta\int_0^t e^{\mu\tau}\left[\int_0^\tau e^{-\beta(\tau-s)}dW_s\right]d\tau\right)$$

$$= e^{-\mu t}\phi(0) + \frac{\mathbb{E}\left[\pi\right]}{\mu}(1-e^{-\mu t}) + \theta e^{-\mu t}\int_0^t e^{(\mu-\beta)\tau}\left[\int_0^\tau e^{\beta s}dW_s\right]d\tau \tag{A2.4}$$

$$= e^{-\mu t}\phi(0) + \frac{\mathbb{E}\left[\pi\right]}{\mu}(1-e^{-\mu t}) + \theta e^{-\mu t}Y(t). \tag{A2.5}$$

In equation A2.4 we recognise a so-called a Double-Integral Process (DIP), described by Touboul and Faugeras [180]: $Y(t) = \int_0^t e^{(\mu-\beta)\tau}\left[\int_0^\tau e^{\beta s}dW_s\right]d\tau$. Note that the variance of the DIP, $\text{Var}\{Y\}_t$, is not bounded in itself, because it resembles a 1D-Brownian motion diffusion process. The mean of Y, however, remains zero. Therefore, the variance of such a DIP is given by its second moment, calculated in reference [180]:

$$\text{Var}\{Y\}_t = \mathbb{E}\left[Y^2\right]_t = 2\int_0^t e^{(\mu-\beta)\tau}\left[\int_0^\tau e^{(\mu-\beta)s}\int_0^s\left(e^{\beta u}\right)^2 du\, ds\right]d\tau$$

$$= \frac{e^{2\mu t}(\mu-\beta)^2 - e^{2(\mu-\beta)t}\mu(\beta+\mu) + 4e^{(\mu-\beta)t} - \beta(\mu+\beta)}{2\beta(\mu-\beta)^2\mu(\beta+\mu)}. \tag{A2.6}$$

As expected, the variance increases with time. However, in equation A2.4 the DIP is quenched by the term $\theta e^{-\mu t}$, such that the variance of ϕ becomes (for reasonably large t):

$$\text{Var}\{\phi\}_t = \theta^2 e^{-2\mu t}\text{Var}\{Y\}_t. \tag{A2.7}$$

In the limit of $t \to \infty$, this results in:

$$\text{Var}\{\phi\} = \frac{\theta^2}{2\beta\mu(\beta+\mu)}. \tag{A2.8}$$

This variance is notably smaller than the variance of the OU process itself, $\theta^2/(2\beta)$. This seems logical, because the dilution via growth also dilutes the fluctuations. We should also note that, because we can relate ϕ to an (double) integrated normal Wiener process, its distribution must be normal as well. This directly opposes measurements and other theoretical work, where protein concentrations were found to be Gamma-distributed.

Interestingly, equation A2.8 does resemble equation S2.57 from Chapter 2 [87], while no linearisation has been done here.

Predicted scaling of $\text{Var}\{\phi\}$ *upon changes of* μ

An interesting question would be, how the noise in protein concentrations changes for different growth rates. The earlier model by Friedman *et al* [47] predicts an increase of

the CV of ϕ with the growth rate under the assumption that with larger growth rates the (relative) burst frequency decreases, while the mean and variance of the burst size increases to maintain the same steady state. Note that the assumption that $\overline{\phi} = \mathbb{E}[\pi]/\mu$ is a constant, corresponds to the assumption that $\mathbb{E}[\pi] \propto \mu$ (as long as $\mathrm{Cov}[\phi, \mu] \approx 0$), which does not seem unreasonable. Then, any increase in CV_{ϕ} would solely be due to an increase in $\mathrm{Var}\{\phi\}$. However, in our framework, *(ii)*, $\mathrm{Var}\{\phi\}$ decreases, rather than increases, with μ (equation A2.8). In conclusion, for our framework to match the model predictions of [47], either the parameter θ or β should change together with μ.

ANALYSING METHOD (*i*): DERIVING A SDE FROM FRIEDMAN *et al*

First, the steady state distribution derived by Friedman *et al* [47] allows us to link parameters from method *(ii)* to method *(i)* using the mean and variance. Let $a = k_1/\gamma_2$, the mean number of mRNA molecules produced per cell cycle, and let $b = k_2/\gamma_1$, the mean number of protein molecules produced per burst.

$$\langle\phi\rangle_{(i)} = ab = \frac{k_1}{\gamma_2}\frac{k_2}{\gamma_1} = \left(\frac{k_1 k_2}{\gamma_1}\right)\frac{1}{\mu} = \mathbb{E}[\pi]/\mu, \qquad \text{(setting } \gamma_2 = \mu) \qquad \text{(A2.9)}$$

$$\langle\phi\rangle_{(ii)} = \mathbb{E}[\pi]/\mu, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A2.10)}$$

$$\sigma^2_{(i)} = ab^2 = \frac{k_1}{\mu}\left(\frac{k_2}{\gamma_1}\right)^2, \qquad\qquad\qquad\qquad\qquad\qquad \text{(A2.11)}$$

$$\sigma^2_{(ii)} = \frac{\theta^2}{2\beta\mu(\beta+\mu)}. \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A2.12)}$$

However, the timescale of fluctuations (how quickly does a fluctuation dissipate) of both methods are not easily compared, because the master equation from Friedman *et al* is only solvable for the stationary distribution, yielding no information about relaxation times. Therefore, we first transform their equation to a Fokker-Planck equation, and then to a Stochastic Differential Equation, in order to try and find the corresponding timescales of fluctuations.

Starting from the equations written by Friedman *et al*, we derive a SDE to be able to calculate auto-correlation times of fluctuations. First, we use the Kramer-Moyer expansion [78] to write the Fokker-Planck equation that is the corresponding approximation of the (*i*):

$$\frac{\partial p(\phi)}{\partial t} = \frac{\partial \gamma_2 \phi p(\phi)}{\partial \phi} + k_1 \int_0^\phi w(\phi - \phi')p(\phi')d\phi' \qquad\qquad \text{(A2.13)}$$

$$\approx \frac{\partial \gamma_2 \phi p(\phi)}{\partial \phi} + k_1 \left(\frac{\partial}{\partial \phi}\left[\int_0^\phi (\phi - \phi')w(\phi - \phi')d\phi' p(\phi)\right] - \frac{\partial^2}{2\partial \phi^2}\left[\int_0^\phi (\phi - \phi')^2 w(\phi - \phi')d\phi' p(\phi)\right]\right)$$

$$\text{(A2.14)}$$

$$= \frac{\partial \gamma_2 \phi p(\phi)}{\partial \phi} - k_1\frac{\partial}{\partial \phi}\left[\left(b - (x + b)e^{-\phi/b}\right)p(\phi)\right] + \frac{k_1}{2}\frac{\partial^2}{\partial \phi^2}\left[\left(2b^2 - (b^2 + (\phi + b)^2 e^{-\phi/b})\right)p(\phi)\right]$$

$$\text{(A2.15)}$$

$$= \frac{\partial\left(\gamma_2\phi - k_1 b\right)p(\phi)}{\partial \phi} + \frac{1}{2}\frac{\partial^2\left(k_1 b^2\right)p(\phi)}{\partial \phi^2}. \qquad \text{(for } \phi \gg b) \qquad \text{(A2.16)}$$

Here, $w(\phi - \phi') = \frac{1}{b}e^{-(\phi-\phi')/b} - \delta(\phi - \phi')$ is the exponentially distributed burst size with mean b. The delta function appears in the function w to ensure normalisation of the

probability (with a total rate of k_1 probability leaks from state ϕ). The corresponding SDE is then, again for $\phi >> b$ [78]:

$$d\phi_t = -\gamma_2 \left(\phi - \frac{k_1 b}{\gamma_2} \right) dt + \sqrt{2k_1}b \; dW_t = -\gamma_2 \left(\phi - \langle \phi \rangle_{(i)} \right) dt + \sqrt{2\sigma_{(i)}^2 \; \mu} \; dW_t. \quad \text{(A2.17)}$$

The last equation now describes the dynamics of ϕ as an OU process that still has the same mean $\langle \phi \rangle = \frac{\mathbb{E}[\pi]}{\mu}$ and variance $\sigma_{(i)}^2$ as framework (*i*), and shows that such a system has a single timescale $\gamma_2 \approx \mu$.

In conclusion, the two approaches are fundamentally different. First, when the OU process is applied to the production rate, this introduces a new, independent timescale, β. Second, only the approximation of method (*i*) is Guassian. (The stationairy distribution found in Friedman *et al* is a Gamma-distribution). Third, the approximation of method (*i*) with a SDE is in itself an OU process, while in method (*ii*) the OU process acts on the rate, instead of ϕ directly.

Deze thesis beschrijft de groei van bacteriën, en dan met name *Escherichia coli* (*E. coli*). *E. coli* is een eencellig organisme dat veel voorkomt in de darmen van zoogdieren, waaronder ook mensen. Daar helpen de bacteriën ons met het verteren van het voedsel dat wij eten. Bacteriën maken eiwitten aan waarmee ze voedsel uit hun omgeving opnemen en omzetten tot biomassa. Op elk moment zijn er ongeveer 1000 verschillende soorten eiwitten in een bacterie aanwezig, die elk hun specifieke taak verrichten. Daardoor kan een bacterie in volume groeien, waarna hij bij een bepaalde grootte in tweeën deelt. De twee dochters groeien vervolgens verder en delen zelf na een tijdje ook weer. Snel groeien heeft een groot voordeel: bacteriën die sneller groeien verdringen diegene die langzamer groeien. Het blijkt dat *E. coli* inderdaad goed kan groeien op veel verschillende voedselbronnen. Dat doen ze door slim te reguleren welke eiwitten ze in welke concentratie produceren. Zo kunnen bacteriën in allerlei verschillende omgevingen efficiënt en snel voedsel opnemen en verwerken.

Toch kunnen bacteriën niet ontkomen aan een bepaalde willekeur. Ze zijn zo klein dat van sommige moleculen maar tientallen door hun cytoplasma drijven. Een chemische reactie, bijvoorbeeld om een nieuw eiwit te maken, vindt pas plaats als twee moleculen op de juiste manier tegen elkaar botsen. Voor *E. coli* heeft deze onzekerheid in het aanmaken van eiwitten grote gevolgen: de concentraties van alle verschillende eiwitten fluctueren door de tijd heen. We noemen die willekeur vaak 'stochasticiteit' of 'ruis'. Uit metingen blijkt dat zelfs de snelheid waarmee het volume van bacteriën toeneemt, niet constant is en tot op zekere hoogte stochastisch/willekeurig is. Zelfs in een sterk gecontroleerde laboratorium omgeving fluctueert de groeisnelheid, terwijl bacteriën juist proberen om die groeisnelheid zo hoog mogelijk te maken!

In hoofdstuk 2 rekenen we met behulp van een wiskundig model uit hoe de willekeur in de timing van het aanmaken van alle eiwitten uiteindelijk ervoor zorgt dat ook de groeisnelheid gaat fluctueren. We moeten dan (o.a.) rekening houden met dat de groei van het volume van de bacterie weer invloed heeft op de concentratie van alle eiwitten; concentratie is immers het aantal deeltjes per volume! Om de berekening mogelijk te maken koppelen we aan elke eiwit-soort een getal: de Growth Control Coefficient (groei-controle coëfficiënt). Dit getal drukt uit hoe sterk een fluctuatie in de concentratie van die eiwit-soort effect heeft op de groeisnelheid. We tonen aan dat de gemeten stochastische dynamica van *E. coli* verklaard kan worden door de gezamenlijke ruis in alle eiwitten. In hoofdstuk 3 onderzoeken we de vraag: als de ruis in eiwitten ervoor zorgt dat de groeisnelheid gaat schommelen, zijn er dan specifieke eiwit soorten die hierin een belangrijke rol spelen? Eerder onderzoek toonde aan dat de concentratie van eiwit-soorten met een lage gemiddelde concentratie, relatief sterk schommelt. De stochasticiteit van eiwit-soorten met gemiddeld een hoge concentratie leek daarentegen juist verwaarloosbaar. Daardoor wordt doorgaans aangenomen dat fluctuaties in de groeisnelheid vooral veroorzaakt worden door de (sterke) schommelingen van de eiwit-soorten met een lage gemiddelde concentratie. Echter, in dit hoofdstuk tonen we juist het tegenovergestelde aan: de eiwitten met een hoge gemiddelde concentratie dragen veel bij aan de ruis in de groeisnelheid. Dat komt omdat niet alleen de amplitude

van de fluctuaties belangrijk is, maar ook hoe gevoelig de groeisnelheid is voor de fluctuaties in die specifieke eiwit-soort. Oftewel: de Growth Control Coefficients van het vorige hoofdstuk zijn ook hier belangrijk.

Tot nu toe hebben we genegeerd dat *E. coli* zijn best doet om zo goed mogelijk te groeien in een bepaalde omgeving en daarvoor 'actief' de concentratie van eiwitten reguleert. In hoofdstukken 4 en 5 onderzoeken we dit samenspel van regulatie en stochasticiteit. We gebruiken daarvoor observaties van een experiment waarbij *E. coli* bacteriën groeien in een constante omgeving. Het experiment vergelijkt ruis in gezonde bacteriën met ruis in bacteriën die door de onderzoekers genetisch aangepast zijn. Deze aangepaste bacteriën kunnen niet goed meer hun eiwit concentraties reguleren, maar groeien gemiddeld genomen nog wel steeds even snel. Er blijkt een verschil te zijn in de ruis-eigenschappen van de gezonde en de 'niet-regulerende' bacteriën: het moet dus wel zo zijn dat de regulatie die een bacterie gebruikt om zich aan te passen aan veranderingen in de omgeving, tegelijkertijd óók invloed heeft op de fluctuaties in eiwit concentraties en de groeisnelheid in een constante omgeving. Met een wiskundig model kunnen we de data verklaren en bevestigen we dat het verschil in ruis inderdaad veroorzaakt wordt doordat de aangepaste bacteriën niet goed kunnen reguleren.

BIBLIOGRAPHY

[1] L. Arike, K. Valgepea, L. Peil, R. Nahku, K. Adamberg, and R. Vilu. "Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli." en. In: *Journal of Proteomics* 75.17 (Sept. 2012), pp. 5437–5448. ISSN: 1874-3919. DOI: 10.1016/j.jprot.2012.06.020.

[2] Mark S. Aronson, Chiara Ricci-Tam, Xinwen Zhu, and Allyson E. Sgro. "Exploiting noise to engineer adaptability in synthetic multicellular systems." en. In: *Current Opinion in Biomedical Engineering*. Molecular & Cellular Engineering; Engineered multi-cellular systems 16 (Dec. 2020), pp. 52–60. ISSN: 2468-4511. DOI: 10.1016/j.cobme.2020.100251.

[3] Lucia Bandiera, Simone Furini, and Emanuele Giordano. "Phenotypic Variability in Synthetic Biology Applications: Dealing with Noise in Microbial Gene Expression." In: *Frontiers in Microbiology* 7 (2016), p. 479. ISSN: 1664-302X. DOI: 10.3389/fmicb.2016.00479.

[4] Arren Bar-Even, Johan Paulsson, Narendra Maheshri, Miri Carmi, Erin O'Shea, Yitzhak Pilpel, and Naama Barkai. "Noise in protein expression scales with natural protein abundance." eng. In: *Nature Genetics* 38.6 (June 2006), pp. 636–643. ISSN: 1061-4036. DOI: 10.1038/ng1807.

[5] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. "Overflow metabolism in Escherichia coli results from efficient proteome allocation." en. In: *Nature* 528.7580 (Dec. 2015), pp. 99–104. ISSN: 1476-4687. DOI: 10.1038/nature15765.

[6] Armin Baumschlager, Stephanie K. Aoki, and Mustafa Khammash. "Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control." In: *ACS Synthetic Biology* 6.11 (Nov. 2017), pp. 2157–2167. DOI: 10.1021/acssynbio.7b00169.

[7] Armin Baumschlager and Mustafa Khammash. "Synthetic Biological Approaches for Optogenetics and Tools for Transcriptional Light-Control in Bacteria." en. In: *Advanced Biology* 5.5 (2021), p. 2000256. ISSN: 2701-0198. DOI: 10.1002/adbi.202000256.

[8] Fredrik Bäckhed, Ruth E. Ley, Justin L. Sonnenburg, Daniel A. Peterson, and Jeffrey I. Gordon. "Host-Bacterial Mutualism in the Human Intestine." en. In: *Science* 307.5717 (Mar. 2005), pp. 1915–1920. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1104816.

[9] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, and Z. N. Oltvai. "Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity." eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 104.31 (July 2007), pp. 12663–12668. ISSN: 0027-8424. DOI: 10.1073/pnas.0609845104.

[10]   Jan Berkhout, Evert Bosdriesz, Emrah Nikerel, Douwe Molenaar, Dick de Ridder, Bas Teusink, and Frank J. Bruggeman. "How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism." eng. In: *Genetics* 194.2 (June 2013), pp. 505–512. ISSN: 1943-2631. DOI: 10.1534/genetics.113.150631.

[11]   Jonathan A. Bernstein, Arkady B. Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N. Cohen. "Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays." en. In: *Proceedings of the National Academy of Sciences* 99.15 (July 2002), pp. 9697–9702. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.112318199.

[12]   S. Berthoumieux, H. de Jong, G. Baptist, C. Pinel, C. Ranquet, D. Ropers, and J. Geiselmann. "Shared control of gene expression in bacteria by transcription factors and global physiology of the cell." en. In: *Molecular Systems Biology* 9.1 (Apr. 2014), pp. 634–634. ISSN: 1744-4292. DOI: 10.1038/msb.2012.70.

[13]   Elena Biselli, Severin Josef Schink, and Ulrich Gerland. "Slower growth of Escherichia coli leads to longer survival in carbon starvation due to a decrease in the maintenance rate." In: *Molecular Systems Biology* 16.6 (June 2020), e9478. ISSN: 1744-4292. DOI: 10.15252/msb.20209478.

[14]   David F. Blair. "How Bacteria Sense and Swim." In: *Annual Review of Microbiology* 49.1 (1995), pp. 489–520. DOI: 10.1146/annurev.mi.49.100195.002421.

[15]   Olivier Borkowski, Francesca Ceroni, Guy-Bart Stan, and Tom Ellis. "Overloaded and stressed: whole-cell considerations for bacterial synthetic biology." en. In: *Current Opinion in Microbiology*. Antimicrobials • Microbial systems biology 33 (Oct. 2016), pp. 123–130. ISSN: 1369-5274. DOI: 10.1016/j.mib.2016.07.009.

[16]   A. Borri, P. Palumbo, and A. Singh. "Noise propagation in metabolic pathways: the role of growth-mediated feedback." In: *2020 59th IEEE Conference on Decision and Control (CDC)*. Dec. 2020, pp. 4610–4615. DOI: 10.1109/CDC42340.2020.9303911.

[17]   Evert Bosdriesz, Douwe Molenaar, Bas Teusink, and Frank J. Bruggeman. "How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization." en. In: *The FEBS Journal* 282.10 (2015), pp. 2029–2044. ISSN: 1742-4658. DOI: 10.1111/febs.13258.

[18]   Clive G. Bowsher and Peter S. Swain. "Identifying sources of variation and the flow of information in biochemical networks." en. In: *Proceedings of the National Academy of Sciences* 109.20 (May 2012), E1320–E1328. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1119407109.

[19]   Matthew J. Brauer, Curtis Huttenhower, Edoardo M. Airoldi, Rachel Rosenstein, John C. Matese, David Gresham, Viktor M. Boer, Olga G. Troyanskaya, and David Botstein. "Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast." In: *Molecular Biology of the Cell* 19.1 (Jan. 2008), pp. 352–367. ISSN: 1059-1524. DOI: 10.1091/mbc.e07-08-0779.

[20]   Hans Bremer and Patrick P. Dennis. "Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates." In: *EcoSal Plus* 3.1 (Oct. 2008). DOI: 10.1128/ecosal.5.2.3.

[21] Frank J. Bruggeman, Nils Blüthgen, and Hans V. Westerhoff. "Noise Management by Molecular Networks." en. In: *PLOS Computational Biology* 5.9 (Sept. 2009), e1000506. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000506.

[22] Frank J Bruggeman, Robert Planqué, Douwe Molenaar, and Bas Teusink. "Searching for principles of microbial physiology." In: *FEMS Microbiology Reviews* 44.6 (Nov. 2020), pp. 821–844. ISSN: 0168-6445. DOI: 10.1093/femsre/fuaa034.

[23] Frank J. Bruggeman and Bas Teusink. "Living with noise: On the propagation of noise from molecules to phenotype and fitness." en. In: *Current Opinion in Systems Biology*. • Regulatory and metabolic networks • Special Section: Single cell and noise 8 (Apr. 2018), pp. 144–150. ISSN: 2452-3100. DOI: 10.1016/j.coisb.2018.02.010.

[24] Andrea K. Bryan, Vivian C. Hecht, Wenjiang Shen, Kristofor Payer, William H. Grover, and Scott R. Manalis. "Measuring single cell mass, volume, and density with dual suspended microchannel resonators." en. In: *Lab on a Chip* 14.3 (Dec. 2013), pp. 569–576. ISSN: 1473-0189. DOI: 10.1039/C3LC51022K.

[25] Long Cai, Nir Friedman, and X. Sunney Xie. "Stochastic protein expression in individual cells at the single molecule level." en. In: *Nature* 440.7082 (Mar. 2006), pp. 358–362. ISSN: 0028-0836. DOI: 10.1038/nature04599.

[26] Zhixing Cao and Ramon Grima. "Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells." en. In: *Proceedings of the National Academy of Sciences* 117.9 (Mar. 2020), pp. 4682–4692. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1910888117.

[27] Yu Chen and Jens Nielsen. "Mathematical modeling of proteome constraints within metabolism." en. In: *Current Opinion in Systems Biology* 25 (Mar. 2021), pp. 50–56. ISSN: 2452-3100. DOI: 10.1016/j.coisb.2021.03.003.

[28] JM Conly and K Stein. "Quantitative and qualitative measurements of K vitamins in human intestinal contents." In: *American Journal of Gastroenterology (Springer Nature)* 87.3 (1992).

[29] Jorge Fernandez-de Cossio-Diaz, Roberto Mulet, and Alexei Vazquez. "Cell population heterogeneity driven by stochastic partition and growth optimality." en. In: *Scientific Reports* 9.1 (June 2019), pp. 1–7. ISSN: 2045-2322. DOI: 10.1038/s41598-019-45882-w.

[30] Elizabeth K. Costello, Christian L. Lauber, Micah Hamady, Noah Fierer, Jeffrey I. Gordon, and Rob Knight. "Bacterial Community Variation in Human Body Habitats Across Space and Time." en. In: *Science* 326.5960 (Dec. 2009), pp. 1694–1697. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1177486.

[31] Daniele De Martino, Anna MC Andersson, Tobias Bergmiller, Călin C. Guet, and Gašper Tkačik. "Statistical mechanics for metabolic networks during steady state growth." en. In: *Nature Communications* 9.1 (July 2018), p. 2988. ISSN: 2041-1723. DOI: 10.1038/s41467-018-05417-9.

[32] Erez Dekel and Uri Alon. "Optimality and evolutionary tuning of the expression level of a protein." en. In: *Nature* 436.7050 (July 2005), pp. 588–592. ISSN: 0028-0836. DOI: 10.1038/nature03842.

[33]   Arnold L Demain. "Microbial biotechnology." en. In: *Trends in Biotechnology* 18.1 (Jan. 2000), pp. 26–31. ISSN: 0167-7799. DOI: 10.1016/S0167-7799(99)01400-6.

[34]   Arnold L. Demain and Jose L. Adrio. "Contributions of Microorganisms to Industrial Biology." en. In: *Molecular Biotechnology* 38.1 (Dec. 2007), p. 41. ISSN: 1559-0305. DOI: 10.1007/s12033-007-0035-z.

[35]   Les Dethlefsen, Margaret McFall-Ngai, and David A. Relman. "An ecological and evolutionary perspective on human–microbe mutualism and disease." en. In: *Nature* 449.7164 (Oct. 2007), pp. 811–818. ISSN: 1476-4687. DOI: 10.1038/nature06245.

[36]   Josef Deutscher. "The mechanisms of carbon catabolite repression in bacteria." en. In: *Current Opinion in Microbiology*. Cell Regulation 11.2 (Apr. 2008), pp. 87–93. ISSN: 1369-5274. DOI: 10.1016/j.mib.2008.02.007.

[37]   Stefano Donati, Timur Sander, and Hannes Link. "Crosstalk between transcription and metabolism: how much enzyme is enough for a cell?" en. In: *WIREs Systems Biology and Medicine* 10.1 (2018), e1396. ISSN: 1939-005X. DOI: 10.1002/wsbm.1396.

[38]   Hugo Dourado and Martin J. Lercher. "An analytical theory of balanced cellular growth." en. In: *Nature Communications* 11.1 (Mar. 2020), pp. 1–14. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14751-w.

[39]   Yann Dublanche, Konstantinos Michalodimitrakis, Nico Kümmerer, M. Foglierini, and L. Serrano. "Noise in transcription negative feedback loops: simulation and experimental analysis." In: *Molecular systems biology* (2006). DOI: 10.1038/msb4100081.

[40]   Mary J. Dunlop, Robert Sidney Cox, Joseph H. Levine, Richard M. Murray, and Michael B. Elowitz. "Regulatory activity revealed by dynamic correlations in gene expression noise." en. In: *Nature Genetics* 40.12 (Dec. 2008), pp. 1493–1498. ISSN: 1061-4036. DOI: 10.1038/ng.281.

[41]   Nils Eling, Michael D. Morgan, and John C. Marioni. "Challenges in measuring and understanding biological noise." en. In: *Nature Reviews Genetics* 20.9 (Sept. 2019), pp. 536–548. ISSN: 1471-0064. DOI: 10.1038/s41576-019-0130-6.

[42]   Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. "Stochastic Gene Expression in a Single Cell." en. In: *Science* 297.5584 (Aug. 2002), pp. 1183–1186. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1070919.

[43]   David W. Erickson, Severin J. Schink, Vadim Patsalo, James R. Williamson, Ulrich Gerland, and Terence Hwa. "A global resource allocation strategy governs growth transition kinetics of Escherichia coli." en. In: *Nature* 551.7678 (Nov. 2017), pp. 119–123. ISSN: 1476-4687. DOI: 10.1038/nature24299.

[44]   Ferric C. Fang and Arturo Casadevall. "Reductionistic and Holistic Science." In: *Infection and Immunity* 79.4 (Apr. 2011), pp. 1401–1404. DOI: 10.1128/IAI.01343-10.

[45]   D A Fell. "Metabolic control analysis: a survey of its theoretical and experimental development." In: *Biochemical Journal* 286.2 (Sept. 1992), pp. 313–330. ISSN: 0264-6021. DOI: 10.1042/bj2860313.

[46] E. Fic, P. Bonarek, A. Gorecki, S. Kedracka-Krok, J. Mikolajczak, A. Polit, M. Tworzydlo, M. Dziedzicka-Wasylewska, and Z. Wasylewski. "cAMP Receptor Protein from Escherichia coli as a Model of Signal Transduction in Proteins – A Review." In: *Microbial Physiology* 17.1 (2009), pp. 1–11. ISSN: 2673-1665, 2673-1673. DOI: 10.1159/000178014.

[47] Nir Friedman, Long Cai, and X. Sunney Xie. "Linking stochastic dynamics to population distribution: an analytical framework of gene expression." eng. In: *Physical Review Letters* 97.16 (Oct. 2006), p. 168302. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.97.168302.

[48] Diego Antonio Fernandez Fuentes, Pablo Manfredi, Urs Jenal, and Mattia Zampieri. "Pareto optimality between growth-rate and lag-time couples metabolic noise to phenotypic heterogeneity in Escherichia coli." en. In: *Nature Communications* 12.1 (May 2021), p. 3204. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23522-0.

[49] Luca Gerosa, Bart R. B. Haverkorn van Rijsewijk, Dimitris Christodoulou, Karl Kochanowski, Thomas S. B. Schmidt, Elad Noor, and Uwe Sauer. "Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data." en. In: *Cell Systems* 1.4 (Oct. 2015), pp. 270–282. ISSN: 2405-4712. DOI: 10.1016/j.cels.2015.09.008.

[50] Michel Godin et al. "Using buoyant mass to measure the growth of single cells." en. In: *Nature Methods* 7.5 (May 2010), pp. 387–390. ISSN: 1548-7105. DOI: 10.1038/nmeth.1452.

[51] Anne Goelzer, Vincent Fromion, and Gérard Scorletti. "Cell design in bacteria as a convex optimization problem." en. In: *Automatica*. Special Issue on Systems Biology 47.6 (June 2011), pp. 1210–1218. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2011.02.038.

[52] Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. "Real-Time Kinetics of Gene Activity in Individual Bacteria." en. In: *Cell* 123.6 (Dec. 2005), pp. 1025–1036. ISSN: 0092-8674. DOI: 10.1016/j.cell.2005.09.031.

[53] David C. Grainger, Hirofumi Aiba, Douglas Hurd, Douglas F. Browning, and Stephen J. W. Busby. "Transcription factor distribution in Escherichia coli : studies with FNR protein." In: *Nucleic Acids Research* 35.1 (Jan. 2007), pp. 269–278. ISSN: 0305-1048. DOI: 10.1093/nar/gkl1023.

[54] Jeffrey Green, Melanie R Stapleton, Laura J Smith, Peter J Artymiuk, Christina Kahramanoglou, Debbie M Hunt, and Roger S Buxton. "Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches." en. In: *Current Opinion in Microbiology*. Cell regulation 18 (Apr. 2014), pp. 1–7. ISSN: 1369-5274. DOI: 10.1016/j.mib.2014.01.003.

[55] Boris Görke and Jörg Stülke. "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." en. In: *Nature Reviews Microbiology* 6.8 (Aug. 2008), pp. 613–624. ISSN: 1740-1534. DOI: 10.1038/nrmicro1932.

[56] Daan H. de Groot, Coco van Boxtel, Robert Planqué, Frank J. Bruggeman, and Bas Teusink. "The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates." en. In: *PLOS Computational Biology* 15.3 (Mar. 2019), e1006858. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006858.

[57] Daan H. de Groot, Julia Lischke, Riccardo Muolo, Robert Planqué, Frank J. Brugge-man, and Bas Teusink. "The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints." en. In: *Cellular and Molecular Life Sciences* 77.3 (Feb. 2020), pp. 441–453. ISSN: 1420-9071. DOI: 10.1007/s00018-019-03380-2.

[58] William H. Grover, Andrea K. Bryan, Monica Diez-Silva, Subra Suresh, John M. Higgins, and Scott R. Manalis. "Measuring single-cell density." en. In: *Proceedings of the National Academy of Sciences* 108.27 (July 2011), pp. 10992–10996. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1104651108.

[59] Chinmaya Gupta, José Manuel López, Robert Azencott, Matthew R. Bennett, Krešimir Josić, and William Ott. "Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations." In: *The Journal of Chemical Physics* 140.20 (May 2014), p. 204108. ISSN: 0021-9606. DOI: 10.1063/1.4878662.

[60] Lucy Ham, David Schnoerr, Rowan D. Brackston, and Michael P. H. Stumpf. "Exactly solvable models of stochastic gene expression." In: *The Journal of Chemical Physics* 152.14 (Apr. 2020), p. 144106. ISSN: 0021-9606. DOI: 10.1063/1.5143540.

[61] Mikihiro Hashimoto, Takashi Nozoe, Hidenori Nakaoka, Reiko Okura, Sayo Akiyoshi, Kunihiko Kaneko, Edo Kussell, and Yuichi Wakamoto. "Noise-driven growth rate gain in clonal cellular populations." en. In: *Proceedings of the National Academy of Sciences* 113.12 (Mar. 2016), pp. 3251–3256. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1519412113.

[62] Fei He, Ettore Murabito, and Hans V. Westerhoff. "Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering." In: *Journal of The Royal Society Interface* 13.117 (Apr. 2016), p. 20151046. DOI: 10.1098/rsif.2015.1046.

[63] Reinhart Heinrich and Stefan Schuster. "Metabolic Control Analysis." en. In: *The Regulation of Cellular Systems*. Ed. by Reinhart Heinrich and Stefan Schuster. Boston, MA: Springer US, 1996, pp. 138–291. ISBN: 978-1-4613-1161-4.

[64] Rutger Hermsen, David W. Erickson, and Terence Hwa. "Speed, Sensitivity, and Bistability in Auto-activating Signaling Circuits." In: *PLoS Comput Biol* 7.11 (Nov. 2011), e1002265. DOI: 10.1371/journal.pcbi.1002265.

[65] Rutger Hermsen, Hiroyuki Okano, Conghui You, Nicole Werner, and Terence Hwa. "A growth-rate composition formula for the growth of E.coli on co-utilized carbon substrates." eng. In: *Molecular Systems Biology* 11.4 (Apr. 2015), p. 801. ISSN: 1744-4292.

[66] Sara Hooshangi and Ron Weiss. "The effect of negative feedback on noise propagation in transcriptional gene networks." In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 16.2 (June 2006), p. 026108. ISSN: 1054-1500. DOI: 10.1063/1.2208927.

[67] J. Michael Hudson and Michael G. Fried. "Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter." en. In: *Journal of Molecular Biology* 214.2 (July 1990), pp. 381–396. ISSN: 0022-2836. DOI: 10.1016/0022-2836(90)90188-R.

[68] Sheng Hui, Josh M. Silverman, Stephen S. Chen, David W. Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R. Williamson. "Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria." eng. In: *Molecular Systems Biology* 11.1 (Jan. 2015), p. 784. ISSN: 1744-4292.

[69] Rafael U. Ibarra, Jeremy S. Edwards, and Bernhard O. Palsson. "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth." en. In: *Nature* 420.6912 (Nov. 2002), pp. 186–189. ISSN: 1476-4687. DOI: 10.1038/nature01149.

[70] John L. Ingraham, Ole Maaløe, Frederick C. Neidhardt, and 1931. *Growth of the bacterial cell*. English. Tech. rep. Sinauer Associates, 1983.

[71] Srividya Iyer-Biswas, Charles S. Wright, Jonathan T. Henry, Klevin Lo, Stanislav Burov, Yihan Lin, Gavin E. Crooks, Sean Crosson, Aaron R. Dinner, and Norbert F. Scherer. "Scaling laws governing stochastic growth and division of single bacterial cells." en. In: *Proceedings of the National Academy of Sciences* 111.45 (Nov. 2014), pp. 15912–15917. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1403232111.

[72] Jakub Jędrak, Maciej Kwiatkowski, and Anna Ochab-Marcinek. "Exactly solvable model of gene expression in a proliferating bacterial cell population with stochastic protein bursts and protein partitioning." In: *Physical Review E* 99.4 (Apr. 2019), p. 042416. DOI: 10.1103/PhysRevE.99.042416.

[73] Jakub Jędrak and Anna Ochab-Marcinek. "Contributions to the 'noise floor' in gene expression in a population of dividing cells." en. In: *Scientific Reports* 10.1 (Aug. 2020), p. 13533. ISSN: 2045-2322. DOI: 10.1038/s41598-020-69217-2.

[74] Yujia Jiang, Fengxue Xin, Jiasheng Lu, Weiliang Dong, Wenming Zhang, Min Zhang, Hao Wu, Jiangfeng Ma, and Min Jiang. "State of the art review of biofuels production from lignocellulose by thermophilic bacteria." en. In: *Bioresource Technology*. SI on Advances in Industrial Bioprocesses and Products- Genetic and Metabolic Engineering Interventions 245 (Dec. 2017), pp. 1498–1506. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2017.05.142.

[75] Ding Jun Jin, Cedric Cagliero, and Yan Ning Zhou. "Growth rate regulation in Escherichia coli." In: *FEMS Microbiology Reviews* 36.2 (Mar. 2012), pp. 269–287. ISSN: 0168-6445. DOI: 10.1111/j.1574-6976.2011.00279.x.

[76] Hidde de Jong, Stefano Casagranda, Nils Giordano, Eugenio Cinquemani, Delphine Ropers, Johannes Geiselmann, and Jean-Luc Gouzé. "Mathematical modelling of microbes: metabolism, gene expression and growth." In: *Journal of The Royal Society Interface* 14.136 (Nov. 2017), p. 20170502. DOI: 10.1098/rsif.2017.0502.

[77] H. Kacser, J. A. Burns, H. Kacser, and D. A. Fell. "The control of flux." en. In: *Biochemical Society Transactions* 23.2 (May 1995), pp. 341–366. ISSN: 0300-5127, 1470-8752. DOI: 10.1042/bst0230341.

[78] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. en. Elsevier, Nov. 1992. ISBN: 978-0-08-057138-6.

[79] Behzad D. Karkaria, Neythen J. Treloar, Chris P. Barnes, and Alex J. H. Fedorec. "From Microbial Communities to Distributed Computing Systems." In: *Frontiers in Bioengineering and Biotechnology* 8 (2020), p. 834. ISSN: 2296-4185. DOI: 10.3389/fbioe.2020.00834.

[80] Yoshiaki Kawamura, Ying Li, Hongsheng Liu, Xinxiang Huang, Zhiyu Li, and Takayuki Ezaki. "Bacterial Population in Russian Space Station "Mir"." en. In: *Microbiology and Immunology* 45.12 (2001), pp. 819–828. ISSN: 1348-0421. DOI: 10.1111/j.1348-0421.2001.tb01321.x.

[81] Andrew S. Kennard, Matteo Osella, Avelino Javer, Jacopo Grilli, Philippe Nghe, Sander J. Tans, Pietro Cicuta, and Marco Cosentino Lagomarsino. "Individuality and universality in the growth-division laws of single E. coli cells." In: *Physical Review E* 93.1 (Jan. 2016), p. 012408. DOI: 10.1103/PhysRevE.93.012408.

[82] Ingrid M. Keseler et al. "EcoCyc: a comprehensive view of Escherichia coli biology." eng. In: *Nucleic Acids Research* 37.Database issue (Jan. 2009), pp. D464–470. ISSN: 1362-4962. DOI: 10.1093/nar/gkn751.

[83] Ingrid M. Keseler et al. "The EcoCyc database: reflecting new knowledge about Escherichia coli K-12." In: *Nucleic Acids Research* 45.D1 (Jan. 2017), pp. D543–D550. ISSN: 0305-1048. DOI: 10.1093/nar/gkw1003.

[84] Juhyun Kim, Alexander Darlington, Manuel Salvador, José Utrilla, and José I Jiménez. "Trade-offs between gene expression, growth and phenotypic diversity in microbial populations." en. In: *Current Opinion in Biotechnology*. Energy Biotechnologyy Environmental Biotechnology 62 (Apr. 2020), pp. 29–37. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2019.08.004.

[85] Mark Kittisopikul and Gürol M. Süel. "Biological role of noise encoded in a genetic network motif." en. In: *Proceedings of the National Academy of Sciences* 107.30 (July 2010), pp. 13300–13305. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1003975107.

[86] Daniel J. Kiviet, Philippe Nghe, Noreen Walker, Sarah Boulineau, Vanda Sunderlikova, and Sander J. Tans. "Stochasticity of metabolism and growth at the single-cell level." en. In: *Nature* 514.7522 (Oct. 2014), pp. 376–379. ISSN: 0028-0836. DOI: 10.1038/nature13582.

[87] Istvan T. Kleijn, Laurens H. J. Krah, and Rutger Hermsen. "Noise propagation in an integrated model of bacterial gene expression and growth." en. In: *PLOS Computational Biology* 14.10 (Oct. 2018), e1006386. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006386.

[88] Istvan T. Kleijn, Amalia Martínez-Segura, François Bertaux, Malika Saint, Holger Kramer, Vahid Shahrezaei, and Samuel Marguerat. "Growth-Rate Dependent And Nutrient-Specific Gene Expression Resource Allocation In Fission Yeast." en. In: *bioRxiv* (Mar. 2021), p. 2021.03.16.435638. DOI: 10.1101/2021.03.16.435638.

[89] Stefan Klumpp and Terence Hwa. "Bacterial growth: global effects on gene expression, growth feedback and proteome partition." In: *Current Opinion in Biotechnology*. Nanobiotechnology • Systems biology 28 (Aug. 2014), pp. 96–102. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2014.01.001.

[90] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. "Molecular crowding limits translation and cell growth." eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.42 (Oct. 2013), pp. 16754–16759. ISSN: 1091-6490. DOI: 10.1073/pnas.1310377110.

[91] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. "Growth-rate dependent global effects on gene expression in bacteria." In: *Cell* 139.7 (Dec. 2009), p. 1366. ISSN: 0092-8674. DOI: 10.1016/j.cell.2009.12.001.

[92] Arthur L. Koch. "Diffusion The Crucial Process in Many Aspects of the Biology of Bacteria." en. In: *Advances in Microbial Ecology*. Ed. by K. C. Marshall. Advances in Microbial Ecology. Boston, MA: Springer US, 1990, pp. 37–70. ISBN: 978-1-4684-7612-5.

[93] Arthur L. Koch. "WHAT SIZE SHOULD A BACTERIUM BE? A Question of Scale." In: *Annual Review of Microbiology* 50.1 (1996), pp. 317–348.

[94] Karl Kochanowski, Luca Gerosa, Simon F. Brunner, Dimitris Christodoulou, Yaroslav V. Nikolaev, and Uwe Sauer. "Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli." en. In: *Molecular Systems Biology* 13.1 (Jan. 2017), p. 903. ISSN: 1744-4292, 1744-4292. DOI: 10.15252/msb.20167402.

[95] Karl Kochanowski, Hiroyuki Okano, Vadim Patsalo, James R. Williamson, Uwe Sauer, and Terence Hwa. "Global coordination of metabolic pathways in Escherichia coli by active and passive regulation." In: *Molecular Systems Biology* 17.4 (Apr. 2021), e10064. ISSN: 1744-4292. DOI: 10.15252/msb.202010064.

[96] Manjunatha Kogenaru, Philippe Nghe, Frank J. Poelwijk, and Sander J. Tans. "Predicting Evolutionary Constraints by Identifying Conflicting Demands in Regulatory Networks." English. In: *Cell Systems* 10.6 (June 2020), 526–534.e3. ISSN: 2405-4712. DOI: 10.1016/j.cels.2020.05.004.

[97] Michał Komorowski, Jacek Miękisz, and Michael P. H. Stumpf. "Decomposing Noise in Biochemical Signaling Systems Highlights the Role of Protein Degradation." In: *Biophysical Journal* 104.8 (Apr. 2013), pp. 1783–1793. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2013.02.027.

[98] Yael Korem Kohanim, Dikla Levi, Ghil Jona, Benjamin D. Towbin, Anat Bren, and Uri Alon. "A Bacterial Growth Law out of Steady State." en. In: *Cell Reports* 23.10 (June 2018), pp. 2891–2900. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2018.05.007.

[99] Laurens H. J. Krah and Rutger Hermsen. "The effect of natural selection on the propagation of protein expression noise to bacterial growth." eng. In: *PLoS computational biology* 17.7 (July 2021), e1009208. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009208.

[100] Mads Kærn, Timothy C. Elston, William J. Blake, and James J. Collins. "Stochasticity in gene expression: from theories to phenotypes." en. In: *Nature Reviews Genetics* 6.6 (June 2005), pp. 451–464. ISSN: 1471-0056. DOI: 10.1038/nrg1615.

[101] H E Kubitschek, W W Baldwin, S J Schroeter, and R Graetzer. "Independence of buoyant cell density and growth rate in Escherichia coli." In: *Journal of Bacteriology* 158.1 (Apr. 1984), pp. 296–299. DOI: 10.1128/jb.158.1.296-299.1984.

[102] Catherine L Lawson, David Swigon, Katsuhiko S Murakami, Seth A Darst, Helen M Berman, and Richard H Ebright. "Catabolite activator protein: DNA binding and transcription activation." en. In: *Current Opinion in Structural Biology* 14.1 (Feb. 2004), pp. 10–20. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2004.01.012.

[103] Guillaume Le Treut, Fangwei Si, Dongyang Li, and Suckjoon Jun. *Single-cell data and correlation analysis support the independent double adder model in both Escherichia coli and Bacillus subtilis*. en. preprint. Systems Biology, Oct. 2020.

[104] Erel Levine and Terence Hwa. "Stochastic fluctuations in metabolic pathways." en. In: *Proceedings of the National Academy of Sciences* 104.22 (May 2007), pp. 9224–9229. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0610987104.

[105] Geraint Lewis. "Not of this world?" en. In: *New Scientist* 236.3155 (Dec. 2017), pp. 24–25. ISSN: 0262-4079. DOI: 10.1016/S0262-4079(17)32404-1.

[106] Nathan E. Lewis et al. "Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models." eng. In: *Molecular Systems Biology* 6 (July 2010), p. 390. ISSN: 1744-4292. DOI: 10.1038/msb.2010.47.

[107] Genyuan Li, Herschel Rabitz, Paul E. Yelvington, Oluwayemisi O. Oluwole, Fred Bacon, Charles E. Kolb, and Jacqueline Schoendorf. "Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs." In: *The Journal of Physical Chemistry A* 114.19 (May 2010), pp. 6022–6032. ISSN: 1089-5639. DOI: 10.1021/jp9096919.

[108] Jie Lin and Ariel Amir. "From single-cell variability to population growth." In: *Physical Review E* 101.1 (Jan. 2020), p. 012401. DOI: 10.1103/PhysRevE.101.012401.

[109] Jie Lin and Ariel Amir. "Disentangling Intrinsic and Extrinsic Gene Expression Noise in Growing Cells." In: *Physical Review Letters* 126.7 (Feb. 2021), p. 078101. DOI: 10.1103/PhysRevLett.126.078101.

[110] Lin Liu and Alexander Bockmayr. "Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling." en. In: *Journal of Theoretical Biology* (May 2020), p. 110317. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2020.110317.

[111] Peijiang Liu, Zhanjiang Yuan, Haohua Wang, and Tianshou Zhou. "Decomposition and tunability of expression noise in the presence of coupled feedbacks." eng. In: *Chaos (Woodbury, N.Y.)* 26.4 (2016), p. 043108. ISSN: 1089-7682. DOI: 10.1063/1.4947202.

[112] O. Maaløe. "An Analysis of Bacterial Growth." In: 1970. DOI: 10.1016/B978-0-12-395541-8.50008-2.

[113] Hédia Maamar, Arjun Raj, and David Dubnau. "Noise in gene expression determines cell fate in Bacillus subtilis." eng. In: *Science (New York, N.Y.)* 317.5837 (July 2007), pp. 526–529. ISSN: 1095-9203. DOI: 10.1126/science.1140818.

[114] M. Madan Babu, Sarah A. Teichmann, and L. Aravind. "Evolutionary dynamics of prokaryotic transcriptional regulatory networks." eng. In: *Journal of Molecular Biology* 358.2 (Apr. 2006), pp. 614–633. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2006.02.019.

[115] Narendra Maheshri and Erin K. O'Shea. "Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression." In: *Annual Review of Biophysics and Biomolecular Structure* 36.1 (2007), pp. 413–434. DOI: 10.1146/annurev.biophys.36.040306.132705.

[116] Arijit Maitra and Ken A. Dill. "Bacterial growth laws reflect the evolutionary importance of energy efficiency." en. In: *Proceedings of the National Academy of Sciences* 112.2 (Jan. 2015), pp. 406–411. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1421138111.

[117] R. A. Majewski and M. M. Domach. "Simple constrained-optimization view of acetate overflow in E. coli." en. In: *Biotechnology and Bioengineering* 35.7 (1990), pp. 732–738. ISSN: 1097-0290. DOI: 10.1002/bit.260350711.

[118] Daniele De Martino, Fabrizio Capuani, and Andrea De Martino. "Growth against entropy in bacterial metabolism: the phenotypic trade-off behind empirical growth rate distributions inE.coli." en. In: *Physical Biology* 13.3 (May 2016), p. 036005. ISSN: 1478-3975. DOI: 10.1088/1478-3975/13/3/036005.

[119] David Martínez-Martín, Gotthold Fläschner, Benjamin Gaub, Sascha Martin, Richard Newton, Corina Beerli, Jason Mercer, Christoph Gerber, and Daniel J. Müller. "Inertial picobalance reveals fast mass fluctuations in mammalian cells." en. In: *Nature* 550.7677 (Oct. 2017), pp. 500–505. ISSN: 1476-4687. DOI: 10.1038/nature24288.

[120] Lucia Marucci et al. "Computer-Aided Whole-Cell Design: Taking a Holistic Approach by Integrating Synthetic With Systems Biology." In: *Frontiers in Bioengineering and Biotechnology* 8 (2020), p. 942. ISSN: 2296-4185. DOI: 10.3389/fbioe.2020.00942.

[121] C. K. Mathews. "The cell-bag of enzymes or network of channels?" eng. In: *Journal of Bacteriology* 175.20 (Oct. 1993), pp. 6377–6381. ISSN: 0021-9193. DOI: 10.1128/jb.175.20.6377-6381.1993.

[122] M. R. Maurizi. "Proteases and protein degradation inEscherichia coli." en. In: *Experientia* 48.2 (Feb. 1992), pp. 178–201. ISSN: 1420-9071. DOI: 10.1007/BF01923511.

[123] Harley H. McAdams and Adam Arkin. "Stochastic mechanisms in gene expression." en. In: *Proceedings of the National Academy of Sciences* 94.3 (Feb. 1997), pp. 814–819. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.94.3.814.

[124] Eyal Metzl-Raz, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich, and Naama Barkai. "Principles of cellular resource allocation revealed by condition-dependent proteome profiling." In: *eLife* 6 (Aug. 2017). Ed. by Patricia J Wittkopp, e28034. ISSN: 2050-084X. DOI: 10.7554/eLife.28034.

[125] Andreas Milias-Argeitis, Marc Rullan, Stephanie K. Aoki, Peter Buchmann, and Mustafa Khammash. "Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth." en. In: *Nature Communications* 7.1 (Aug. 2016), p. 12546. ISSN: 2041-1723. DOI: 10.1038/ncomms12546.

[126] Allen P. Minton. "The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media*." en. In: *Journal of Biological Chemistry* 276.14 (Apr. 2001), pp. 10577–10580. ISSN: 0021-9258. DOI: 10.1074/jbc.R100005200.

[127] Saurabh Modi, Cesar Augusto Vargas-Garcia, Khem Raj Ghusinga, and Abhyudai Singh. "Analysis of Noise Mechanisms in Cell-Size Control." en. In: *Biophysical Journal* 112.11 (June 2017), pp. 2408–2418. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2017.04.050.

[128] Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. "Shifts in growth strategies reflect tradeoffs in cellular economics." en. In: *Molecular Systems Biology* 5.1 (Jan. 2009), p. 323. ISSN: 1744-4292, 1744-4292. DOI: 10.1038/msb.2009.82.

[129] Jacques Monod. "The Growth of Bacterial Cultures." In: *Annual Review of Microbiology* 3.1 (1949), pp. 371–394. DOI: 10.1146/annurev.mi.03.100149.002103.

[130] Rafael Moreno-Sánchez, Emma Saavedra, Sara Rodríguez-Enríquez, and Viridiana Olín-Sandoval. "Metabolic Control Analysis: A Tool for Designing Strategies to Manipulate Metabolic Pathways." en. In: *Journal of Biomedicine and Biotechnology* 2008 (July 2008), e597913. ISSN: 2314-6133. DOI: 10.1155/2008/597913.

[131] Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. "Constrained Allocation Flux Balance Analysis." eng. In: *PLoS computational biology* 12.6 (2016), e1004913. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004913.

[132] Brian Munsky, Gregor Neuert, and Alexander van Oudenaarden. "Using Gene Expression Noise to Understand Gene Regulation." en. In: *Science* 336.6078 (Apr. 2012), pp. 183–187. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1216379.

[133] Kamalendu Nath and Arthur L. Koch. "Protein Degradation in Escherichia coli: I. MEASUREMENT OF RAPIDLY AND SLOWLY DECAYING COMPONENTS." en. In: *Journal of Biological Chemistry* 245.11 (June 1970), pp. 2889–2900. ISSN: 0021-9258. DOI: 10.1016/S0021-9258(18)63072-8.

[134] Nela Nikolic, Frank Schreiber, Alma Dal Co, Daniel J. Kiviet, Tobias Bergmiller, Sten Littmann, Marcel M. M. Kuypers, and Martin Ackermann. "Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations." en. In: *PLOS Genetics* 13.12 (Dec. 2017), e1007122. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1007122.

[135] Edward J. O'Brien, Jose Utrilla, and Bernhard O. Palsson. "Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments." en. In: *PLOS Computational Biology* 12.6 (June 2016), e1004998. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004998.

[136] Hiroyuki Okano, Rutger Hermsen, Karl Kochanowski, and Terence Hwa. "Regulation underlying hierarchical and simultaneous utilization of carbon substrates by flux sensors in Escherichia coli." en. In: *Nature Microbiology* 5.1 (Jan. 2020), pp. 206–215. ISSN: 2058-5276. DOI: 10.1038/s41564-019-0610-7.

[137] Diego A. Oyarzún, Jean-Baptiste Lugagne, and Guy-Bart V. Stan. "Noise Propagation in Synthetic Gene Circuits for Metabolic Control." In: *ACS Synthetic Biology* 4.2 (Feb. 2015), pp. 116–125. DOI: 10.1021/sb400126a.

[138] Johan Paulsson. "Summing up the noise in gene networks." en. In: *Nature* 427.6973 (Jan. 2004), pp. 415–418. ISSN: 0028-0836. DOI: 10.1038/nature02257.

[139] Juan M. Pedraza and Alexander van Oudenaarden. "Noise Propagation in Gene Networks." en. In: *Science* 307.5717 (Mar. 2005), pp. 1965–1969. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1109090.

[140] Carlos Pedrós-Alió, Marianne Potvin, and Connie Lovejoy. "Diversity of planktonic microorganisms in the Arctic Ocean." en. In: *Progress in Oceanography*. Overarching perspectives of contemporary and future ecosystems in the Arctic Ocean 139 (Dec. 2015), pp. 233–243. ISSN: 0079-6611. DOI: 10.1016/j.pocean.2015.07.009.

[141] Robert Planqué, Josephus Hulshof, Bas Teusink, Johannes C. Hendriks, and Frank J. Bruggeman. "Maintaining maximal metabolic flux by gene expression control." en. In: *PLOS Computational Biology* 14.9 (Sept. 2018), e1006412. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006412.

[142] Anastasia Politou and Piero Andrea Temussi. "Revisiting a dogma: the effect of volume exclusion in molecular crowding." en. In: *Current Opinion in Structural Biology*. Folding and binding/Nucleic acids and their protein complexes 30 (Feb. 2015), pp. 1–6. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2014.10.005.

[143] Miguel Prado Casanova. "Noise and Synthetic Biology: How to Deal with Stochasticity?" en. In: *NanoEthics* 14.1 (Apr. 2020), pp. 113–122. ISSN: 1871-4765. DOI: 10.1007/s11569-020-00366-4.

[144] Arjun Raj and Alexander van Oudenaarden. "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences." In: *Cell* 135.2 (Oct. 2008), pp. 216–226. ISSN: 0092-8674. DOI: 10.1016/j.cell.2008.09.050.

[145] Jonathan M. Raser and Erin K. O'Shea. "Noise in Gene Expression: Origins, Consequences, and Control." In: *Science* 309.5743 (2005), pp. 2010–2013. ISSN: 0036-8075.

[146] J. Christian J. Ray, Michelle L. Wickersheim, Ameya P. Jalihal, Yusuf O. Adeshina, Tim F. Cooper, and Gábor Balázsi. "Cellular Growth Arrest and Persistence from Enzyme Saturation." en. In: *PLOS Computational Biology* 12.3 (Mar. 2016), e1004825. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004825.

[147] Hannes Risken. "Fokker-Planck Equation." en. In: *The Fokker-Planck Equation: Methods of Solution and Applications*. Ed. by Hannes Risken. Springer Series in Synergetics. Berlin, Heidelberg: Springer, 1996, pp. 63–95. ISBN: 978-3-642-61544-3.

[148] Nitzan Rosenfeld, Jonathan W. Young, Uri Alon, Peter S. Swain, and Michael B. Elowitz. "Gene Regulation at the Single-Cell Level." en. In: *Science* 307.5717 (Mar. 2005), pp. 1962–1965. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1106914.

[149] Nitzan Rosenfeld, Jonathan W. Young, Uri Alon, Peter S. Swain, and Michael B. Elowitz. "Gene regulation at the single-cell level." English. In: *Science* 307.5717 (2005), pp. 1962–5. DOI: 10.1126/science.1106914.

[150] JR Roth, JG Lawrence, and TA Bobik. "COBALAMIN (COENZYME B12): Synthesis and Biological Significance." In: *Annual Review of Microbiology* 50.1 (1996), pp. 137–181. DOI: 10.1146/annurev.micro.50.1.137.

[151] Marc Rullan, Dirk Benzinger, Gregor W. Schmidt, Andreas Milias-Argeitis, and Mustafa Khammash. "An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation." en. In: *Molecular Cell* 70.4 (May 2018), 745–756.e6. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2018.04.012.

[152] R. Saravanamuthu. *Industrial Exploitation of Microorganisms*. en. I. K. International Pvt Ltd, May 2010. ISBN: 978-93-80026-53-4.

[153]   John T Sauls, Dongyang Li, and Suckjoon Jun. "Adder and a coarse-grained approach to cell size homeostasis in bacteria." en. In: *Current Opinion in Cell Biology*. Cell architecture 38 (Feb. 2016), pp. 38–44. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2016.02.004.

[154]   M. A. Savageau. "Biochemical systems analysis. A study of function and design in molecular biology." In: *ADDISON WESLEY PUBL.* 1976.

[155]   Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. "The quantitative and condition-dependent Escherichia coli proteome." en. In: *Nature Biotechnology* 34.1 (Jan. 2016), pp. 104–110. ISSN: 1546-1696. DOI: 10.1038/nbt.3418.

[156]   Jörn M. Schmiedel, Lucas B. Carey, and Ben Lehner. "Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise." en. In: *Nature Communications* 10.1 (July 2019), pp. 1–12. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11116-w.

[157]   M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa. "Emergence of robust growth laws from optimal regulation of ribosome synthesis." en. In: *Molecular Systems Biology* 10.8 (Aug. 2014), pp. 747–747. ISSN: 1744-4292. DOI: 10.15252/msb.20145379.

[158]   Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. "Interdependence of Cell Growth and Gene Expression: Origins and Consequences." en. In: *Science* 330.6007 (Nov. 2010), pp. 1099–1102. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1192588.

[159]   Matthew Scott and Terence Hwa. "Bacterial growth laws and their applications." en. In: *Current Opinion in Biotechnology*. Nanobiotechnology and Systems Biology 22.4 (Aug. 2011), pp. 559–565. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2011.04.014.

[160]   Vahid Shahrezaei and Samuel Marguerat. "Connecting growth with gene expression: of noise and numbers." In: *Current Opinion in Microbiology*. Environmental microbiology • Extremophiles 25 (June 2015), pp. 127–135. ISSN: 1369-5274. DOI: 10.1016/j.mib.2015.05.012.

[161]   Vahid Shahrezaei and Peter S. Swain. "Analytical distributions for stochastic gene expression." en. In: *Proceedings of the National Academy of Sciences* 105.45 (Nov. 2008), pp. 17256–17261. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0803850105.

[162]   Abhyudai Singh, Brandon S Razooky, Roy D Dar, and Leor S Weinberger. "Dynamics of protein noise can distinguish between alternate sources of gene-expression variability." In: *Molecular Systems Biology* 8 (Aug. 2012), p. 607. ISSN: 1744-4292. DOI: 10.1038/msb.2012.38.

[163]   Abhyudai Singh and Mohammad Soltani. "Quantifying Intrinsic and Extrinsic Variability in Stochastic Gene Expression Models." In: *PLoS ONE* 8.12 (Dec. 2013). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0084301.

[164]   M. Smith, K. R. Ghusinga, and A. Singh. "Comparison of feedback strategies for noise suppression in protein level." In: *2019 American Control Conference (ACC)*. July 2019, pp. 1513–1518. DOI: 10.23919/ACC.2019.8815616.

[165] Christian Solem, Dina Petranovic, Brian Koebmann, Ivan Mijakovic, and Peter Ruh-dal Jensen. "Phosphoglycerate Mutase Is a Highly Efficient Enzyme without Flux Control in Lactococcus lactis." In: *Journal of Molecular Microbiology and Biotechnology* 18.3 (2010), pp. 174–180. ISSN: 1464-1801, 1660-2412. DOI: 10.1159/000315458.

[166] Wieland Steinchen, Victor Zegarra, and Gert Bange. "(p)ppGpp: Magic Modulators of Bacterial Physiology and Metabolism." In: *Frontiers in Microbiology* 11 (2020), p. 2072. ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.02072.

[167] Lee Susman, Maryam Kohram, Harsh Vashistha, Jeffrey T. Nechleba, Hanna Salman, and Naama Brenner. "Individuality and slow dynamics in bacterial growth home-ostasis." en. In: *Proceedings of the National Academy of Sciences* 115.25 (June 2018), E5679–E5687. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1615526115.

[168] Peter S. Swain, Michael B. Elowitz, and Eric D. Siggia. "Intrinsic and extrinsic con-tributions to stochasticity in gene expression." en. In: *Proceedings of the National Academy of Sciences* 99.20 (Oct. 2002), pp. 12795–12800. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.162041399.

[169] Mariola Szenk, Ken A. Dill, and Adam M. R. de Graff. "Why Do Fast-Growing Bac-teria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis." en. In: *Cell Systems* 5.2 (Aug. 2017), pp. 95–104. ISSN: 2405-4712. DOI: 10.1016/j.cels.2017.06.005.

[170] Sattar Taheri-Araghi, Serena Bradde, John T. Sauls, Norbert S. Hill, Petra A. Levin, Johan Paulsson, Massimo Vergassola, and Suckjoon Jun. "Cell-size control and home-ostasis in bacteria." eng. In: *Current biology: CB* 25.3 (Feb. 2015), pp. 385–391. ISSN: 1879-0445. DOI: 10.1016/j.cub.2014.12.009.

[171] Vakil Takhaveev and Matthias Heinemann. "Metabolic heterogeneity in clonal mi-crobial populations." en. In: *Current Opinion in Microbiology*. Antimicrobials * Micro-bial systems biology 45 (Oct. 2018), pp. 30–38. ISSN: 1369-5274. DOI: 10.1016/j.mib.2018.02.004.

[172] Cheemeng Tan, Philippe Marguet, and Lingchong You. "Emergent bistability by a growth-modulating positive feedback circuit." en. In: *Nature Chemical Biology* 5.11 (Nov. 2009), pp. 842–848. ISSN: 1552-4469. DOI: 10.1038/nchembio.218.

[173] Yuichi Taniguchi, Paul J. Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X. Sunney Xie. "Quantifying E. coli proteome and tran-scriptome with single-molecule sensitivity in single cells." eng. In: *Science (New York, N.Y.)* 329.5991 (July 2010), pp. 533–538. ISSN: 1095-9203. DOI: 10.1126/science.1188308.

[174] Mukund Thattai and Alexander van Oudenaarden. "Intrinsic noise in gene regu-latory networks." en. In: *Proceedings of the National Academy of Sciences* 98.15 (July 2001), pp. 8614–8619. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.151588598.

[175] Mukund Thattai and Alexander van Oudenaarden. "Stochastic Gene Expression in Fluctuating Environments." en. In: *Genetics* 167.1 (May 2004), pp. 523–530. ISSN: 0016-6731, 1943-2631. DOI: 10.1534/genetics.167.1.523.

[176]   Philipp Thomas, Guillaume Terradot, Vincent Danos, and Andrea Y. Weie*. "Sources, propagation and consequences of stochasticity in cellular growth." en. In: *Nature Communications* 9.1 (Oct. 2018), pp. 1–11. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06912-9.

[177]   Sorin Tănase-Nicola, Patrick B. Warren, and Pieter Rein ten Wolde. "Signal Detection, Modularity, and the Correlation between Extrinsic and Intrinsic Noise in Biochemical Networks." In: *Physical Review Letters* 97.6 (Aug. 2006), p. 068102. DOI: 10.1103/PhysRevLett.97.068102.

[178]   Sorin Tănase-Nicola and Pieter Rein ten Wolde. "Regulatory Control and the Costs and Benefits of Biochemical Noise." en. In: *PLOS Computational Biology* 4.8 (Aug. 2008), e1000125. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000125.

[179]   Mona K. Tonn, Philipp Thomas, Mauricio Barahona, and Diego A. Oyarzún. "Stochastic modelling reveals mechanisms of metabolic heterogeneity." en. In: *Communications Biology* 2.1 (Mar. 2019), pp. 1–9. ISSN: 2399-3642. DOI: 10.1038/s42003-019-0347-0.

[180]   Jonathan Touboul and Olivier Faugeras. "A characterization of the first hitting time of double integral processes to curved boundaries." en. In: *Advances in Applied Probability* 40.2 (June 2008), pp. 501–528. ISSN: 0001-6678, 1475-6064. DOI: 10.1239/aap/1214950214.

[181]   Benjamin D. Towbin, Yael Korem, Anat Bren, Shany Doron, Rotem Sorek, and Uri Alon. "Optimality and sub-optimality in a bacterial growth law." en. In: *Nature Communications* 8.1 (Jan. 2017), pp. 1–8. ISSN: 2041-1723. DOI: 10.1038/ncomms14123.

[182]   Saburo Tsuru, Junya Ichinose, Akiko Kashiwagi, Bei-Wen Ying, Kunihiko Kaneko, and Tetsuya Yomo. "Noisy cell growth rate leads to fluctuating protein concentration in bacteria." en. In: 6.3 (June 2009), p. 036015. ISSN: 1478-3975. DOI: 10.1088/1478-3975/6/3/036015.

[183]   G. E. Uhlenbeck and L. S. Ornstein. "On the Theory of the Brownian Motion." In: *Physical Review* 36.5 (Sept. 1930), pp. 823–841. DOI: 10.1103/PhysRev.36.823.

[184]   A. E. Vasdekis, H. Alanazi, A. M. Silverman, C. J. Williams, A. J. Canul, J. B. Cliff, A. C. Dohnalkova, and G. Stephanopoulos. "Eliciting the impacts of cellular noise on metabolic trade-offs by quantitative mass imaging." en. In: *Nature Communications* 10.1 (Feb. 2019), p. 848. ISSN: 2041-1723. DOI: 10.1038/s41467-019-08717-w.

[185]   Andreas E. Vasdekis and Abhyudai Singh. "Microbial metabolic noise." en. In: *WIREs Mechanisms of Disease* 13.3 (2021), e1512. ISSN: 2692-9368. DOI: 10.1002/wsbm.1512.

[186]   V. K. Viswanathan. "Sizing up microbes." In: *Gut Microbes* 3.6 (Nov. 2012), pp. 483–484. ISSN: 1949-0976. DOI: 10.4161/gmic.22637.

[187]   Noreen Walker. "A single-cell study on stochasticity growth and gene expression." Dissertation (TU Delft). Apr. 2016.

[188]   Noreen Walker, Philippe Nghe, and Sander J. Tans. "Generation and filtering of gene expression noise by the bacterial cell cycle." In: *BMC Biology* 14 (2016), p. 11. ISSN: 1741-7007. DOI: 10.1186/s12915-016-0231-z.

[189] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. "Robust Growth of Escherichia coli." In: *Current Biology* 20.12 (June 2010), pp. 1099–1103. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.04.045.

[190] Zihao Wang, Zhenquan Zhang, and Tianshou Zhou. "Analytical results for non-Markovian models of bursty gene expression." In: *Physical Review E* 101.5 (May 2020), p. 052406. DOI: 10.1103/PhysRevE.101.052406.

[191] Wehrens, M. and Tans, S.J. "Dynamical regulation in single cells." en. PhD thesis. Delft University of Technology, Apr. 2019.

[192] Martijn Wehrens, Ferhat Büke, Philippe Nghe, and Sander J. Tans. "Stochasticity in cellular metabolism and growth: Approaches and consequences." en. In: *Current Opinion in Systems Biology*. • Regulatory and metabolic networks • Special Section: Single cell and noise 8 (Apr. 2018), pp. 131–136. ISSN: 2452-3100. DOI: 10.1016/j.coisb.2018.02.006.

[193] Andrea Y. Weie*, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. "Mechanistic links between cellular trade-offs, gene expression, and growth." en. In: *Proceedings of the National Academy of Sciences* 112.9 (Mar. 2015), E1038–E1047. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1416533112.

[194] Guillaume Witz, Erik van Nimwegen, and Thomas Julou. "Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism." In: *eLife* 8 (Nov. 2019). Ed. by Michael T Laub, Gisela Storz, and Alan Leonard, e48063. ISSN: 2050-084X. DOI: 10.7554/eLife.48063.

[195] Luise Wolf, Olin K Silander, and Erik van Nimwegen. "Expression noise facilitates the evolution of gene regulation." In: *eLife* 4 (June 2015). Ed. by Ido Golding, e05856. ISSN: 2050-084X. DOI: 10.7554/eLife.05856.

[196] Meike T. Wortel, Evert Bosdriesz, Bas Teusink, and Frank J. Bruggeman. "Evolutionary pressures on microbial metabolic strategies in the chemostat." en. In: *Scientific Reports* 6 (July 2016), p. 29503. ISSN: 2045-2322. DOI: 10.1038/srep29503.

[197] Pablo Yarza, Pelin Yilmaz, Elmar Pruesse, Frank Oliver Glöckner, Wolfgang Ludwig, Karl-Heinz Schleifer, William B. Whitman, Jean Euzéby, Rudolf Amann, and Ramon Rosselló-Móra. "Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences." en. In: *Nature Reviews Microbiology* 12.9 (Sept. 2014), pp. 635–645. ISSN: 1740-1534. DOI: 10.1038/nrmicro3330.

[198] Conghui You, Hiroyuki Okano, Sheng Hui, Zhongge Zhang, Minsu Kim, Carl W. Gunderson, Yi-Ping Wang, Peter Lenz, Dalai Yan, and Terence Hwa. "Coordination of bacterial proteome with metabolism by cyclic AMP signalling." en. In: *Nature* 500.7462 (Aug. 2013), pp. 301–306. ISSN: 0028-0836. DOI: 10.1038/nature12446.

[199] Jonathan W. Young, James C. W. Locke, Alphan Altinok, Nitzan Rosenfeld, Tigran Bacarian, Peter S. Swain, Eric Mjolsness, and Michael B. Elowitz. "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy." en. In: *Nature Protocols* 7.1 (Jan. 2012), pp. 80–88. ISSN: 1750-2799. DOI: 10.1038/nprot.2011.432.

[200] Ji Yu, Jie Xiao, Xiaojia Ren, Kaiqin Lao, and X. Sunney Xie. "Probing Gene Expression in Live Cells, One Protein Molecule at a Time." In: *Science* 311.5767 (Mar. 2006), pp. 1600–1603. DOI: 10.1126/science.1119623.

[201] Thomas A. Zangle and Michael A. Teitell. "Live-cell mass profiling: an emerging approach in quantitative biophysics." en. In: *Nature Methods* 11.12 (Dec. 2014), pp. 1221–1228. ISSN: 1548-7105. DOI: 10.1038/nmeth.3175.

[202] Fuzhong Zhang, Sarah Rodriguez, and Jay D Keasling. "Metabolic engineering of microbial pathways for advanced biofuels production." en. In: *Current Opinion in Biotechnology*. 22/6 Chemical biotechnology and Pharmaceutical biotechnology 22.6 (Dec. 2011), pp. 775–783. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2011.04.024.

[203] Dongling Zheng, Chrystala Constantinidou, Jon L. Hobman, and Stephen D. Minchin. "Identification of the CRP regulon using in vitro and in vivo transcriptional profiling." In: *Nucleic Acids Research* 32.19 (Oct. 2004), pp. 5874–5893. ISSN: 0305-1048. DOI: 10.1093/nar/gkh908.

[204] Manlu Zhu and Xiongfeng Dai. "Growth suppression by altered (p)ppGpp levels results from non-optimal resource allocation in Escherichia coli." In: *Nucleic Acids Research* 47.9 (May 2019), pp. 4684–4693. ISSN: 0305-1048. DOI: 10.1093/nar/gkz211.

[205] Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. "Economics of membrane occupancy and respiro-fermentation." In: *Molecular Systems Biology* 7 (June 2011), p. 500. ISSN: 1744-4292. DOI: 10.1038/msb.2011.34.

[206] J. van Zuylen. "The microscopes of Antoni van Leeuwenhoek." en. In: *Journal of Microscopy* 121.3 (1981), pp. 309–328. ISSN: 1365-2818. DOI: 10.1111/j.1365-2818.1981.tb01227.x.

ACKNOWLEDGEMENTS

It is spring now and while the world begins anew, it is time for me to write the final sentences of this thesis. Here I'd like to thank all those involved in my life, scientifically or as friends; everything has become a blur. Thanks for making me feel accepted and appreciated.

First: the Kruyt Building, the fifth-became-sixth-floor: I had a great time in the Theoretical Biology group, so a big thanks to you all! Although I am no biologist originally and way too often skipped our social lunch-time, I still felt welcome the whole time. It was great to re-establish some stereotypes (Paulien saying: "Ha! The physicists play chess! But of course!"), but hopefully I also managed to slightly lessen the gap between the disciplines. Some people from the group require explicit thanks. Rob, your crystal clear science and conscious application of mathematics lured me in, and that is great! I know you for a long time now, but my respect has only been growing. You run the group as if it's 'peanuts', you're in control of everything and everyone and you're still having fun: You manage to inspire so many students and staff. Paulien, your mind-boggling course was maybe the first time I understood concepts and conclusions that were not entirely based on dry mathematical derivations. I am grateful for your undoubting passion for science and your (scientific and personal) integrity. It has also been great to work together (a bit) for some courses. Also special thanks for all the epic anecdotes of a lifetime of science during lunch breaks! Bram and Hilje, for me you are perfect prototypes of not only scientists but also of people. Your transformative periods when finalizing your PhDs was eye-opening for me and I'm honoured to have been seated right beside you! Arpit, without our writing, running, swimming and 'PhD-planning' sessions, I would still be writing my second chapter...Thanks for countless times 25 minutes. For you I hope you will go back to India someday, but I'll be forever grateful to have met you! Thea, thanks that I could always rely on your wisdom whenever needed. Sam, thanks for helping me out in the last months and I still want to hear you play the saxophone! My student, Bram, thanks for your enormous hunger for details and eternal dive to the depths, I've had a great time together! Floor, thanks for slightly lifting the experimental veil for me! Istvan, you were the one that started this all, so thanks for all the laughs, the work and the physicist vs physicist chess games. Martijn, you surely saved me...without your data, endless talks and brainstorm sessions, I would not have managed to finish this. Thanks for your trust in me and the great work we did together.

Next, of course Rutger deserves mountains of praise. Rutger, I am still thoroughly confused and impressed how you can have so many daily tasks and still remain many steps ahead of me at all times (although you'd deny this every time!). Your patience and countless cartoons while explaining me all the things I should already know, did eventually add up to me feeling like a scientist! Your never-failing politeness (also in your clearly and concisely written e-mails, even when in a hurry) set an example for me as a person, also outside science. Thanks for your incredibly sharp insights, your superman-like gaze that looks and understand the depths of seemingly all things and somehow turns the most difficult mathematics into trivialities.

To the members of the reading committee: thanks for checking all my work, I understand the mountains of equations can be quite dry from time to time. I am a big fan of all your work and am humbled to call you colleges.

Leden van mijn introductie-groepje van de bachelor: Damaz, Niels, Huibert, Frits, en Ugur...het is nog steeds vreemd hoeveel jullie voor mijn studie hebben betekent; jullie zijn allemaal geniaal en ik heb het gevoel dat ik bijna alle vakken van jullie heb geleerd!

En dan de hele scharen aan muzikanten waarop ik zo erg (nagenoeg irrationeel en onredelijk) trots op ben. Mijn bandgenoten Stan, Rikke, Tim, Joshua, Remy en Meidi...Waar zijn we in deze jaren allemaal beland...Het was pure vrijheid om met jullie naar de poorten van de figuurlijke hel te varen (en misschien net iets te vaak daar voorbij), vervolgens onze levenskeuzes gevaarlijk in twijfel te trekken, en dat dan de volgende week te herhalen. Jullie zijn misschien wel de enige die weten hoe slecht ik in sommige dingen ben en dat toch volledig lijken te accepteren. Ik respecteer jullie geweldig en ik ben dankbaar dat jullie met deze verwarde wetenschapper-slash-meneer-met-een-accordeon op pad willen gaan het leven tegemoet. Ook al is het soms moeilijk...laten we alsjeblieft altijd familie blijven! Ook een diepe dankbetuiging voor andere muzikanten in de folkscene. Femke en Lennart, het voelt nog steeds onwerkelijk om muzikanten te kennen met een visie die zo dicht bij die van mij ligt! Milan, bedankt dat je me het aardappel-pad op hebt gesleurd! Alex en heel het CeltCast team, het vertrouwen dat jullie in mij hebben, heeft ook zijn doorslag gevonden in dit boekje. Dank jullie wel voor al jullie steun.

Thanks to all my friends that explored some parts of the world with me and kept me sane in the last years. Stefan and Jenia (and now Leda and Nona as well) thanks for being so incredibly steady people and thanks for having the greatest hospitality ever. Sven, thanks for your enthusiasm for the weirdest of things and thanks for dragging me along! Quentin, Amber, Maartje, and Ellen: thanks for being my friends at those times I needed you all the very most.

Pap en mam, ik voel me altijd gesteund door jullie in wat voor keuzes ook. Papa vindt het natuurlijk leuk dat ik nu toch een beetje een wetenschapper bent, en ik denk dat hij me ook wel vergeeft dat ik niet toch Latijn of Nederlands ben gaan studeren. Pap, je bent altijd enorm inspirerend en ik denk dat ik door jou alles leuk vind en overal enthousiast over ben...en dat is een heel fijn iets om mee te krijgen tijdens je opvoeding. Mam, jij was stiekem altijd al de undercover bèta in ons huis lang voordat we wisten dat bètas meer dan een soort mythologische wezens waren. Bedankt dat ik van jou heb geleerd beslissingen te maken, maar daar toch altijd dieper over na te blijven denken. Ik hoop dat ik ook nog meer van jouw doortastendheid mag leren. Bedankt dat je zo'n zacht hart hebt en zo vergevingsgezind en begripvol bent. Eva bedankt voor het zijn van mijn grote zus voor altijd! Je begreep al mijn struggles en ik waardeer je enorm natuurlijk als grote zus, maar ook als een wetenschapper; door jou bleef ik zien hoe cool en onmogelijk moeilijk alfa studies zijn. En lieve Robin, bedankt voor al je eeuwige steun en vertrouwen, alle duizend koffies op het balkon en alle tijdsloze en oneindig diepe muziekjes tussendoor! Alles lijkt makkelijker en leuker zolang jij maar in de buurt bent. Toch heb je me echt op m'n slechts gezien en je lacht nog steeds...Niks is meer vreemd, sneller leven!

Laurens Krah (August 13th, 1992) grew up in Zaltbommel and attended the Stedelijk Gymnasium in 's-Hertogenbosch. In 2010 he enrolled in a physics bachelor at Utrecht University, and, quite accidentally, also in a mathematics bachelor. Interdisciplinarity soon became a main theme and Laurens participated in the honours programme, resulting in contacts with students from other science disciplines: chemistry, computer science and biology. After failing the very last math course (by 0.01 points), the consequential half year of study delay was mostly filled with biology related courses. After finishing both bachelors *cum laude*, the drive to apply physics and mathematics to interdisciplinary subjects remained. Laurens therefore joined the MSc programme 'Theoretical Biology' and the Qbio ('Quantitative Biology') honours programme. There, he discovered the fascinating mathematics used to explore the mind-boggling stochasticity that is inherent to almost all biological systems. Via the Qbio honours programme, Laurens was awarded a PhD position at Utrecht University under supervision of Dr. Rutger Hermsen, where the stochastic gene expression in bacterial cells has kept him puzzled since September 2016. His work on stochastic gene expression is described in this thesis. Currently, Laurens remains active as a teaching assistant and lecturer at Utrecht University.

Besides his scientific journey, Laurens has actively been on the road as an musician. During his PhD, he co-wrote and recorded 4 full-length albums, 2 EPs and 7 singles with four different bands (together good for well over 5 million streams world-wide). Additionally he played countless shows across Europe.

Laurens Krah (photo by Victor Roos)