LIÈGE université
Gembloux
Agro-Bio Tech

Cellular Entropy: A Key to control Gene expression and Unlock Continuous Bioproduction

Henrion Lucas

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE - GEMBLOUX AGRO-BIO TECH

Cellular Entropy: A Key to control Gene expression and Unlock Continuous Bioproduction

Henrion Lucas

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteur: Prof. Dr. Frank Delvigne

Année civile: 2024 - 2025

Abstract

The fundamental characteristics of living organisms, including growth, adaptation, homeostasis, organization, metabolism, and reproduction, are encoded in their DNA. Despite the crucial control exerted on genes, clonal cells exhibit heterogeneous behavior even when exposed to homogeneous environments. This phenomenon is justified by the theory of bet-hedging, which suggests that it allows populations to spread risks and opportunities. However, this heterogeneity poses significant challenges in biotechnology, reducing bioprocess performance and robustness, and has far-reaching implications for human health, including antibiotic resistance, cancer treatment, and potentially even its onset. The causes of this heterogeneity are multifaceted, including the diversity and uneven distribution of essential intracellular macromolecules necessary for gene expression. Our study reveals that, despite the complexity of these factors, intracellular dynamics alone are insufficient to explain the observed heterogeneity at the population level. We quantified population heterogeneity using Shannon entropy, a metric borrowed from information theory, and identified three diversification regimes characterized by increasing heterogeneity. A key predictor of these regimes is the impact of gene activation on cell growth, the switching cost. We found that as this switching cost increases, also known as the burden or production load, heterogeneity within the population rises accordingly. Interestingly, when the switching cost becomes significant, cells can synchronize their gene expression into periodic bursts, oscillations.

From these observations, the rest of the work is structured around two primary questions:

I. Why are burdensome genes associated with greater population entropy and; II. How can a gene circuit that does not have the properties of an oscillator nevertheless exhibit an oscillatory behavior?

To understand the relationship between population heterogeneity and the burden imposed by gene activation, we introduced the concept of burden entropy compensation. This mechanism suggests that the burden associated with gene activation is offset by the overgrowth of cells with lower activation of the burdensome gene. This overgrowth spreads the population thus increasing entropy, but safeguards the population from being washed out in continuous cultivation devices.

We focused on the T7 expression system in *E. coli* BL21 and found that periodic addition of the inducer can reduce entropy, but this control method is countered by the emergence of mutants with weakened promoters, lower burden, and more homogeneous induction. We then demonstrated that reducing the switching cost by lowering the quality of the main carbon source, and thus the maximum growth rate, homogenized gene expression across the population without sacrificing induction strength. This counterintuitive outcome highlights the importance of considering phenotypic heterogeneity using a system biology approach, rather than solely focusing on intra-

cellular sources of noise.

The significance of this system approach was further confirmed when we attempted to explain why highly burdensome gene circuits can exhibit bursty expression. We developed a mathematical model that incorporates the interplay between cellular stress response and environmental changes. The model revealed that, for highly burdensome gene circuits under certain dilution rate conditions, expression can be bursty and exhibit oscillatory behavior. The model predictions were validated experimentally in chemostat cultures of *Bacillus subtilis*, where sporulation synchronized across the population and appeared as oscillations. These oscillations result from a feedback loop where if cells are growing then the concentration in glucose drops below a threshold that triggers sporulation, and if cells sporulate, then the concentration rises above this threshold thus preventing sporulation.

Crucially, our results show that this unwanted instability can be eliminated, and entropy reduced, by carefully choosing process parameters that align with the biological characteristics of the cultivated organism. Our findings establish a clear link between gene expression burden and cellular entropy, demonstrating that entropy can be reduced to benefit overall population productivity, potentially unlocking the wider utilization of continuous cultivation.

Résumé

Les caractéristiques fondamentales du vivant, telles que la croissance, l'adaptation, l'homéostasie, l'organisation, le métabolisme et la reproduction, sont codées dans l'ADN. Cependant, malgré le contrôle crucial exercé sur les gènes, des cellules clonales exposées à un environnement homogène présentent un comportement hétérogène. Ce phénomène est justifié par la théorie du "bet-hedging" qui propose que les risques et les opportunités sont répartis au sein de la populations (ne pas mettre tous ses œufs dans le même panier). Cependant, cette hétérogénéité s'avère problématique pour des applications de processus biotechnologiques, car elle en réduit les performances et la robustesse. De plus, cette hétérogénéité a des implications importantes pour la santé humaine, notamment en ce qui concerne la résistance aux antibiotiques, le traitement du cancer et même, potentiellement, son apparition. Les causes de cette hétérogénéité sont multiples. La nature stochastique de l'activation des gènes en est l'une des principales. Elle résulte de la diversité et de la distribution inégale des macromolécules requises aux deux réactions essentielles à l'expression d'un gène: la transcription et la traduction. Malgré la complexité de ces facteurs, notre travail révèle que, dans les cultures continues, les dynamiques intracellulaires seules ne suffisent pas à expliquer l'hétérogénéité observée au niveau de la population. Grâce à un indice emprunté à la théorie de l'information, l'entropie de Shannon, nous avons pu quantifier le dégrée d'hétérogénéité au sein d'une population bactérienne et identifier trois régimes de diversification caractérisés par une hétérogénéité croissante. Un élément clé permettant de prédire ces régimes de diversification est la réduction de croissance associée à l'activation du gène, le coût associé au changement de phénotype. Nous avons observé qu'à mesure que ce coût, analogue à la charge métabolique ou à la charge de production, augmente, la population devient de plus en plus hétérogène. Nous avons également découvert que, lorsque le coût de l'activation des gènes sur la croissance devient important, les cellules peuvent synchroniser leur expression, laissant apparaître certains phénotypes par vagues périodiques. À partir de ces observations initiales, ce travail a été structuré autour des deux questions suivantes :

I. Pourquoi l'expression génique coûteuse pour l'organisme est-elle associée à une population avec beaucoup d'entropie ? II. Comment un circuit génique qui ne possède pas les propriétés d'un oscillateur peut-il néanmoins exhiber un comportement oscillatoire ?

Pour aborder le lien entre l'hétérogénéité de la population et le coût du changement phénotypique imposé par l'activation d'un gène, nous décrivons le mécanisme de compensation de cette charge par l'entropie. Ce mécanisme suggère que la perte de croissance associée à l'activation des gènes est compensée par une sur-croissance de cellules ayant une activation plus faible du gène dont l'expression est coûteuse. La présence de ces cellules augmente l'hétérogénéité de la population, son entropie, et lui permet ainsi d'éviter d'être lessivée lors d'une culture continue. En nous concen-

trant sur l'activation coûteuse du système d'expression T7 dans *Escherichia coli* BL21, nous avons trouvé que l'addition périodique de l'inducteur permet d'homogénéiser la population. Cependant, ce contrôle est contrecarré par l'émergence de mutants qui présentent une expression du gène affaiblie, et donc un coût métabolique moindre, mais plus homogène. Le coût sur la croissance associé à l'expression du gène étant la différence entre les taux de croissance des cellules induites et non induites, nous avons montré qu'il peut être réduit en diminuant la qualité de la principale source de carbone, ce qui réduit le taux de croissance maximal des cellules non induites. Cette stratégie permet d'obtenir une expression du gène plus homogène au sein de la population, sans pour autant réduire l'induction et, par conséquent, sans entraîner de perte de production. Ce résultat contre-intuitif fait sens lorsque l'on considère l'hétérogénéité phénotypique avec une approche de biologie du système et non uniquement en fonction des sources de bruit intracellulaire. L'importance de cette approche système a été confirmée par la suite, lors de l'investigation de la question concernant le caractère oscillant que peuvent avoir certains circuits génétiques dont l'expression est couteuse. Nous avons développé un modèle mathématique qui intègre la relation entre la réponse au stress cellulaire et les changements environnementaux. La résolution analytique de ce modèle a montré que, effectivement, pour les circuits géniques coûteux et sous certaines conditions, l'expression du gène peut se synchroniser et avoir un comportement oscillant. La résolution de ce modèle adapté pour la sporulation de *Bacillus subtilis* a montré que c'était en effet le cas pour une culture continue sous certains taux de dilutions. Ces oscillations résultent d'un mécanisme de rétroaction où, si les cellules sont en croissance, la concentration en glucose chute en dessous d'un seuil qui déclenche la sporulation, et si les cellules sporulent, la concentration dépasse ce seuil, empêchant ainsi la sporulation. Ces prédictions ont été confirmée expérimentalement. Surtout, ces résultats montrent aussi que cette instabilité indésirable pourrait être éliminée, et l'entropie réduite, en choisissant soigneusement les paramètres du processus (par exemple, le taux de dilution) en fonction des caractéristiques biologiques de l'organisme cultivé.

Pris dans leur ensemble, nos résultats établissent un lien clair entre la charge d'expression génique et l'entropie cellulaire. Nous décrivons comment cette entropie apparaît dans une culture continue, un système censé promouvoir une population stable. Finalement, nous démontrons que l'entropie cellulaire peut être réduite au bénéfice de la productivité globale de la population, ouvrant potentiellement la voie à davantage d'applications de cultures continues.

Acknowledgments

I first wish to thank my promotor Frank Delvigne. If asked, he will probably deny it, but I have begun and finished this thesis because he trusted I could do so. For his trust, energy and guidance I truly thank him. Beside Frank, I also wish to express my gratitude to my PhD committee, Marc Ongena, Patrick Fickers, Moreno Galleni, Alexander Grünberger and Abhyudai Singh.

The work I have done during these four years is a team work. Within this team, I have learned and debated during four years. These discussions have shaped my work and for this I thank them all. We have spent countless hours discussing nerdy stuff like: How can we even read a .fcs file? What the hell is git? Should the antibiotic X be dissolved in 20 or 80 % ethanol? But we also had a lot of highly pointless talks that truly contributed into making this thesis a pleasant experience.
Whilst far too many of these never ending talks were sparked by Juan Andrès Martinez, I nevertheless thank him for his guidance and for having taught me to think before rushing into the lab. Thank you Vincent Vandenbroucke for not running away when you saw how dirty my first python codes were and how misaligned by PowerPoint presentations still are. Thank you, Max and Hannah Sehrt for sharing, paper planes, unwanted hugs, confusion and a taste for German insults. Thank you Matheo Delvenne for sticking with me during all of my studies, during this thesis and for making sure good coffee prevailed. Thank you Melanie, Laurie, Fabian, Fabian, Fabian, Tiphaine, Fanny, Romain, Ghazal, Satoshi, Dingrong, Thai and Mika for making these four years pass like a blink of and eye. Thank you Andrew and Sam for teaching me so much and allowing to try and fail safely. I also wish to thank the entire MiPi lab and especially Marina, Marguerite and Danielle for your support over these four years.

Micro-organisms can sometime be a pain in the neck to interact with, and so am I. For this I wish to thank Aenor and my friends for tolerating, supporting and helping me. A special thanks to the friends I shared a beautiful and cold house with, Guillaume, Florian, Lisa, Tiphaine, Christophe, Florine. With each and all of you, the weird experiments do not stop once leaving the lab. Last but not least, I wish to thank my parents, Antoine, Oscar, Quentin, Granny, my Grandad that exposed me very early on to the pleasure of 'bidouiller', and my entire family for their support.

Contents

List of Figures

List of Tables

1

Chapter 1: Microbial populations diversification and control

1. Roots and benefits of cellular noise

1.1. Diversity as a survival strategy

Microbial cells exhibit a remarkable ability to inhabit diverse environments across the globe. This adaptability stems from the wide array of phenotypes these species can manifest. Among them are cells capable of plastic degradation [1], thriving in high-temperature conditions (Figure 1.1), or utilizing hydrogen as a source of energy [2]. These unique capabilities, phenotypes, are encoded within the genetic sequences known as genes. But expressing a phenotype always comes at a cost. It consumes amino acids, energy, or simply leads to the emergence of a phenotype that might not be fitted to the current environment. To ensure the combination of growth and resilience that forms the cell fitness always remains optimum, cells rely on a complex array of sensors and regulators to turn on and off genes. The diversity of phenotype a cell can exhibit upon the regulation of its genome is then known as the phenotypic plasticity.

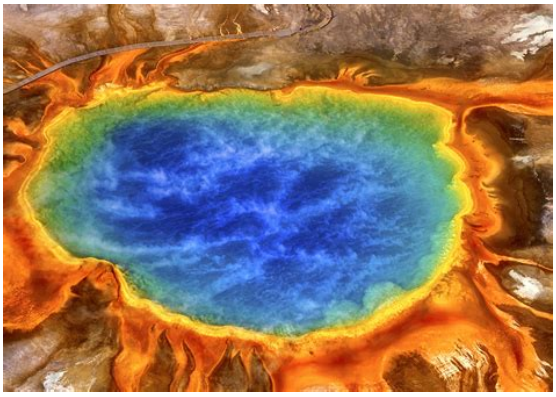

Figure 1.1: Microorganisms can exhibit remarkable adaptations to survive in challenging environments, such as extremely high temperatures and acidic conditions like in the hot springs of Yellowstone. The phenotype is the visible expression of genes, allowing the organism to thrive in conditions that would be hostile to many other forms of life. From Charles O'Rear/Corbis via Smithsonian

Interestingly, it has been observed that, despite the complexity of sensing and regulatory mechanisms, a clonal microbial population does not uniformly respond to clear environmental stimuli. Instead, gene expression varies across the population, fluctuating over time. Thus, cell populations exhibit phenotypes with varying degrees of strength and heterogeneity over time. These two characteristics are combined under the umbrella wording of population dynamics.

Consider the accumulation of glycogen in *Saccharomyces cerevisiae*. *S. cerevisiae*, also known as baker's yeast, has evolved to sense the availability of glucose, a carbon source, and convert a portion of it into a source of storage. This storage, glycogen is synthesized notably through the activation of the *glc3* gene when the abundance of glucose decreases [3]. Later on, it can then be broken back down into glucose to avoid famine if the sugar is to be exhausted. This phenotype sacrifices growth for resilience and serves as a prime example of phenotypic plasticity. When exposed to a low concentration of glucose,

not all clonal cells adopt this phenotype. Some continue to grow rapidly while others slow down and begin storing glycogen. Another example of an uneven response to a defined environment is the antibiotic persistence of *Escherichia coli* [4] [5]. As a gram-negative bacterium, *E. coli* can adapt the composition of pore-forming proteins (porins) in its outer membrane. Large pores allow for the rapid uptake of nutrients, promoting growth but increasing susceptibility to toxic compounds such as antibiotics that can more easily enter and kill the bacteria. Conversely, *E. coli* can express smaller porins characteristic of a phenotype defined by slower growth but better resistance to toxic compounds. This phenotype hampers growth and fitness in an environment where no toxic compounds are present. However, it is still observed alongside the fast-growth phenotype when no such compounds are found.

These two examples illustrate the concept of bet-hedging, a theory that justifies the presence of unfit phenotypes in a population as a strategy to mitigate risks in an unpredictable environment [6]. Imagine a scenario where glucose suddenly disappears or antibiotics are introduced into the environment. If all cells express a phenotype solely in response to the environment, the entire cell population would swiftly transition from a well-adapted state to an unfit one. This would be problematic as the population needs time to sense the new conditions, express mRNA, and translate it into proteins that form the phenotype. This process takes time and this time could prove fatal to the entire population. To prevent this predicament, maintaining a small proportion of the population in a less-fit state becomes sensible. These cells safeguard the entire population in case of sudden environmental changes. The heterogeneity in gene expression has been extensively studied across the genome of *E. coli*, revealing an intriguing correlation between gene noisiness and functionality [7]. Genes involved in maintenance functions, such as DNA repair and ribosome synthesis, exhibit relatively homogeneous expression in a population. In contrast, genes related to alternative carbon source utilization or membrane composition display more variability in expression, suggesting an evolutionary constraint on noisiness. This evolutionary constraint implies that there exists an optimal level of noisiness in gene expression. Kussel and Leibler proposed a model supporting this idea, indicating that the noisiness and responsiveness in the expression of a gene depends on: the cost of the sensing machinery associated to the responsive switching and the frequency and speed of environmental change [8]. For example, if an environment is changing quite fast, cells can afford a more costly sensing machinery than if the environment changes slowly. Also, once an environment has changed, if it remains constant for a while, a costly sensing machinery does not really make sense and a constant stochastic switching is favored [8].
As a nice confirmation of the selection pressure on gene expression stochastic component, the distribution in the lag time on growth associated to the growth on glucose to lactose in *E. coli* was changed [9]. It was observed that when transitioning to a media where glucose was exhausted to a media rich in lactose, *E. coli* can take up to 10 hours to resume growth. Interestingly some cell re-started growing faster and a wide

distribution in lag time was observed, suggesting noisy gene expression. By selecting this more responsive population many times, the authors changed the induction profile to a narrower distribution, i.e., a more responsive phenotype switching. Sequencing revealed mutations in the *Lac* operon changing the regulatory protein LacI structure.

1.2. Molecular mechanisms behind biological noise

Why gene expression can be noisy has already been discussed and appears to be under evolutionary pressure. The how is equally complex and is multifactorial.

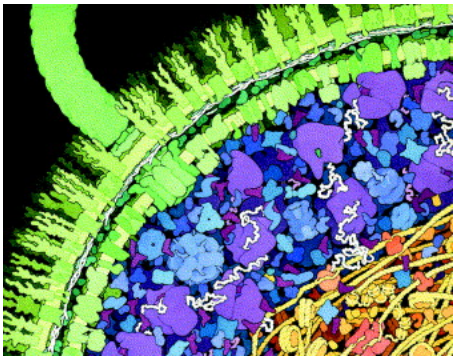

Figure 1.2: The intracellular environment is often more gel-like than a completely fluid space. It contains a diverse array of macromolecules in relatively low concentrations. Illustration by David S. Goodsell

The intracellular environment of a cell is crowded (Figure 1.2), characterized by the presence of numerous macromolecules present in uneven and sometimes very low numbers [10] [11]. The expression of a phenotype depends on the activation of one or many genes, themselves characterized by at least two reactions, transcription and traduction. All these reactions require the interaction of these scarce macromolecules making them probabilistic in time [12]. This is referred to as intrinsic noise. But intrinsic noise is not the sole source of stochasticity in gene expression. This intracellular crowding and consequently the poor diffusion of macromolecules means they inner content is not equally partitioned. This heterogeneity in the number of these macromolecules thus adds a second layer of noise called extrinsic noise. The number of gene copies may fluctuate among cells, with this disparity becoming more pronounced when the gene of interest resides on a plasmid. Furthermore, the "state" of individual cells within the population may differ; some may be in the midst of division cycles while others are newly formed offspring (Figure 1.3). Given that genes often form intricate networks wherein they reciprocally regulate each other, this inherent variability propagates throughout the genome. In essence, a myriad of factors influences the expression of a phenotype. Considering this complexity, it comes as no surprise that gene expression can manifest considerable variation between cells.

We now know that the degree of noise in gene expression depends on the functionality of the gene. Yet, intrinsic and extrinsic noise appear to impact the expression of all genes, so why are some genes more noisy than others? One factor that could impact the noisiness of a gene is the binding affinity of its regulator. A weaker affinity might produce a more stochastic expression, while a strong affinity could promote a

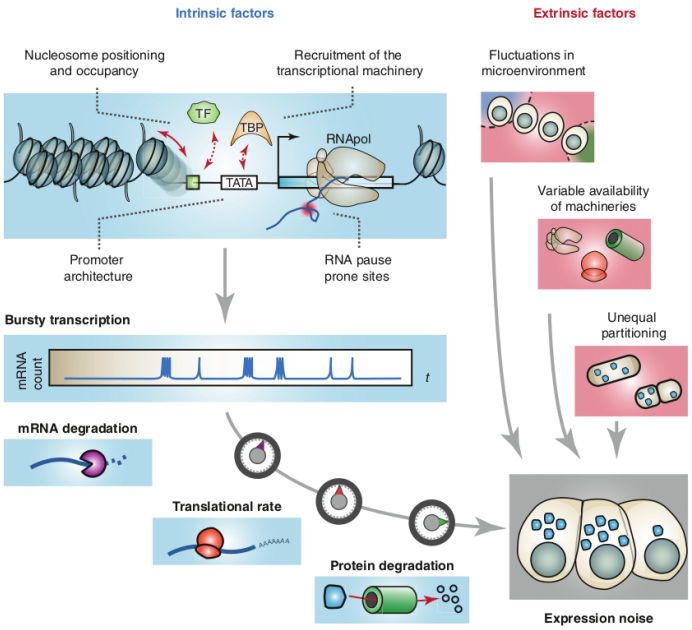

Figure 1.3: Extrinsic and intrinsic noise are inherent to biological systems and together yield differences in expression level between isogenic cells. Figure taken from Chalancon and colleges [13]

very responsive one. The shape of the regulator also affects the noise. For example, the regulator of the lac operon, LacI, is a dimeric protein that, when lactose is not present, binds to two operators above the lac promoter. By doing so, it forms a loop that hides the promoter, thus repressing expression. When lactose is present, it binds to the dimeric protein, changes shape, and cannot hide the lac promoter anymore. Interestingly, even in the absence of lactose, some cells transcribe the lac operon. This leaky expression can actually be reduced by changing one amino acid in the LacI regulator [14]. Again, suggesting the role of the regulator and the DNA region it binds to, it was shown that changing the operator sequence LacI binds to, can also diminish the leakiness of the system [15].

Another key factor contributing to the heterogeneity in gene expression is the complex gene network they belong to. Genes rarely function in isolation, as they are intricately connected to each other in the genome. Environmental transitions often require the expression of complex phenotypes, which necessitates the coordinated expression of multiple genes. A single transcription factor can activate multiple genes, and the resulting expression of these genes can trigger the expression of additional genes, forming a cascade. This creates a ripple effect (Figure 1.4), where noise in the activation of one gene can propagate throughout the network, influencing the expression of connected genes [16].

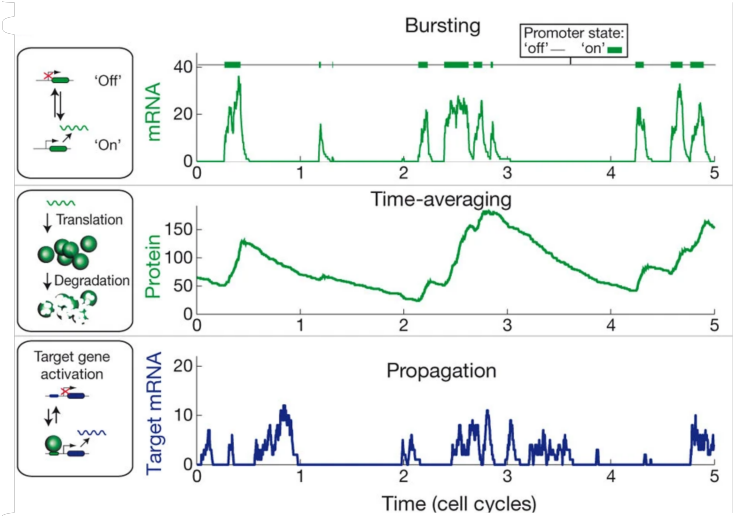

Figure 1.4: Intrinsic and extrinsic noise lead to bursty mRNA production that then gets smoothed out into a varying protein concentration. If this protein is a regulator embedded in a gene network, then its time varying concentration propagates the noise on the expression of other proteins. Figure from Eldar and colleges [16]

1.3. Gene network motives and noise

Throughout its life cycle, a cell encounters a diverse range of environments and transitions through various developmental phases. The bacterium *E. coli*, for instance, possesses 4,200 distinct genes, whereas human mammalian cells have approximately 30,000. To maintain optimal fitness in a given environment, a cell must interpret environmental cues and translate them into a phenotype comprising the expression or repression of one or multiple proteins. These genes require not only precise temporal and conditional expression but also varying levels of activation intensity. Furthermore, they must be activated in a sequential order and deactivated promptly, if necessary. All these properties are governed by the gene regulation network (GNR). In its simplest form, the GNR involves a transcription factor (TF) binding to a gene's promoter region to activate it. However, gene expression is often linked to feedback loops. Two prominent networks in *E. coli* are the negative autoregulation (NAR) and positive autoregulation (PAR) networks, where gene expression coincides with the production of a regulatory protein that inhibits or activates expression, respectively. NAR regulates approximately 40 % of genes [17], a significant percentage of genes in *E. coli*. The two networks, NAR and PAR, exhibit distinct dynamical properties that influence their ability to handle noise. NAR is characterized by a more graded response, which allows it to shorten response times and buffer out noise [17]. This property is often described as providing homeostasis, maintaining a stable equilibrium in the system. In contrast, PAR amplifies noise and generates a specific form of phenotypic heterogeneity in gene

expression, known as bimodality. In this scenario, the expression of a gene across a population tends to be binary, with cells being fully committed to a phenotype or not expressing it at all. This occurs because the concentration of the transcription factor must cross a threshold, triggering a positive feedback loop that locks the cell into an "on" state. If the threshold is not reached, the cells remain in the "off" state.
Most tf regulate up to three different genes [18] but some of them such as CRP can regulate 230 genes in *E. coli*. This leads to significant cross talks in the expression of genes and explain why noise can trickle down throughout the genome.

2. The need for predictable and controllable cells

In the previous section, the distinction between responsive and stochastic activation in gene expression was made, the rationale behind the ratio of both was explained, and the causes of noise were described. While this noise in gene expression makes sense from an evolutionary perspective, it can give rise to unwanted outcomes. Noise and unpredictable behavior arising from it have been observed at all levels of life, from microorganisms to mammalian and plant cells. Therefore, understanding cell dynamics is crucial in all topics linked to life science. In this section, I will briefly explain the significance of understanding cellular dynamics from a biomedical and bioprocess perspective.

Biomedical research is a field that aims to understand the chemical and biological principles that underlie medical treatments, with the ultimate goal of improving therapies. To achieve this, it is essential to understand the phenotypic dynamics, i.e. the behavior, of both bacterial and human cells. As previously explained, not all antibiotic resistance arises from acquired genome-encoded resistance. *E. coli* can change its membrane permeability to give rise to a slow growth/high resistance to antibiotics phenotype commonly named persistence [19]. Similarly, *Salmonella enterica* has been shown to manage the initial shock induced by antibiotics through variability in large pore-forming proteins (ompc) within the population [5]. The surviving subpopulation contains fewer of these porins and has a lower outer membrane permeability. This heterogeneity within the population allows it to survive long enough for mutation and selection to occur, resulting in a phenotype with a more active efflux pump that can then resume growth. In that sense, phenotypic heterogeneity within the population acts as a springboard for mutations and selection to occur.
It is suspected that for certain cancer types, a similar phenomenon occurs. A study [20] investigated the role of mismatch repair (MMR) deficiencies in Endometrioid endometrial cancer (EEC). MMR is crucial for the genome stability of cell because as it names implies, it corrects errors that occur during the DNA replication. Failing MMR drastically increases the mutation rate, and consequently, the appearance of cancerous cells. The study found that while only 5% of EEC cases had mutations impairing MMR, a significant 20% of cases exhibited MMR deficiency without any mutations.

This suggests that for certain cancer types, including EEC, a significant proportion of cases may be linked to epigenetic factors rather than solely mutations. This challenges the traditional view of cancer as primarily a condition caused by mutations and suggest a role for phenotypic heterogeneity in the onset of cancer. However, what is clear is that this cell-to-cell difference in phenotype within clonal cells complicates cancer treatment by adding to the existing challenge of genetic diversity arising from mutations. Similarly, to how Salmonella uses phenotypic heterogeneity to survive adverse conditions long enough to allow a resistant mutant to emerge, cancer cells form an expanding cone of phenotypes over time that makes treatment more and more arduous [21]. This expanding diversity enhances the cancer's resistance to drugs, as well as making it more aggressive, as phenotypes more prone to migration and metastasis appear.

Bioprocess resilience and productivity are also affected by this heterogeneity. While mutations can lead to loss of function over time, what primarily hampers bioprocess productivity within the initial hours is the noise in gene activation [22]. This incomplete activation means only a fraction of the population produces the desired product, while the rest utilizes resources without contributing. The cultivation of *Bacillus subtilis*, a soil bacterium used in agriculture for biocontrol, illustrates this challenge. Whether producing beneficial components like lipopeptides or being utilized directly in the field after sporulation, the desired outcomes are often hindered by a mix of both states during cultivation. Another example is the production of pluripotent cells for organ reconstruction. Ideally, these cells should remain pluripotent until needed for differentiation, but in practice, maintaining this state is challenging and limits production efficiency [23].

Additionally, heterogeneity can be exacerbated by the production process itself. Many processes rely on batch or fed-batch fermentation, where the environment transitions from a hyper-nutritious state to a poor one. This shift in conditions means that the environment is only briefly optimal for microbial cells, and the cells must constantly adapt to the new conditions. Adapting always comes at a cost, and the population expends significant resources to adapt to the process. This cost, combined with the lack of control over the induction strength of the gene of interest within the population, can result in poor product titer.

Scaling up has often been regarded as a good solution to increase production. But without understanding cell-to-cell heterogeneity, the incorrect assumption about uniform induction levels often lead to non-linear production increases. Indeed, in larger reactors, spacial heterogeneity in environmental conditions adds up to the temporal instability. As mixing time increases with volume, individual cells encounter different conditions (pH, oxygen content, nutrient concentration) over time, forming distinct environmental lifelines (Figure 1.5). And while there are examples where these fluctuating conditions have been shown to improve productivity [24], because environmental conditions can vary on sub-minute time scales, cells are constantly adapting

to conditions that are no longer relevant by the time they have expressed the adapted phenotype.

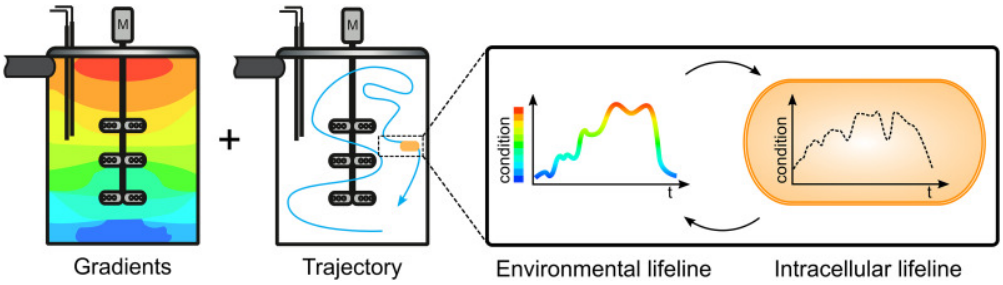

Figure 1.5: Representation of the microenvironments within a large scale bioreactor and the subsequent cells lifelines. Illustration taken from Blöbaum et al., [25]

Returning to the previously discussed yeast glycogen accumulation phenotype, in a large-scale bioreactor, yeast cells will encounter fluctuating areas of high and depleted glucose concentrations. In sufficiently large reactors, some cells may experience prolonged periods of glucose depletion, triggering them to switch to a low-growth phenotype that accumulates glycogen. However, this phenotype is likely ill-suited because the cell may subsequently encounter regions of high glucose availability and miss the opportunity to utilize it efficiently.
Continuous cultivation offers a promising alternative to batch processes, providing a more stable environment that allows cells to reach a steady state [26]. This approach enables longer cultivation times, which can be achieved in smaller reactors with greater control, reducing environmental heterogeneity and resulting in cost and energy savings. However, even in these more stable environments, phenotypic heterogeneity persists, emphasizing the need for effective control strategies. Synthetic biology is one such strategy, where scientists design and assemble well-characterized genetic building blocks to create controllable and predictable genetic switches.

3. Designing predictable and controllable genetic switches with synthetic biology

Gene circuits, like electrical circuits, would ideally operate with precision and predictability, allowing for seamless switching between 'on' and 'off' states. However, as previously discussed, biological systems are inherently complex and noisy, with gene circuits often deeply intertwined with other cellular components. This interconnectedness can lead to unpredictable behavior, as noise and variability propagate throughout the system. While some degree of stochasticity in gene expression may be unavoidable, it's also clear that certain gene architectures and components have evolved to amplify noise and variability [27, 6].

Synthetic biology aims to use well-defined building blocks, such as regulators and promoters, to construct functional gene circuits. These building blocks are assembled in a specific way to create a behavior that is as close as possible to the intended one. This modular approach allows for the design, construction, and optimization of gene circuits in a systematic and predictable manner. This is similar to how electronic circuits are built using a library of well-defined components, such as transistors and resistors, to create a desired functionality. One such engineered circuit, the toggle switch, utilized a tandem of two promoters, each controlling the expression of the repressor protein of the second (Figure 1.6). In addition to this dual-lock architecture, each promoter is inducible by a chemical inducer that cannot be metabolized by the cell. This makes the toggle switch highly controllable, multistable, and is thus a good example of how synthetic biology can create highly predictable and controllable circuits.

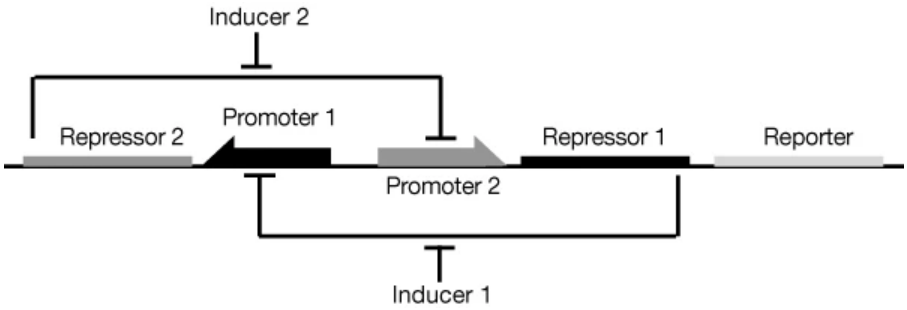

Figure 1.6: The toggle switch is a gene circuit that features a unique design, where two promoters are connected in a feedback loop. Each promoter drives the expression of the repressor of the other promoter. This clever configuration allows the circuit to be toggled between two stable states by adding either inducer, effectively "flipping" the switch between the two positions. This figure originates from the original publication featuring this gene circuit [28]

An alternative strategy involves developing host-aware systems that adjust gene expression strength based on cell resource availability. Since microbial cells have limited resources, diverting them towards production can hinder growth. To optimize product formation while preserving cell resources, parameters such as promoter strength, ribosome binding site (RBS), and gene of interest copy number can be fine-tuned. However, this approach is time-consuming and inflexible, requiring multiple design-build-test-learn cycles and redesign of elements for each change in growth conditions or product requirements. Researchers have pushed for the development of strategies that can adapt to dynamic environments [29]. Host-aware circuits offer a promising solution, as they are portable and do not require cell monitoring or external feedback control.

To replicate the complexity of natural cellular systems, synthetic biology has sought to integrate genetic circuitry in organisms like yeast and bacteria, enabling them to

undergo differentiation. These systems aim to induce the irreversible segregation of a homogeneous population into subgroups with distinct phenotypes, i.e., functionality. In yeast, the CRE-Lox system, originally used to remove antibiotic resistance markers, has been re-purposed to trigger differentiation. When the CRE recombinase is produced, the lox recognition sites are combined, activating a gene [30]. However, only a subset of the population is directed to differentiate, resulting in the emergence of a productive population with reduced growth rates, outcompeted by an undifferentiated population that prioritizes growth. This leads to a co-culture-like scenario, where both populations are genetically distinct. Furthermore, periodic re-differentiation is necessary, requiring the re-expression of CRE.

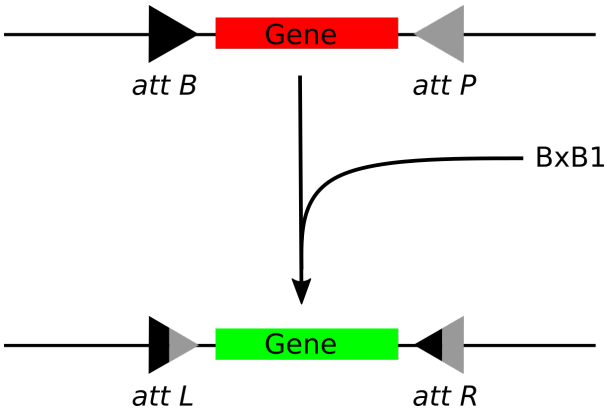

Figure 1.7: The Bxb1 integrase recognizes and recombines *attP* and *attB*. In doing so, it flips the sequence in between, making it code for a gene of interest. This irreversible switch permits differentiation in *E. coli*.

In *E. coli*, a system utilizing the BxB1 recombinase has been proposed. This recombinase recognizes two small regions known as *attP* and *attB* and recombines them, forming *attL* and *attR* sites (Figure 1.7), each containing half of the attP and attB sequences. Essentially, this means that whatever DNA sequence was originally situated between the two recognition sites is now oriented in the opposite direction. In other words, a DNA sequence that is non coding in one direction and that thus does not lead to a burden, can be inverted on demand in a fraction of a population. This sequence then codes for a burdensome gene that reduces growth. This productive population will then be periodically overtaken by the undifferentiated population that maintains genome stability. It thus acts as a reservoir for to be differentiated cells. The implementation of cell differentiation techniques benefit from cell machine interphase as it is important to monitor the proportion of differentiated cells at all time. Failing to differentiate enough cell is a waste of production potential whilst overshooting and differentiating all the population will lead the appearance of mutants without naive cells to outgrow them.

On top of generating new gene circuit outputs, synthetic biology has also been used to exploit new inputs. Optogenetic systems rely on light at specific wavelength to energizes engineered transcription factors to trigger the expression of a gene. There are many advantages to use light. Light can be added and "removed" to a culture in a highly precise manner. Micro-array mirrors can even be used to shine individual cells. Also, light is cheap compared to exotic inducer such as tetracycline.
All these synthetic biology approach are rather new. The first optogenetic system was described in 1995 [31], the toggle switch in 2000 [28], and the field is rapidly expending. Quantitative biology is at the heart of the all the research carried out in this expending field where using well-defined and controllable part is as important as relying on quantitative approach to characterize the behavior of these gene circuits.

4. Capturing and controlling gene expression dynamics with single cell analysis tools

Population analysis can be conducted at two levels: bulk and single-cell. Bulk measurements, such as optical density, total fluorescence, or oxygen consumption, are widely used for characterizing populations. However, single-cell measurements are essential in specific scenarios where understanding population heterogeneity is crucial. For instance, when comparing the efficacy of two induction systems, bulk measurements may suggest that strain B exhibits superior induction strength compared to strain A Figure 1.8. This might lead to prioritizing strain B for further optimization. However, single-cell analysis reveals a more nuanced outcome. While the average induction strength of strain B remains higher, single-cell analysis uncovers a subset of cells in strain A with significantly greater production potential than strain B. In this scenario, understanding the origin of this heterogeneity is crucial. Optimizing strain A to harness the potential of these high-producing cells might ultimately prove more beneficial than focusing solely on strain B.

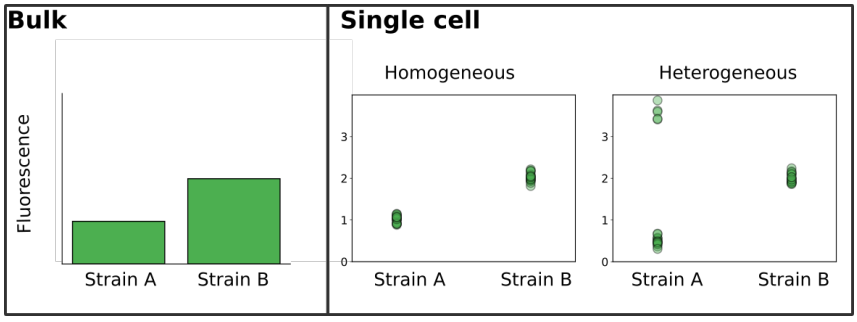

Figure 1.8: Bulk measurements can lead to incorrect assumption because they do not capture the diversity of expression within the population.

4.1. Microfluidic cultivation

A microfluidic is a specialized cultivation technique that involves introducing cells into chambers created on a polydimethylsiloxane (PDMS) chip, which can be mounted on a microscope for analysis. The chip provides a homogeneous and stable environment, with media continuously perfused in nanoliter-scale chambers that individually act as reactors Figure 1.9. The transparency of PDMS allows for the observation of cells within the chamber using a microscope. By repeatedly observing a chamber, a time-lapse is created, enabling the tracking of individual cell growth and gene expression, particularly when using fluorescent proteins. This feature enables the reconstruction of a cell's history and that of its entire lineage. The continuous perfusion of media provides a defined and stable environment, making it useful for characterizing strain behavior and observing cell-to-cell phenotypic heterogeneity despite a homogeneous environment. This tool has been used to characterize multiple inducible systems in *E. coli*, namely the induction strength, heterogeneity among the cell population, basal expression level, and growth reduction associated with the induction [32].

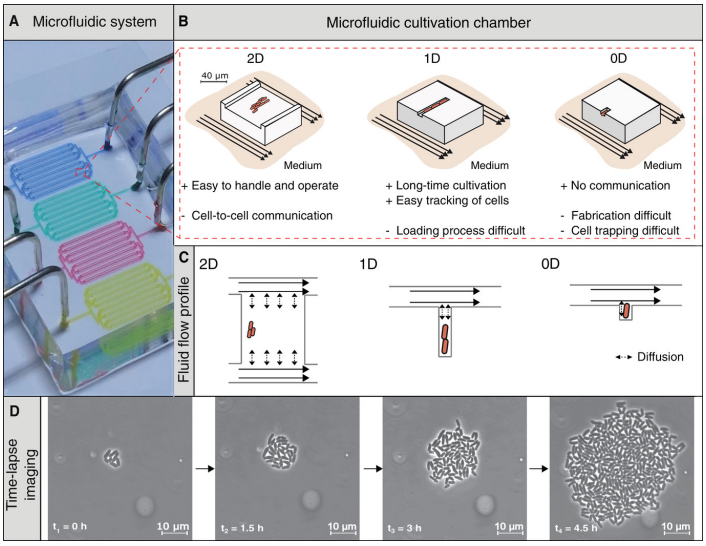

Figure 1.9: **A**. MSCC chip, containing several cultivation arrays with monolayer cultivation chambers. **B**. Different cultivation chamber geometries (2D to 0D) with their advantages and disadvantages and **C**. fluid flow schematic of these chambers. **D**. Live-cell imaging of typical microbial cell colony cultivated at constant environmental conditions. Figure from Täuber et al., 2020 [33]

A crucial factor driving the adoption of microfluidics is the associated data treatment. Recently, AI-based segmentation pieces of software have been developed, enabling effective cell segregation [34, 35, 36]. Generally, larger cells are easier to segment.

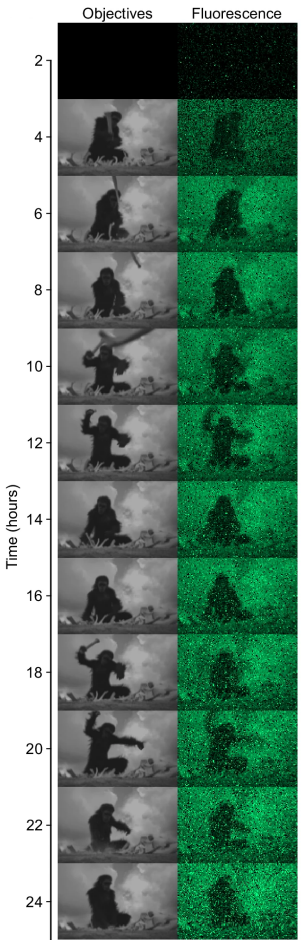

Figure 1.10: Control objectives more complex than single static threshold can be achieved as exemplified by Lugagne et al,. On the left a scene from *Space Odyssey*, on the right the GFP expression across the cell population.

These software tools prove valuable not only for post-experiment data treatment but can also be utilized in real-time during experiments to regulate processes, a field called cybergenetic. In this context, microfluidics can serve as an interface between the cell and the machine, where cells growing within are periodically monitored, and actuation can be performed based on predefined control rules. For instance, an inducer can be added whenever a specific proportion of the population falls below an induction threshold, which can be visualized using fluorescent proteins. Lugagne and his colleges have exemplified this by keeping a toggle switch in an unstable place with both proteins being expressed. Normally, a toggle switch only has two stable points. Without active control cells are either locked in one state or the other. By using a close loop control, the toogle switch was kept in an unstable point where both output were expressed [37].
Perrino and his colleagues highlighted the significance of precisely controlling gene expression, enabling them to investigate the relationship between synecluine concentration and its aggregation, a phenomenon hypothesized to contribute to the development of Parkinson's disease [38]. Recent advancements in control objectives, as demonstrated in Lugagne et al.'s work [39], have enabled the development of more sophisticated control strategies, paving the way for the generation of complex outcomes such as organs. In this work, a micro-mirror array to activate an optogenetic circuit of at the single cell level was coupled with AI for segmentation and decision-making to dynamically control the expression of GFP across thousands of cells (Figure 1.10). By leveraging this optogenetic system and combining it with single-cell fluorescence analysis, the authors developed an AI-driven model that enabled rapid, on-the-fly determination of the optimal light intensity for each of the thousands of analyzed cell to achieve the desired fluorescence pattern fig. 1.10. The level of control and observation microfluidic enables makes it a crucial tool for quantitative biology to understand interspecies ecology [40], bio-process robustness [41] and the functionality of genes [39].

4.2. Flow cytometry

Flow cytometry (FC) is a powerful single-cell analysis technique that characterizes cells as they pass through an interrogation point. As illustrated on fig. 1.11, this point typically consists of one or more lasers, detectors, and filters that capture and analyze the emitted signals. The silhouette formed by a cell's passage before a laser correlates with its size, while the reflected light provides insights into the cell's external structure. Additionally, the integration of fluorescent proteins, either expressed endogenously or attached to specific antibodies, enables the quantitative assessment of gene expression levels. This allows for the identification of pathogenic bacteria or cancerous cells, thereby amplifying the technique's utility in diverse research domains, including immunology and oncology.
Flow cytometry (FC) is a high-throughput technique that can analyze thousands of cells per minute, generating straightforward results files with numerical values for each cell and detector. The data requires minimal post-processing, making it relatively easy to utilize compared to microfluidics. However, FC also has several limitations. For example, it is prone to clogging and struggles to handle complex matrices containing high amounts of particles. Furthermore, each cell is only analyzed once, making it impossible to track individual cells over time.

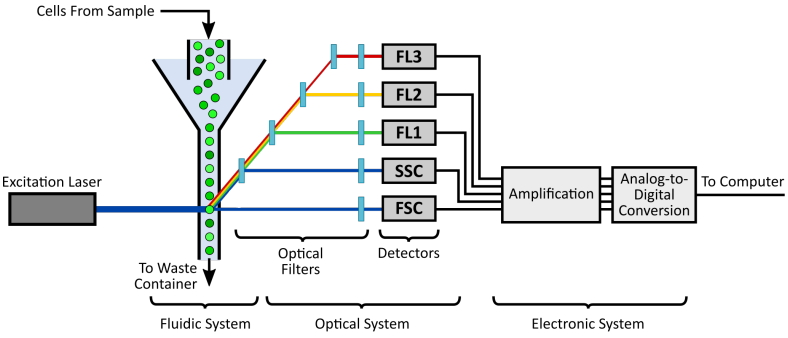

Figure 1.11: The sample is first structured using a laminar flow to separate the cells and have them pass one by one in front of the laser(s). This initial separation is only achievable if the concentration of cells is not too high. Then, each cell passes in front of a laser, and the optical signal (fluorescence, diffraction, reflection) is recorded and associated with the cell. Image credit to AAT Bioquest.

Furthermore, FC is limited in its ability to analyze cells in their native environment, as it requires a sample to be drawn from the environment, diluted, and then analyzed. This can lead to changes in the phenotype of cells during the analysis process. In contrast, microfluidics allows for the analysis of cells within their native environment without delay, providing a more accurate representation of cellular behavior.

Similar to microfluidics, FC has been incorporated in cell machine interphase. These interfaces, sometime referred to as reactive flow cytometry (sources), enable the anal-

ysis of cells in real-time, allowing for the control of cellular processes. In this setup, a sample is taken from a cultivation device, diluted, and then analyzed using FC. The resulting data can be quickly processed and utilized to exert control on the reactor through various means, such as the addition of a chemical inducer or, in optogenetics, by shining light onto the cells.
One such cell machine, ReacSight, utilized an open-source pipetting robot (Opentrons) to draw and dilute a sample from small reactors for subsequent FC [42]. The FC analysis statistics are then compared to a pre-set control threshold, and if the conditions are not met, actuation is performed. The device has been successfully employed to maximize protein production by dynamically regulating the expression based on the secretion stress in yeast. Induction was optimized using an optogenetic system, while simultaneously monitoring the stress level of the yeast [43]. The platform's flexibility was used to incorporate an additional step to quantify the actual amount of secreted product. Magnetic beads harboring an antibody that binds to the secreted proteins were incubated with the sample, washed and recover before being introduced in the FC to see the fluorescence resulting from the binding. This enabled the measurement of the amount of GFP secreted into the environment while at the same time regulating the production based on the stress level of the cells.
Another FC based cell-machine interface is the Segregostat. The Segregostat played a central role as the primary tool used in the subsequent research and is thus extensively described in the next section.

4.2.1. The Segregostat

The Segregostat (Figure 1.12) is a cell machine interface developed by Frank Delvigne's lab in the Microbial Process and Interaction Lab at Gembloux Agro Bio-Tech (University of Liege) [44]. It utilizes flow cytometry (FC) at its core to gather single-cell data from a lab-scale bioreactor. These data can be used not only for analysis but also for regulation purposes. This section describes how the Segregostat operates as well as the advantages and limitations of the technique.

The sampling process begins by extracting a sample from a lab-scale bioreactor to a dilution and analysis chamber using a capillary tube and a peristaltic pump. The capillary tube is designed to minimize dead volume, with a length of less than 30 cm and an inner diameter of 0.2 mm. To ensure that only fresh cells are analyzed, an initial flushing step is performed to remove any cells from the previous sampling. The analysis chamber is then cleaned by filling it with filtered phosphate-buffered saline (PBS) and emptying it multiple time. Then a sample is drawn again and diluted using a series of peristaltic pumps that sequentially add PBS and partially empty the chamber. The dilution sequence is determined based on the cell concentration recorded in the previous FC analysis. This sequence varies over time with the cellular concentration in the reactor, but unfortunately, it also varies over time due to tubing wear, making it unreliable for concluding long-term changes in cellular concentration. Once the sample is diluted

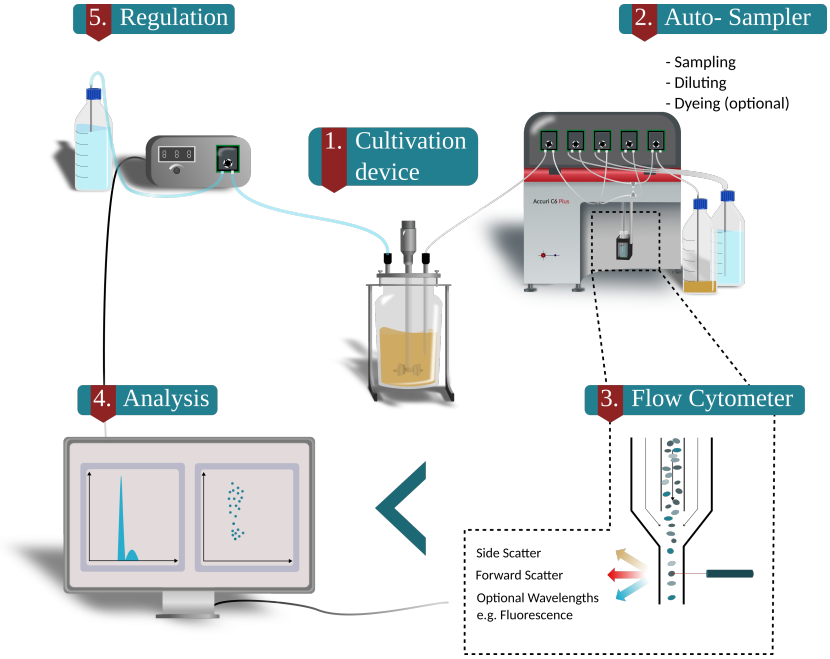

Figure 1.12: The Segregostat is a versatile cell machine interface. Here, the sample is drawn from a bioreactor, but because the sampling system relies on tubes and pumps, the device can be connected to any kind of cultivation setup. The FC rapid analysis time of tens of thousands of cells allows for the rapid capture of population complexity. These numerical data can then be used to compare the population to a target structure and apply regulation. This schematic representation of the Segregostat was produced by Maximilian Sehrt.

and the FC analysis is complete, the data undergoes cleaning to remove zero values, and statistical analyses are performed using an Octave script. These statistics, along with the raw data from the entire FC analysis, are then saved for further analysis. The process duration from analysis to analysis varies depending on the dilution sequence and the analysis parameters but roughly takes 10 minutes. The Segregostat is a versatile and adaptable cell machine interface that can be easily integrated into a wide range of cultivation setups, as long as a capillary for sampling can be introduced. While it offers many benefits, it also inherits some limitations from its underlying technology, including the risk of clogging and the inability to track individual cells.

The Segregostat can operate in either an open-loop configuration, where it solely monitors the cultivation process, or a closed-loop configuration, where the gathered data are used to trigger an actuation. In its current implementation, the actuation process employs a bang-bang controller, where a peristaltic pump is activated at a specific speed and time to deliver a precise amount of inducer into the reactor. However, future implementations could involve various types of actuation, such as optogenetic stimuli, pH changes, or alterations in dilution rates. As with all single cell analysis

tools, the amount of data gathered during the process is important and extends beyond general parameters such as means or standard deviation. Both in close and open loop configuration modes, data treatment is key to offer a good visual representation of the dynamics and to quantitatively appreciate how the population structures itself over time. Some of these principles that where adopted within our team come from the information theory field and are presented in the following section.

5. Utilizing information theory to enable population control

This section is an adapted version of the article:

> 1.Henrion, L., Delvenne, M., Bajoul Kakahi, F., Moreno-Avitia, F. & Delvigne, F. Exploiting Information and Control Theory for Directing Gene Expression in Cell Populations. Front. Microbiol. 13, 869509 (2022).

5.1. Abstract

Microbial populations can adapt to adverse environmental conditions either by appropriately sensing and responding to the changes in their surroundings or by stochastically switching to an alternative phenotypic state. Recent data point out that these two strategies can be exhibited by the same cellular system, depending on the amplitude/frequency of the environmental perturbations and on the architecture of the genetic circuits involved in the adaptation process. Accordingly, several mitigation strategies have been designed for the effective control of microbial populations in different contexts, ranging from biomedicine to bioprocess engineering. Technically, such control strategies have been made possible by the advances made at the level of computational and synthetic biology combined with control theory. However, these control strategies have been applied mostly to synthetic gene circuits, impairing the applicability of the approach to natural circuits. In this review, we argue that it is possible to expand these control strategies to any cellular system and gene circuits based on a metric derived from this information theory, i.e., mutual information (MI). Indeed, based on this metric, it should be possible to characterize the natural frequency of any gene circuits and use it for controlling gene circuits within a population of cells.

5.2. Introduction

The parallel advances made at the level of cell culturing procedures (i.e., microfluidics [45] and cell machine interfaces [37]), as well as the manipulation of gene circuits ([46] [47] [48]), have paved the way for the design of efficient cell population control procedures. It is now possible to act either on the cell population ([49];[50];[51]) or

on individual cells within the population ([52];[53]) for directing gene expression and cellular functions. In this review article, we will focus more precisely on a generic approach that could be used to control gene expression in individual cells among population. A critical aspect that must be taken into account before being able to manipulate gene expression in cell population is related to the inherent noise of cellular systems [54]. This noise induces cell-to-cell variability in gene expression, and a potential control procedure must be designed by taking into account the inherent functionality exhibited by noise on the cellular system ([47];[55]). Indeed, it is known that biological noise is a mechanism exploited by cell population in order to increase its fitness in front of fluctuating environmental conditions ([56];[19]). As an example, in natural ecosystems, microbial populations are often exposed to unpredictable environmental changes such as nutrient starvation, exposure to antibiotics, temperature variations, and many other sources of stress [55] that can fluctuate periodically or randomly. Cellular systems have then evolved accordingly by adapting different cellular components in order to accommodate such fluctuations involving different timescales. If environmental conditions change slowly and regularly, a responsive switching strategy leads to increased fitness for the cell population [57]. On the other hand, if environmental conditions are fast and erratic, a random switching mechanism, leading to pre-adapted cells, is more suited for optimizing population fitness. The study of phenotypic diversification mechanisms involved in antibiotic persistence in bacteria has pointed out that cellular systems can take benefit from both stochastic and responsive switching [57]. It is clear that, for designing an efficient population control procedure, stochastic switching must be minimized and responsive switching must be favored. Such responsive mechanisms typically involve gene circuits, able to record environmental changes and to respond accordingly.

A spectacular realization of the inference of periodic environmental changes by gene circuits is the implementation of circadian (oscillation with a period of 24 h; [58]) or ultradian (oscillation with a period < 24 h; [59]) rhythms by cellular systems. Many other gene circuit architectures or motifs are known to be able to infer extracellular signals and trigger appropriate biological responses ([60]; [61]). Even if we have now access to a classification of the motifs and their possible dynamics ([62]), it is still a challenge to infer the dynamics when several motifs are combined to each other or when the response interferes with many other cellular components. Indeed, in some case, gene circuit architectures can involve overlaps between different stress response pathways, allowing cells to anticipate environmental changes ([63]; [64]). This anticipatory switching arises in ecosystems where different environmental changes exhibit a strongly correlated time profile. As an example, Escherichia coli has evolved in order to be able to grow inside and outside a host (i.e., a mammals; [65]). When invading the host, *E. coli* is exposed to heat shock where temperature increases from 20 to 37°C. This heat shock is then followed by oxygen limitation as bacteria are reaching the gastrointestinal tract. The gene circuits involved in heat shock response and oxygen

limitation have been found to share common inputs and outputs in *E. coli*, elevation of heat leading to the adaptation to oxygen limitation in order to anticipate correlated environmental changes. Given all these elements, it is then difficult to infer the mode of switching, i.e., stochastic, responsive, or anticipatory (or a combination of them) based on the gene circuit architecture. Accordingly, we propose in this work, a generalizable approach aiming at stimulating the responsive component of switching for directing gene expression in cell population. Such approach could be made possible through the use of a universal metrics aiming at quantifying the information transfer efficiency in cells and leading to the design of robust cell machine interfaces.

5.3. Using information theory for determination the optimal stimulation frequency leading to coordinated gene expression in cell population

Cells are intrinsically programmed in order to react to external stimuli and to adapt appropriately by switching to different phenotypic states ([66]; [67]). It is then unrealistic to try to keep these cells into a specific phenotypic state, even if this would be a nice outcome for several applications, such as the optimization of cell factories for bioprocessing [68]. Indeed, these phenotypic states are linked to specific environmental states through selection pressure and the resulting fitness advantage, environmental condition being under constant evolution [56]. A more realistic alternative is to control cell switching itself, which is now technically feasible through the use of cell machine interfaces ([69]; [50]; [51]). In order to make this control strategy successful, two specific aspects must be taken into account, i.e., the efficiency in information transmission through the targeted gene circuits and the timing at which cells commit to phenotypic switching. These two aspects will be illustrated through a case study recently addressed, i.e., the synchronization for the activation of the gene circuit responsible for the induction of the arabinose operon in *E. coli* [51]. The relevance of this case study is also justified by the fact that the arabinose operon has been long used as a biological case study for the characterization of the functionality of biological noise in cell population [70] and also by the fact that the genes belonging to the arabinose operon are widely used for synthetic biology applications and notably for the synchronization of cell response ([71]; [72]). Finally, the arabinose operon is known to exhibit strong cell-to-cell variability both in the timing for activation ([70]; [73]) and also the level of expression of the corresponding genes [74], suggesting that the underlying cell switching mechanisms involves a mix of responsive and stochastic components. This make this system very interesting to be considered for possible coordination at the population level. The first step in the cell-to-cell coordination for the activation of the arabinose operon is to know the possible effector for the underlying gene circuit. The activation of the arabinose operon is under the control of a feedforward loop ([75]; Figure 1.13 A) combining the glucose depletion signal (through the accumulation of cAMP inside cells) and the presence of arabinose (through the activation of the transcription factor

AraC).

Under glucose-limiting conditions, it is then possible to activate or deactivate this gene circuit based on arabinose pulsing ([51]; Figure 1.13 A). At this stage, the first drawback exhibited by biological noise can be observed. Indeed, upon arabinose pulsing, cells will commit to the activation of the feedforward circuit leading to the synthesis of the different proteins involved in arabinose assimilation. However, due to biological noise, timing in commitment will exhibit cell-to-cell heterogeneity ([76]; [77]). Timing in cellular commitment to alternative phenotypes depends on the accumulation of regulatory proteins at the single cell level. Transcription and translation processes in individual cells are prone to biological noise([78]; [79]). These processes can be simulated based on the resolution of the chemical master equation or, more practically, based on the Gillespie algorithm [78]. These simulations have been shown to lead to very realistic pictures for mRNA and protein synthesis in individual cells [61] and pointed out that these processes follow Poisson statistics. Accordingly, the transition of cells between two adjacent phenotypic states (for example, the two states, GFP negative and GFP positive, drawn in Figure 1.13 A) can also be represented by a Poisson process. One key property of the Poisson process is that the timing between two consecutive events (e.g., the time between the synthesis of two mRNAs from the same DNA sequence in a single cell) follows an exponential distribution. Based on this statement, the residence time distribution of cells in a given phenotypic state can be represented by an exponential distribution [80].

It is thus very critical to take into account this residence time distribution for coordinating cell switching at the population level. One way to overcome this use is to rely on the use of a cell machine interface allowing the on-line monitoring of the switching process at the level of individual cells in the population and to react accordingly. This principle has been notably adopted for developing the Segregostat ([50]; [51]). This system is based on the use of on-line flow cytometry for recording the cell switching rate and to trigger environmental switching accordingly. The fact that the frequency of environmental perturbation must be set based on the phenotypic switching frequency has been previously deduced from numerical simulation [56]. Similarly, another study has pointed out that the control of gene circuits is dramatically reduced above a critical stimulation frequency [81]. Under these conditions, the frequency of the extracellular signal is effectively transmitted, leading to a cell population with synchronized gene expression (Figure 1.13 B). Effective entrainment of cell population can be assessed based on the oscillatory gene expression profile exhibiting a frequency close to the one of the input stimulations. Such oscillations were experimentally observed during Segregostat experiments carried out for controlling the activation of the arabinose operon [51]. It is also important to point out that in this case, square waves are used as stimulatory input. This strategy is also called pulse width modulation (PWM; [82]; [83]). We will see in the next section that this strategy has been used several times for controlling different cellular systems (listed in Table 1.3). In the present case, population

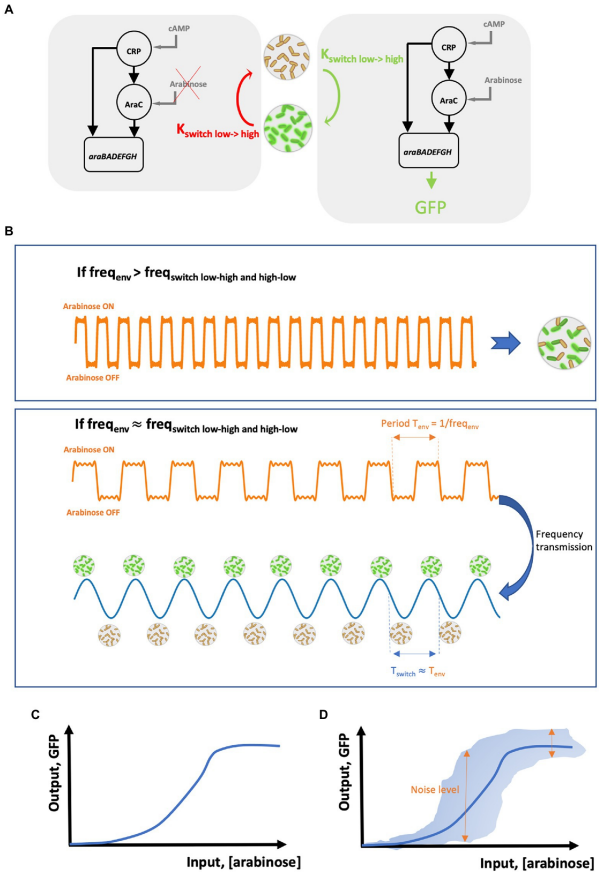

Figure 1.13: A Scheme of the feedforward loop motif involved in the regulation of the arabinose operon. On the left, arabinose is not available and the AraC branch cannot be induced. Accordingly, cell switching does not take place and, eventually, previously induced cells are relaxed back to the un-induced (low) state at a rate K_{switch} low->high. On the right, arabinose is available and the AraC branch, together with the cAMP-CRP branch, is activated leading to the induction of the genes araBADEFGH involved in arabinose metabolism. Under these conditions, cells from the low-state switch actively to the high state at a rate Kswitch high->low. **B** Proper coordination/synchronization of gene expression can be achieved based on periodic stimulations (or environmental fluctuations) made at a specific frequency freqenv. If freqenv is too high by comparison with the frequency for cell switching freqswitch, then cells are not coordinated and exhibit strong variability in gene expression. However, when freqenv is set close to freqswitch, coordination in gene expression is possible leading to synchronized gene expression. **C** Typical shape of a Hill relationship between an input (here the concentration of arabinose in the medium) and its resulting output (here, detected based on the synthesis of GFP based on a $P_{araB} : GFP$ transcriptional reporter). **D** Impact of biological noise (represented by double arrows) on the probability for delivering an output based on a given input.

oscillates according to a frequency corresponding to the one of the input square waves. This is represented in Figure 1.13 B based on the period Tenv (Tenv = 1/ frequency) of the input square wave stimulation, which is transmitted to the population and lead to oscillation in gene expression with the same period Tswitch = Tenv. All these observations point out that cells are able to deduce changes in their surroundings based on diverse sensory mechanisms. We do not want here to discuss about the biological diversity of these mechanisms, but rather to quantify the efficiency at which a cell is able to infer extracellular perturbation. A universal way to quantify information transmission through biochemical network can be derived from Shannon theory or information theory [84]. In order to be able to understand the importance of information theory in biochemical signal processing, it is important to introduce the concept of input output (i/o) or dose response relationship. For many gene circuits, this i/o relationship can be represented by a sigmoidal curve (Figure 1.13 C), also called Hill equation [85]. For example, Hill equation can be used to infer the response of the feedforward loop involved in the regulation of the arabinose operon [86]. This i/o correlation tells us what will be the output of the gene circuit according to a given input. However, we have seen that cell switching mechanism involves a random component in addition to the responsive one. This random component can be represented by the error bars on the i/o correlation (Figure 1.13 D).

Accordingly, one input can drive different output trajectories, leading to cell-to-cell heterogeneity. It can be seen that some input leads to a very heterogeneous response, making cell unable to properly infer the state of the environment. This is exactly where information theory can be useful, i.e., by providing a metric for quantifying the amount of information transmitted by the gene network for some specific input environmental conditions. This metric, mutual information (MI), corresponds to the logarithm of the number of distinct, input-dependent, states that can be reached by cells [85] and is quantified in bits. For example, a gene network exhibiting a MI of one bit means that only two physiologically distinct states can be resolved by cells based on the input conditions. Generally speaking, most of the gene circuits are corrupted by noise and can carry only a limited amount of information, and most of the studies carried out so far in this area have pointed out that MI equals to only 1 2 bits for different gene networks and model organisms ([87]; [60]; [84]; [88]; [89]). This leads to the conclusion that only these states have to be targeted when designing a cell population control strategy. The next section will be dedicated to the description of some realization in the field of cell population control (also termed cybergenetics), pointing out that the above-mentioned methodology could help at this level by providing a general framework aiming at developing further cell population control procedures.

5.4. The contribution of control theory and the nascent field of cybergenetics

The fact that cell population can be controlled based on pulsatile inputs has been reported a long time ago. Indeed, long before the advent of single-cell technologies, [90] observed that it was possible to synchronize cell cycle in *E. coli* cells by periodically pulsing phosphate in a phosphate-limited chemostat. This pioneering work has led to the establishment of a robust modeling framework for the understanding of the impact of external conditions on the synchronization of cell cycle for many types of organisms ([91]; [92]; [93]). However, these studies have been carried out based on an open-loop control approach and the application of regular pulses with varying frequencies and amplitudes. More recently, the application of control theory to the manipulation of cellular systems, i.e., cybergenetics, has set the ground for a more rational design of cell population control procedures ([49]; [94]). Cybergenetics is an entirely new and exciting field of research at the interface between control engineering and synthetic biology, and emerged with the recent advances made in genetic engineering combined with the works initially derived from cybernetics [95]. A distinction can be made between "internal cybergenetics" (also called in vivo and involving genetic controllers directly embedded in cells) and "external cybergenetics" (also called in silico controllers; [52]; [96]; [97]). In the context of this review, we will be focused more on the latter technology, since it involves cell machine interface and pulsatile inputs used as actuators.

Remarkably, although different systems have been used (e.g., various model organisms, type of gene circuits to be controlled, and single cell techniques), all the data accumulated point out that it is possible to effectively control gene circuits at the level of individual cells by applying external periodic signals (Table 1.3). Evidences have been provided suggesting that pulses of inducers tend to decrease noise in biochemical network, leading to synchronized gene expression ([98]; [99]). This effect can be explained based on the dose response relationship (Figure 1.13 D) where input concentrations at the extremities of the dynamic range lead to a homogenous response at the population level. In contrast, input concentrations at the center of the dynamic range produce a heterogeneous population. This strategy, known as PWM, seems to be generalizable for the effective control of diverse gene circuits in diverse cellular systems. Most of the experiments involving the control of gene expression in cellular systems have been performed in microfluidic devices (Table 1.3). This type of cultivation device allows the acquisition of single cell data with a high spatio-temporal resolution, but with a low experimental throughput due to the time and computational power required for image analysis [100] and with possible technical biases by comparison with conventional cultivation devices ([101]; [102]). Nonetheless, there is a growing interest in using standard cultivation devices (e.g., flasks, bioreactors, etc.) for studying and controlling cell populations [103]. In this case, single-cell analyses can

Approximated period T for the input stimulationA	Type of perturbationB	Controlled trait	Organism	Culture system	Single-cell analysis tool	References
83 min*	Pulse of phosphate in phosphate-poor medium	Cell cycle	*E. coli*	Bioreactor 240 ml	None	Goodwin, 1969
98 min*	Pulses of methionine to induce (MET3pr-CLN2)	A long description of the controlled trait	*S. cerevisiae*	Microfluidics	Microscopy	Charvin et al., 2009
110 min*	Nutrient availability, i.e., poor-rich	Cell cycle	*S. cerevisiae*	Microfluidics	Microscopy	Tian et al., 2012
16 min	Red-far red	gal1-	*S. cerevisiae*	Microplates with 96 wells	Flow cytometry	Milias-Argeitis et al., 2011, 2016
28 min	Pulses of sorbitol-enriched medium in normal medium	Osmostress (value of 1,500 a.u.)	*S. cerevisiae*	Microfluidics	Microscopy	Uhlendorf et al., 2012
142 min	Glucose-galactose	pgal1-GFP	*S. cerevisiae*	Microfluidics	Microscopy	Fiore et al., 2016
990 min	Tetracycline	Tetracycline-inducible system p(CMV-TET)-d2EYFP	Mammalian (CHO)	Microfluidics	Microscopy	Fracassi et al., 2016
120 min	Green light intensity (in relation with red light intensity)	CcaS/CcaR gene expression system at a varying level of expression	*E. coli*	Bioreactor 20 ml	Automated flow cytometry	Milias-Argeitis et al., 2016

Approximated period T for the input stimulationA	Type of perturbationB	Controlled trait	Organism	Culture system	Single-cell analysis tool	References
6 min	Computed open-loop green-red light	ccaSR-based system	*E. coli*	Microfluidics	Microscopy	Chait et al., 2017
150 min 210 min (depending on the amplitude/concentration)	IPTG—aTc	Maintain a toggle switch at an unstable intermediary level	*E. coli*	Microfluidics	Microscopy	Lugagne et al., 2017
10 min	High and low red intensity blue light	Transcription	*S. cerevisiae*	Microfluidics	Microscopy	Rullan et al., 2018
For GFP: 5 s off/80s on 8 s off/80 s on 11 s off 80 s on. For isobutanol and butanol: 6 s off and 15 s on	Blue light	-GFP expression -production of isobutanol and 2-methyl-1-butanol	*S. cerevisiae*	For GFP: Microplates 24 wells For isobutanol and butanol: Bioreactor 0.5 L	None	Zhao et al., 2018
5-60 min	Pulses of galactose to stabilize synecluine concentration at different values	a-synuclein formationl	*S. cerevisiae*	Microfluidics	Microscopy	Perrino et al., 2019
90 min / 105min / 180 min	Methionine concentration in cultivation medium	Yeast cell cycle coordination	*S. cerevisiae*	Microfluidics	Microscopy	Perrino et al., 2021
600 min	Arabinose pulses	Arabinose operon induction	*E. coli*	Bioreactor 1 L (continuous mode)	Online flow-cytometry	Nguyen et al., 2021

Table 1.3: Summary of control

be performed based on automated flow cytometry, leading to the rapid accumulation of data at the population level. In this context, the use of cell machine interface relying on flow cytometry can lead to the automated determination systematic determination of the optimal stimulation frequency for the effective synchronization of gene expression at the population level [51].

5.5. Perspective: Exploiting intrinsic frequency of gene circuits

Taken altogether, the elements assembled in the previous sections point out that a lot of different gene circuits architectures can exhibit periodic behavior [and not only the motifs reported behaving as natural oscillators, such as the repressilator [104] or the oscillator motif ([71]; [72])] if stimulated at the appropriate frequencies [81]. Development made in information theory and in cybergenetics provides the computational framework and the experimental tools in order to generalize this concept to many biological systems. Impressive achievements can be expected from these field of research such as the control of complex cell regulatory program (e.g., control of cell cycle program; [105]) and the control of microbial community's composition ([106]; [107]), with applications in various field from bioproduction ([108]; [103]) to biomedicine ([82]; [109]).

6. Thesis objectives

The primary objectives of this thesis are twofold. The first is to enhance our general understanding of microbial dynamics, i.e., how is a gene expressed in a population (strength and heterogeneity) over time. To achieve this objective, we primarily focused our attention on continuous cultivation devices, examining why and how the expression of one phenotype evolves in a homogeneous environment.

The second objective of this thesis is to develop methods for operating control over these microbial dynamics with the aim of achieving more predictable and productive outcomes from continuous cultivation. Initially, it might have been reasonable to assume that this thesis would be divided into two parts, with understanding the dynamics coming first and the control aspect following. However, as the reader will notice throughout this manuscript, and as I have observed during the four years of my thesis, both objectives are intertwined and feed into each other.
Understanding and controlling microbial dynamics in continuous cultivation has significant applied implications. The COVID-19 crisis has highlighted the importance of biomanufacturing, and greater productivity would have driven costs down, increased production volumes of lifesaving antibodies and speed up the vaccine roll-out. Additionally, more efficient bioprocesses can accelerate the emergence of the bioeconomy and serve as an enabler for various research initiatives that have demonstrated the possibility of replacing fossil-based products with bio-based alternatives.

This thesis aims to be as general as possible and will therefore focus on multiple microorganisms. The ultimate goal being to infer population dynamics rules and feed into the fundamental aspects of science that aim to understand how living organisms behave, with all the applied implications that follow.

Bibliography

[1] Zhiqiang Gan and Houjin Zhang. "PMBD: a Comprehensive Plastics Microbial Biodegradation Database". In: *Database* 2019 (Jan. 1, 2019), baz119. ISSN: 1758-0463. DOI: 10.1093/database/baz119. URL: https://academic.oup.com/database/article/doi/10.1093/database/baz119/5625879 (visited on 10/02/2024).

[2] Fungmin Eric Liew et al. "Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale". In: *Nature Biotechnology* 40.3 (Mar. 2022), pp. 335–344. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/s41587-021-01195-w. URL: https://www.nature.com/articles/s41587-021-01195-w (visited on 10/01/2024).

[3] Donald W Rowen, Marilyn Meinke, and David C Laporte. "GLC3 and GHAI of Saccharomyces cerevisiae Are Allelic and Encode the Glycogen Branching Enzyme". In: *MOL. CELL. BIOL.* ().

[4] Markus Arnoldini et al. "Bistable Expression of Virulence Genes in Salmonella Leads to the Formation of an Antibiotic-Tolerant Subpopulation". In: *PLoS Biology* 12.8 (Aug. 19, 2014). Ed. by Nathalie Balaban, e1001928. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1001928. URL: https://dx.plos.org/10.1371/journal.pbio.1001928 (visited on 09/18/2024).

[5] María Antonia Sánchez-Romero and Josep Casadesús. "Contribution of phenotypic heterogeneity to adaptive antibiotic resistance". In: *Proceedings of the National Academy of Sciences* 111.1 (Jan. 7, 2014), pp. 355–360. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1316084111. URL: https://pnas.org/doi/full/10.1073/pnas.1316084111.

[6] Martin Ackermann. "A functional perspective on phenotypic heterogeneity in microorganisms". In: *Nature Reviews Microbiology* 13.8 (Aug. 2015), pp. 497–508. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/nrmicro3491. URL: https://www.nature.com/articles/nrmicro3491.

[7] Olin K. Silander et al. “A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli”. In: *PLoS Genetics* 8.1 (Jan. 19, 2012). Ed. by Ivan Matic, e1002443. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1002443. URL: https://dx.plos.org/10.1371/journal.pgen.1002443.

[8] Edo Kussell and Stanislas Leibler. “Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments”. In: *Science* 309.5743 (Sept. 23, 2005), pp. 2075–2078. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1114383. URL: https://www.science.org/doi/10.1126/science.1114383.

[9] Stefany Moreno-Gámez et al. “Wide lag time distributions break a trade-off between reproduction and survival in bacteria”. In: *Proceedings of the National Academy of Sciences* 117.31 (Aug. 4, 2020), pp. 18729–18736. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2003331117. URL: https://pnas.org/doi/full/10.1073/pnas.2003331117.

[10] R.John Ellis. “Macromolecular crowding: obvious but underappreciated”. In: *Trends in Biochemical Sciences* 26.10 (Oct. 2001), pp. 597–604. ISSN: 09680004. DOI: 10.1016/S0968-0004(01)01938-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968000401019387.

[11] R. Grima. “Intrinsic biochemical noise in crowded intracellular conditions”. In: *The Journal of Chemical Physics* 132.18 (May 14, 2010), p. 185102. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.3427244. URL: https://pubs.aip.org/jcp/article/132/18/185102/905604/Intrinsic-biochemical-noise-in-crowded.

[12] Maike M. K. Hansen et al. “Macromolecular crowding creates heterogeneous environments of gene expression in picolitre droplets”. In: *Nature Nanotechnology* 11.2 (Feb. 2016), pp. 191–197. ISSN: 1748-3387, 1748-3395. DOI: 10.1038/nnano.2015.243. URL: https://www.nature.com/articles/nnano.2015.243 (visited on 10/01/2024).

[13] Guilhem Chalancon et al. “Interplay between gene expression noise and regulatory network architecture”. In: *Trends in Genetics* 28.5 (May 2012), pp. 221–232. ISSN: 01689525. DOI: 10.1016/j.tig.2012.01.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S016895251200015

[14] Pietro Gatti-Lafranconi et al. “A single mutation in the core domain of the lac repressor reduces leakiness”. In: *Microbial Cell Factories* 12.1 (Dec. 2013), p. 67. ISSN: 1475-2859. DOI: 10.1186/1475-2859-12-67. URL: https://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-12-67.

[15] Robert Daber and Mitchell Lewis. “A Novel Molecular Switch”. In: *Journal of Molecular Biology* 391.4 (Aug. 2009), pp. 661–670. ISSN: 00222836. DOI: 10.1016/j.jmb.2009.06.039. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022283609007591.

[16] Avigdor Eldar and Michael B. Elowitz. “Functional roles for noise in genetic circuits”. In: *Nature* 467.7312 (Sept. 2010), pp. 167–173. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature09326. URL: https://www.nature.com/articles/nature09326.

[17] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. “Negative Autoregulation Speeds the Response Times of Transcription Networks”. In: *Journal of Molecular Biology* 323.5 (Nov. 2002), pp. 785–793. ISSN: 00222836. DOI: 10.1016/S0022-2836(02)00994-4. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022283602009944.

[18] Denis Thieffry et al. “From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli”. In: *BioEssays* 20.5 (Dec. 6, 1998), pp. 433–440. ISSN: 02659247. DOI: 10.1002/(SICI)1521-1878(199805)20:5<433::AID-BIES10>3.0.CO;2-2. URL: https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1521-1878(199805)20:5<433::AID-BIES10>3.0.CO;2-2.

[19] E. Kussell et al. “Bacterial persistence: a model of survival in changing environments”. In: *Genetics* 169 (2005), 1807–1814. DOI: 10.1534/genetics.104.035352.

[20] Casey M. Cosgrove et al. “Epigenetic silencing of MLH1 in endometrial cancers is associated with larger tumor volume, increased rate of lymph node positivity and reduced recurrence-free survival”. In: *Gynecologic Oncology* 146.3 (Sept. 2017), pp. 588–595. ISSN: 00908258. DOI: 10.1016/j.ygyno.2017.07.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0090825817309800.

[21] Guido Lenz et al. “The Origins of Phenotypic Heterogeneity in Cancer”. In: *Cancer Research* 82.1 (Jan. 1, 2022), pp. 3–11. ISSN: 0008-5472, 1538-7445. DOI: 10.1158/0008-5472.CAN-21-1940. URL: https://aacrjournals.org/cancerres/article/82/1/3/674851/The-Origins-of-Phenotypic-Heterogeneity-in.

[22] Frank Delvigne and Philippe Goffin. “Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations”. In: *Biotechnology Journal* 9.1 (Jan. 2014), pp. 61–72. ISSN: 1860-6768, 1860-7314. DOI: 10.1002/biot.201300119. URL:

https://onlinelibrary.wiley.com/doi/10.1002/biot.201300119.

[23] Chunhui Xu et al. “Feeder-free growth of undifferentiated human embryonic stem cells”. In: *TECHNICAL REPORT* ().

[24] Edgar Velastegui et al. “Is heterogeneity in large-scale bioreactors a real problem in recombinant protein synthesis by Pichia pastoris?” In: *Applied Microbiology and Biotechnology* 107.7 (Apr. 2023), pp. 2223–2233. ISSN: 0175-7598, 1432-0614. DOI: 10.1007/s00253-023-12434-2. URL: https://link.springer.com/10.1007/s00253-023-12434-2.

[25] Luisa Blöbaum, Cees Haringa, and Alexander Grünberger. “Microbial lifelines in bioprocesses: From concept to application”. In: *Biotechnology Advances* 62 (Jan. 2023), p. 108071. ISSN: 07349750. DOI: 10.1016/j.biotechadv.2022.108071. URL: https://linkinghub.elsevier.com/retrieve/pii/S0734975022001677.

[26] Aaron Novick and Leo Szilard. “Description of the Chemostat”. In: *Science* 112.2920 (Dec. 15, 1950), pp. 715–716. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.112.2920.715. URL: https://www.science.org/doi/10.1126/science.112.2920.715 (visited on 09/13/2024).

[27] O. K. et al. Silander. “A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli”. In: *PLoS Genet.* 8 (2012), e1002443.

[28] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. “Construction of a genetic toggle switch in Escherichia coli”. In: *Nature* 403.6767 (Jan. 2000), pp. 339–342. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/35002131. URL: https://www.nature.com/articles/35002131.

[29] Alice Boo, Tom Ellis, and Guy-Bart Stan. “Host-aware synthetic biology”. In: *Current Opinion in Systems Biology* 14 (Apr. 2019), pp. 66–72. ISSN: 24523100. DOI: 10.1016/j.coisb.2019.03.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S245231001930006X.

[30] Chetan Aditya et al. “A light tunable differentiation system for the creation and control of consortia in yeast”. In: *Nature Communications* 12.1 (Oct. 5, 2021), p. 5829. ISSN: 2041-1723. DOI: 10.1038/s41467-021-26129-7. URL: https://www.nature.com/articles/s41467-021-26129-7 (visited on 11/06/2023).

[31] Georg Nagel et al. “Functional expression of bacteriorhodopsin in oocytes allows direct measurement of voltage dependence of light induced H^+ pumping”. In: *FEBS Letters* 377.2 (Dec. 18, 1995), pp. 263–266. ISSN: 0014-5793, 1873-3468. DOI: 10.1016/0014-5793(95)01356-3. URL: https://febs.onlinelibrary.wiley.com/doi/10.1016/0014-5793%2895%2901356-3.

[32] Dennis Binder et al. “Comparative Single-Cell Analysis of Different E. coli Expression Systems during Microfluidic Cultivation”. In: *PLOS ONE* 11.8 (Aug. 15, 2016). Ed. by Paul D. Riggs, e0160711. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0160711. URL: https://dx.plos.org/10.1371/journal.pone.0160711.

[33] Sarah Täuber, Eric von Lieres, and Alexander Grünberger. “Dynamic Environmental Control in Microfluidic Single-Cell Cultivations: From Concepts to Applications”. In: *Small* 16.16 (Apr. 2020), p. 1906670. ISSN: 1613-6810, 1613-6829. DOI: 10.1002/smll.201906670. URL: https://onlinelibrary.wiley.com/doi/10.1002/smll.201906670.

[34] Kevin J. Cutler et al. “Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation”. In: *Nature Methods* 19.11 (Nov. 2022), pp. 1438–1448. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-022-01639-4. URL: https://www.nature.com/articles/s41592-022-01639-4.

[35] Carsen Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. In: *Nature Methods* 18.1 (Jan. 2021), pp. 100–106. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-020-01018-x. URL: https://www.nature.com/articles/s41592-020-01018-x.

[36] Marius Pachitariu and Carsen Stringer. “Cellpose 2.0: how to train your own model”. In: *Nature Methods* 19.12 (Dec. 2022), pp. 1634–1641. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-022-01663-4. URL: https://www.nature.com/articles/s41592-022-01663-4.

[37] J.-B. et al. Lugagne. “Balancing a genetic toggle switch by real-time feedback control and periodic forcing”. In: *Nat. Commun.* 8 (2017), p. 1671.

[38] Giansimone Perrino et al. “Quantitative Characterization of Synuclein Aggregation in Living Cells through Automated Microfluidics Feedback Control”. In: *Cell Reports* 27.3 (Apr. 2019), 916–927.e5. ISSN: 22111247. DOI: 10.1016/j.celrep.2019.03.081. URL: https://linkinghub.elsevier.com/retrieve/pii/S2211124719304206.

[39] Jean-Baptiste Lugagne, Caroline M. Blassick, and Mary J. Dunlop. “Deep model predictive control of gene expression in thousands of single cells”. In: *Nature Communications* 15.1 (Mar. 8, 2024), p. 2148. ISSN: 2041-1723. DOI: 10.1038/s41467-024-46361-1. URL: https://www.nature.com/articles/s41467-024-46361-1.

[40] Alma Dal Co et al. “Short-range interactions govern the dynamics and functions of microbial communities”. In: *Nature Ecology & Evolution* 4.3 (Feb. 10, 2020), pp. 366–375. ISSN: 2397-334X. DOI: 10.1038/s41559-019-

1080-2. URL: https://www.nature.com/articles/s41559-019-1080-2.

[41] Phuong Ho et al. "Microfluidic Reproduction of Dynamic Bioreactor Environment Based on Computational Lifelines". In: *Frontiers in Chemical Engineering* 4 (Mar. 31, 2022), p. 826485. ISSN: 2673-2718. DOI: 10.3389/fceng.2022.826485. URL: https://www.frontiersin.org/articles/10.3389/fceng.2022.826485/full.

[42] F. et al. Bertaux. "Enhancing bioreactor arrays for automated measurements and reactive control with ReacSight". In: *Nat. Commun.* 13 (2022), p. 3363.

[43] Sebastián Sosa-Carrillo et al. "Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches". In: *Nature Communications* 14.1 (May 25, 2023), p. 3028. ISSN: 2041-1723. DOI: 10.1038/s41467-023-38807-9. URL: https://www.nature.com/articles/s41467-023-38807-9.

[44] H. et al. Sassi. "Segregostat: a novel concept to control phenotypic diversification dynamics on the example of Gram-negative bacteria". In: *Microb. Biotechnol.* 12 (2019), pp. 1064–1075.

[45] Alexander Grünberger, Wolfgang Wiechert, and Dietrich Kohlheyer. "Single-cell microfluidics: opportunity for bioprocess development". In: *Current Opinion in Biotechnology* 29 (Oct. 2014), pp. 15–23. ISSN: 09581669. DOI: 10.1016/j.copbio.2014.02.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0958166914000391.

[46] W. W. Wong and J. C. Liao. "The design of intracellular oscillators that interact with metabolism". In: *Cell Mol. Life Sci.* 63 (2006), pp. 1215–1220.

[47] Joe H. Levine, Yihan Lin, and Michael B. Elowitz. "Functional Roles of Pulsing in Genetic Circuits". In: *Science* 342.6163 (Dec. 6, 2013), pp. 1193–1200. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1239999. URL: https://www.science.org/doi/10.1126/science.1239999 (visited on 09/18/2024).

[48] M. O. Din et al. "Interfacing gene circuits with microelectronics through engineered population dynamics". In: *Sci. Adv.* 6 (2020), eaaz8344.

[49] A. Milias-Argeitis et al. "Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth". In: *Nat. Commun.* 7 (2016), p. 12546. DOI: 10.1038/ncomms12546.

[50] H. Sassi et al. "Segregostat: a novel concept to control phenotypic diversification dynamics on the example of gram-negative bacteria". In: *Microb. Biotechnol.* 12 (2019), 1064–1075. DOI: 10.1111/1751-7915.13442.

[51] T. M. Nguyen et al. “Reducing phenotypic instabilities of a microbial population during continuous cultivation based on cell switching dynamics”. In: *Biotechnol. Bioeng.* 118 (2021), 3847–3859. DOI: 10.1002/bit.27860.

[52] J.-B. Lugagne et al. “Balancing a genetic toggle switch by real-time feedback control and periodic forcing”. In: *Nat. Commun.* 8 (2017), p. 1671. DOI: 10.1038/s41467-017-01498-0.

[53] M. Rullan et al. “An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation”. In: *Mol. Cell* 70 (2018), 745.e6–756.e6. DOI: 10.1016/j.molcel.2018.04.012.

[54] Y. Pilpel. “Noise in biological systems: pros, cons, and mechanisms of control”. In: *Methods Mol. Biol.* 759 (2011), 407–425. DOI: 10.1007/978-1-61779-173-4_23.

[55] M. Ackermann. “A functional perspective on phenotypic heterogeneity in microorganisms”. In: *Nat. Rev. Microbiol.* (2015), pp. 497–508. DOI: 10.1038/nrmicro3491.

[56] M. Thattai and A. van Oudenaarden. “Stochastic gene expression in fluctuating environments”. In: *Genetics* 167 (2004), pp. 523–530. DOI: 10.1534/genetics.167.1.523.

[57] E. Kussell and S. Leibler. “Phenotypic diversity, population growth, and information in fluctuating environments”. In: *Science* 309 (2005), 2075–2078. DOI: 10.1126/science.1114383.

[58] R. M. Voigt et al. “Circadian rhythm and the gut microbiome”. In: *Int. Rev. Neurobiol.* 131 (2016), pp. 193–205. DOI: 10.1016/bs.irn.2016.07.002.

[59] A. Isomura and R. Kageyama. “Ultradian oscillations and pulses: coordinating cellular responses and cell fate decisions”. In: *Development* 141 (2014), 3627–3636. DOI: 10.1242/dev.104497.

[60] T. J. Perkins and P. S. Swain. “Strategies for cellular decision-making”. In: *Mol. Syst. Biol.* 5 (2009), p. 326. DOI: 10.1038/msb.2009.83.

[61] G. Balazsi, A. van Oudenaarden, and J. J. Collins. “Cellular decision making and biological noise: from microbes to mammals”. In: *Cell* 144 (2011), pp. 910–925. DOI: 10.1016/j.cell.2011.01.030.

[62] O. Shoval and U. Alon. “Snap shot: network motifs”. In: *Cell* 143 (2010), 326–326.e1. DOI: 10.1016/j.cell.2010.09.050.

[63] I. Tagkopoulos, Y.-C. Liu, and S. Tavazoie. “Predictive behavior within microbial genetic networks”. In: *Science* 320 (2008), 1313–1317. DOI: 10.1126/science.1154456.

[64] P. L. Freddolino and S. Tavazoie. “Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior”. In: *Annu. Rev. Cell Dev. Biol.* 28 (2012), pp. 363–384. DOI: 10.1146/annurev-cellbio-092910-154129.

[65] A. Mitchell et al. “Adaptive prediction of environmental changes by microorganisms”. In: *Nature* 460 (2009), 220–224. DOI: 10.1038/nature08112.

[66] M. Acar, J. T. Mettetal, and A. van Oudenaarden. “Stochastic switching as a survival strategy in fluctuating environments”. In: *Nat. Genet.* 40 (2008), pp. 471–475. DOI: 10.1038/ng.110.

[67] F. Schreiber and M. Ackermann. “Environmental drivers of metabolic heterogeneity in clonal microbial populations”. In: *Curr. Opin. Biotechnol.* 62 (2020), 202–211. DOI: 10.1016/j.copbio.2019.11.018.

[68] D. Binder et al. “Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity”. In: *Metab. Eng.* (2017), pp. 145–156. DOI: 10.1016/j.ymben.2017.06.009.

[69] F. Delvigne et al. “Taking control over microbial populations: current approaches for exploiting biological noise in bioprocesses”. In: *Biotechnol. J.* (2017), p. 549. DOI: 10.1002/biot.201600549.

[70] J. A. Megerle et al. “Timing and dynamics of single cell gene expression in the arabinose utilization system”. In: *Biophys. J.* 95 (2008), 2103–2115. DOI: 10.1529/biophysj.107.127191.

[71] J. Stricker et al. “A fast, robust and tunable synthetic gene oscillator”. In: *Nature* 456 (2008), 516–519. DOI: 10.1038/nature07389.

[72] O. Mondragón-Palomino et al. “Entrainment of a population of synthetic genetic oscillators”. In: *Science* 333 (2011), 1315–1319. DOI: 10.1126/science.1205369.

[73] N. Nikolic et al. “Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations”. In: *PLoS Genet.* 13 (2017), e1007122. DOI: 10.1371/journal.pgen.1007122.

[74] P. Sagmeister et al. “Tunable recombinant protein expression with E. coli in a mixed-feed environment”. In: *Appl. Microbiol. Biotechnol.* 98 (2014), 2937–2945. DOI: 10.1007/s00253-013-5445-1.

[75] S. Mangan, A. Zaslaver, and U. Alon. “The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks”. In: *J. Mol. Biol.* 334 (2003), 197–204. DOI: 10.1016/j.jmb.2003.09.049.

[76] E. Yurkovsky and I. Nachman. “Event timing at the single-cell level”. In: *Brief Funct. Genomics* 12 (2013), pp. 90–98. DOI: 10.1093/bfgp/els057.

[77] K. R. Ghusinga, J. J. Dennehy, and A. Singh. “First-passage time approach to controlling noise in the timing of intracellular events”. In: *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017), 693–698. DOI: 10.1073/pnas.1609012114.

[78] M. Thattai and A. van Oudenaarden. “Intrinsic noise in gene regulatory networks”. In: *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001), pp. 8614–8619. DOI: 10.1073/pnas.151588698.

[79] P. S. Swain, M. B. Elowitz, and E. D. Siggia. “Intrinsic and extrinsic contributions to stochasticity in gene expression”. In: *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002), 12795–12800. DOI: 10.1073/pnas.162041399.

[80] T. M. Norman et al. “Stochastic switching of cell fate in microbes”. In: *Annu. Rev. Microbiol.* 69 (2015), 381–403. DOI: 10.1146/annurev-micro-091213-112852.

[81] C. Tan, F. Reza, and L. You. “Noise-limited frequency signal transmission in gene circuits”. In: *Biophys. J.* 93 (2007), pp. 3753–3761. DOI: 10.1529/biophysj.107.110403.

[82] E. A. Davidson, A. S. Basu, and T. S. Bayer. “Programming microbes using pulse width modulation of optical signals”. In: *J. Mol. Biol.* (2013), pp. 4161–4166. DOI: 10.1016/j.jmb.2013.07.036.

[83] J. E. Purvis and G. Lahav. “Encoding and decoding cellular information through signaling dynamics”. In: *Cell* 152 (2013), 945–956. DOI: 10.1016/j.cell.2013.02.005.

[84] R. Cheong et al. “Information transduction capacity of noisy biochemical signaling networks”. In: *Science* (2011), pp. 354–358. DOI: 10.1126/science.1204553.

[85] A. Levchenko and I. Nemenman. “Cellular noise and information transmission”. In: *Curr. Opin. Biotechnol.* 28 (2014), 156–164. DOI: 10.1016/j.copbio.2014.05.002.

[86] S. Mangan and U. Alon. “Structure and function of the feed-forward loop network motif”. In: *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003), 11980–11985. DOI: 10.1073/pnas.2133841100.

[87] P. Mehta et al. “Information processing and signal integration in bacterial quorum sensing”. In: *Mol. Syst. Biol.* 5 (2009), p. 325. DOI: 10.1038/msb.2009.79.

[88] A. S. Hansen and E. K. O’Shea. “Limits on information transduction through amplitude and frequency regulation of transcription factor activity”. In: *Elife* 4 (2015), e06559. DOI: 10.7554/eLife.06559.

[89] S. Sarkar, D. Tack, and D. Ross. “Sparse estimation of mutual information landscapes quantifies information transmission through cellular biochemical reaction networks”. In: *Commun. Biol.* 3 (2020), p. 203. DOI: 10.1038/s42003-020-0901-9.

[90] B. C. Goodwin. “Synchronization of Escherichia coli in a chemostat by periodic phosphate feeding”. In: *Eur. J. Biochem.* 10 (1969), 511–514. DOI: 10.1111/j.1432-1033.1969.tb00718.x.

[91] P. Ruoff et al. “The Goodwin model: simulating the effect of light pulses on the circadian sporulation rhythm of Neurospora crassa”. In: *J. Theor. Biol.* 209 (2001), 29–42. DOI: 10.1006/jtbi.2000.2239.

[92] D. Gonze and W. Abou-Jaoudé. “The goodwin model: behind the hill function”. In: *PLoS One* 8 (2013), e69573. DOI: 10.1371/journal.pone.0069573.

[93] D. Gonze and P. Ruoff. “The goodwin oscillator and its legacy”. In: *Acta Biotheor.* 69 (2021), 857–874. DOI: 10.1007/s10441-020-09379-8.

[94] A. Banderas et al. “Autonomous and assisted control for synthetic microbiology”. In: *Int. J. Mol. Sci.* (2020), p. 9223. DOI: 10.3390/ijms21239223.

[95] N. Wiener. *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press, 1961.

[96] C. Carrasco-López et al. “Optogenetics and biosensors set the stage for metabolic cybergenetics”. In: *Curr. Opin. Biotechnol.* (2020), pp. 296–309. DOI: 10.1016/j.copbio.2020.07.012.

[97] E. Pedone et al. “Cheetah: a computational toolkit for cybergenetic control”. In: *ACS Synth. Biol.* 10 (2021), 979–989. DOI: 10.1021/acssynbio.0c00463.

[98] J. Uhlendorf et al. “Long-term model predictive control of gene expression at the population and single-cell levels”. In: *Proc. Natl. Acad. Sci. U. S. A.* 109 (2012), pp. 14271–14276. DOI: 10.1073/pnas.1206810109.

[99] D. Benzinger and M. Khammash. “Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation”. In: *Nat. Commun.* (2018), p. 3521. DOI: 10.1038/s41467-018-05882-2.

[100] C. Dusny and A. Schmid. “Microfluidic single-cell analysis links boundary environments and individual microbial phenotypes”. In: *Environ. Microbiol.* (2014), pp. 1839–1856. DOI: 10.1111/1462-2920.12667.

[101] C. Dusny et al. “Technical bias of microcultivation environments on single-cell physiology”. In: *Lab Chip* (2015), pp. 1822–1834. DOI: 10.1039/C4LC01270D.

[102] C. Westerwalbesloh et al. “Coarse-graining bacteria colonies for modelling critical solute distributions in picolitre bioreactors for bacterial studies on single-cell level”. In: *Microb. Biotechnol.* 10 (2017), pp. 845–857. DOI: 10.1111/1751-7915.12708.

[103] E. M. Zhao et al. “Optogenetic amplification circuits for light-induced metabolic control”. In: *ACS Synth. Biol.* 10 (2021), pp. 1143–1154. DOI: 10.1021/acssynbio.0c00642.

[104] M. B. Elowitz and S. Leibler. “A synthetic oscillatory network of transcriptional regulators”. In: *Nature* (2000), pp. 335–338. DOI: 10.1038/35002125.

[105] G. Perrino et al. “Automatic synchronisation of the cell cycle in budding yeast through closed-loop feedback control”. In: *Nat. Commun.* 12 (2021), p. 2452. DOI: 10.1038/s41467-021-22689-w.

[106] G. Fiore et al. “In-silico analysis and implementation of a multicellular feedback control strategy in a synthetic bacterial consortium”. In: *ACS Synth. Biol.* (2017), pp. 507–517. DOI: 10.1021/acssynbio.6b00220.

[107] M. J. Liao et al. “Rock-paper-scissors: engineered population dynamics increase genetic stability”. In: *Science* 365 (2019), 1045–1049. DOI: 10.1126/science.aaw0542.

[108] C. Briat and M. Khammash. “Perfect adaptation and optimal equilibrium productivity in a simple microbial biofuel metabolic pathway using dynamic integral control”. In: *ACS Synth. Biol.* (2018), pp. 419–431. DOI: 10.1021/acssynbio.7b00188.

[109] M. O. Din et al. “Interfacing gene circuits with microelectronics through engineered population dynamics”. In: *Sci. Adv.* (2020), eaaz8344. DOI: 10.1126/sciadv.aaz8344.

2

Chapter 2: Fitness cost as an essential parameter that drives population dynamics

1. Initial case study: Characterizing *E. coli* arabinose induction dynamics

E. coli main and preferred source of carbon and energy is glucose. But when this premium carbon source is exhausted and that other carbon sources are perceived in the environment, the bacteria can switch to the additional consumption of some of those. This is typically done by expressing a cluster of genes coding for the transporter(s) and enzyme(s) responsible for its assimilation. To avoid unnecessary metabolic cost, these genes are inducible and often requires two signal. The first is cAMP. When glucose gets exhausted, cAMP accumulates in the intracellular environment and serves as a general signal to lift the catabolic repression associated to the availability of glucose [1, 2]. This mechanism ensures that the bacteria is not going to invest energy into the tools to metabolize an additional carbon source when the preferred one is widely available.

The second signal is the carbon source itself. For the specific case study of arabinose, the pentose presence lifts the inhibition exerted by the regulatory protein AraC on the operon promoter. When arabinose is not present, AraC loops the DNA by binding to two operators upstream of the promoter. This loop makes the binding of the RNA polymerase on the promoter impossible, thus impeaching the transcription (Figure 2.1). This inhibition can be lifted by the presence of cAMP and the binding of arabinose to AraC [3]. Both work together to change the conformation of the regulator protein, thus preventing DNA looping.

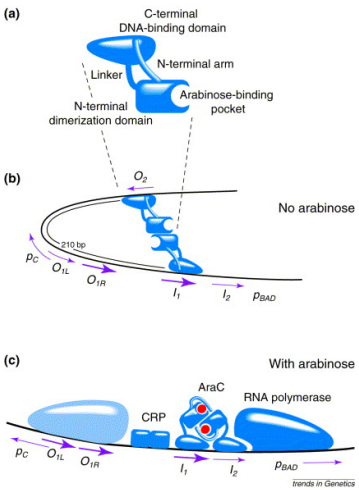

Figure 2.1: The regulatory protein AraC binds to two regions (operators), preventing the interaction between the RNA polymerase and the promoter. When bonded with arabinose, the conformation of AraC changes and the transcription inhibition is lifted. From Schleif et al., 2000 [3]

The arabinose promoter thus requires the shortage of glucose and the presence of arabinose to activate the expression of the arabinose operon comprising the transport protein and the three enzymes necessary for its metabolization. This promoter has been observed to be tight. This means that in the absence of arabinose, no basal expression is observed. Additionally, the arabinose promoter is strong and arabinose is a widely available sugar. All this combined justify its utilization to drive the expression of heterologous proteins both for biomolecular protocol (CRISPR) [4] and in bioprocesses [5] [6]. It was also previously observed in our lab that this system could be entrained using the Segregostat, and it was thus decided that understanding its induction dynamics would serve as a starting point in our quest to better understand gene expression dynamics across microbial populations [7].

1.1. Disassociating induction and relaxation phases to isolate population dynamic components

The level of gene expression within a microbial population depends on both induction and dilution processes. Cells gradually dilute certain properties as they grow and divide, distributing their intracellular contents between the daughter cells. This phenomenon resembles memory loss and occurs more rapidly with higher growth rates. In addition to division, other factors such as protein degradation contribute to the memory clear-out; however, since protein degradation is relatively slow and considering the high dilution rate used in our continuous cultivation, we assume that protein concentrations decline primarily due to growth. We monitor the induction levels of the arabinose promoter utilizing a plasmid-based biosensor developed by Silander and its colleagues [8]. In this medium copy plasmid, GFP mut2 expression is controlled by the native arabinose promoter in *E. coli*. This construction coupled with our online flow cytometry platform ensures that the activation of this gene circuit can be closely monitored at a single cell resolution (Figure 2.2). During all batch cultivation, only glucose is available. When switching to a chemostat, arabinose is co-fed with glucose, and the cells switch from an un-induced state (basal fluorescence level) to an induced one. At this point, the population induction stabilizes and adopts a log-normal distribution. In contrast to the chemostat, in a Segregostat, only glucose is continuously fed, and arabinose is conditionally provided as a pulse. In this setup, once 50% of the cells are below a given fluorescence value, arabinose is added as a pulse.

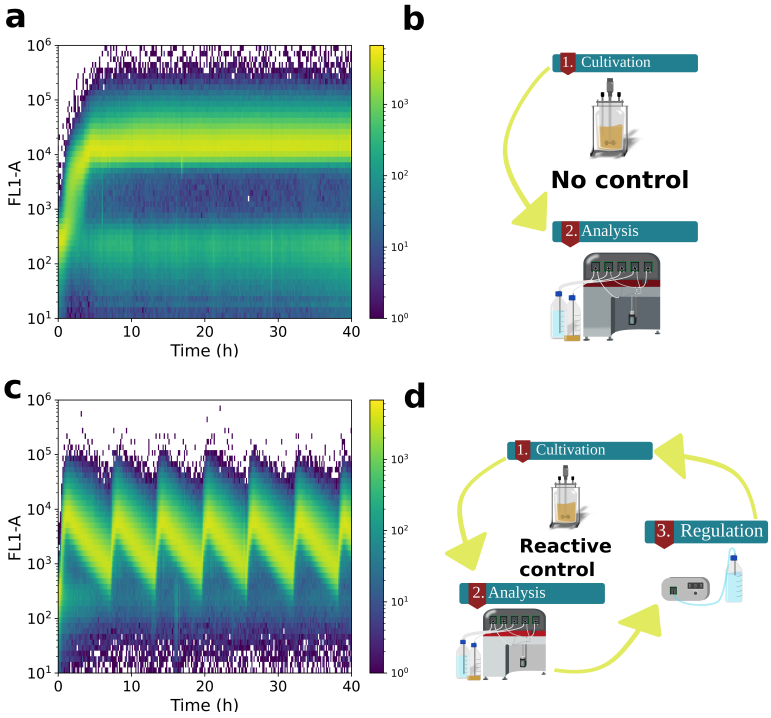

Figure 2.2: **a** Time scatter plot representing a chemostat cultivation of *E. coli* W3110 $P_{araB} : GFP$. **b** There, a sample is periodically drawn from a lab scale bioreactor and analyzed with flow cytometry. **c** When the regulation is activated (**d**), the population oscillates in response to the periodic addition of arabinose.

This kind of simple strategy is called a bang-bang controller. It allows for the perfect desynchronization of induction and relaxation phases. During the relaxation phase, we confirmed that de-induction resulted exclusively from dilution, as the decay followed the growth rate of our cells (Figure 2.3). It is important to stress that because flow cytometry only takes population snapshots, this analysis is impossible in a chemostat where cells constantly switch up and down.

Analyzing the induction phase presents challenges, as one would ideally like to fully characterize the induction profile for each inducer concentration. However, induction itself results in the metabolization of the inducer, potentially biasing any resulting measurements. A potential method for addressing this concern involves employing microfluidic cultivations, wherein fresh medium continuously replaces the old one, thereby eliminating any feedback effects. Despite its advantages, working with microfluidics can be labor-intensive and time-consuming. At the start of this study, these tasks proved challenging; however, significant advancements in image segmentation techniques facilitated by artificial intelligence have greatly improved the process.

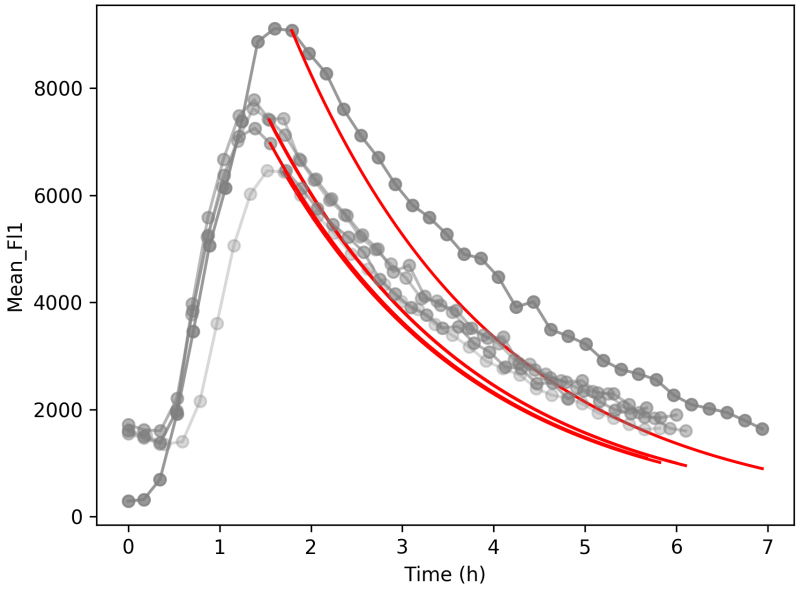

Figure 2.3: In gray, the mean FU recorded during a relaxation phase in a Segregostat cultivation (Figure 2.2). In red, decay of the fluorescence simulated based on an exponential decay with $\mu = D$.

We pursued an alternative strategy involving the construction of a strain incapable of metabolizing arabinose. Utilizing lambda red phage enhancement of homologous recombination combined with Cas9 negative selection, we generated the *E. coli* w3110 $\Delta araBAD$ strain. When equipped with the $P_{araBAD::gfp}$ plasmid from Silander's work, this modified strain signals the moment of induction without triggering the synthesis of enzymes responsible for consuming arabinose. Upon verification of this behavior, we proceeded to investigate the relationship between arabinose concentration and induction (Figure 2.4). Specifically, we introduced a predetermined amount of arabinose into a shake flask during the latter stages of the exponential phase, ensuring glucose limitation, and subsequently analyzed the ratio of induced cells via flow cytometry. This analysis revealed that the relationship linking inducer concentration with the proportion of induced cells resembles a Hill function characterized by the parameters K, the concentration at which half of the cells are induced, and n, the steepness of the relationship.

We now know that ratio of cells that activate the arabinose promoter depends on the pentose concentration and that when cells are not induced they dilute their phenotype *via* growth.

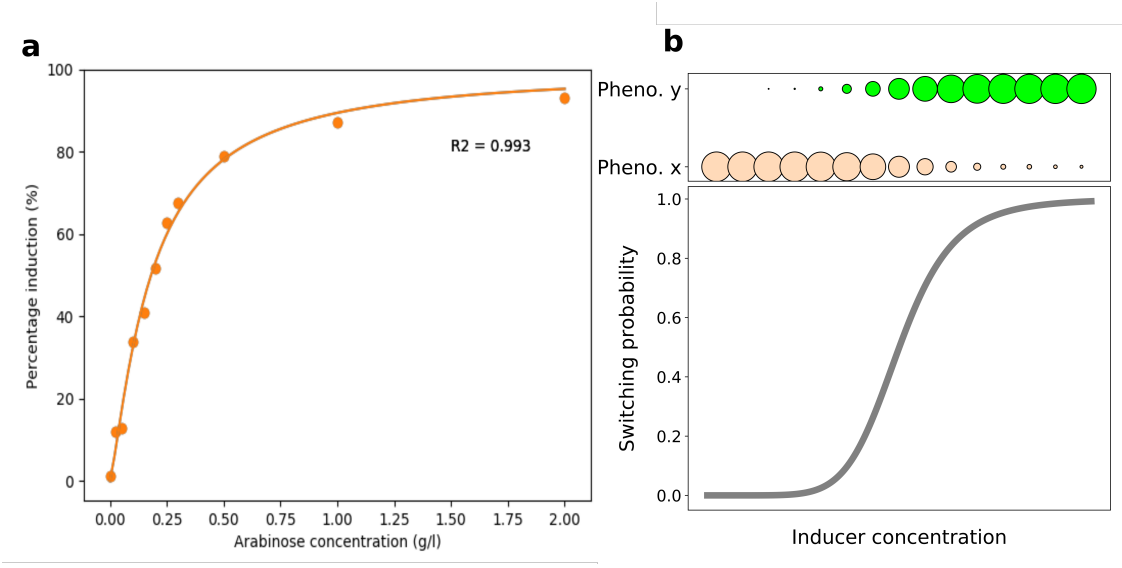

Figure 2.4: a The proportion of induced cell follows a Hill function dependent on the arabinose concentration. **b** Low concentrations yield a clear response not to get induced, high concentration represent a good opportunity and intermediates yield an unclear commitment.

1.2. A new mathematical framework to model population dynamics

This section is adapted from the description of the FlowStocK model from the following article:

> 2.Henrion, L. et al. Fitness cost associated with cell phenotypic switching drives population diversification dynamics and controllability. Nat Commun 14, 6128 (2023).

With the induction and relaxation dynamics at its core, we built a stochastic model named FlowStocKs to evaluate if both of these components are sufficient to reproduce the induction dynamics observed in chemostat and Segregostat cultivation of *E. coli* $P_{araBAD:gfp}$. FlowStocKS can be considered as:

- biologically segmented as it considers single cells
- abiotically unsegmented as it assumes a homogeneous environment
- an unstructured cell model as it does not take intracellular kinetics or metabolic fluxes into consideration

This model comprises two modules: the growth module and the switch module. The

system is resolved using a Markov chain with discrete time, where the growth module is described by a set of ordinary differential equations (ODE). These ODEs detail how single cells grow (2.1) and (2.2) and consume (2.3) their substrate (S), in accordance with the Monod-type equations.

$$\mu = \mu_{\max} \frac{S}{S + Ks} \tag{2.1}$$

$$\frac{dX}{dT} = (\mu - D)X \tag{2.2}$$

$$\frac{dS}{dT} = -\mu XY - DS + DS_{feed} \tag{2.3}$$

In the growth module, single cells are simulated to grow until they double in size, at which point they divide into two daughter cells. Additionally, to simulate continuous cultivations, cells are randomly flushed out of the system based on a probability (P_{out}) set by the dilution rate and the time step (T_{step}) used in the simulation (2.4).

$$P_{out} = DT_{step} \tag{2.4}$$

The growth parameters ($\mu_{\max}$, KI, Y, Ks) are given by the switch module and define the phenotype of each cell. To initiate the switching process, a cell must first cross a time threshold by accumulating $Tstep$, and this commitment process is governed by a switching probability (P). This probability is determined by a response function, taking the inducer concentration (i) as input (2.5). The P function is a classical sigmoid function characterized by a steepness (n) and a 50 $\%$ switching probability at concentration (K). This function, that we will hereby refer to as the response function, is the one presented in figure x.

$$P_i = \frac{[i]^n}{[i]^n + Ki} \tag{2.5}$$

The accumulation of (T_{step}) is analogous to the build-up of enzymes that are necessary for the expression of a different phenotype, commonly known in many processes e.g., substrate consumption switching and the diauxic shift time. Once the time threshold τ is reached, the cell switches phenotype. In our experimental set-up, we use GFP-based reporters to track the phenotype switch. Thus, once the time threshold is reached a GFP is produced as a burst following a production rate set by a gamma distribution. Similarly to Taniguchi et al., we assumed that active degradation of GFP is neglectable and accordingly dilution is determined by cell division.

Let us detail how FlowStocKs is used to model the phenotypic profile of the arabinose induction in *E. coli*. As previously mentioned, *E. coli* is able to consume arabinose, but it prefers to do so once glucose is exhausted. Fortunately, all continuous

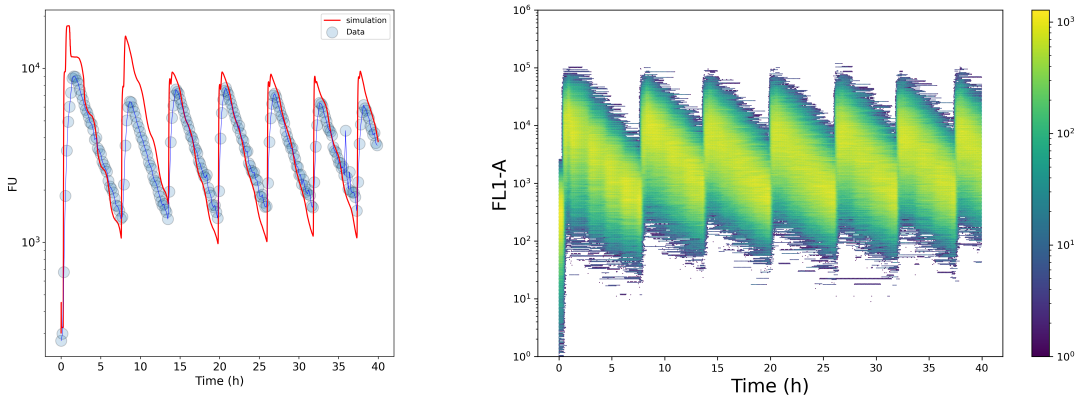

Figure 2.5: Comparison of experimental and simulated mean fluorescence intensity in Segregostat mode (left). The reconstituted time scatter plot (right) of the simulation exhibits a similar distribution to the experimental data, as shown in Figure 2.2.

cultivations performed in this work were done on the basis of a carbon limitation. Thus, in this context, the bacteria starts assimilating arabinose once it perceives it and has produced the necessary proteins for its intake and consumption. The transcription, translation, and maturation time are part of the delay, and in this scenario, we fixated it to 10 minutes, as it was observed that between the addition of arabinose in the reactor and the first increase in fluorescence, 10 minutes pass by.
In a chemostat, arabinose is fed continuously, and thus, given this delay and the inability of naive cells to consume the pentose, it first accumulates. This high concentration then acts as an initial trigger, and most cells get induced, in the process, producing GFP. Once they do so, the consumption rate of arabinose increases, and because its influx in chemostat is stable, its concentration in the reactor goes down. This lower concentration means that not all cells get perpetually induced and thus the GFP content across the population stabilizes. In Segregostat, arabinose is not continuously fed but pulsed whenever the ratio of induced cells is above 50 %. Because the arabinose promoter is tight, i.e., if no arabinose is present there is no expression, the pulsing of arabinose yields a clear induction that then stops once the pentose gets exhausted from the reactor. We have shown that the biosensor we used in our reactor does not yield any growth default once activated. Thus, the GFP content goes down with growth with similar rates regardless of the induction strength.

The FlowStocK model was used to replicate the dynamics observed in Segregostat, with parameters n and k derived from a pre-defined response function (Figure 2.4). The model's parameters, including $\mu_{\max}$ and Ks for arabinose and glucose, were experimentally determined. The GFP production rate was fitted. The model successfully reproduced the dynamics observed in Segregostat, and its single-cell resolution

enabled the reconstruction of a time scatter plot similar to that obtained from experimental data (Figure 2.5). In this simulation, the production of GFP does not impose a fitness cost on the cell growth, as *E. coli* grows similarly well on arabinose and glucose. However, the activation of a gene circuit can sometimes result in a significant growth reduction. To extend FlowStocK to these systems, we added a non-competitive growth inhibition term ((2.6)) and assumed that the inhibitor concentration equals the GFP content to capture the effect of gene expression strength on the population's phenotypic structure.

$$\mu = \mu_{max} \frac{S}{S + Ks} \frac{Ki}{[I] + Ki} \tag{2.6}$$

2. Fitness cost associated with cell phenotypic switching drives population diversification dynamics and controllability

So far, we have successfully modeled the gene expression dynamics across a population using the induction and relaxation dynamics. This has enabled us to gain a deeper understanding of the arabinose-inducible system in *E. coli*. However, to further challenge our model and extend our analysis, we have decided to broaden our scope to multiple systems in different organisms. This has enabled us to identify the fitness cost associated to the switching of one phenotype to another an important driver of population heterogeneity and controllability. At this point, it is important to empathize that we define the fitness cost as a reduction in growth rate. Fitness is a notion that is highly dependent of the context and approximating it to just growth rate makes no sense in a natural environment. But as all analysis below have been performed in a continuous cultivation device, and thus, the capacity of a phenotype to outgrow another is deemed closed to its fitness.
This section is an adapted version of the article:

2.Henrion, L. et al. Fitness cost associated with cell phenotypic switching drives population diversification dynamics and controllability. Nat Commun 14, 6128 (2023).

2.1. Abstract

Isogenic cell populations can cope with stress conditions by switching to alternative phenotypes. Even if it can lead to increased fitness in a natural context, this feature is typically unwanted for a range of applications (e.g.,bioproduction, synthetic biology, and biomedicine) where it tends to make cellular response unpredictable. However, little is known about the diversification profiles that can be adopted by a cell population. Here, we characterize the diversification dynamics for various systems (bacteria and yeast) and for different phenotypes (utilization of alternative carbon sources, general stress response and more complex development patterns). Our results suggest that the diversification dynamics and the fitness cost associated with cell switching are coupled. To quantify the contribution of the switching cost on population dynamics, we design a stochastic model that let us reproduce the dynamics observed experimentally and identify three diversification regimes, i.e., constrained (at low switching cost), dispersed (at medium and high switching cost), and bursty (for very high switching cost). Furthermore, we use a cell-machine interface called Segregostat to demonstrate that different levels of control can be applied to these diversification regimes, enabling applications involving more precise cellular responses.

2.2. Introduction

Cell populations can respond to environmental changes, and to the frequency of these changes, by adjusting their phenotypes through the activation of dedicated gene circuits [9] [10]. This phenotypic plasticity holds significant importance in microbial ecology, where the fitness of a cell population depends on a cost benefit ratio between the sensing machinery needed for the activation and deactivation of a given gene circuit [11] and its activity. Therefore, controlling the phenotype of cells has a lot of importance in various fields of research, such as bioproduction and synthetic biology, where coordinated gene expression is typically desired [7] [12]. Generating and controlling cell collective behavior is considered as a hallmark of synthetic biology [13],[14],[15], and is now enabled by the parallel advances made at the level of cell cultivation procedures (i.e., microfluidics [16] and cell-machine interfaces [17]), as well as the manipulation of synthetic gene circuits [18] [19]. Effective control of gene expressions and their underlying cellular functions can be achieved in cell populations [20],[21],[22] or individual cells within a population [17],[23]. Different approaches can be used to coordinate/synchronize gene expression in cell populations. On the one hand, specific gene circuits can be designed in order to generate natural oscillations [15],[24]. On the other hand, external forcing can be used for coordinating cellular responses [25],[26],[27]. According to this last approach, a given stimulus (e.g., chemical inducer [17] and light [21],[28]) is repeatedly applied at a given frequency and amplitude in order to entrain gene expression within a cell population. In this case, the effective transfer of information from the extracellular environment

to the effector sites within cellular systems is of critical importance and can be corrupted by biological noise [29] [30]. In silico experiments pointed out that specific environmental fluctuation frequencies could significantly reduce stochasticity in cell switching [19],[30]. This switching behavior, which we will now refer to as a diversification regime, significantly impacts the population structure and lies at the core of phenotypic control. However, the main factors affecting these diversification regimes are not known, when applied at a given frequency and amplitude in order to entrain gene expression within a cell population. For this purpose, chemostat runs are complemented by experiments conducted in Segregostat. Segregostat relies on a cell-machine interface to generate environmental perturbations that are compatible with the diversification rate of the considered cell population (Figure 2.6) [22]. This rational environmental forcing allows for the observation of several diversification cycles in one experimental run (Figure 2.6). We apply this technology to look at the dynamics of cell populations with cellular functions leading to different fitness costs, i.e., utilization of alternative carbon sources (*Escherichia coli*), general stress response (*E. coli* and *Saccharomyces cerevisiae*), sporulation (*Bacillus subtilis*) and activation of a T7-based expression system (*E. coli*). Based on the fitness cost associated with the cell switching mechanism (referred to as switching cost or fitness cost in this study), three different population diversification regimes, with different levels of sensitivity to environmental perturbations, are observed.

2.3. Characterization of population diversification dynamics based on automated flow cytometry

In the context of this work, the temporal diversification of cell populations has been followed based on chemostat cultivation of GFP reporter-bearing strains and automated flow cytometry (FC) (Figure 2.6 a). We define the phenotype as the cellular content in GFP and the characteristics associated with the activation of the observed gene circuit. By coupling FC to GFP, we are able to visualize the diverse range of phenotypes and study the dynamics of population diversification based on snapshot data. To better describe these dynamics, we have developed a methodology to compute the fluxes of cells from one phenotype to another and the resulting degree of heterogeneity of the population, i.e., in our case, based on the measurement of information entropy (Figure 2.6 b) and the flux of cells (Figure 2.6 c). Entropy is a measurement derived from information theory allowing to compute the degree of heterogeneity of a population [31]. Briefly, GFP distributions obtained from automated FC measurements are binned and the resulting phenotype distribution is used to compute the population entropy (Supplementary Note 1 and Figure 6.3). Based on the same binning strategy and by evaluating the enrichment or depletion of bins between consecutive measurements, we can also determine the flux of cells. Consequently, we can assess the dynamics of population diversification by monitoring the changes in population diversity (H (t)), as well as understanding how the phenotype shift evolves over time (F (t)). In addition

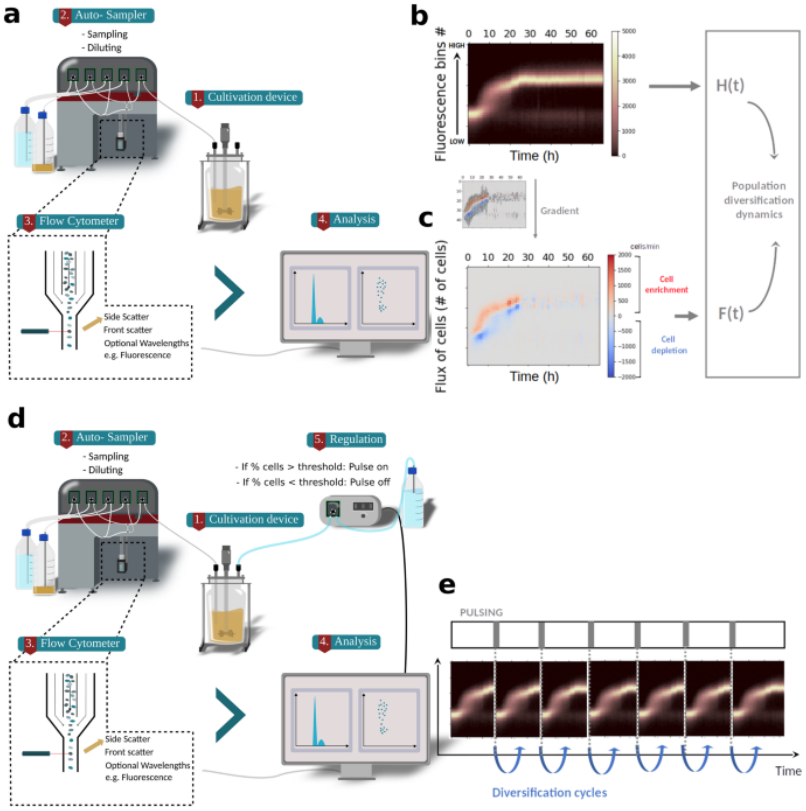

Figure 2.6: **a** Chemostat culture is monitored based on automated FC. **b** The fluorescence distribution acquired by FC is assembled into a time scatter plot. This time scatter plot is then further reordered into 50 fluorescence bins in order to compute the evolution of the entropy H of the population (Supplementary Note 1). **c** The binned data are further processed by applying a gradient to compute the fluxes of cells into the phenotypic space, leading to the quantification of the total fluxes of cells per time interval F(t). Both H(t) and F(t) will be exploited for characterizing the phenotypic diversification dynamics of diverse cell populations. **d** Scheme of Segregostat set-up. Pulses of nutrients are added in function of the ratio between GFP negative and GFP positive cells, as recorded by automated FC. **e** Expected evolution of a Segregostat experiment where, upon controlled environmental forcing, several diversification cycles can be generated.

to quantifying cell-to-cell heterogeneity within the population, H will also be used to calculate the information transmission to a cell population when subjected to environmental forcing. The benefit of this proxy is its independence from the mean of the distribution, by contrast with other noise proxies (e.g., Fano factor) that are known to be overestimated when the mean value increases [8],[32]. Using entropy to analyze chemostat experiments, however, provides limited information about diversification dynamics. Indeed, the main diversification process takes place during the transition between the batch and continuous phases of the culture. Therefore, we used a cell machine interface allowing us to produce several diversification cycles in a single experiment. This device is called Segregostat and comprises a continuous cultivation device connected to an in-house online FC platform [33] (Figure 2.6 d). This device enables the generation of several diversification cycles per experiment, leading to a better characterization of the population switching dynamics (Figure 2.6 e). Practically, the cells analyzed based on automated FC are clustered into a GFP negative and a GFP positive group. Depending on the gene circuits used, a pulse of inducer is applied when the minimum ratio between the two phenotype clusters (i.e., either 50 % or 20 % of the total amount of cells in the desired state, depending on the cellular system considered) is not reached. Based on this experimental and theoretical framework, we characterized the population diversification profile of six different gene circuits in three distinct cellular systems. Our approach involved linking a GFP reporter to each gene circuit, enabling us to leverage the analytical power of FC (20,000 cells per analysis) to study the population diversification over time.

2.4. Coordinated gene expression in a cell population can be obtained based on environmental fluctuations triggered by a cell-machine interface.

The methodology described in the previous section was applied to map the diversification of cell populations in chemostat and Segregostat cultivations. We began our analysis by considering two gene circuits involved in simple cellular processes in *E. coli*, i.e., the activation of the arabinose (Figure 2.7 a) and lactose (Figure 2.7 b) operons, respectively. Extended analysis of these systems is provided in Figure 6.1, and reproducibility of the data is displayed in Figure 6.2. This type of cellular process is quite simple since it involves two inputs, i.e., the absence of glucose or its limitation, and the presence of either lactose or arabinose as an alternative carbon source. Since these cultures were conducted in continuous mode, it was quite easy to ensure glucose limitation. Furthermore, the gene circuitry behind the activation/deactivation of these two operons is well documented in the literature, making them ideal case studies. The maximal growth rate of these alternative carbon sources is close to that of glucose. Further, despite the production of GFP, the induced phenotype still has a maximal growth rate above the dilution rate imposed in continuous cultivation. Thus, a feature of these two systems is that their activation does not result in a reduction in growth

rate in these conditions, which we will from now on refer to as a fitness reduction or switching cost. An extended analysis of the switching cost for the different systems can be found in Supplementary Note 2 and Supplementary Figs. 6.5 and 6.4. We thus decided to investigate other systems involved in more complex cellular processes known to lead to a higher switching cost. We first chose to consider the general stress response in *E. coli* and selected the promoter of the *bolA* gene as a representative σS-dependent system (Figure 2.7 c). To extend our analysis to another biological system, we also selected a gene circuit involved in the accumulation of glycogen in yeast, i.e., P_{glc3}, as a representative reporter of bet-hedging in *S. cerevisiae* (Figure 2.7 d). Both genes are involved in very complex regulons, making it difficult to find an external trigger. However, these general stress response reporter systems are known to share common features in the sense that their expression is anticorrelated with the growth of individual cells, making them very useful for analyzing cell collective behavior such as bet-hedging. We then decided to use the external glucose concentration as the main actuator for these two systems. Glucose-limited chemostats were then run as reference conditions.

For Segregostat experiments, glucose was pulsed instead of lactose or arabinose, allowing it to generate feast-to-famine environmental transitions. Segregostat cultivation of all four cellular systems investigated led to entrainment and sustained oscillation of gene expression (Figure 2.7 a-d). Based on the analysis of the entropy $H(t)$ and the flux of cells $F(t)$ over time, we observed that for all systems, entrainment phases were accompanied by an increased flux of cells switching to the alternative phenotype and a corresponding decrease of entropy $H(t)$ at the time of pulsing (Figure 2.7 where the analysis is shown for the $P_{araB} : GFP$ system, Supplementary Note 1, Supplementary Fig. 6.1). However, entropy increases during the relaxation phase (GFP dilution upon cell division). In addition to these four gene circuits, two more were analyzed: a T7-based expression system in *E. coli* and a circuit involved in sporulation in *B. subtilis*. These two systems exhibit a very high switching cost, leading to very specific population diversification profiles that will be described in the next section.

2.5. Phenotypic switching associated with extreme fitness cost gives rise to a bursty diversification regime

We investigated the diversification dynamics of two other systems known for their high impact on cellular fitness; the T7-based expression system in *E. coli* BL21 ($pET28 : GFP$)—a typical heterologous protein production platform—and the sporulation regulon in *B. subtilis* ($P_{spoIIE} : GFP$). The T7 expression system is inducible by lactose and $P_{spoIIE} : GFP$ is expressed when glucose is limiting, thus acting as an early trigger for sporulation. The $pET28 : GFP$ system is a T7-based expression vector inducible with lactose. When lactose is pulsed, cells are turning green, and GFP is diluted by growth when lactose pulsing is turned off (glucose is added continuously according to a classical continuous mode of cultivation at a dilution rate D=0.5 h^{-1}

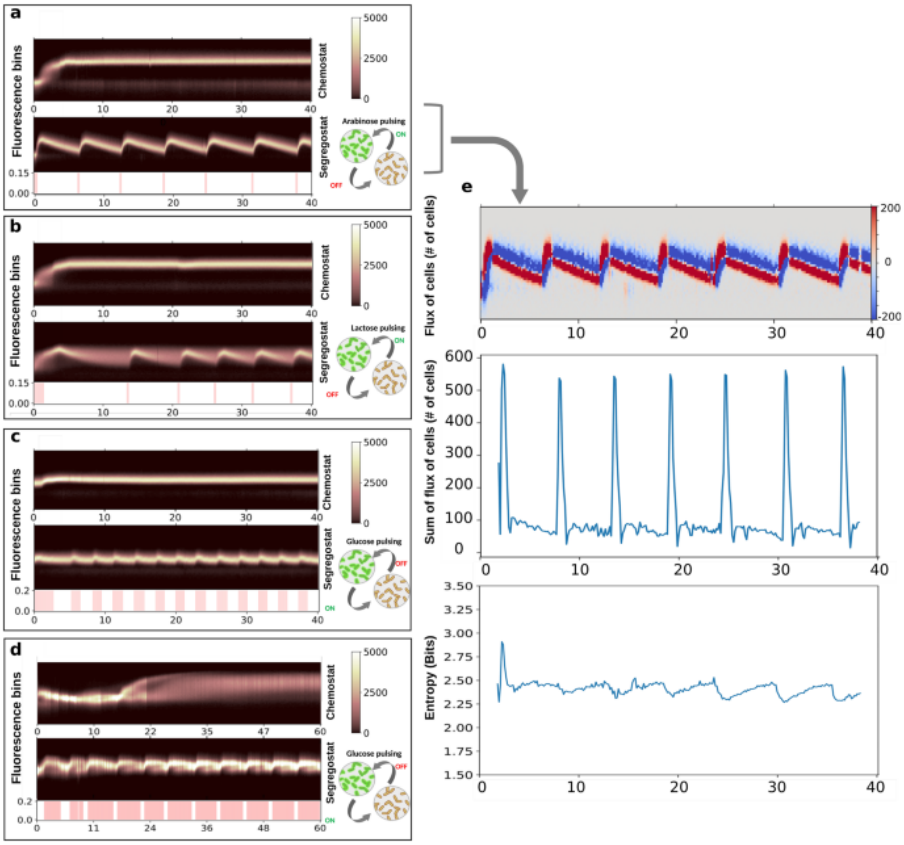

Figure 2.7: Time scatter plots (in hours) binned into 50 cell clusters (fluorescence bins) for cultivations made in chemostat and Segregostat for **a** the $P_{araB} : GFP$ system in *E. coli*. **b** the $P_{lacZ} : GFP$ system in *E. coli*. **c** the $P_{bolA} : GFP$ system in *E. coli*. **d** the $P_{glc3} : GFP$ system in *S. cerevisiae*. Animated movies for the time evolution of the FC raw data for each system are available. For Segregostat experiments, environmental forcing has been performed based on nutrient pulsing (the type of nutrient shown in the drawings for cell switching). **e** Computation of the flux of cells and the entropy (higher values mean more heterogeneous) for the $P_{araB} : GFP$ system cultivated in Segregostat mode (Supplementary Note 1, Supplementary Fig. 6.1). Reproducibility of the data has been assessed based on two biological replicates (n = 2) for each system. All FC measurements contain 20,000 analyzed cells (Supplementary Note 1 and Supplementary Fig. 6.2). The color bar alongside the scatterplot represents the number of cells. Source data are provided as a Source Data file.

D=0.5 h^{-1}. The $P_{spoIIE} : GFP$ system is induced upon glucose limitation. In our case, this system is cultivated at a very low dilution rate (0.1 h^{-1}) in order to generate glucose limitation and stimulate cell switching to sporulation. When too many cells are switching, additional glucose is pulsed in order to keep the population under control. For both, the phenotypic switch leads to a drastic growth reduction. Surprisingly, even in chemostat (with lactose present for the T7 system), FC profiles reveal bursts of diversification and a subsequent high entropy at the population level. These bursts are the result of a subpopulation of cells deciding to switch and being washed out from the continuous cultivation device due to the associated fitness cost. Then, upon environmental forcing based on Segregostat cultivation (lactose pulse, if more than 50% of cells are not induced with the T7 system and glucose is pulsed if more than 20% of cells express $P_{spoIIE} : GFP$), the number of bursts is reduced, and the fluxes of cells involved in the process are increased, leading to a substantial but temporary reduction of the entropy for the population. These results point out that Segregostat transiently reduces the average entropy of gene circuits with a high fitness cost despite very complex dynamics. It has been suggested in the literature that the stochasticity in cell switching is associated with its associated fitness cost and is important for the survival of the whole population. This feature is well illustrated in this case, where a phenotype switch induces a dramatic loss of growth rate, leading to the wash-out of these cells in continuous cultivation conditions.

However, this stochasticity can be reduced by applying environmental perturbations at a rate matching the phenotypic switching rate of cells. In the context of the T7 system, this approach led to the periodic maximization of cells in the GFP-positive state, suggesting that it could be used for mitigating metabolic burden and maximizing productivity in continuous bioprocesses. Indeed, the flux of cells to the high fluorescence bins is more regular for Segregostat conditions (Figure 2.8 a).

2.6. Coordinated gene expression does not necessarily lead to a more homogeneous cell population

Based on the analysis of the population diversification profiles exhibited by the six systems investigated, both similarities and differences can be observed. All six systems exhibit coordinated gene expression upon environmental forcing in the Segregostat device. This feature can be notably quantified based on F(t). On the other hand, the same systems display different H(t) profiles (Figure 2.9). We computed the average entropy H(t) for the six systems upon cultivation in chemostat and observed systems exhibiting a low basal entropy of around 2.2 bits (i.e., $P_{araB} : GFP$, $P_{lacZ} : GFP$ and $P_{bolA} : GFP$) and systems exhibiting a higher basal entropy of more than 3 bits (i.e., $P_{glc3} : GFP$, $P_{T7} : GFP$ and $P_{spoIIE} : GFP$) (Figure 2.9 a). For each system, the impact of Segregostat control system on the entropy was then investigated. We use entropy as an indicator of control because one would expect a controlled system to exhibit a more homogeneous phenotype and, therefore, a lower H(t). How-

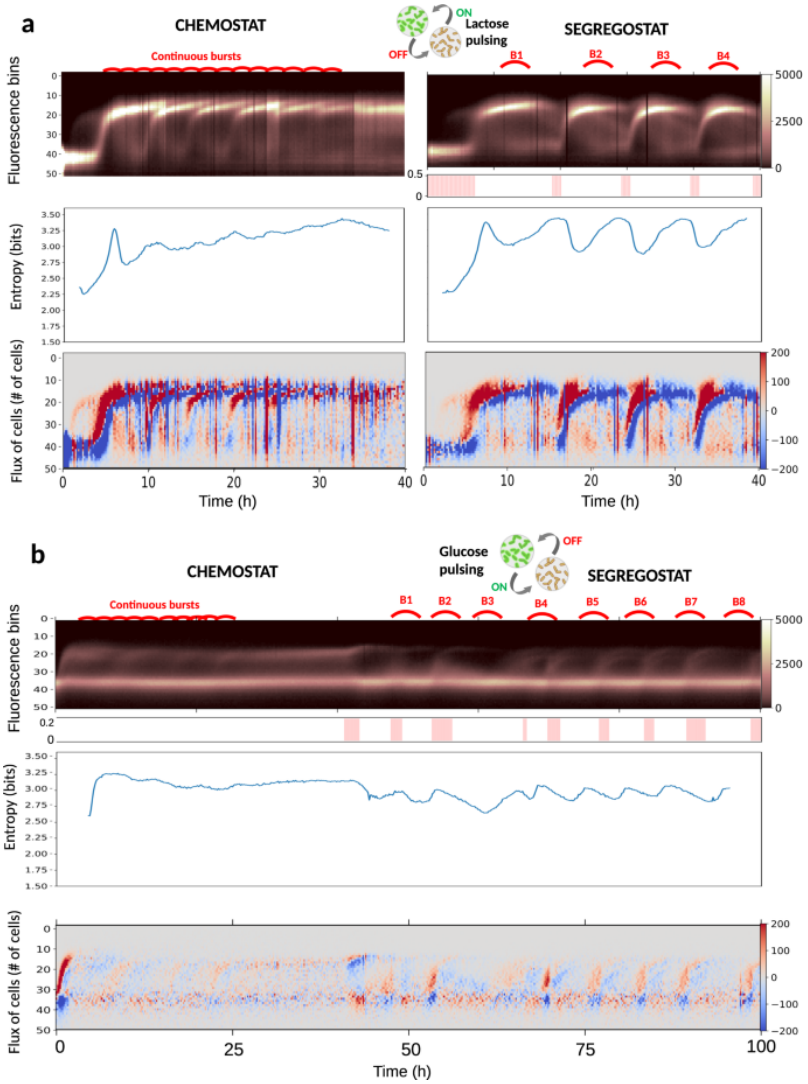

Figure 2.8: Temporal diversification profile for the $P_{T7} : GFP$ system in *E. coli* **a** and the $P_{spoIIE} : GFP$ system **b** cultivated in chemostat (from 0 to 40 h for $P_{spoIIE} : GFP$) and Segregostat modes (from 40 to 100 h for $P_{spoIIE} : GFP$). Below Segregostat cultivation modes, vertical lines indicate times of pulsing with the concentration (in g/L) displayed on the y-axis. Only the continuous phase of the cultivation is shown on the graphs. Bursts of diversification are highlighted in red on the fluorescence bins data. In both cases, the entropy H (t) and the fluxes of cells in the phenotypic space F(t) have been computed from the binned fluorescence data. The reproducibility of the data has been assessed based on two biological replicates (n = 2) for each system. For each time interval, 20,000 cells have been analyzed by FC (Supplementary Note 1 and Supplementary Fig. 6.2). The color bars represent the number of cells for the fluorescence time scatter plots and the number of cells per minute for the F(t) scatter plots. Source data are provided as a Source Data file.

ever, based on this definition, we observed that not all systems were controllable with Segregostat, as some of them showed no decrease in entropy despite their coordinated gene expression. In particular, (1) the systems exhibiting a low basal entropy ($P_{araB} : GFP$, $P_{lacZ} : GFP$, and $P_{bolA} : GFP$) showed no decrease in entropy. $P_{lacZ} : GFP$ even showed an increased entropy instead, probably due to the leakiness of the promoter during the relaxation phase of the diversification cycles (Figure 2.9 b, c). 2) The systems exhibiting a high basal entropy ($P_{glc3} : GFP$, $P_{T7} : GFP$P and $P_{spoIIE} : GFP$) showed a homogenization of the population. (Figure 2.9 b, c). We thus propose this criterion as a classification of the cellular systems, where trying to control a homogeneous cell population in chemostats produces no benefits and can even lead to an increased heterogeneity, while heterogeneous systems in chemostats can, by contrast, be made much more homogeneous (Figure 2.9 c). Based on this criterion, it can be observed that the $P_{glc3} : GFP$ system in *S. cerevisiae* exhibits a higher level of controllability than the other systems. Indeed, in this case, Segregostat cultivation led to a drastic decrease in H(t) by comparison with the reference condition in chemostat. This effect was then verified in microfluidics. Stress response pathways in yeast are known to be involved in bet-hedging strategies, leading to a trade-off between growth and expression of stress-related genes [34, 35]. The *glc3* gene belongs to this category. Accordingly, cells activating glc3 exhibit reduced growth. This phenomenon has been characterized based on microfluidics single-cell cultivation (MSCC) [36] experiments allowing to exposure of yeast cells to tightly defined glucose concentrations (Supplementary Fig. 6.4).

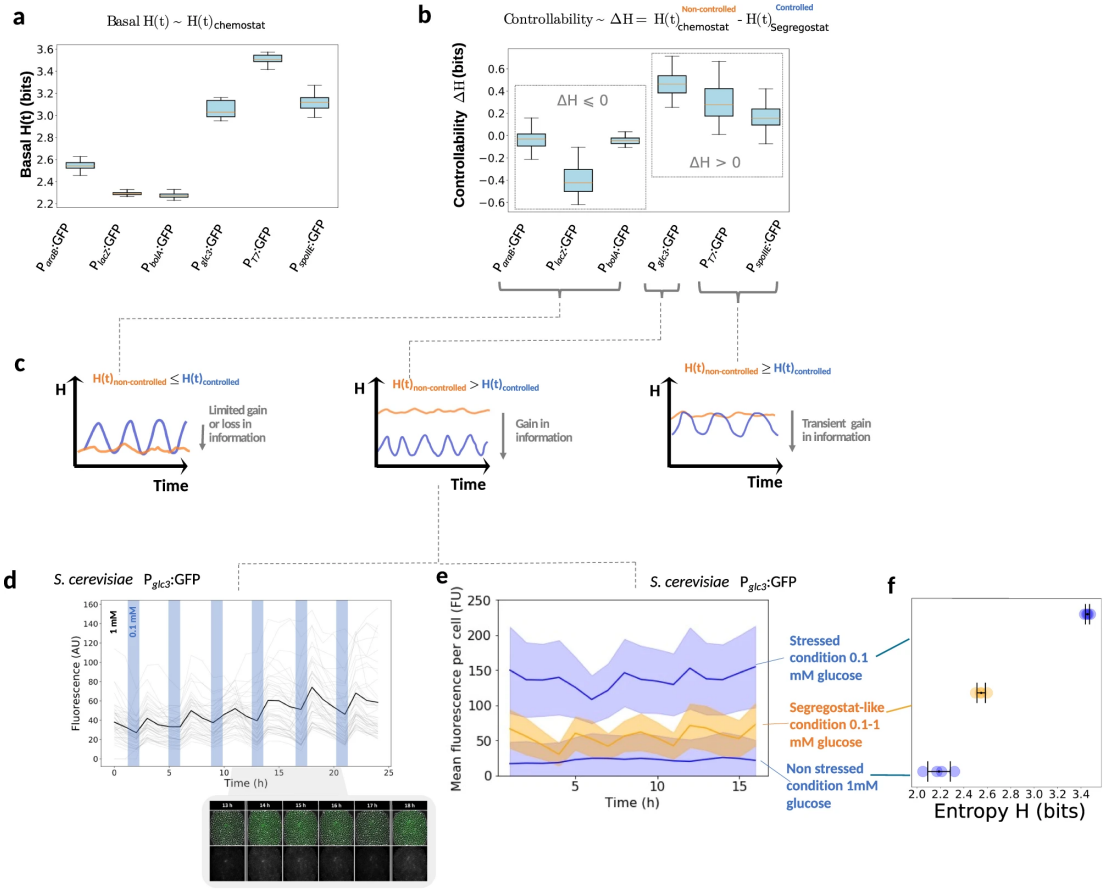

Figure 2.9: a Basal entropy (represented as a boxplot with whiskers where the interquartile range as extremities of the box, the median is the horizontal line in the box, and the whiskers as data extremes) recorded based on automated FC during chemostat experiments for the six biological systems investigated (values of entropy are computed from the population diversification profiles over the entire cultivation). Experiments have been done in duplicates (n biological replicate = 2) for each cellular system and cultivation conditions (chemostat or Segregostat) and exhibit a high reproducibility, see Supplementary Fig. 6.2. **b** Computation of the controllability (gain in entropy from chemostat to Segregated conditions represented as a boxplot with same structure than for plot A) for the six biological systems investigated. These average gains have been obtained by subtracting the mean value of H(t) recorded in Segregostat (considered as the controlled condition, i.e., cell population under environmental forcing) from the basal entropy. **c** Different diversification profiles, with different levels of controllability, can be observed based on the comparison of the H(t) profiles between non-controlled (chemostat) and controlled (Segregostat) conditions. **d** Single-cell traces of yeast $P_{glc3} : GFP$ cells cultivated in a dMSCC device fluctuating between 1 and 0.1 mM of glucose (T1mM = 3 h; T0.1mM = 0.8 h). Between 10 and 40 cells have been tracked in four different cultivation chambers over two biological replicates (mean fluorescence is shown in bold). Pictures of a micro-colony taken at regular time intervals are shown (Supplementary Movie 3).

e Comparison of the mean fluorescence profile obtained in dMSCC with the ones obtained in classical MSCC at high (1 mM) and low (0.1 mM) glucose concentrations. The shaded region around the lines represents the standard deviation computed from the measurement (between 10 and 40 cells have been tracked in four different cultivation chambers over 4 biological replicates (n = 4) for each condition). **f** Mean values of the entropy over the whole microfluidics experiments run at different glucose concentrations and standard deviation across chambers (n microfluidic chambers = 4 observed over one experiment). Each mean value over a cultivation run in a chamber is represented as a dot, and the mean of all replicates with the standard deviation across them as a black dot with error bars. Source data are provided as a Source Data file.

Unlike with FC analyses where only population snapshots are captured, these experiments let us monitor cell traces and thus analyze the fitness cost (growth reduction upon switching) associated to the switching. At low glucose concentration (<0.2 mM), a single cell fully activates the stress reporter and stops growing. At a higher glucose concentration, the growth of the microcolonies is faster, but some stochastic switching events can be clearly observed, with cells suddenly expressing the fluorescence reporter and stopping their growth. In order to confirm the beneficial impact of Segregostat condition on the $P_{glc3} : GFP$, we used dynamic microfluidic single-cell cultivation (dMSCC) [37] where we applied environmental fluctuations between 0.1 and 1 mM of glucose at the frequency recorded in Segregostat conditions. These fast and sharp transitions are an idealized scenario of Segregostat cultivation, but similarly, we observed a very homogenous gene expression pattern with cells turning green in perfect synchrony (Figure 2.9 d). The growth of all cells was comparable; the stress level was kept at a low-level thanks to the fluctuating environmental conditions. If we compare the experiments run in MSCC and dMSCC, it can be observed that cultivated cells under fluctuating glucose concentration lead to an intermediate scenario, both at the level of the mean fluorescence profile (Figure 2.9 e) and the global entropy (Figure 2.9 f). It seems that when phenotypic switching is associated with a loss of fitness, there is more stochasticity in the diversification pattern followed by the population. However, when the nutrient level is changed at a given frequency, switching and growth can be kept under control, leading to a drastic reduction of the phenotypic heterogeneity of the cell population. Another parameter that can explain the differences observed between the $P_{glc3} : GFP$ system and the $P_{T7} : GFP$ and $P_{spoIIE} : GFP$ systems (Figure 2.9 b, c) is the rate of switching.

It can be reasonably assumed that gene expression and cell growth are slower in Eukaryotes than in Procaryotes [38] , potentially explaining why it is easier to maintain the H(t) profile at a very low level in Segregostat conditions for the $P_{glc3} : GFP$ system since cells are switching more slowly in this case. The concept of controllability refers to the extent to which environmental forcing provided based on Segregostat cultivation reduces or not the global entropy of the population. So far, we

have used information entropy as a proxy for the quantification of the heterogeneity of cell populations. However, information entropy can also be used to evaluate the mutual information (MI), i.e., the reduction in entropy of the cell population when appropriate external stimulations are applied (Supplementary Note 3, Supplementary Figs. 6.6-6.9). We then computed MI for a system exhibiting low ($P_{araB} : GFP$) and high ($P_{glc3} : GFP$) controllability (Supplementary Note 3, Supplementary Fig. 6.10). Based on this analysis, we concluded that the MI is already maximal in chemostat conditions for the $P_{araB} : GFP$P system, explaining the low controllability measured in this case.

2.7. Fitness cost drives the appearance of different dynamical regimes with different levels of controllability

According to the datasets acquired from the different biological systems, we observed three types of diversification regimes named respectively constrained, dispersed and bursty (Figure 2.10). We proceeded to an in-depth analysis of the potential factors influencing the three observed modes of diversification. The key distinguishing factor between the systems displaying low and high basal entropy and controllability lies in the fitness cost linked to phenotypic switching. When the fitness cost associated with the switching is low or non-existent, we observe a constrained diversification regime where the population switches upon environmental change and adopts a homogeneous distribution. In that case, the stimulation of the population with controlled environmental pulsing does not homogenize it (since information transmission is already maximal in the non-controlled conditions). When there is a fitness cost associated with the switch, what we call a dispersed diversification regime can be observed. In that case, cells react to environmental changes but then adopt a more heterogeneous population structure.

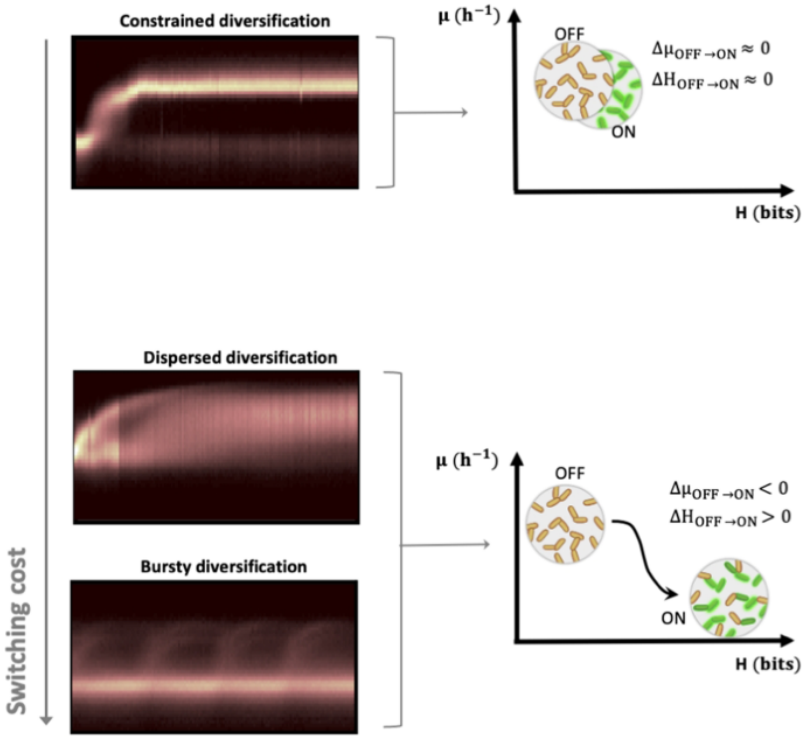

Figure 2.10: The constrained diversification regime is observed at low switching costs. According to the regime, all cells switch from the OFF state to the ON state according to a relatively homogeneous diversification process, and the population exhibits a low H(t) there is no increase in entropy upon activation of cells. The dispersed and bursty regimes are observed at high switching costs. Accordingly, cells switch from the OFF to the ON state in a quite heterogeneous way. The resulting stochastic switching is hypothesized to play a role in the stabilization of the cell population under continuous cultivation as a way to mitigate the fitness cost associated with switching. Upon diversification, the population exhibits a higher entropy.

In this case, the application of controlled environmental perturbations allowed a substantial reduction in population heterogeneity. For the bursty diversification regime (higher fitness cost), cells switch in bursts, leading to a very heterogeneous population structure. The application of controlled environmental perturbations reduces the number of bursts and increases the number of cells involved in these bursts, leading to a transient decrease in population heterogeneity. It seems that the regime depends on the fitness cost associated with the phenotypic switching event. In order to verify that fitness cost is indeed the driver for the diversification pattern adopted by cell populations, we conducted in silico experiments. For this purpose, we considered the kinetic parameters obtained from the inference of the $P_{araB} : GFP$ system in yeast and conducted stochastic simulations based on FlowStocKS (Supplementary Note 4).

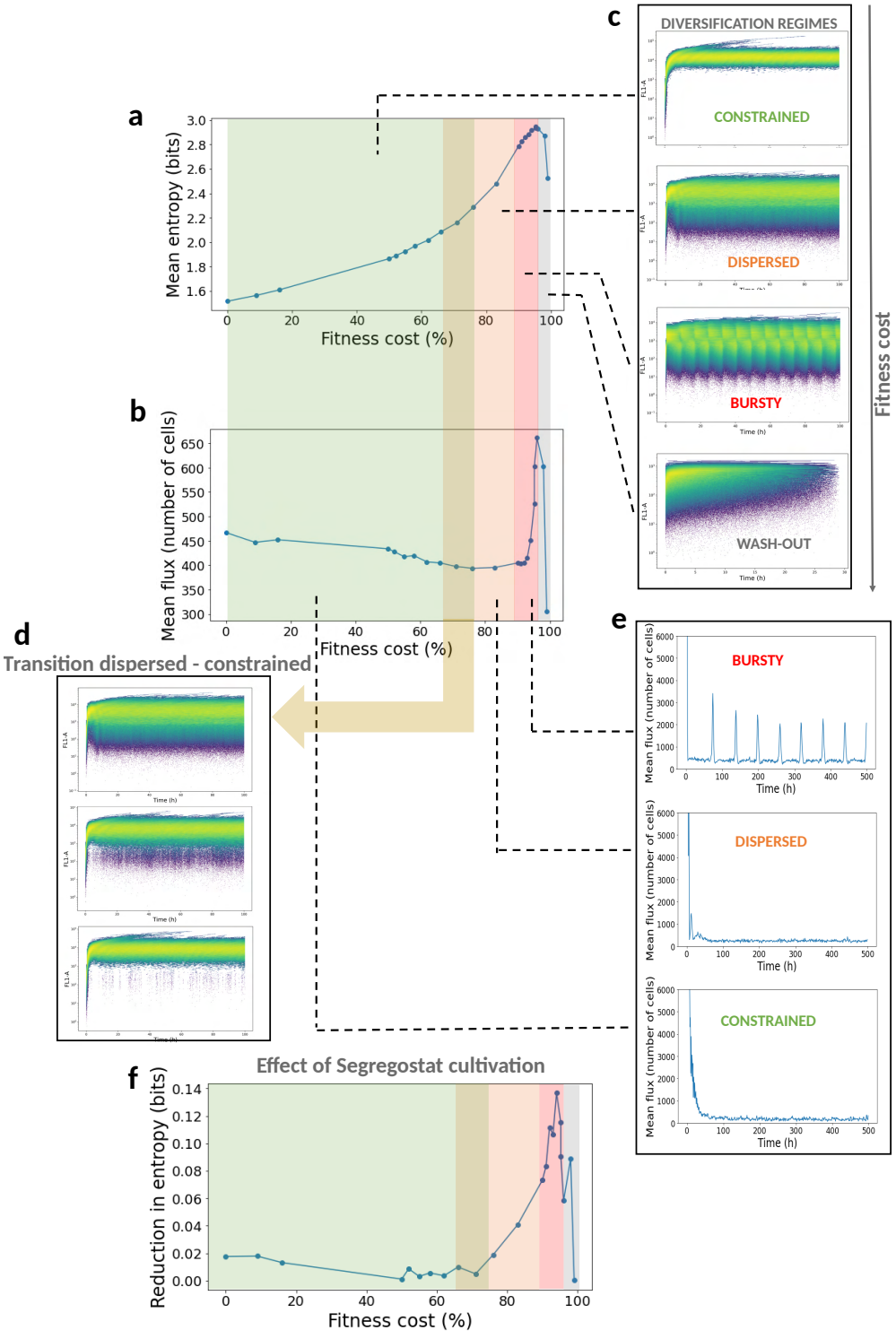

Figure 2.11: a Evolution of the mean entropy recorded from the H(t) Figure 2.9

profile in a chemostat in function of the fitness cost (here expression as the percentage of reduction of the initial growth rate prior phenotypic switching). These values are equivalent to the basal entropy shown in **b** Evolution of the mean flux of cells recorded during chemostat experiments in the function of the fitness cost associated with phenotypic switching (see Figure 2.6 for more details about the computation of the flux of cells). **c** Selected simulated time scatter fluorescence plots illustrating the different diversification regimes observed at different fitness costs (the whole simulation dataset can be found in Supplementary Movie 4). **d** Selected simulated time scatter fluorescence plots illustrating the progressive transition between the dispersed and constrained diversification regimes. **e** Selected simulated time evolution for the fluxes of cells recorded for different values of fitness cost. The bursty regime is characterized by the spontaneous generation of flux of cells (bursts) in chemostat cultivations. **f** Reduction of the entropy upon environmental forcing in the function of the fitness cost associated with phenotypic switching. The entropy values have been computed by subtracting the mean entropy value recorded for the chemostat experiment from the corresponding ones obtained in Segregostat and are equivalent to the controllability shown in Figure 2.9

We conducted 32 different chemostat simulations by varying only the value for the fitness cost and computed the entropy H (Figure 2.11 a) and the fluxes of cells (F) involved in phenotypic switching (Figure 2.11 b). Solely based on the fitness cost associated with the switching, we were able to reproduce the three types of diversification regimes experimentally observed during the experiments (Figure 2.11 c). Complete wash-out of the cells was observed for extreme fitness cost (>99% reduction in growth rate). We then wondered if we could observe clear transitions between the different regimes. Such transition was observed between the bursty and the dispersed regime based on the computation of the flux of cells F. Indeed, while the transition between the dispersed and constrained regime is progressive (Figure 2.11 d), the bursty regime is marked by the appearance of a strong variation in flux of cell which is not observed for the other two regimes (Figure 2.11 e). FlowStocKS was also able to reproduce the behavior of the population under Segregostat cultivation, and the reduction in entropy upon environmental forcing was computed (Figure 2.11 f). Again, reduction in entropy depended on the associated fitness cost and thus was observed for the dispersed and bursty regimes, in accordance with our experimental observations.

2.8. Discussion

We used Segregostat to better characterize cell population diversification dynamics by generating successive diversification cycles during the same experimental run. The basic principle behind this technology is to revert the environmental conditions when a fraction of cells (50 % or 10 % of the total population, depending on the investigated system) crosses a predefined fluorescence threshold. This approach allows to maintain a cell population in a dynamic switching state during the experiment. Based on the analysis of the MI, i.e., the amount of information transferred from the extracellular conditions to the cell systems [39],[31],[40], we determined that for the $P_{araB} : GFP$ system, the chemostat drives a similar amount of information to the Segregostat. On the other hand, we observed a drastic reduction in entropy when entraining stress-related systems, such as the $P_{glc3} : GFP$ system in yeast, in Segregostat. In this case, we determined that the high entropy observed in the chemostat was related to a trade-off between growth and gene expression [41] [35],[42] [43], which was further confirmed based on a microfluidics experiment. To better relate the switching cost to the resulting population structure, we then considered two additional systems where phenotypic switching induced a huge fitness cost i.e., the sporulation system in *B. subtilis* and the T7-based expression system in *E. coli*. Based on all the data accumulated by automated FC for six different biological systems, we found that cell populations diversified according to three distinct regimes associated with increasing fitness costs, i.e., constrained, dispersed, and bursty. The most noticeable difference between these regimes is the level of entropy of the cell population, the entropy being a measure derived from information theory giving a robust estimate of population dispersion [31],[44]. The lowest entropy values were associated with the constrained regime and the highest ones were seen for dispersed and bursty regimes. The other difference was observed in the cultivation of cell populations under fluctuating environmental conditions. In Segregostat, a reduction of entropy compared to chemostat cultivation was associated with the dispersed and bursty regimes but not the constrained one. Taken altogether, the data suggested that on top of affecting the population heterogeneity, the phenotypic switching mechanism changes the controllability of the system. All these observations were confirmed based on stochastic simulations (FlowStocKS), suggesting that the proposed diversification framework could be generalized for characterizing diversification dynamics for any kind of cellular system. Harnessing phenotypic heterogeneity of microbial populations has been the subject of much research, leading to the design of various technologies aiming at homogenizing gene expression in cell populations [45]. We have shown that the level of diversification of microbial populations cultivated in continuous bioreactors depends mainly on the fitness cost. Since many applications involve gene circuits whose activation leads to a substantial burden for the cellular system [46],[47], active diversification processes have to be expected in a number of cases [48]. For example, bursty diversification profiles have been ob-

served for two cellular systems exhibiting high switching costs. According to this regime, marked cycles of diversification can be observed even in chemostat cultures. These cycles are due to the rapid switching (burst) of a fraction of the population that lower the average growth rate of the population and are washed out of the system. These cells are then replaced by the next burst of diversification, starting from a sub-population of non-diversified cells. This type of temporal profile has been previously observed, but it is based on spatially organized cells equipped with synthetic circuits [49], [50], [51]. Here, we show that it is possible to reproduce such a complex but organized diversification profile with cells in suspension in a bioreactor and that the complex dynamics behind phenotypic heterogeneity are linked to the fitness cost associated to the switch.

Bibliography

[1] Ehab M. Ammar, Xiaoyi Wang, and Christopher V. Rao. "Regulation of metabolism in Escherichia coli during growth on mixtures of the non-glucose sugars: Arabinose, lactose, and xylose". In: *Scientific Reports* 8.1 (2018). Publisher: Springer US, pp. 1–11. DOI: 10.1038/s41598-017-18704-0. URL: http://dx.doi.org/10.1038/s41598-017-18704-0.

[2] Hongying Lin et al. "Change of extracellular cAMP concentration is a sensitive reporter for bacterial fitness in high-cell-density cultures of Escherichia coli". In: *Biotechnology and Bioengineering* 87.5 (2004), pp. 602–613. DOI: 10.1002/bit.20152.

[3] Robert Schleif. "Regulation of the l-arabinose operon of Escherichia coli". In: *Trends in Genetics* 16.12 (Dec. 2000), pp. 559–565. ISSN: 01689525. DOI: 10.1016/S0168-9525(00)02153-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168952500021533.

[4] Y. Jiang et al. "Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system". In: *Appl Environ. Microbiol.* 81 (2015), pp. 2506–2514.

[5] Philipp Moritz Fricke et al. "A tunable l-arabinose-inducible expression plasmid for the acetic acid bacterium Gluconobacter oxydans". In: *Applied Microbiology and Biotechnology* 104.21 (Nov. 2020), pp. 9267–9282. ISSN: 0175-7598, 1432-0614. DOI: 10.1007/s00253-020-10905-4. URL: https://link.springer.com/10.1007/s00253-020-10905-4 (visited on 10/03/2024).

[6] Suchada Chanprateep Napathorn et al. "Polyhydroxybutyrate Production Using an Arabinose-Inducible Expression System in Comparison With Cold Shock Inducible Expression System in Escherichia coli". In: *Frontiers in Bioengineering and Biotechnology* 9 (May 3, 2021), p. 661096. ISSN: 2296-4185. DOI: 10.3389/fbioe.2021.661096. URL: https://www.frontiersin.org/articles/10.3389/fbioe.2021.661096/full (visited on 10/03/2024).

[7] T. M. Nguyen et al. “Reducing phenotypic instabilities of a microbial population during continuous cultivation based on cell switching dynamics”. In: *Biotechnol. Bioeng.* 118 (2021), 3847–3859. DOI: 10.1002/bit.27860.

[8] O. K. et al. Silander. “A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli”. In: *PLoS Genet.* 8 (2012), e1002443.

[9] M. Thattai and A. van Oudenaarden. “Stochastic gene expression in fluctuating environments”. In: *Genetics* 167 (2004), pp. 523–530. DOI: 10.1534/genetics.167.1.523.

[10] J. Nguyen, J. Lara-Gutiérrez, and R. Stocker. “Environmental fluctuations and their effects on microbial communities, populations and individuals”. In: *FEMS Microbiol. Rev.* 45 (2021), fuaa068.

[11] U. Alon. *An Introduction to Systems Biology*. CRC Press, 2020.

[12] A. et al. Wagemakers. “Entraining synthetic genetic oscillators”. In: *Chaos* 19 (2009), p. 033139.

[13] O. Mondragón-Palomino et al. “Entrainment of a population of synthetic genetic oscillators”. In: *Science* 333 (2011), pp. 1315–1319.

[14] R. Wang and L. Chen. “Synchronizing genetic oscillators by signaling molecules”. In: *J. Biol. Rhythms* 20 (2005), pp. 257–269.

[15] L. Potvin-Trottier et al. “Synchronous long-term oscillations in a synthetic gene circuit”. In: *Nature* 538 (2016), pp. 514–517.

[16] A. Grunberger, W. Wiechert, and D. Kohlheyer. “Single-cell microfluidics: opportunity for bioprocess development”. In: *Curr. Opin. Biotechnol.* 29 (2014), 15–23. DOI: 10.1016/j.copbio.2014.02.008.

[17] J.-B. et al. Lugagne. “Balancing a genetic toggle switch by real-time feedback control and periodic forcing”. In: *Nat. Commun.* 8 (2017), p. 1671.

[18] Joe H. Levine, Yihan Lin, and Michael B. Elowitz. “Functional Roles of Pulsing in Genetic Circuits”. In: *Science* 342.6163 (Dec. 6, 2013), pp. 1193–1200. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1239999. URL: https://www.science.org/doi/10.1126/science.1239999 (visited on 09/18/2024).

[19] M. Thattai and A. van Oudenaarden. “Stochastic gene expression in fluctuating environments”. In: *Genetics* 167 (2004), pp. 523–530.

[20] T. M. et al. Nguyen. “Reducing phenotypic instabilities of a microbial population during continuous cultivation based on cell switching dynamics”. In: *Biotechnol. Bioeng.* (2021). DOI: 10.1002/bit.27860.

[21] A. Milias-Argeitis et al. “Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth”. In: *Nat. Commun.* 7 (2016), p. 12546.

[22] H. et al. Sassi. “Segregostat: a novel concept to control phenotypic diversification dynamics on the example of Gram-negative bacteria”. In: *Microb. Biotechnol.* 12 (2019), pp. 1064–1075.

[23] M. Rullan et al. “An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation”. In: *Mol. Cell* 70 (2018), 745–756.e6.

[24] J. et al. Stricker. “A fast, robust and tunable synthetic gene oscillator”. In: *Nature* 456 (2008), pp. 516–519.

[25] S. et al. Pérez-García. “Synchronization of gene expression across eukaryotic communities through chemical rhythms”. In: *Nat. Commun.* 12 (2021), p. 4017.

[26] F. et al. Bertaux. “Enhancing bioreactor arrays for automated measurements and reactive control with ReacSight”. In: *Nat. Commun.* 13 (2022), p. 3363.

[27] D. Benzinger and M. Khammash. “Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation”. In: *Nat. Commun.* 9 (2018), p. 3521.

[28] J. et al. Melendez. “Real-time optogenetic control of intracellular protein concentration in microbial cell cultures”. In: *Integr. Biol.* 6 (2014), pp. 366–372.

[29] R. Cheong et al. “Information transduction capacity of noisy biochemical signaling networks”. In: *Science* 334 (2011), pp. 354–358.

[30] C. Tan, F. Reza, and L. You. “Noise-limited frequency signal transmission in gene circuits”. In: *Biophys. J.* 93 (2007), pp. 3753–3761.

[31] C. G. Bowsher and P. S. Swain. “Environmental sensing, information transfer, and cellular decision-making”. In: *Curr. Opin. Biotechnol.* 28 (2014), pp. 149–155.

[32] A. Sanchez and I. Golding. “Genetic determinants and cellular constraints in noisy gene expression”. In: *Science* 342 (2013), pp. 1188–1193.

[33] H. Sassi et al. “Segregostat: a novel concept to control phenotypic diversification dynamics on the example of gram-negative bacteria”. In: *Microb. Biotechnol.* 12 (2019), 1064–1075. DOI: 10.1111/1751-7915.13442.

[34] S. F. Levy, N. Ziv, and M. L. Siegal. “Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant”. In: *PLoS Biol.* 10 (2012), e1001325.

[35] N. Ziv et al. “Resolving the complex genetic basis of phenotypic variation and variability of cellular growth”. In: *Genetics* 206 (2017), pp. 1645–1657.

[36] A. et al. Grunberger. “A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level”. In: *Lab Chip* 12 (2012), pp. 2060–2068.

[37] Sarah Täuber, Eric von Lieres, and Alexander Grünberger. "Dynamic Environmental Control in Microfluidic Single-Cell Cultivations: From Concepts to Applications". In: *Small* 16.16 (Apr. 2020), p. 1906670. ISSN: 1613-6810, 1613-6829. DOI: 10.1002/smll.201906670. URL: https://onlinelibrary.wiley.com/doi/10.1002/smll.201906670.

[38] R. Milo and R. Phillips. *Cell Biology by the Numbers*. Garland Science, 2016.

[39] F. Tostevin and P. R. ten Wolde. "Mutual information between input and output trajectories of biochemical networks". In: *Phys. Rev. Lett.* 102 (2009), p. 218101.

[40] L. Henrion et al. "Exploiting information and control theory for directing gene expression in cell populations". In: *Front. Microbiol.* 13 (2022), p. 869509.

[41] N. Ziv, M. L. Siegal, and D. Gresham. "Genetic and nongenetic determinants of cell growth variation assessed by high-throughput microscopy". In: *Mol. Biol. Evol.* 30 (2013), pp. 2568–2578.

[42] M. et al. Basan. "A universal trade-off between growth and lag in fluctuating environments". In: *Nature* 584 (2020), pp. 470–474.

[43] T. J. Perkins and P. S. Swain. "Strategies for cellular decision-making". In: *Mol. Syst. Biol.* 5 (2009), p. 326.

[44] A. Levchenko and I. Nemenman. "Cellular noise and information transmission". In: *Curr. Opin. Biotechnol.* 28 (2014), pp. 156–164.

[45] D. et al. Binder. "Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity". In: *Metab. Eng.* 42 (2017), pp. 145–156.

[46] F. Ceroni, B. A. Blount, and T. Ellis. "Sensing the right time to be productive". In: *Cell Syst.* 3 (2016), pp. 116–117.

[47] F. et al. Ceroni. "Burden-driven feedback control of gene expression". In: *Nat. Methods* 15 (2018), pp. 387–393.

[48] O. Borkowski et al. "Overloaded and stressed: whole-cell considerations for bacterial synthetic biology". In: *Curr. Opin. Microbiol.* 33 (2016), pp. 123–130.

[49] M. O. et al. Din. "Synchronized cycles of bacterial lysis for in vivo delivery". In: *Nature* 536 (2016), pp. 81–85.

[50] M. J. Liao et al. "Rock-paper-scissors: engineered population dynamics increase genetic stability". In: *Science* 365 (2019), pp. 1045–1049.

[51] P. Bittihn et al. "Genetically engineered control of phenotypic structure in microbial colonies". In: *Nat. Microbiol.* 5 (2020), pp. 697–705.

3

Chapter 3: Biological oscillations without genetic oscillator or external forcing

The switching cost associated with the activation of a given gene circuit appears to be a significant contributor to the population diversification dynamics observed during continuous cultivation. One notable example of such diversification dynamics is the bursty pattern, where cells diversify in waves and exhibit an oscillatory behavior. This finding was surprising, as traditional textbook knowledge suggests that the stable and limiting conditions imposed by a chemostat promote a stable population outcome known, as a steady state. In this study, we explore the reasons behind the departure from this stable state and provide explanations for the observed oscillatory behavior.

Pre-print Biological oscillations without genetic oscillator or external forcing Vincent Vandenbroucke, Lucas Henrion, Delvigne Frank bioRxiv 2024.09.20.614027; doi: https://doi.org/10.1101/2024.09.20.614027

1. Abstract

Oscillators play a crucial role in biological systems, enabling complex behaviors such as cell division and circadian cycles. As a result, researchers have been studying both natural and synthetic genetic oscillators. Interestingly, oscillating cellular behaviors are often linked to biological oscillators. In a previous study [1], we discovered sustained oscillations related to phenotypic switching in different cellular systems and gene circuits, without any external forcing. These oscillations were associated with the induction of slow-growing phenotypes. In this paper, we aim to understand when and why such phenotypic instabilities can arise, leading to oscillations in cell populations. To achieve this, we developed a simple mathematical model of a stress phenotype and solved it analytically to determine the range of operating conditions in continuous culture for which the gene expression should not reach a stable equilibrium. This range of conditions is expected to exhibit an unstable biological behavior, which we verified in *Bacillus subtilis* cultures. Our findings demonstrate that oscillations can occur in the absence of genetic oscillators or external forcing, departing from the classical Monod model. While we explain this phenomenon for the stress response in continuous culture, it is likely that similar unexpected behavior can occur in any long-term culture where there is environmental feedback connecting the inducer and the cellular system considered.

2. Introduction

Oscillating behaviors are ubiquitous through biology, appearing at multiple scales from single cells to multicellular organisms and controlling a sizable portion of their metabolism[2][3]. Their study yielded the discovery of natural gene circuits acting as

biological oscillators, e.g., the circadian clock of cyanobacteria[4] or the *E.coli* cell cycle[5]. These discoveries were complemented by the creation of synthetic oscillators[6], such as the repressillator, involving three repressor proteins to generate spontaneous oscillations in gene expression[7]. Oscillations in gene expression can also be generated based on external forcing. So far, biological oscillations can be produced based either on specific gene circuits[7], external forcing [8], or a combination of both [9].

Previous examples in the field have thus shown that, typically, when cells cultured in stable conditions exhibit oscillating gene expression, a biological oscillator is the cause[7]. However, we previously observed that highly burdensome gene circuits can oscillate when grown in a simple chemostat [1], a continuous cultivation device with stable operating conditions designed to promote stable conditions for the cells to grow in[10]. Notably, the two gene circuits that exhibited this behavior, the T7 expression system in *E.coli* BL21, and the sporulation of *B.subtilis*, do not have gene network architectures prone to generating oscillations. To understand the cause of these unstable behaviors, we focus our attention on the latter sytem because it can be reduced to 1) two defined phenotypes, i.e., growth and no growth, 2) one trigger, glucose, also acting as the carbon source.

To this end, we first developed a simple mathematical model of the stress response under a stable nutrient influx (a chemostat), where stress is induced when the nutrient concentration is too low. This model shows that even under stable operating conditions, there is a range of dilution rates for which no stable population profile can be found. To verify these predictions, we performed experiments on *B.subtilis* and observed that sporulation can exhibit oscillating profiles, observed as successive bursts of sporulation in the population.

3. A mathematical model predicts instabilities in gene expression for a specific range of dilution rate

As a first step toward the understanding of the observed oscillations, we designed an ODE model of stress response taking into account the dynamics of two phenotypes i.e., the non-stressed biomass X1 and the stressed biomass X2 (Eq. 6.5-6.11).

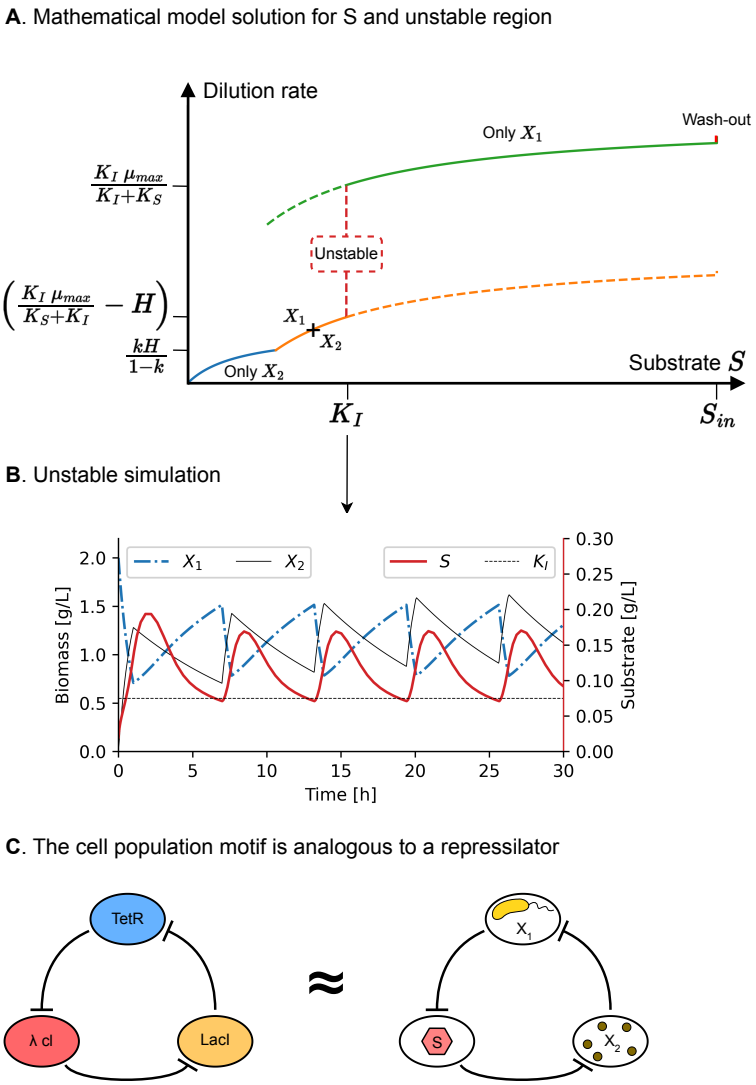

Figure 3.1: Representation of the solution of the mathematical model at steady state. **A**. There are two sides to the set of equations. At high dilution rates, only the non-stressed phenotype is present, as the sugar concentration is above the threshold K_I. At low dilution rates, the stressed phenotype is present as the sugar concentration is below K_I. In between those two cases, however, there is a region where no steady-state exists and oscillations are possible.**B** As long as there is a delay in stress response, oscillations appear in silico in the unstable region. Here, a system with similar values to *B.subtilis* sporulation is represented with $D = 0.1$ h^{-1} and a delay of 0.5 h, but very small changes in parameters can have a very large impact on the results. **C** The system, here represented schematically for bacillus sporulation (right), acts as a repressilator system (left)[7]. When sporulation or stress is induced, the amount of unstressed cells decreases. This causes the sugar concentration to increase and inhibit sporulation, letting the unstressed cells take over anew.

The stressed phenotype may not grow as fast as its counterpart, and its maximum growth rate was described as $\mu_{max,2} = k\mu_{max,1}$ where $k \in [0, 1]$. A step function $H_1(S)$ was used to account for the phenotypic switch from X_1 to X_2 when the cells are stressed (Eq. 6.11). Accordingly, the transition rate is considered a constant H below K_I and none existing above. Since the stressed phenotype X_2 grows slower and the proteins responsible for stress are mainly diluted by growth, the opposite transition from X_2 to X_1 was neglected.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = (S_{in} - S)D - \frac{\mu X_1}{Y_{X1}} - \frac{k\mu X_2}{Y_{X2}} \tag{3.1}$$

$$\frac{\mathrm{d}X_1}{\mathrm{d}t} = (\mu - D)X_1 - H_1(S)X_1 \tag{3.2}$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = (k\mu - D)X_2 + H_1(S)X_1 \tag{3.3}$$

$$\mu = \mu_{max}\frac{S}{K_S + S} \tag{3.4}$$

$$H_1(S) = \begin{cases} H & \text{if } S \leq K_I \\ 0 & \text{if } S > K_I \end{cases} \tag{3.5}$$

Where S_{in} is the sugar concentration in feed, D is the dilution rate, Y_{Xi} are the yields of each phenotype, and K_S is the sugar concentration at half maximum of the growth rate. Solving this model analytically (see SI Appendix) showed that there is a range of dilution rates where equilibrium between the different state variables cannot be reached and the system is unstable (Fig. 3.1 A). Within this region, the substrate concentration S cannot get to the substrate concentration predicted by the traditional Monod curve, as that is below the switching concentration K_I. It cannot reach the expected equilibrium where cells switch to the stressed phenotype either, as that would be above K_I. Instead, the cells must constantly respond to their environment and S should approach K_I. This constant response is a negative feedback loop for the growth of the population, and, combined with a delayed response to changing conditions, the outcome is an oscillator analogous to the repressilator[7] (Fig. 3.1 B and C). The oscillations, their appearance, and their persistence are expected to depend on the importance of the delay and the accuracy of the step function to approximate the switch to a stressed phenotype, but the range of dilution rate in which they appear can be verified experimentally.

4. Experimental confirmation of instability region under multiple dilution rates in chemostat

Thus, the model predictions were challenged under continuous cultivation conditions, a type of cultivation commonly used to promote stable conditions and set the residual substrate concentration S through the dilution rate. As a case study, *Bacillus subtilis* and one of its stress responses, sporulation, were selected. In this case, X_1 is the non-stressed biomass and X_2 is the stressed biomass, i.e., the spores. Sporulation is triggered when the residual glucose concentration drops below a concentration K_I and can be monitored with automated flow cytometry (FC) using a GFP transcriptional reporter, i.e., P_{spoIIE}*::GFP*. Given the values of H, K_I, μ_{max} and K_S determined in batch cultivation (SI Appendix) and the absence of growth rate for the stressed phenotype, the model predicts that when D is high and the residual glucose concentration is above K_I, only the vegetative cells X_1 are present. If D is low and the residual glucose concentration should be equal or below K_I, both X_1 and X_2 can be observed with the abundance of X_2, the spores, periodically increasing. As predicted (Fig. 3.2), at high dilution rates, where the residual glucose concentration exceeds the switching threshold, cells do not activate the sporulation cascade, and the entire biomass consists of X_1. Upon decrease of the dilution rate, the residual concentration falls below the switching threshold, and bursts of gene expression from X_1 to X_2 appear. This temporal instability in gene expression persists for at least five retention times and gradually decreases in amplitude.

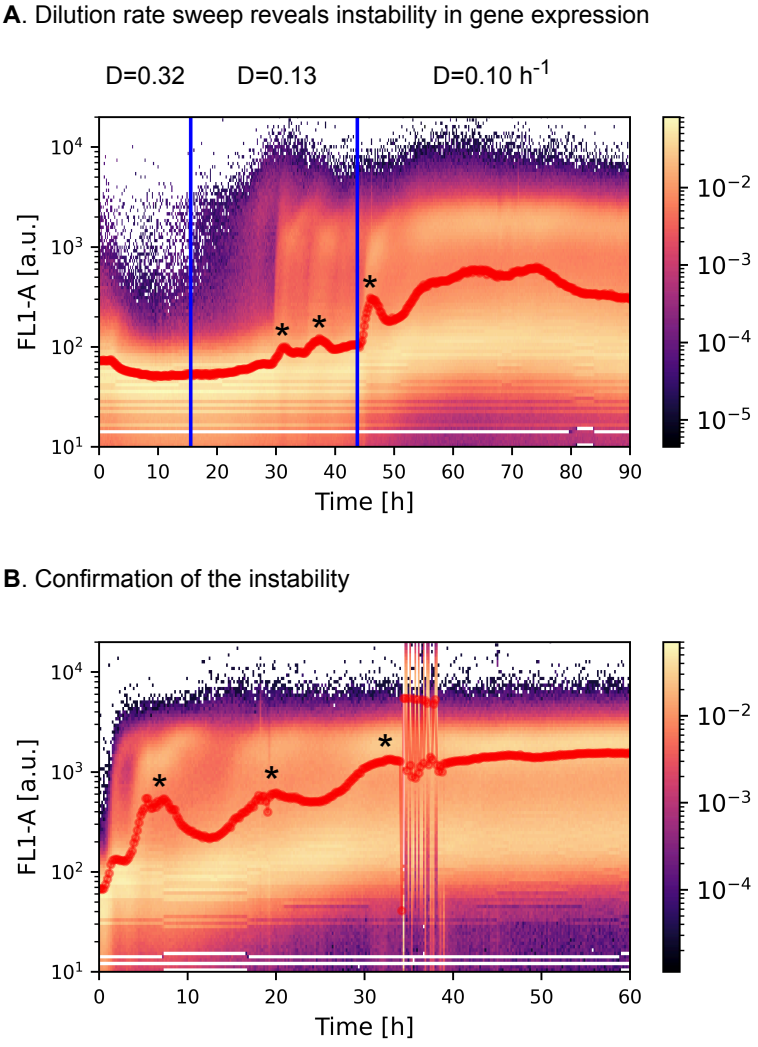

Figure 3.2: Mapping of cell population dynamics based on automated flow cytometry. The line indicates the 75th percentile of the fluorescence data. Large oscillations are highlighted with a '$*$'. **A**. According to the model, the gene expression dynamics depend on the dilution rate. At a high dilution rate ($D = 0.32$ h^{-1}) cells are not sporulating but once the dilution rate goes down ($D = 0.13$ h^{-1}) cells sporulate as burst that can be seen based on the activation of the P_{spoIIE}::*GFP*. The bursts gradually fade away when switching to a dilution rate of 0.1 h^{-1}. **B**. To confirm the observed instability in gene expression at a low dilution rate, we performed another chemostat ($D = 0.11$ h^{-1}) and observed similar outcome.

5. Conclusion and Perspectives

The results displayed in this work show how the output of a gene network without oscillating properties can nevertheless oscillate without the need for environmental forcing. This finding challenges the long-held assumption that microbial populations grown in chemostat exhibit a stable phenotype, instead suggesting that the population's characteristics can fluctuate over time. These oscillations are the result of a strong environmental feedback that could also be observed for many other phenotypic switches associated with an important reduction in growth. Eventually, oscillations fade away in the case of *Bacillus*. This deviation from the model prediction is speculated to originate from an over simplistic view of the switching that does not follow exactly a step function but more a hill one that allow for some heterogeneity in the response. Further, if a single sharp transition in phenotype causes unstable oscillations, there is no telling how multiple gene network interacting both with each other and the environment might create emergent behaviors on top of the ones predicted by current synthetic biology.

If cells can show such instabilities under stable operating conditions, then other culture conditions such as fed-batch (SI Appendix), turbidostat, or periodically refreshed cultures may be subject to the same or worse effects. In fact, similar oscillations involving stress related genetic switches have already been observed in a microfluidic where the feed was periodically interrupted [11].

Once those effects are brought to light, it may be necessary to take them into account when studying the biological system of interest. By doing so, researchers can gain a more comprehensive understanding of population dynamics and develop more effective strategies to control and manipulate cellular behavior.

6. Material and methods

Additional details of the materials and methods are available in the SI Appendix, and the relevant data is available in the Supplementary Dataset.

Three sets of experiments were performed for this paper: first, the growth parameters of a *Bacillus subtilis 168* with the P_{spoIIE}*::GFP* construction were determined with a triplicate batch experiment. During this experiment, on-line flow cytometry was employed to measure the sporulation status as indicated by the *spoIIE* reporter, and samples were regularly taken to measure the OD and glucose concentration over time. Glucose was measured with a glucose assay kit (D-glucose enzymatic assay kit; Megazyme, Bray, Ireland). These results were used to approximate reasonably well the parameters needed, although the fast variations in sugar towards the end of the culture made K_I and K_S very uncertain.

Second, two chemostat experiments with successive dilution rates going from 0.3 h^{-1} to 0.1 h^{-1} were performed to observe the limits of the instability region. Just like with the batches, online flow cytometry was used to measure the sporulation state of the

population. Further, samples from the 0.3 h^{-1} dilution rate were used to provide a better approximation of K_S.

Finally, an additional chemostat at 0.11 h^{-1} was performed to showcase the difference in equilibrium time between different starting points (here, a batch, in previous experiments, a continuous culture where some cells had already sporulated).

Bibliography

[1] Lucas Henrion et al. “Fitness Cost Associated with Cell Phenotypic Switching Drives Population Diversification Dynamics and Controllability”. In: *Nature Communications* 14.1 (Oct. 2023), p. 6128. ISSN: 2041-1723. DOI: 10.1038/s41467-023-41917-z. (Visited on 11/01/2023).

[2] Zhengda Li and Qiong Yang. “Systems and synthetic biology approaches in understanding biological oscillators”. In: *Quantitative Biology* 6.1 (Mar. 2018), pp. 1–14. ISSN: 2095-4689, 2095-4697. DOI: 10.1007/s40484-017-0120-7. URL: https://onlinelibrary.wiley.com/doi/10.1007/s40484-017-0120-7 (visited on 09/13/2024).

[3] Vakil Takhaveev et al. “Temporal segregation of biosynthetic processes is responsible for metabolic oscillations during the budding yeast cell cycle”. In: *Nature Metabolism* 5.2 (Feb. 27, 2023). Publisher: Springer Science and Business Media LLC, pp. 294–313. ISSN: 2522-5812. DOI: 10.1038/s42255-023-00741-x. URL: https://www.nature.com/articles/s42255-023-00741-x (visited on 09/17/2024).

[4] Susan E. Cohen and Susan S. Golden. “Circadian Rhythms in Cyanobacteria”. In: *Microbiology and Molecular Biology Reviews* 79.4 (Dec. 2015), pp. 373–385. ISSN: 1092-2172, 1098-5557. DOI: 10.1128/MMBR.00036-15. URL: https://journals.asm.org/doi/10.1128/MMBR.00036-15 (visited on 09/13/2024).

[5] Ilaria Iuliani et al. “Direct single-cell observation of a key *Escherichia coli* cell-cycle oscillator”. In: *Science Advances* 10.29 (July 19, 2024), eado5398. ISSN: 2375-2548. DOI: 10.1126/sciadv.ado5398. URL: https://www.science.org/doi/10.1126/sciadv.ado5398 (visited on 09/13/2024).

[6] Jesse Stricker et al. “A fast, robust and tunable synthetic gene oscillator”. In: *Nature* 456.7221 (Nov. 2008). Publisher: Springer Science and Business Media LLC, pp. 516–519. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature07389. URL: https://www.nature.com/articles/nature07389 (visited on 09/17/2024).

[7] Michael B. Elowitz and Stanislas Leibler. “A Synthetic Oscillatory Network of Transcriptional Regulators”. In: *Nature* 403.6767 (Jan. 2000), pp. 335–338. ISSN: 1476-4687. DOI: 10.1038/35002125. (Visited on 08/07/2021).

[8] Jean-Baptiste Lugagne et al. “Balancing a genetic toggle switch by real-time feedback control and periodic forcing”. In: *Nature Communications* 8.1 (Nov. 17, 2017), p. 1671. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01498-0. URL: https://www.nature.com/articles/s41467-017-01498-0 (visited on 09/13/2024).

[9] Octavio Mondragón-Palomino et al. “Entrainment of a Population of Synthetic Genetic Oscillators”. In: *Science* 333.6047 (Sept. 2, 2011), pp. 1315–1319. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1205369. URL: https://www.science.org/doi/10.1126/science.1205369 (visited on 09/13/2024).

[10] Aaron Novick and Leo Szilard. “Description of the Chemostat”. In: *Science* 112.2920 (Dec. 15, 1950), pp. 715–716. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.112.2920.715. URL: https://www.science.org/doi/10.1126/science.112.2920.715 (visited on 09/13/2024).

[11] Rosa Martinez-Corral et al. “Bistable emergence of oscillations in growing *Bacillus subtilis* biofilms”. In: *Proceedings of the National Academy of Sciences* 115.36 (Sept. 4, 2018). Publisher: Proceedings of the National Academy of Sciences. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1805004115. URL: https://pnas.org/doi/full/10.1073/pnas.1805004115 (visited on 09/17/2024).

4

Chapter 4: Lowering the switching cost related to the activation of burdensome gene circuits promotes cell population homogeneity and productivity

In chapter 2, the link between the switching cost (also sometime refer to as fitness cost) associated to the activation of a gene circuit and the heterogeneity in its activation was underlined. This is of particular applied interest because:

- As we define switching cost as the growth rate reduction associated with the activation of a gene circuit, it is analogous to production load. Consequently, a significant production load leads to substantial cell-to-cell differences in the expression of the gene of interest.
- This observation was made in the context of continuous cultivation, a process that, despite its numerous advantages, is not yet widely adopted due to the heterogeneity in expression it is associated with.

In the following work, we address the reasons why the expression of burdensome genes in continuous cultivation results in significant heterogeneity. Additionally, we explore control methods to promote a more homogeneous outcome, with the hope that this homogenization will actually lead to a more productive bioprocess.

Pre-print Lowering the switching cost related to the activation of burdensome gene circuits promotes cell population homogeneity and productivity
Lucas Henrion, Vincent Vandenbroucke, Juan Andres Martinez Alvarez, Julian Kopp, Samuel Telek, Andrew Zicler, Frank Delvigne bioRxiv 2024.10.14.618176; doi: https://doi.org/10.1101/2024.10.14.618176

1. Abstract

The activation of gene circuits can impose a significant burden on cells, leading to heterogeneous expression and reduced productivity. In this work, we focused on the T7 production system in *E. coli* BL21, a prime example of a burdensome gene circuit, to investigate the main cause for this gene expression heterogeneity and methods to mitigate it. Based on continuous cultivation analyzed and control by automated flow cytometry, we quantified the trade-off between cellular growth and gene expression and tracked the cell-to-cell heterogeneity in gene expression (measured as entropy). We concluded that the growth reduction associated to the activation of the burdensome gene circuit, i.e., the switching cost, is at the origin of the population heterogeneity. The loss of growth rate imposed by the burdensome activation of the gene is compensated at the population level by the overgrowth of less induced cells that safeguard the population by generating entropy. We tried to homogenize the population by pulsing the inducer with increasing frequency but found that the population escapes control through promoter mutation, leading to a genotype exhibiting reduced gene expression,

but also, reduced entropy. To engineer a more homogeneous population without sacrificing gene expression, we decreased the switching cost associated to the induction by lowering the quality of the main carbon source. This strategy successfully led to a more homogeneous and productive population. Our approach allows for a precise quantification of the trade-off between growth and gene expression in cell population cultivated under dynamic conditions and highlights the importance of the switching cost for designing efficient approaches of cell population control.

2. Introduction

Even in a uniform environment, genetically identical cells still display significant differences in gene expression [1, 2, 3]. This heterogeneity stems from two primary sources, i.e., intrinsic noise, which is an inherent characteristic of biochemical reactions, and extrinsic noise, which arises from variations in the content of macromolecules between cells [4]. This noisy expression can have functional consequences at the population level by generating phenotypically distinct sub-populations further improving population fitness by providing robustness against sudden environmental changes [5, 6, 7]. The beneficial impact of phenotypic heterogeneity on population fitness has been reported several times for microbial populations evolving in a natural context [8, 9, 10]. However, in bio-production, cell-to-cell heterogeneity is still perceived negatively [11, 12]. Indeed, within cell populations engineered for bio-production of recombinant molecules, important cell to cell variability in production and viability was observed [13, 14, 15, 16]. To better control the outcome of gene expression, synthetic biology approaches have been used to engineer more robust and predictable gene networks [17, 18] yielding more homogeneous induction. Despite all these efforts, the impact of cell diversification dynamics on bio-process performances is still an open question. In this context, our previous work suggests that the characteristics of the gene circuit alone are insufficient to explain the observed degree of heterogeneity in gene expression [19]. Instead, analysis of the heterogeneity in gene expression across diverse gene circuits in multiple model organisms revealed that heterogeneity increases with the switching cost associated to the activation of the gene [19]. In that work, which focused on the study of cellular heterogeneity in continuous cultivations, switching cost was defined as the reduction in growth rate associated with the activation of a gene circuit. This definition is analogous to production load or metabolic burden and highlights the importance of this previous observation in the context of bio-processes.

To control gene expression, it was proposed that the inducer could be pulsed instead of being added continuously in the cultivation device [20, 21, 22]. The efficiency of this approach was previously confirmed with a optogenetics system where the expression of a fluorescent protein was more homogeneous across the population when applying light as pulses instead of continuous supplied [23]. It was suggested that pul-

satile induction homogenizes expression because the transcription factors expression itself is bursty [24]. Previously, we also observed the benefit of pulsing the inducer instead of feeding it continuously, but only for gene circuits exhibiting a important burden [19]. This current work aims to shed light on why the expression of burdensome genes is more heterogeneous than non-burdensome genes and to propose new approaches for promoting the homogeneous activation of such genes. To this end, the activation of the T7 system in *E. coli* BL21 (DE3) was selected as a case study for our analysis since its activation results in a significant burden [19] and is characterized by an unstable induction in continuous cultivations [25].

3. The increase in cell population entropy reflects a trade-off between growth and gene expression at the single cell level

To understand the origin of the heterogeneous expression of burdensome genes during continuous cultivation, the T7 expression system in *E. coli* BL21 containing the $pET28 : GFP$ plasmid was periodically stimulated to characterize the induction and relaxation phases. The activation of the T7 system was followed by the expression of EGFP, the production both mimicking the expression of a heterologous protein in a bio-process while being easily quantified at a single cell resolution with a flow cytometer (FC). For this purpose, the strain was grown in a Segregostat, where glucose is supplied continuously while lactose (the inducer) is conditionally added in pulses (Fig. 4.1 **a**). These pulses are triggered when FC analysis indicates that less than 50% of the population exceeds a pre-set fluorescence threshold that is much greater than the basal auto-fluorescence of the strain (here 1,000 f.u.). The fluorescence data, collected in the FL1-A channel of the FC, can be combined to create a visual representation of expression levels over time. This is achieved by concatenating the data from multiple analysis cycles, resulting in a time density plot that provides a comprehensive overview of the expression patterns (Fig. 4.1 **b**). The fluorescence values in the time density plot can be binned, each bin containing a proportion of the population. This binning allows for the quantification of heterogeneity using a metric derived from information theory, Shannon entropy. Unlike the standard deviation or Fano factor, the entropy is not biased by the mean and does not assume a normal distribution. Measured in bits, entropy increases with the degree of cell-to-cell heterogeneity in gene expression. The second type of data that can be extracted from the time density plot is the growth rate of cells as a function of gene expression intensity. This can be achieved by analyzing the fluorescence decrease of each quartile during a relaxation phase (SI Note 1). This decrease in florescence is driven by growth and thus this approach quantifies the trade-off between gene expression (fluorescence level) and growth (Fig. 4.1 **c**). For the T7 system, gene activation can lead to a dramatic loss of growth rate, termed switching cost, of more than 80% of the initial growth rate.

With this quantitative metric of heterogeneity, entropy, we analyzed the dynamics of cell population in the Segregostat. As expected, the mean fluorescence rises during induction (following lactose addition) and falls during the relaxation phase, until the next pulse (Fig. 4.1 **d**). More intriguingly, the observation of entropy reveals that the population becomes more homogeneous during induction and spreads out during relaxation (Fig. 4.1 **d**). Further investigation demonstrated that the increased heterogeneity during relaxation was driven by the trade-off between growth and gene expression (Fig. 4.1**c**). Specifically, cells that are less induced grow faster than those that are more induced, their GFP content decreases faster, and the population gets more heterogeneous. To sum up; the activation of a burdensome gene threatens the population survival in the reactor since if all cells were to have the same degree of activation, the population would be washed out. But the cells with a lower expression of the gene compensate for this burden by outgrowing the others and rescuing growth. This phenomenon is termed Burden-Entropy (B-E) compensation (Fig. 4.1 **e**).

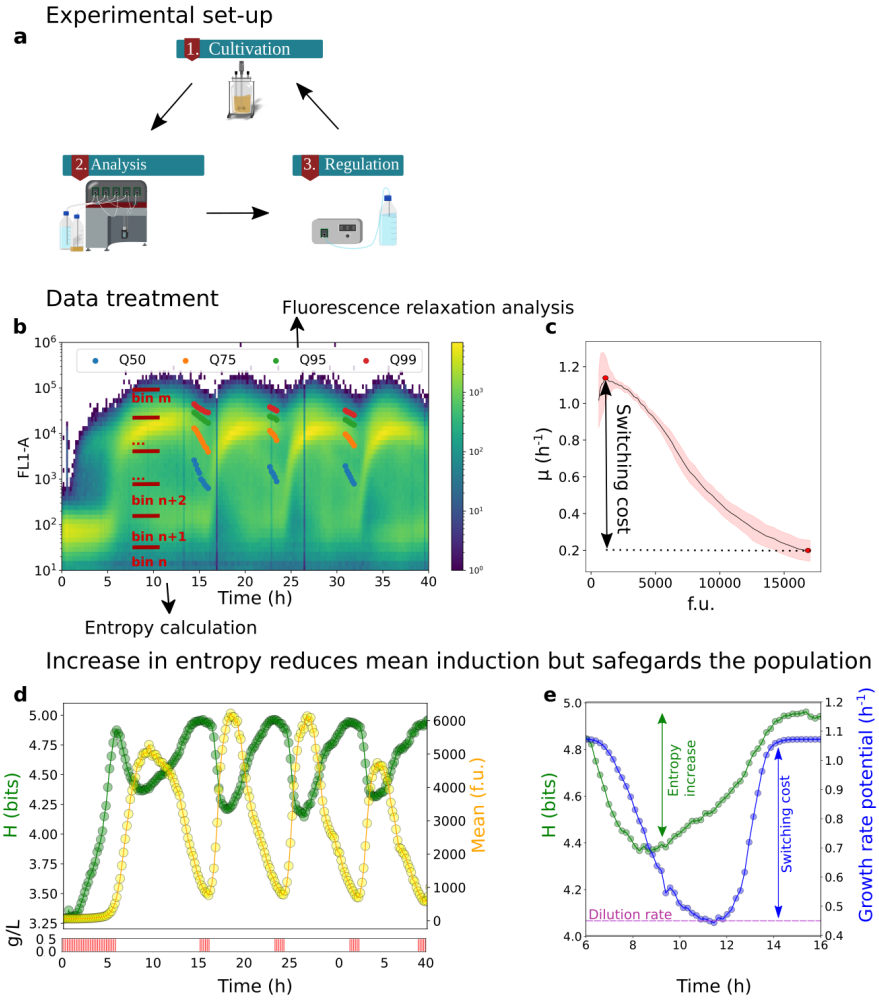

Figure 4.1: Heterogeneity in the expression of a burdensome gene circuit derives from a burden entropy compensation mechanism: **a** The Segregostat is a cell-machine interface allowing to draw automatically samples from a cultivation device (in our case, a continuous bioreactor) and trigger FC analyses. The collected data are then compared to a pre-set fluorescence threshold that triggers the regulation. **b** The collected data (20,000 cells per analysis) is plotted as a time density plot with the fluorescence on the Y axis (here FL1-A to measure GFP abundance) and the time in hours on the x-axis (two biological replicates SI Figure 6.13, n=2). The y-channel values are binned to compute entropy (H) as a quantitative proxy for heterogeneity. The fluorescence relaxation can be analyzed by measuring the decrease in quartile values.**c** Looking at three relaxation phases and assuming the GFP gets solely diluted by growth, the growth rate associated to the GFP content can be computed (slope of the decay the log10 values of fluorescence quartiles).

This reveals that a higher initial fluorescence corresponds to a much lower growth rate. **d** The mean fluorescence increases after each pulsing (0.5 g of lactose per pulse in a 1 l reactor with a OD of 5 as shown by red v-lines below the main graph). H decreases when pulsing and goes up during relaxation. **e** The activation of the burdensome gene reduces the maximum growth rate achievable by the population because of the switching cost. The maximum growth rate achievable by the population was computed from the median fluorescent content and corresponding growth rate gathered from the switching cost curve (c) it is associated with. The population growth rate is nevertheless limited by the dilution rate. Before the population growth rate drops below the dilution rate and the population gets washed-out, the cells with a lower induction overgrow the other highly induced ones and compensate for their low growth rate. We can observe here that growth is restored upon population diversification (measured based on entropy).

4. Reducing cell population entropy based on forced induction cycles gives rise to mutational escape

These initial data show that the trade-off between growth and gene expression promotes the heterogeneity of the population. But there also point out that applying a sharp induction phase by pulsing lactose can transiently reduce entropy. We therefore hypothesized that by progressively increasing the induction frequency, the population could be uniformly driven towards higher fluorescence (f.u.) values and reduced entropy (H). Ultimately, this strategy should result in a washout when the population becomes sufficiently homogeneous, such that the median fluorescence (fu) leads to a growth rate of the population lower than the dilution rate (D) imposed by the cultivation device. According to this hypothesis, we considered three cell population outcomes depending on the induction frequency (Figure 4.2 **a**).

We hypothesized that at low pulsing frequencies, the B-E compensation is the main mechanism influencing the population dynamics, rescuing the growth rate loss associated with the harsh induction. As a result, the growth rate of the population is determined by the dilution rate imposed on the population and the population remains highly heterogeneous. Then, as the periodic homogenization of the population increases with higher pulsing frequencies, B-E has less time to enable population diversification, and thus entropy decreases to the benefit of the mean induction. At a critical point, if the share of the population with high induction is too high. There is not enough heterogeneity to compensate the burden imposed on the population, and the system is washed out.

We then experimentally tested our hypotheses in a continuous culture where increasing frequencies of lactose pulses were applied (Figure 4.2 **b**). We applied three successive cycles of increasing pulsing frequency to the cell population. Unlike the predictions (Figure 4.2 **a**), the cell population never washed-out.

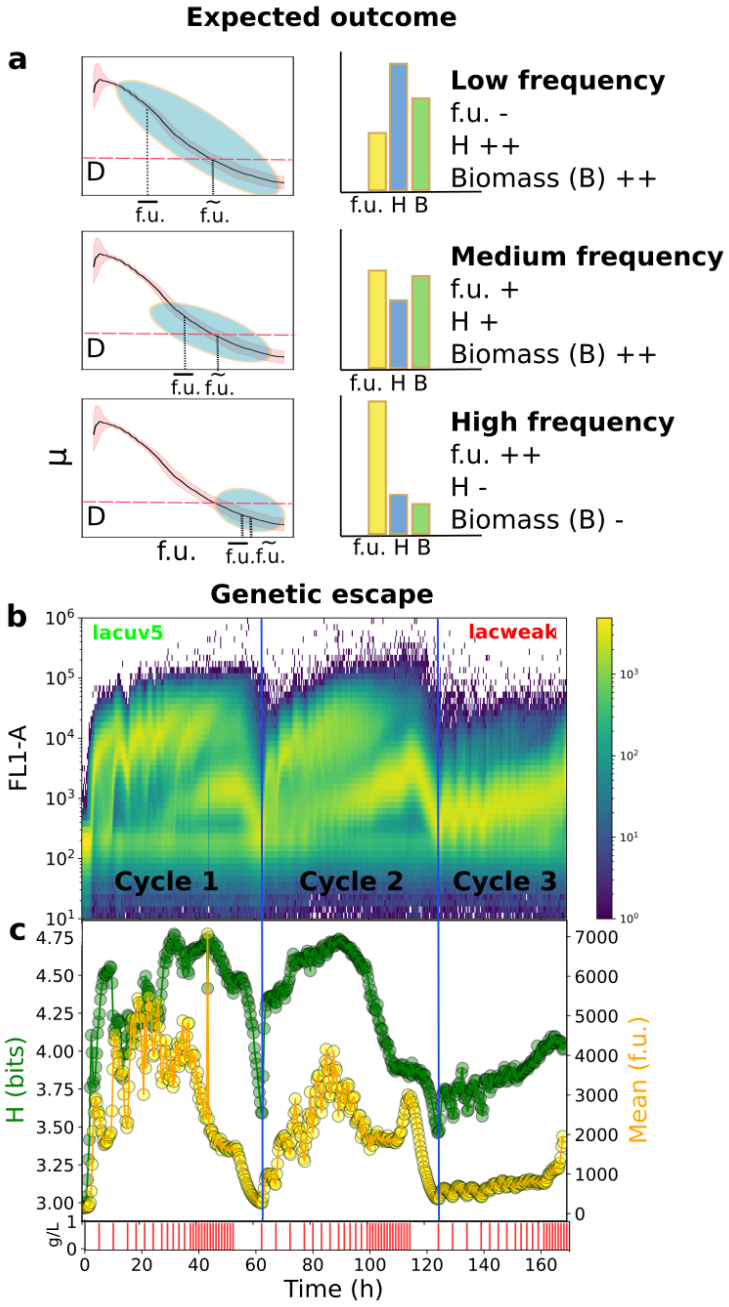

Figure 4.2: High pulsing frequencies fail at lowering entropy while increasing induction:**a** The population growth rate corresponds to the growth rate of a cell with a median fluorescence content (fu). Following the B-E compensation theory, at low pulsing frequency, a wide diversity of induction (blue shade on the graph) is present with a median induction (fu) above the mean (fu).

When increasing the pulsing frequency, the median should not increase but the mean induction should go up, resulting in a lower heterogeneity (entropy). It was speculated that as the pulsing frequency increases further, ultimately the population would be washed out because the burden would be too high ((f.u.) so high that the population growth rate would fall below the dilution rate). **b** Unexpectedly, the time density plot (20,000 cells per analysis) shows that an intermediate population appears and that the population does not washout throughout the three cycles of increasingly high pulsing frequency (two biological replicates SI Figure 6.14, n=2). **c** While at first the median induction increased (during the first cycle), it quickly starts going down whilst the entropy increases. Once the intermediate population becomes more prevalent, the entropy is much lower but the mean induction as well.

During the first pulsing cycle, the mean fluorescence increased upon increasing the lactose pulsing frequency, but the entropy remained quite constant over time (Figure 4.2 **c**). The entropy dropped only for the highest lactose pulsing frequency, but mainly due to the appearance of a new sub-population exhibiting a lower gene expression. The appearance of this sub-population was not expected, and thus before applying a second cycle, the population was left un-stimulated during five retention time to clear out any potential memory effect. Despite this time, as soon the pulsing sequence resumed, the intermediate sub-population appeared again and increased in abundance over time before taking over during the third cycle (Figure 4.2 **b**). This sub-population appeared to have lost in induction strength but was more homogeneous. Sequencing revealed that the strong lacuv5 promoter that drives the expression of the polymerase T7 had mutated back to the weaker lac wild-type promoter. This mutation had previously been described and is known to lower the expression of the polymerase, leading to a reduction of the switching cost associated with gene expression [26]. Again, this data suggests that the trade-off between growth rate and gene expression greatly impacts the entropy exhibited by the whole cell population.

5. Reducing the switching cost promotes population stability and productivity

The initial effort to promote higher population induction by reducing heterogeneity led to mutational escape, an irreversible outcome that has been reported several times as a major cause of decrease in productivity for various bio-processes [15]. However, this outcome showed another path to promote a more homogeneous population, reducing the growth rate loss associated with the activation of the gene circuit, i.e, the switching cost. In essence, the mutation reduced the burden as it resulted in a weaker induction. Thus, the mechanism of burden entropy compensation is weakened because there the gap growth rate gap between the induced an un-induced cells is lower. To promote a more homogeneous population without losing growth potential, we thus need to reduce this growth rate gap between the induced cells and the others without reduc-

ing the promoter strength. Reducing the dilution rate imposed on the system should not promote a more homogeneous population because it does not impact the switching cost and thus the competition within the reactor would remain the same. Thus, we propose to reduce the maximum achievable growth rate of the cells. Since the switching cost is the difference between the maximum growth rate of the non induced cells and the growth rate of the induced one, this option is the only way to reduce the switching cost without reducing gene expression strength. We screened the maximum growth rates of *E. coli* on various alternative carbon sources (SI Figure 6.15). Among the eight carbon sources tested, arabinose and xylose were selected. Arabinose reduced the maximum growth rate by 15% and xylose by 50% compared to glucose, while both maintaining growth rates above the dilution rate used in continuous cultivations of this work. We then assessed the impact of this strategy on population induction by conducting chemostat experiments, continuously co-feeding lactose with glucose (1 g of lactose per 5 g of the other carbon source), arabinose and, finally, xylose (Figure 4.2 **a**). As hypothesized, cell population induction exhibited lower entropy when xylose was the main carbon source by comparison with glucose (Figure 4.2 **b**). Additionally, unlike the previous pulse-based approach to homogenize the population, this strategy reduces the trade-off with production, resulting in a significantly higher average induction with xylose than with glucose, as reported based on the mean fluorescence level (Figure 4.2 **c**).

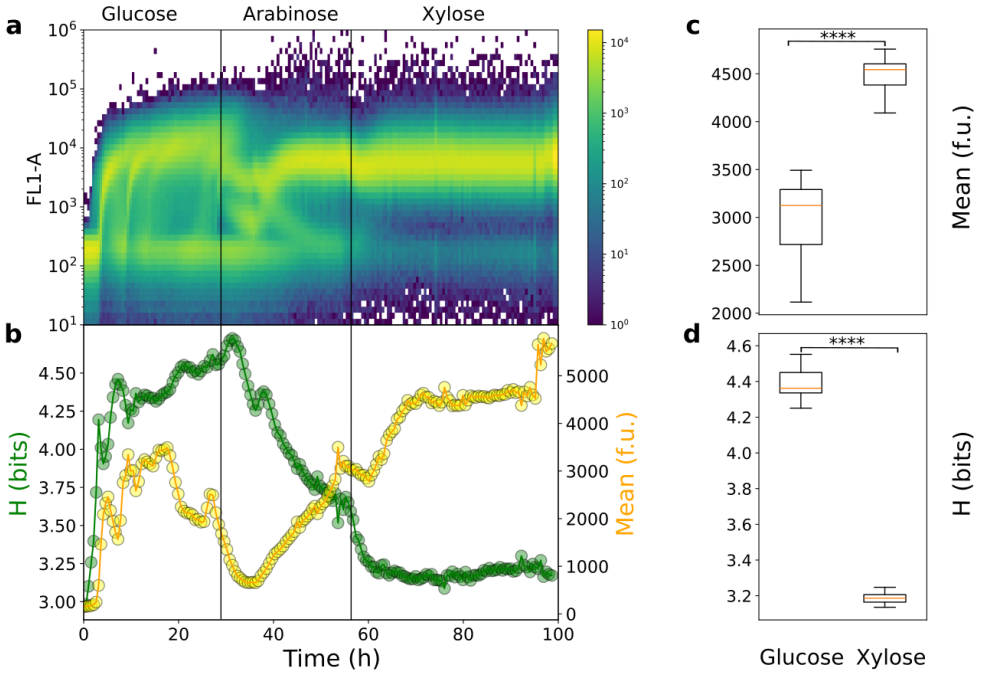

Figure 4.3: Reducing the global growth rate of the cell population lowers entropy and increases induction: **a** In chemostat cultures, the times-scatter plot (20,000 cells per analysis) shows that the population is highly heterogeneous when grown with glucose as the main carbon source. However, when switching to arabinose, and then to xylose, the population is more homogeneously induced (three biological replicates SI Figure 6.14, n=3). **b** The mean induction increases simultaneously with the decrease in entropy, as shown in further analysis. **c** Reducing the switching cost leads to a significantly (p<0.001) more productive population. **d** This increased productivity stems from a cell-to-cell heterogeneity in gene expression (measured as information entropy) that is significantly (p<0.001) smaller when xylose is used as the main carbon source compared to glucose.

6. Discussion

The interplay between growth and gene expression is a crucial factor influencing cell population behavior. However, the nature of relationship between this trade-off and the degree of heterogeneity (entropy) within cell populations remains unclear.

Our previous research has identified a link between the growth-gene expression trade-off and cell population heterogeneity, entropy. We found that the switching cost associated with activating burdensome gene circuits is a major driver of cell population dynamics. Specifically, we discovered a burden entropy compensation mechanism that promotes the stability of cell populations upon the activation of burdensome gene

circuits. In essence, the loss of growth resulting from burden is compensated by an increase in cellular entropy, leading to the emergence of cells with reduced gene activation levels and, consequently, reduced burden within the population. Our latest results show that the cell population exhibited a robust growth rate, but at the cost of reduced expression of the target gene expression and important cell to cell difference in the level of expression (Figure 4.4 **a**). To address this loss of potential, we explored the possibility of reducing cell population entropy and increasing T7 system activation by increasing the frequency of induction. This control method turned out to be a dead end, we observed mutational escape when increasing the frequency of induction, resulting in the emergence of mutant cells with reduced gene expression (Figure 4.4 **b**). This rapid escape is likely due to the increase in mutation rate associated with the degree of burden experienced by cells [15].

An alternative strategy for mitigating the switching cost is to slow down the maximum growth rate of the population (Figure 4.4 **c**), i.e., limiting the ability of cells with a lower induction to overgrow the producing population. In this study, we employed less effective carbon sources to this end, such as arabinose and xylose. Notably, xylose, which led to the lowest growth rate, was particularly effective in reducing the impact of the switching cost on cell population entropy while increasing the level of T7 system induction throughout the cultivation. Interestingly, the population does not appear to escape control by mutation as fast as it did when we tried controlling it by applying high frequencies of induction. This is likely a result of this softer control method where we did not periodically force harsh inductions but instead reduced that competition between highly induced and less induced cells. Our approach is complementary to the use of "host-aware" gene circuits, which are designed to mitigate in vivo metabolic burden generated by the synthesis and accumulation of various bioproducts [27]. Microbial strains engineered to exhibit reduced growth rates could also be explored in the future. Taken together, our approach contradicts the conventional bioprocess design principles, which typically aim to encourage cells to work harder and faster. Our findings reveal that B-E compensation is a critical factor in maintaining the stability of cell populations expressing a burdensome gene. Moreover, a deeper understanding of this phenomenon can lead to the development of innovative strategies for designing more robust continuous bio-process technologies.

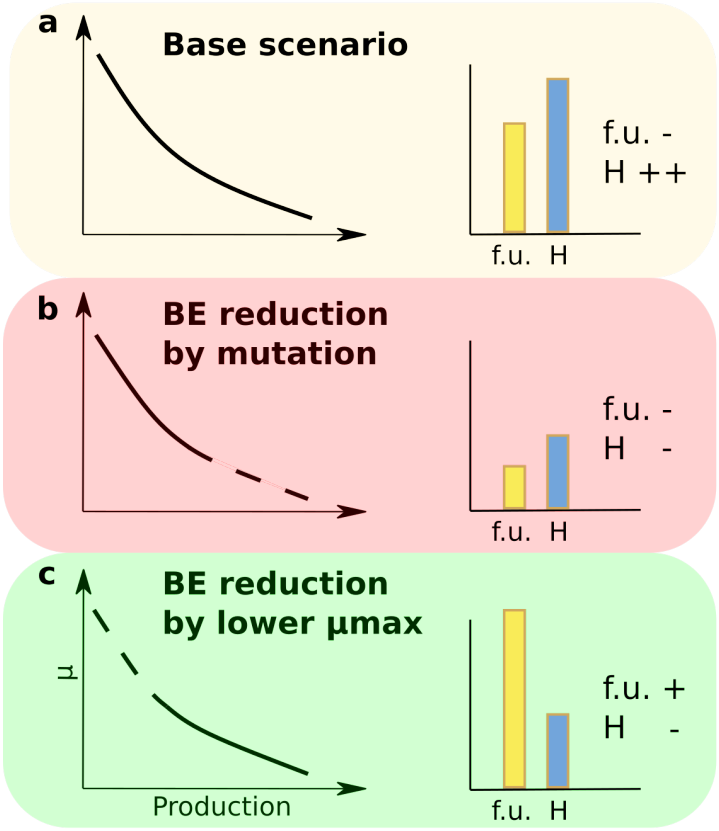

Figure 4.4: Representation of the growth-production trade-off curve on the population induction: **a** In a traditional continuous cultivation, *E. coli* is grown with glucose, a carbon source providing a high growth rate, and tasked with the production of a burdensome protein. The important switching cost leads to a high Burden Entropy compensation (BE) that harms population homogeneity and productivity. **b** Trying to reduce this heterogeneity by applying periodic pulses can lead to mutational escape, resulting in a population that is more homogeneous but less productive. **c** An alternative way of reducing the switching cost is to reduce the maximum growth rate of the cells, which can be done by lowering the quality of the main carbon source. This enables the creation of a more productive and homogeneous population.

7. Material and methods

Strain and growth media

Experiments performed in this work focused on *E. coli* BL21 DE3 with a $pET28 : GFP$ plasimd (Plasmid 60733) deposited on Addgene by Matthew Bennett. In all cultivations, the strain was grown in a minimal mineral media containing (in g/l): K_2HPO_4: 14.6; $NaH_2PO_4 \cdot 2H_2O$: 3.6; Na_2SO_4: 2; $(NH_4)_2SO_4$: 2.47; NH_4Cl: 0.5; $(NH_4)_2H - \text{citrate}$: 1; glucose: 5, thiamine: 0.01, tryptophan: 0.05. The medium is supplemented with a trace element solution totaling 11 ml/l assembled from the following solutions (in g/l), 3/11 of $FeCl_3 \cdot 6H_2O$: 16.7 , 3/11 of EDTA: 20.1, 2/11 of $MgSO_4$: 120 and 3/11 of a metallic trace element solution. The metallic trace element solution contains (in g/L): $CaCl_2 \cdot 2H_2O$: 0.74; $ZnSO_4 \cdot 7H_2O$: 0.18; $MnSO_4 \cdot H_2O$: 0.1; $CuSO_4 \cdot 5H_2O$: 0.1, $CoSO_4 \cdot 7H_2O$: 0.21. Both the trace element solution and tha kanamycin for plasmid maintenance (50 $\mu g/l$) filter sterilized (0.22 μm). For bioreactor cultivations, antifoam was added (KS911, 2 drops per liter).

Bioreactor cultivations and FC analysis

Bioreactor cultivations were performed in a 1 L Bionet F1 lab-scale bioreactor. The batch was inoculated at an initial optical density (OD) of 0.5 from an overnight preculture (baffled flask, 1 L total volume, 0.1 L working volume). The continuous cultivation was started once oxygen levels began rising at the end of the batch, signaling the transition from exponential to stationary phase. All cultivations were performed at 37°C, 1 VVM, pH 7, and 1000 revolutions per minute (RPM).

In the chemostat, the feed contained 5 g/L of the main carbon source (glucose, arabinose, or xylose) and 1 g/L of lactose, the inductor. In the Segregostat, the feed only contained the carbon source (5 g/L glucose) and the inductor was added as a conditional pulse (0.5 g of lactose per pulse). In the experiments where the pulsing frequency varied, each pulse consisted of 1 g of lactose and was automatically triggered from a pre-set .csv file. The automated flow cytometry platform was already described, but briefly: every 12 minutes, a sample is drawn from the bioreactor using a capillary tube and a peristaltic pump. The sample is introduced into an analysis chamber that also contains the flow cytometer. The sample is diluted based on the number of events per microliter (μL) recorded at the previous flow cytometry (FC) analysis by successively flushing and filling the chamber with phosphate-buffered saline (PBS). Then, a FC analysis of 40,000 events is automatically performed, followed by the chamber being cleaned by flushing it multiple times with PBS, and the cycle starts again. In the Segregostat, if more than 50% of the cells are below a fluorescence threshold (1000 arbitrary units on the FL1-A channel), a pulse is triggered.

Sequencing

At the end of the third cycle of increasingly high pulsing frequency where the appearance of a lesser induced subpopulation, a sample of 1 ml was taken from the reactor and plated to recover 10 random individual colonies. Each colonies was then re-grown in LB with kanamycin, the genome extracted and the region of the lac promotor driving the T7 polymerase was amplified with the following primers:

FW Sequencing
CGCCGTTAACCACCATCAAAC
RV Sequencing
CGCAACTCGTGAAAGGTAG

The same procedure was done from a pre-culture used to start the cultivation to verify that the reactor was not inoculated with a mutant to begin with. All PCR products were then sent for sequencing (Sanger sequencing, Eurofins). All colonies from the pre-culture had the lac uv5 strong promotor and but 30 % of the one from the end of the cultivations had a lac wt promotor.

Maximum growth rate determination

The maximum growth rate under glucose, arabinose, xylose, fructose, and mannose was determined by performing batch cultivations of *E. coli* BL21 (DE3) $pET28 : GFP$ in mineral media containing 5 g/l of the carbon source and 5 g/l of MOPS buffer to maintain a pH of 7. The batch cultivations were conducted in a flower plate (37°C; 1000 RPM) with a cultivation volume of 1 ml, using a Biolector (Beckman Coulter) instrument, where the backscatter signal was recorded for 5 wells per condition.

Bibliography

[1] Dennis Binder et al. “Comparative Single-Cell Analysis of Different E. coli Expression Systems during Microfluidic Cultivation”. In: *PLOS ONE* 11.8 (Aug. 15, 2016). Ed. by Paul D. Riggs, e0160711. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0160711. URL: https://dx.plos.org/10.1371/journal.pone.0160711.

[2] G. Balazsi, A. van Oudenaarden, and J. J. Collins. “Cellular decision making and biological noise: from microbes to mammals”. In: *Cell* 144 (2011), pp. 910–925. DOI: 10.1016/j.cell.2011.01.030.

[3] Avigdor Eldar and Michael B. Elowitz. “Functional roles for noise in genetic circuits”. In: *Nature* 467.7312 (Sept. 2010), pp. 167–173. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature09326. URL: https://www.nature.com/articles/nature09326.

[4] Michael B. Elowitz et al. “Stochastic gene expression in a single cell”. In: *Science* 297.5584 (2002), pp. 1183–1186. DOI: 10.1126/science.1070919.

[5] Martin Ackermann. “A functional perspective on phenotypic heterogeneity in microorganisms”. In: *Nature Reviews Microbiology* 13.8 (Aug. 2015), pp. 497–508. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/nrmicro3491. URL: https://www.nature.com/articles/nrmicro3491.

[6] M. Thattai and A. van Oudenaarden. “Stochastic gene expression in fluctuating environments”. In: *Genetics* 167 (2004), pp. 523–530. DOI: 10.1534/genetics.167.1.523.

[7] E. Kussell and S. Leibler. “Phenotypic diversity, population growth, and information in fluctuating environments”. In: *Science* 309 (2005), pp. 2075–2078.

[8] María Antonia Sánchez-Romero and Josep Casadesús. “Contribution of phenotypic heterogeneity to adaptive antibiotic resistance”. In: *Proceedings of the National Academy of Sciences* 111.1 (Jan. 7, 2014), pp. 355–360. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1316084111. URL: https://pnas.org/doi/full/10.1073/pnas.1316084111.

[9] Markus Arnoldini et al. “Bistable Expression of Virulence Genes in Salmonella Leads to the Formation of an Antibiotic-Tolerant Subpopulation”. In: *PLoS Biology* 12.8 (Aug. 19, 2014). Ed. by Nathalie Balaban, e1001928. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1001928. URL: https://dx.plos.org/10.1371/journal.pbio.1001928 (visited on 09/18/2024).

[10] Stefany Moreno-Gámez et al. “Wide lag time distributions break a trade-off between reproduction and survival in bacteria”. In: *Proceedings of the National Academy of Sciences* 117.31 (Aug. 4, 2020), pp. 18729–18736. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2003331117. URL: https://pnas.org/doi/full/10.1073/pnas.2003331117.

[11] Frank Delvigne and Philippe Goffin. “Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations”. In: *Biotechnology Journal* 9.1 (Jan. 2014), pp. 61–72. ISSN: 1860-6768, 1860-7314. DOI: 10.1002/biot.201300119. URL: https://onlinelibrary.wiley.com/doi/10.1002/biot.201300119.

[12] Mary E Lidstrom and Michael C Konopka. “The role of physiological heterogeneity in microbial population behavior”. In: *Nature Chemical Biology* 6.10 (Oct. 2010), pp. 705–712. ISSN: 1552-4450, 1552-4469. DOI: 10.1038/nchembio.436. URL: https://www.nature.com/articles/nchembio.436 (visited on 09/18/2024).

[13] Florian David et al. “Single cell analysis applied to antibody fragment production with Bacillus megaterium: development of advanced physiology and bioprocess state estimation tools”. In: *Microbial Cell Factories* 10.1 (Apr. 15, 2011), p. 23. ISSN: 1475-2859. DOI: 10.1186/1475-2859-10-23. URL: https://doi.org/10.1186/1475-2859-10-23 (visited on 09/18/2024).

[14] Tian Xia, Mark A. Eiteman, and Elliot Altman. “Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of Escherichia coli strains”. In: *Microbial Cell Factories* 11 (2012), pp. 1–9. DOI: 10.1186/1475-2859-11-77.

[15] Peter Rugbjerg et al. “Diverse genetic error modes constrain large-scale bio-based production”. In: *Nature Communications* 9.1 (Feb. 20, 2018), p. 787. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03232-w. URL: https://www.nature.com/articles/s41467-018-03232-w (visited on 09/18/2024).

[16] Peter Rugbjerg and Morten O. A. Sommer. “Overcoming genetic heterogeneity in industrial fermentations”. In: *Nature Biotechnology* 37.8 (Aug. 2019), pp. 869–876. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/s41587-019-0171-6. URL: http://www.nature.com/articles/s41587-019-0171-6 (visited on 03/23/2023).

[17] Dmitry Nevozhay et al. “Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression”. In: *Proceedings of the National Academy of Sciences* 106.13 (Mar. 31, 2009), pp. 5123–5128. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0809901106. URL: https://pnas.org/doi/full/10.1073/pnas.0809901106 (visited on 09/18/2024).

[18] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. “Construction of a genetic toggle switch in Escherichia coli”. In: *Nature* 403.6767 (Jan. 2000), pp. 339–342. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/35002131. URL: https://www.nature.com/articles/35002131.

[19] Lucas Henrion et al. “Fitness Cost Associated with Cell Phenotypic Switching Drives Population Diversification Dynamics and Controllability”. In: *Nature Communications* 14.1 (Oct. 2023), p. 6128. ISSN: 2041-1723. DOI: 10.1038/s41467-023-41917-z. (Visited on 11/01/2023).

[20] D. Benzinger and M. Khammash. “Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation”. In: *Nat. Commun.* (2018), p. 3521. DOI: 10.1038/s41467-018-05882-2.

[21] T. M. et al. Nguyen. “Reducing phenotypic instabilities of a microbial population during continuous cultivation based on cell switching dynamics”. In: *Biotechnol. Bioeng.* (2021). DOI: 10.1002/bit.27860.

[22] J. A. Martinez et al. “Controlling microbial co-culture based on substrate pulsing can lead to stability through differential fitness advantages”. In: *PLoS Comput. Biol.* 18 (2022), e1010674.

[23] D. Benzinger and M. Khammash. “Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation”. In: *Nat. Commun.* 9 (2018), p. 3521.

[24] Joe H. Levine, Yihan Lin, and Michael B. Elowitz. “Functional Roles of Pulsing in Genetic Circuits”. In: *Science* 342.6163 (Dec. 6, 2013), pp. 1193–1200. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1239999. URL: https://www.science.org/doi/10.1126/science.1239999 (visited on 09/18/2024).

[25] Stefan Kittler et al. “The Lazarus Escherichia coli Effect: Recovery of Productivity on Glycerol/Lactose Mixed Feed in Continuous Biomanufacturing”. In: *Frontiers in Bioengineering and Biotechnology* 8 (Aug. 13, 2020), p. 993. ISSN: 2296-4185. DOI: 10.3389/fbioe.2020.00993. URL: https://www.frontiersin.org/article/10.3389/fbioe.2020.00993/full (visited on 09/18/2024).

[26] Soon-Kyeong Kwon et al. “Comparative genomics and experimental evolution of Escherichia coli BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction”. In: *Scientific Reports* 5.1 (Nov. 4, 2015), p. 16076. ISSN: 2045-2322. DOI: 10.1038/srep16076. URL: https://www.nature.com/articles/srep16076 (visited on 01/04/2024).

[27] Alice Boo, Tom Ellis, and Guy-Bart Stan. “Host-aware synthetic biology”. In: *Current Opinion in Systems Biology* 14 (Apr. 2019), pp. 66–72. ISSN: 24523100. DOI: 10.1016/j.coisb.2019.03.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S245231001930006X.

5

Chapter 5: General discussion and outlook

1. Population dynamics needs a system biology approach to be studied and controlled

The objectives of this thesis were twofold. The fist, was to improve our understanding on the population heterogeneity drivers, specifically in the context of continuous bioproduction, and the second was to propose control strategies. Over this work, it appeared that to fully understand the complex dynamics of gene expression heterogeneity across a microbial population, a holistic approach is essential. The level of gene expression and the degree of heterogeneity cannot be solely attributed to the gene circuit or environmental conditions. In their work on the noisiness of numerous promoters in *E. coli*, Silander and colleagues found that *bolA* and *araB* are highly noisy [1]. Our results [2], however, are in direct contradiction with these observations. The discrepancy between the two works likely originates from the different growing conditions used: batch cultivation in a shake flask for Silander's study, versus continuous cultivation with glucose as the limiting element in our work. A system biology approach to understand microbial population dynamics highlights the importance of considering the complex interplay between gene properties, environmental conditions, and population dynamics. By adopting this approach, we can better appreciate the context-dependent nature of cellular systems and the limitations of reductionist approaches that focus solely on the gene network architecture.

To sum up the conclusion to the first objective; cells interpret environmental cues and respond to them in both a responsive and stochastic manner by expressing multiple phenotypes. This ratio between stochastic and responsive is under evolutionary pressure that shapes, among others, the gene network architecture, gene copy number, and even regulatory protein and region structure. This ratio is also a question of signal interpretability. In *E. coli*, arabinose induction is highly responsive when arabinose is absent (no induction) or present at a concentration above 1 g/l, but every concentration in between result in an important share of stochastic switching. Thus, this response function is characterized by its leakiness, dynamic range, and sharpness.

Then, the choice of cells operating in the population is further impacted by competition, certain phenotypes may have a greater growth than others. Labeled as burden entropy compensation in this work, it has been shown to heterogenize the population further in the context of continuous cultivation. But the complex interaction between the population phenotypic composition and the environment does not stop there. As the environment is shared among the population, usually the choice of one cell to adopt a phenotype will change the chance of the following cells to do so. This environmental feedback was exemplified in the case of sporulation in *Bacillus subtilis*, where it leads to oscillatory behavior under certain operating conditions.

The activation of the T7 expression system in *E. coli* exemplifies further the multifactorial nature of phenotypic heterogeneity and how it shapes population dynamics. This system was described in this work as a prime example of burden entropy com-

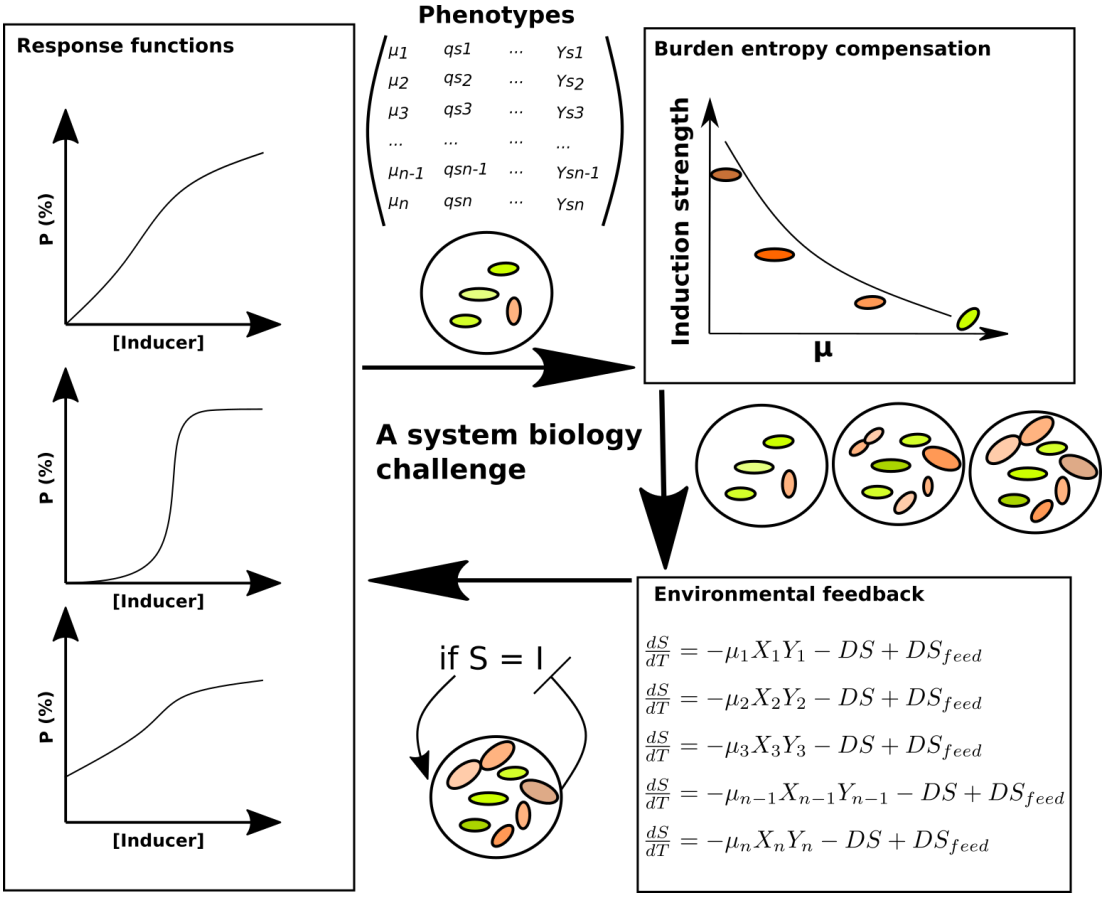

Figure 5.1: Multiple components come into play to form the population dynamics over time. The response function of a gene circuit is characterized by a given leakiness and sharpness, which together with the perceived inducer concentration shape the population induction. This initial heterogeneity can be amplified by the difference between these phenotypes (growth, survival, sporulation, etc.). Additionally, the population interacts with the inducer (environmental feedback) by, for example, consuming it. All three of these components (induction dictated by the response function, burden entropy compensation, and environmental feedback) happen simultaneously, with each having an impact on the population dynamics with varying strengths.

pensation, but in this system as well and similarly to the sporulation of *B. subtilis*, bursts of expression can be observed. These bursts suggest again a role for environmental feedback in the dynamics of this system as well. It can only be speculated that there, glucose again is the culprit of these oscillations and not lactose. Glucose is the main carbon source but also impacts the induction dynamics because its shortage is required to lift the catabolic repression and trigger a strong induction to lactose. If the induction is strong enough, it could be envisioned that the associated growth rate loss is accompanied by an increase in glucose concentration, thus inhibiting induction and forming an oscillatory behavior very similar to the one described for the sporulation in *B. subtilis*. Ultimately, this work proposes that to understand the population dynamics, one has to consider the following questions:

1. How does a cell population react to an inducer? In other words, what is the dose-response relationship that links a gene circuit to its inducer (leakiness, sharpness, dynamic range)?

2. How do the resulting phenotypes structure themselves? Is a phenotype going to have the capacity to overtake others?

3. How will the environmental conditions be modified once taking into account the presence and abundance of these phenotypes?

Understanding the parameters at play behind gene expression dynamics across a population is key if one wants to infer new control methods, the second objective of this work. Methods such as the one proposed in this work, where it was shown that using a lower-quality carbon source for the population actually increases the homogeneity of the induction of the T7 system only make sense once the reasons behind the heterogeneity become apparent. Again, understanding population dynamics is a multifactorial issue that cannot be tackled solely with a process or biology approach, but rather a system biology one (Figure 5.1).

It is also within this framework that whether phenotypic heterogeneity is bad or good should be answered. Although probably asked in many ways by many scientists, this question was debated within this research team when asked by Pr. Grünberger. The answer that appeared to win was, it depends. The other answer was, that it depends on who you ask this question to. Phenotypic heterogeneity is a survival mechanism that in a natural context improves the population fitness [3]. However, within the bioprocess community, it is often viewed as a loss that hampers bioprocess productivity [4]. Many have attempted to eliminate it, but our work reveals that it actually safeguards the population when tasked with expressing burdensome genes. Nevertheless, we show that while it should not be eliminated, it can be reduced at the benefit of increased productivity.

2. Open source and standardization

Humans have cultivated cells for a very long time but the comprehension of the population characteristics at a single cell level is young. Single-cell analysis tools provide invaluable insight into the functionality of this noise and its causes, but the obtained data are also challenging to treat. Artificial intelligence based segmentation is facilitating the conversion of images (such as the one provided by microfluidics) into numerical values but even then, the question of how to treat these data is not straightforward [5] [6]. Fortunately, these tools (AI based segmentation, flow cytometry toolbox, data treatment workflow) are being shared more freely, in part thanks to the trend of open science.

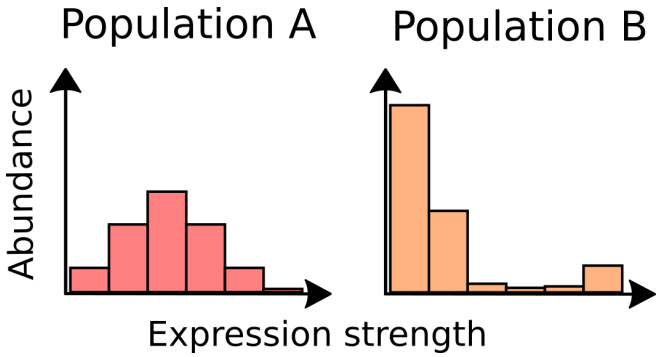

Figure 5.2: With the standard deviation it would be easy to assume that population B is more heterogeneous than population A. Yet, most of the cells are concentrated within certain area of gene expression in population B, thus more homogeneous.

With new tools and data, new ways of interpreting those are also needed. Comparing two populations, one of which is normal and the other one not, with the standard deviation can lead to wrong conclusions (Figure 5.2). With biological systems this is an important issue if one wants to quantify heterogeneity, since gene expression is often skewed or even binomial. Additionally, the standard deviation is not very robust to outliers. This sheds light on the importance to develop new proxies to quantify the heterogeneity in the population, such as among others, entropy. Shannon Entropy, offers the advantage of being distribution independent, robust to outliers and thus has been extensively used throughout this work to provide a quantitative analysis of heterogeneity. Using entropy as a noise measurement comes with the added benefit that it is linked with another important metric, mutual information. This metric that describes, how much information, is conveyed by a given environmental condition is analogous to the dose response curve.

But even using the same metrics, there is also the question of data interpretability between research groups. As an example, plate readers and flow cytometers give numerical values in arbitrary units. Two brands of such devices will give values in a

totally different range. But work has also been done at this level. Universal calibration are being proposed to convert plate reader arbitrary unit output into protein counts [7]. As suggested [7], absolute protein counts could probably be implemented in flow cytometer by immobilizing the proteins on beads.

It feels like the quest to understand living organisms has barely started, let alone to engineer and control them. Through this work, I have discovered many tools that enabled me to benefit from the work of others. Of course, having access to articles in open source is important, but tools that compile the information such as BioNumbers massively facilitate others work. Repositories like Addgene permit easy exchange of genetic material others have built, and codes used in research dropped on GitHub smooths out the learning curve for beginners in modeling and data treatment as well as providing new ideas to others.

The complexity in collecting and treating data on DNA construction also happens the development of standard DNA parts. A common principle in engineering is to rely on characterized and standard parts. As stated in the first line of Tom Knight's essay (Idempotent Vector Design for Standard Assembly of Biobricks) on this matter "The lack of standardization in assembly techniques for DNA sequences forces each DNA assembly reaction to be both an experimental tool for addressing the current research topic, and an experiment in and of itself.". In an effort to solve this issue and as a prime example of how standardization and open science can help the community, Knight proposed the concept of BioBrick, a library of parts that can be easily assembled to form genetically engineered *E. coli*. The BioBricks are now commonly used during the iGEM competitions and have also been used in many other scientific works [8].

3. Engineering long term stability by reducing genotypic escape

A crucial component of population dynamics that has not been addressed in this work is genotypic escape. While it is recognized that in a clonal population, phenotypic heterogeneity is the first to shape the population dynamics, cells eventually mutate and mutation spreads. It is especially important to tackle this issue because the mutation rate increases with burden, i.e., production load [9]. Interestingly, our work shows that burden, on top of promoting genotypic escape, also promotes phenotypic heterogeneity. In this work, the importance of phenotypic heterogeneity to preserve the population fitness was outlined. It was also shown that reducing this crucial heterogeneity could actually be quickly met by mutations reducing productivity. While eventually, it was found that the population could be homogenized without applying harsh periodic inductions, it is a certainty that over time, a mutation leading to the population functionality loss will arise.

As an outlook, maybe phenotypic heterogeneity could be engineered to promote

population stability at the genetic level as well. There, the noisiness in the expression of a recombinase, such as the Bxb1 recombinase in *E. coli* or the CRE-lox system in yeast, could generate partial differentiation. This would trigger the segregation of a population into two groups, growers and producers. Such differentiation has already been successfully built before and has yielded encouraging results [10] [11]. Because the mutation rate is greater when the production load is important, and cells with a high production load are also the easiest to out-compete, any mutations that occur in a system without differentiation quickly invades the population. With a differentiable system, mutants will outgrow the producers but probably not the growers. There, the growers out-compete even the producers that mutated and thus safeguard the production potential. In this scenario, phenotypic heterogeneity in the activation of a gene circuit is leveraged by, for example, inducing the population with very low concentrations of inducer, promoting noisy responses, to periodically generate new producers from the growers.

Going further, the differentiation system could be engineered in a way that once triggered, the producers would lose out any ability to grow. This terminal differentiation has already been proposed and would totally eliminate any possibility of a producer to manage to mutate and overtake the rest of the population [11]. In a way, since these cells lose their growth capacity, it could be argued that they are no longer alive, and that such terminal differentiation is a self-sustained cell-free expression system. These cells, incapable of growing, are not subjected to a burden entropy compensation mechanism nor are they to mutations. This strategy would then eliminate two bottlenecks that limit the adoption of continuous bio-manufacturing, genotypic and phenotypic heterogeneity.

Bibliography

[1] O. K. et al. Silander. “A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli”. In: *PLoS Genet.* 8 (2012), e1002443.

[2] Lucas Henrion et al. “Fitness Cost Associated with Cell Phenotypic Switching Drives Population Diversification Dynamics and Controllability”. In: *Nature Communications* 14.1 (Oct. 2023), p. 6128. ISSN: 2041-1723. DOI: 10.1038/s41467-023-41917-z. (Visited on 11/01/2023).

[3] Martin Ackermann. “A functional perspective on phenotypic heterogeneity in microorganisms”. In: *Nature Reviews Microbiology* 13.8 (Aug. 2015), pp. 497–508. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/nrmicro3491. URL: https://www.nature.com/articles/nrmicro3491.

[4] Frank Delvigne and Philippe Goffin. “Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations”. In: *Biotechnology Journal* 9.1 (Jan. 2014), pp. 61–72. ISSN: 1860-6768, 1860-7314. DOI: 10.1002/biot.201300119. URL: https://onlinelibrary.wiley.com/doi/10.1002/biot.201300119.

[5] J. Stricker et al. “A fast, robust and tunable synthetic gene oscillator”. In: *Nature* 456 (2008), 516–519. DOI: 10.1038/nature07389.

[6] Marius Pachitariu and Carsen Stringer. “Cellpose 2.0: how to train your own model”. In: *Nature Methods* 19.12 (Dec. 2022), pp. 1634–1641. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-022-01663-4. URL: https://www.nature.com/articles/s41592-022-01663-4.

[7] Eszter Csibra and Guy-Bart Stan. “Absolute protein quantification using fluorescence measurements with FPCountR”. In: *Nature Communications* 13.1 (Nov. 3, 2022), p. 6600. ISSN: 2041-1723. DOI: 10.1038/s41467-022-34232-6. URL: https://www.nature.com/articles/s41467-022-34232-6.

[8] Christina D Smolke. “Building outside of the box: iGEM and the BioBricks Foundation”. In: *Nature Biotechnology* 27.12 (Dec. 2009), pp. 1099–1102. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt1209-1099. URL: https://www.nature.com/articles/nbt1209-1099 (visited on 10/04/2024).

[9] Peter Rugbjerg et al. “Diverse genetic error modes constrain large-scale bio-based production”. In: *Nature Communications* 9.1 (Feb. 20, 2018), p. 787. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03232-w. URL: https://www.nature.com/articles/s41467-018-03232-w (visited on 09/18/2024).

[10] Chetan Aditya et al. “A light tunable differentiation system for the creation and control of consortia in yeast”. In: *Nature Communications* 12.1 (Oct. 5, 2021), p. 5829. ISSN: 2041-1723. DOI: 10.1038/s41467-021-26129-7. URL: https://www.nature.com/articles/s41467-021-26129-7 (visited on 11/06/2023).

[11] Rory L. Williams and Richard M. Murray. “Integrase-mediated differentiation circuits improve evolutionary stability of burdensome and toxic functions in E. coli”. In: *Nature Communications* 13.1 (Nov. 10, 2022), p. 6822. ISSN: 2041-1723. DOI: 10.1038/s41467-022-34361-y. URL: https://www.nature.com/articles/s41467-022-34361-y (visited on 10/03/2024).

Published content and contributions

First authorship

Article

Henrion, L., Delvenne, M., Bajoul Kakahi, F., Moreno-Avitia, F. & Delvigne, F. Exploiting Information and Control Theory for Directing Gene Expression in Cell Populations. Front. Microbiol. 13, 869509 (2022).

Henrion, L., Martinez, J.A., Vandenbroucke, V. et al. Fitness cost associated with cell phenotypic switching drives population diversification dynamics and controllability. Nat Commun 14, 6128 (2023). https://doi.org/10.1038/s41467-023-41917-z

Biological oscillations without genetic oscillator or external forcing Vincent Vandenbroucke, Lucas Henrion, Delvigne Frank bioRxiv 2024.09.20.614027;
doi: https://doi.org/10.1101/2024.09.20.614027 **Pre-print**

Lowering the switching cost related to the activation of burdensome gene circuits promotes cell population homogeneity and productivity Lucas Henrion, Vincent Vandenbroucke, Juan Andres Martinez Alvarez, Julian Kopp, Samuel Telek, Andrew Zicler, Frank Delvigne bioRxiv 2024.10.14.618176;
doi: https://doi.org/10.1101/2024.10.14.618176 **Pre-print**

Secondary authorship

Article

Martinez, J. A. et al. Controlling microbial co-culture based on substrate pulsing can lead to stability through differential fitness advantages. PLoS Comput Biol 18, e1010674 (2022).

Book chapter

Delvigne, F., Henrion, L., Vandenbroucke, V., Martinez, J.A. (2023). Avoiding the All-or-None Response in Gene Expression During E. coli Continuous Cultivation Based on the On-Line Monitoring of Cell Phenotypic Switching Dynamics. In: Kopp, J., Spadiut, O. (eds) Inclusion Bodies. Methods in Molecular Biology, vol 2617. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2930-7_7

6

Supplementary information

1. Supplementary information chapter 2

Supplementary Notes

Supplementary Note 1: Determination of the entropy (H) of the population from automated FC data and reproducibility of these data

Determination of the entropy (H) of the population from automated FC data and reproducibility of these data. Information theory has been used in this work to characterize the response of cell populations to environmental perturbations. This framework involves the computation of entropy H, which can be regarded as a measurement of uncertainty about the response of the cell population (output) in function of the environmental stimulation (input). This input-output relationship, which is the basis of information theory, will be detailed in Supplementary Note 3. In this note, we will concentrate on the description of the entropy H as a measurement of the level of heterogeneity of the population. Entropy can be considered as a measure of uncertainty about the outcome of a draw from a probability distribution1. As an example, if we pick randomly a cell in our population, how much are we surprised to pick one cell with a given GFP level?
In our case, we can measure the entropy based on fluorescence distribution acquired with automated FC based on the following equation:

$$H = -\sum_{i=1}^{n} m_i p(x_i) \log_2(p_i) \tag{6.1}$$

With m being the number of states observed (i.e., GFP classes in our cases) and p being the probability to observe this state. The probability to observe different classes of fluorescence can be easily determined based on automated FC. The computation of H has been exemplified based on fictive population of cells clustered in three fractions according to the level of GFP exhibited by cells (Supplementary Figure 1). As an example, the computation for the first population distribution (Supplementary Figure 1a) is performed as

$H : -(0.1 \log_2(0.1)) + (0.6 \log_2(0.6)) + (0.3 \log_2(0.6)) = 1.29 bits.$

Based on this first example, the entropy of the population can be either increased (Supplementary Figure 1b), the maximum entropy value being reached when cells are equally distributed into the 3 different clusters. On the opposite, H can be decreased and set to zero when all the cells exhibit the same fluorescence range (Supplementary Figure 1c). This approach has been applied to automated FC data (animated movies of the automated FC data for all the biological systems considered are available at GitLab [https://gitlab.uliege.be/mipi/published-software/mipi-model-and-simulation-database/] for computing the evolution of H(t) for different types of cell population (Supplementary Figure 2 for three out of six of the cell systems investi-

gated)). In this case, we applied 50 bins for the computation of H(t). Reproducibility of the corresponding experiments conducted in Segregostat is provided in Supplementary Figure 3.

Supplementary Note 2: Determination of the fitness cost associated with phenotypic switching

For the systems exhibiting high switching cost, the quantification of this parameter was a bit difficult because cells switched stochastically from the OFF state to the ON state. This was particularly true for the P glc3 :GFP system in yeast where only a small fraction of cells decided to switch under normal cultivation conditions. This effect can be observed based on MSCC experiments run at different glucose concentration (Supplementary Figure 4). However, the determination of the switching cost was easier in the case of the $pET28 : GFP$ (T7 polymerase) system in *E. coli* BL21(DE3), this system being inducible upon addition of lactose. Accordingly, we cultivated systems exhibiting low ($P_{araB} : GFP$ and $P_{lacZ} : GFP$ in *E. coli* W3110) and high ($pET28 : GFP$ system in *E. coli* BL21(DE3)) in a multiplate cultivation device (Biolector) (Supplementary Figure 5). Based on these data, it can be observed that the effect of gene circuit activation itself and the change of carbon source are coupled most of the time. However, we can also see that, in most of the cases, the activation of the gene circuit itself drives the switching cost. It can be seen that the *E. coli* W3110 strain exhibit slightly reduced growth when cultivated on lactose and arabinose, independently of the presence of a fluorescent reporter. On the opposite, we can see that the *E. coli* BL21(DE3) strains exhibit a huge reduction in growth upon cultivation on lactose, mostly due here to the activation of the chromosomal insert of the T7 RNA polymerase.

Supplementary Note 3: Experimental determination of the response function for the $P_{araB} : GFP$ system in *E. coli* and the $P_{glc3} : GFP$ system in *S. cerevisiae* and computation of the mutual information (MI)

Information theory relies on the characterization of the input-output relationship for various systems, and has been applied recently to the analysis of signal propagation in biological systems 23 . Basically, MI allows to quantify how much we can know about an input (e.g., change in environmental condition) from the output (i.e., in our case, the fluorescence distribution of the population). The first step for the computation of MI is to calculate the entropy of the population exposed to defined environmental conditions. These conditional distributions represent the response function of our cellular systems, and more precisely the $P_{araB} : GFP$ system in *E. coli* (Supplementary Figure 6) and $P_{glc3} : GFP$ system in *S. cerevisiae* (Supplementary Figure 7).

Characterization of the response function for the *S. cerevisiae* $P_{glc3} : GFP$ system

The response function of the $P_{glc3} : GFP$ in *S. cerevisiae* was determined by growing culture at different dilution rates in chemostat (Supplementary Figure 7). For this

purpose, the dilution rate of a chemostat was progressively increased to release the stress response of the population. This procedure is known as accelerostat (A-stat). In our case, the pump flow rate was modified with a step change every 2 hours, resulting in a progressive increase of the dilution rate of 0.002 h^{-1} per hour. This incremental range was chosen in order to ensure pseudo steady-state for each increment. The entire process was followed by automated FC for mapping the GFP distribution of the cell population (Supplementary Figure 7).

Computation of mutual information (MI) based on the response function of a cell population

Knowing the response function, and the corresponding conditional GFP distribution, it is possible to compute the MI of a specific cellular system. This computation will be exemplified for the $P_{glc3} : GFP$ reporter in *S. cerevisiae*. The environmental input for this system is the glucose uptake rate determined based on the value of the dilution rate, as well as based on glucose and biomass measurement. The conditional fluorescence distributions were then acquired for different substrate uptake rates and H was computed accordingly (Supplementary Figure 8a). If all the fluorescence distributions are summed up, the corresponding entropy value is the total entropy of the system. MI is then computed according to (Supplementary Figure 8b):

$$MI(y, x) = H(x) - H(x, y) \tag{6.2}$$

With H(x) being the total entropy for output x and H(x,y) being the conditional entropy computed from the conditional distribution of the output x (x, being GFP distribution and y being the sugar uptake rate). When doing so, it is important to adjust the number of bins used for computing the entropy. In our case, this number was set to 50 bins and leads to a precise computation of MI without increasing the computing power (Supplementary Figure 9). In order to explain the differences in controllability between the different systems investigated, we computed the mutual information (MI) between the environmental conditions and the activation of the target gene circuit for the $P_{araB} : GFP$ and $P_{glc3} : GFP$ systems. MI is a proxy derived from information theory 26,27 and involves the computation of the entropy of the cell population, as defined in the previous section. In short, MI tells us how much we can learn about the input (i.e., in our case the environmental stimulus used for entraining the cell population) from the output (i.e., in our case the distribution of GFP in cell population, the dispersion being quantified based on the entropy). Thus, in our case, MI is a proxy for information transfer efficiency between the inducer concentration and the cell population induction. The total entropy for each system was evaluated by summing up all the conditional probabilities obtained by exposing cell populations to different cultivation conditions (Supplementary Figure 10). MI was obtained by subtracting the time-dependent entropies to the total entropies recorded for each system. For the P araB :GFP system, MI is already relatively high in the

chemostat, leaving little room for improvement in the Segregostat (Supplementary Figure 10). It means that a reduction in entropy between the two cultivation modes has to be expected when the amount of information conveyed in chemostat condition is low (e.g., when cells cultivated in chemostat do not sense the inducer and, accordingly, do not activate the corresponding gene circuit). This is exactly what happened during the chemostat culture of the P glc3 ::GFP system (Supplementary Figure 10d). MI analysis had pointed out that there was still room for additional reduction in entropy (Supplementary Figure 10), and this was observed in Segregostat where glucose pulses reduced the average entropy (Supplementary Figure 10).

Supplementary Tables

SI table 1. Sequences and primers used in this work.

Primer name	Sequence (5'-3')
Fw_sgRNA_20N	agctagctcagtcctaggtataatactagtCGCCCGCAGGATATTCGTCG gttttagagctagaaatagcaagttaaaa
Fw_fragment1	gttaaacgagtatcccggcagca
Rv_fragment1	cgtttcactccatccaaaaaaacgg
Fw_fragment2	tggagtgaaacgtgactgtataaaaccacagccaa
Rv_fragment2	catcggcctcgtagacggtaac

Table 6.1: Primer sequences to make $\Delta araBAD$ strain

SI table 2. Operating conditions used for cultivations

Strain	Process conditions	Control conditions
E. coli W3110 $P_{araB} : GFP$	D = 0.45 h^{-1} pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min^{-1} [Glucose feed] = 5 g/L [Arabinose feed]$_{\text{Chemostat}}$ = 1.5 g/L [Arabinose feed]$_{\text{Segregostat}}$ = 0 g/L	Chemostat: none Segregostat: Regulation threshold = 50 % cells below 1000 F.U. Regulation = Pulse 0.15 g arabinose
E. coli W3110 $P_{lacZ} : GFP$	D = 0.45 h^{-1} pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min^{-1} [Glucose feed] = 5 g/L [Lactose feed]$_{\text{Chemostat}}$ = 1.5 g/L [Lactose feed]$_{\text{Segregostat}}$ = 0 g/L	Chemostat: none Segregostat: Regulation threshold = 50 % cells below 1000 F.U. Regulation = Pulse 0.15 g Lactose
E. coli BL21 $pET28 : GFP$	D = 0.45 h^{-1} pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min^{-1}	Chemostat: none Segregostat: Regulation threshold = 50 % cells

	[Glucose feed] = 5 g/L $[\text{Lactose feed}]_{\text{Chemostat}}$ = 1.5 g/L $[\text{Lactose feed}]_{\text{Segregostat}}$ = 0 g/L	below 1000 F.U. Regulation = Pulse 0.5 g Lactose
E. coli W3110 $P_{bolA} : GFP$	D = 0.45 h^{-1} pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min^{-1} [Glucose feed] = 5 g/L	Chemostat: none Segregostat: Regulation threshold = 50 % cells above 2000 F.U. Regulation = Pulse 0.2 g glucose
S. cerevisiae $P_{glc3} : GFP$	D = 0.1 h^{-1} pH = 5 T = 30 °c Aeration = 1 L/min Agitation = 1000 min^{-1} [Glucose feed] = 5 g/L	Chemostat: none Segregostat: Regulation threshold = 50 % cells above 5000 F.U. Regulation = Pulse 0.2 g glucose
B. subtilis 168 $P_{spoIIE} : GFP$	D = 0.1 h^{-1} pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min^{-1} [Glucose feed] = 5 g/L	Chemostat: none Segregostat: Regulation threshold = 20 % cells above 1000 F.U. Regulation = Pulse 0.2 g glucose

Table 6.2: Operating conditions

SI table 3: Parameters used for running FlowStocKS simulations

Parameter	Description	Value	Unit	Source
$\mu_{\text{max\_glucose}}$	Maximal growth rate on glucose	0.54	h^{-1}	10
$Ks_{\text{max\_glucose}}$	Affinity constant for glucose	0.034	g/L	10
Y_{glucose}	Substrate to biomass yield	0.5	g/g	11
n	Hill coefficient	-2		This study
k	50 % probability switching concentration	0.05		This study
GFP_{prod}	Mean value of GFP production burst given a gamma discribution of scale 2 and growth inhibition	$1e^5$	fu	This study

KI	Growth inhibition	From 0 to $1e^{99}$	fu	This study
delay (τ)	Delay	0.4	h	This study

Supplementary figures

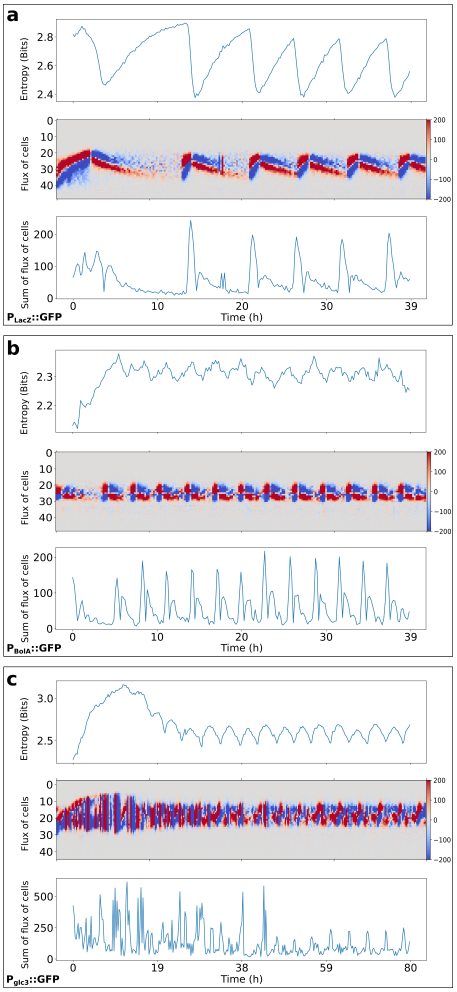

Figure 6.1: Computation of the entropy and flux of cells for different cellular systems. **a** $P_{lacZ} : GFP$ (*E. coli*), **b** $P_{bolA} : GFP$ (*E. coli*) and **c** $P_{glc3} : GFP$ (*S. cerevisiae*).

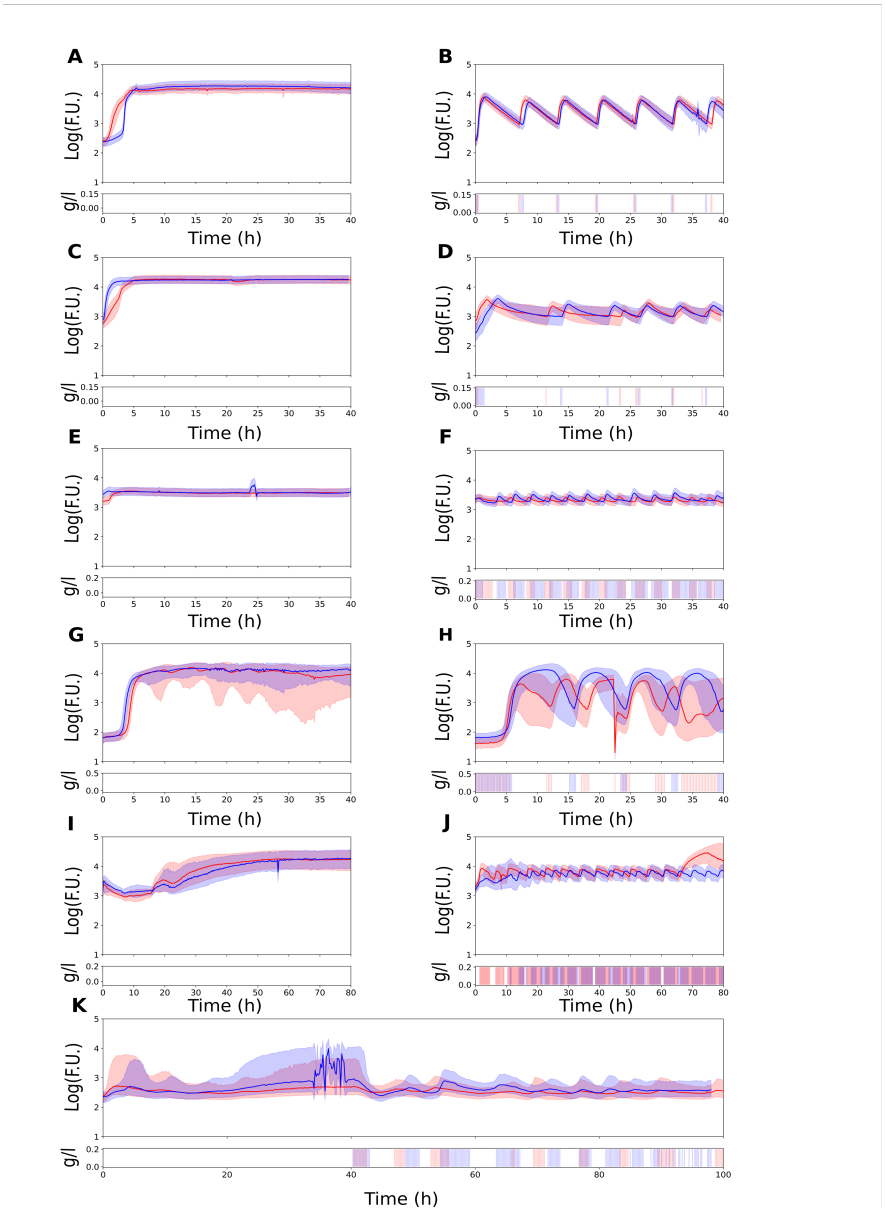

Figure 6.2: Assessment of the reproducibility of the Segregostat experiments. Evolution of the median fluorescence (plain line) and interquartile range (shadowed area around the plain line) between two biological replicates for **a** *E. coli* $P_{araB} : GFP$ in chemostat and **b** in Segregostat.**c** *E. coli* $P_{lacZ} : GFP$ in chemostat and **d** in Segregostat.**e** *E. coli* $P_{bolA} : GFP$ in chemostat and **f** in Segregostat.**g** *E. coli* $P_{lacZ} : GFP$P in chemostat and **h** in Segregostat.**i** *S. cerevisiae* $P_{glc3} : GFP$ in chemostat and **j** in Segregostat.**k** *B. subtilis* $P_{spoIIE} : GFP$P in chemostat for 40 h followed by Segregostat.

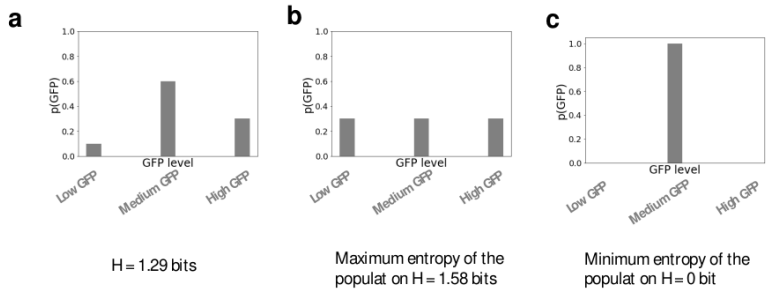

Figure 6.3: Examples of computation of the entropy H for a cell population clustered in three bins (computation according to Equation 1). Each bin corresponds to a subpopulation of cells with a givenfluorescence range i.e., low, medium or high, leading to different level of entropy H with **a** medium level of H, **b** maximum level of H for a three states system and **c** minimum level of entropy H.

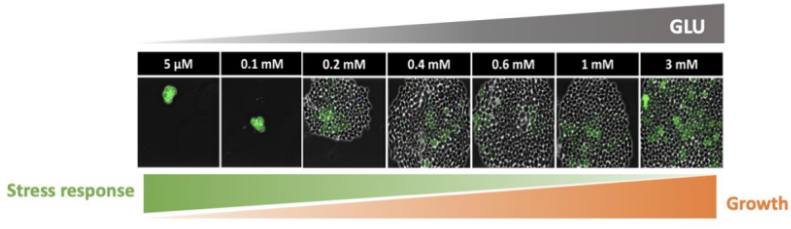

Figure 6.4: Pictures of yeast $P_{glc3} : GFP$ microcolonies cultivated in a MSCC device at different glucose concentrations

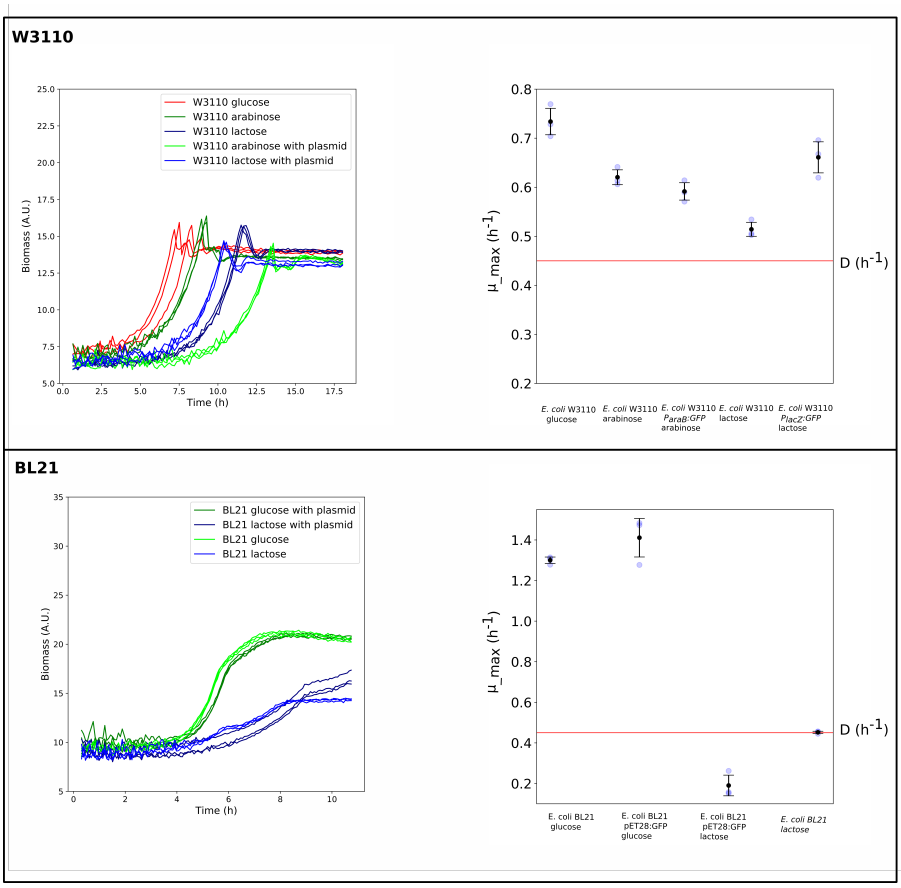

Figure 6.5: Experimental evaluation of the switching costs: Growth of different strains, the maximal growth rate of each triplicate is presented as a blue dot (plots on the right) and the means and standard deviations as a black dot with error bar. (Top) Determination of the maximal growth rate from Biolector cultivation with automated biomass measurement (n=3) for *E. coli* W3110, *E. coli* W3110 $P_{araB} : GFP$ and *E. coli* W3110 $P_{lacZ} : GFP$P on glucose, arabinose and lactose as carbon sources. (Bottom) Determination of the maximal growth rate from Biolector cultivation with automated biomass measurement (n=3) for *E. coli* BL21 (DE3) $pET28 : GFP$ on glucose and lactose and *E. coli* BL21 (DE3) on glucose and lactose. The red horizontal line is the dilution rate (0.45 h^{-1}) used for the continuous cultivations.

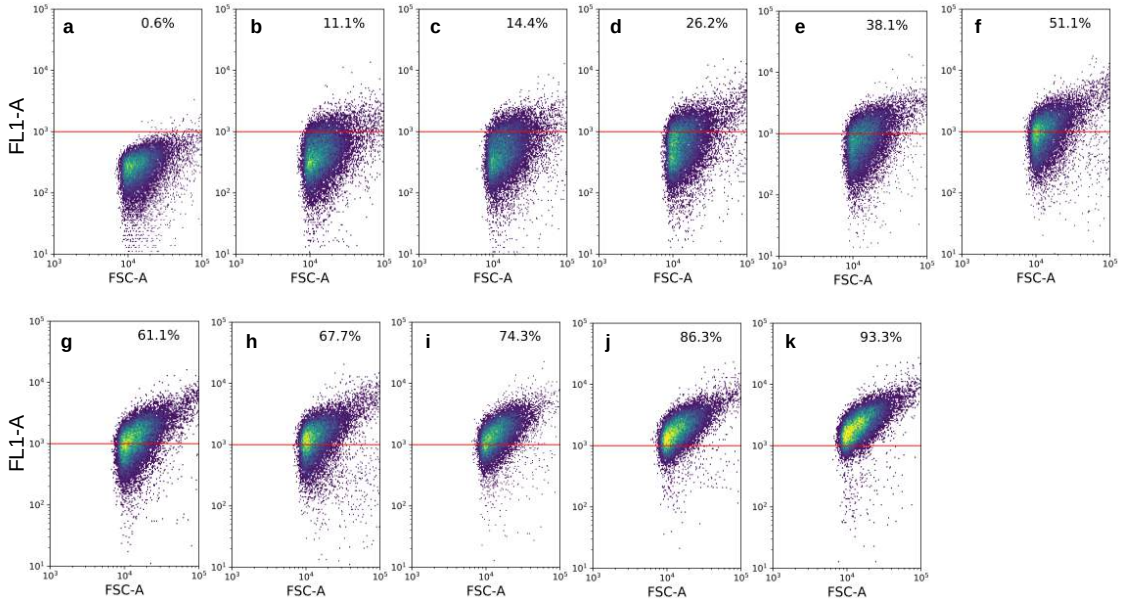

Figure 6.6: Analysis of the input-output relationship for *E. coli* $P_{araB} : GFP$: Scatter plots of cell size (FSC-A) versus GFP fluorescence (FL1-A) for *E. coli* W3110 Δ *araBAD* $P_{araB} : GFP$ exposed to arabinose concentrations of **a** 0, **b** 0.025, **c** 0.05, **d** 0.1, **e** 0.15, **f** 0.20, **g** 0.25, **h** 0.30, **i** 0.50, **j** 1.00 and **k** 2.00 g/L. The red line stands for the fluorescence threshold (i.e., 1000 F.U.) used for computing the GFP positive fraction of cells. Each FC analysis comprises 20,000 analyzed cells. The experiment was repeated twice with the similar results.

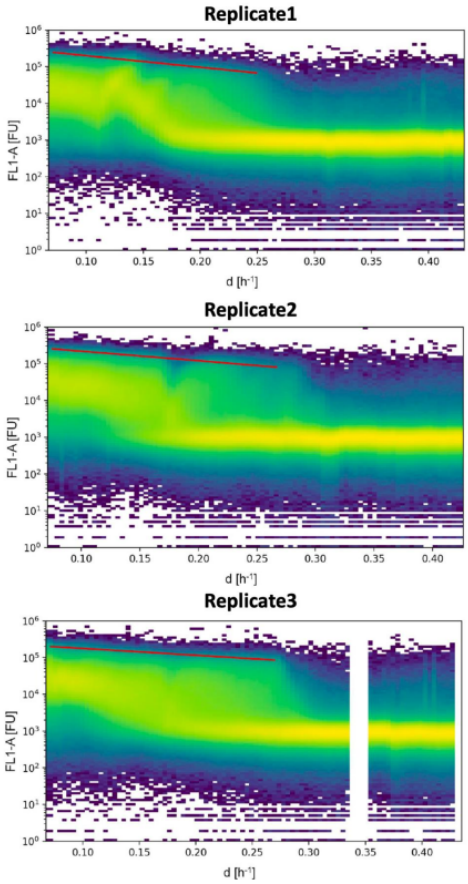

Figure 6.7: Analysis of the input-output relationship for *S. cerevisiae* $P_{glc3} : GFP$: A-stat experiment monitored based on automated FC. The GFP level distribution (FL1-A channel) was determined for each dilution rate (D). The red line highlights the progressive release of the stress response at the population level based on the deactivation of the $P_{glc3} : GFP$ reporter. Each FC analysis comprises 20,000 analyzed cells, the cultivation was performed in triplicate with similar outcomes (the three biological replicates are displayed).

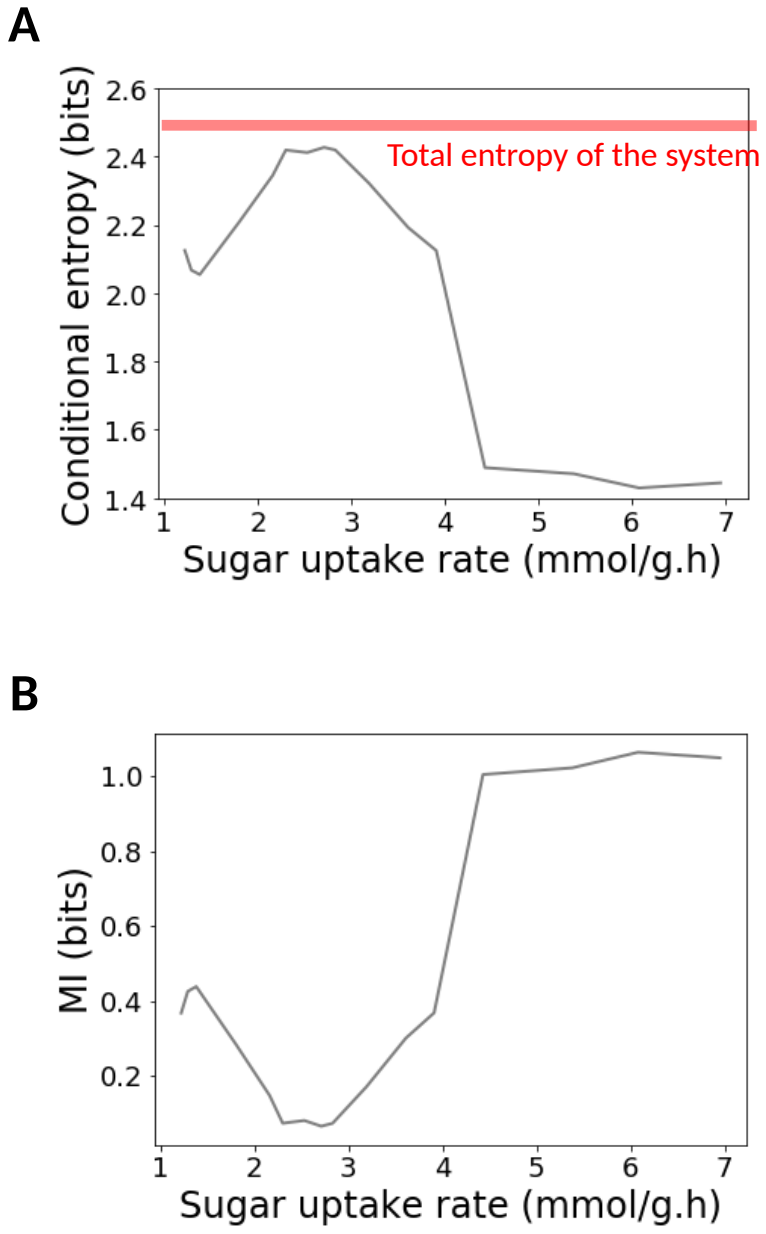

Figure 6.8: Evaluation of the mutual information (MI):**a** Evolution of the conditional entropy for the $P_{glc3} : GFP$ in *S. cerevisiae* exposed to different uptake rates. **b** MI can be deduced by subtracting the value of the total entropy of the system (2.49 bits in this case) by the corresponding conditional entropy.

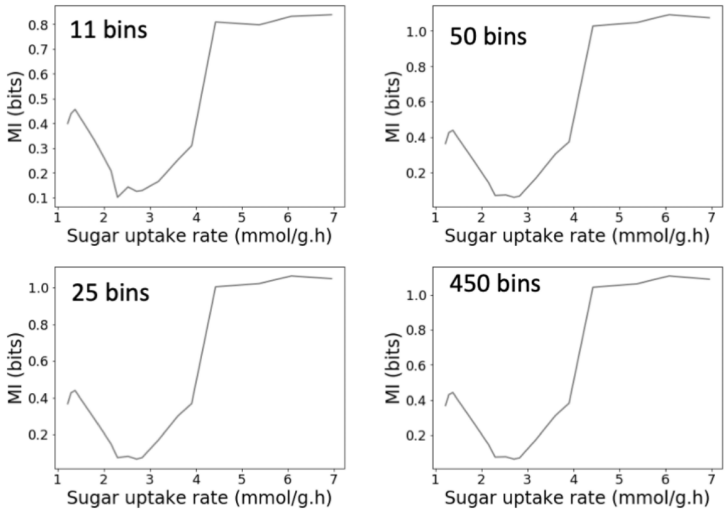

Figure 6.9: Impact of the binning procedure (number of bins considered) on the estimation of MI for the $P_{glc3} : GFP$ in *S. cerevisiae* (data extracte from A-stat experiments).

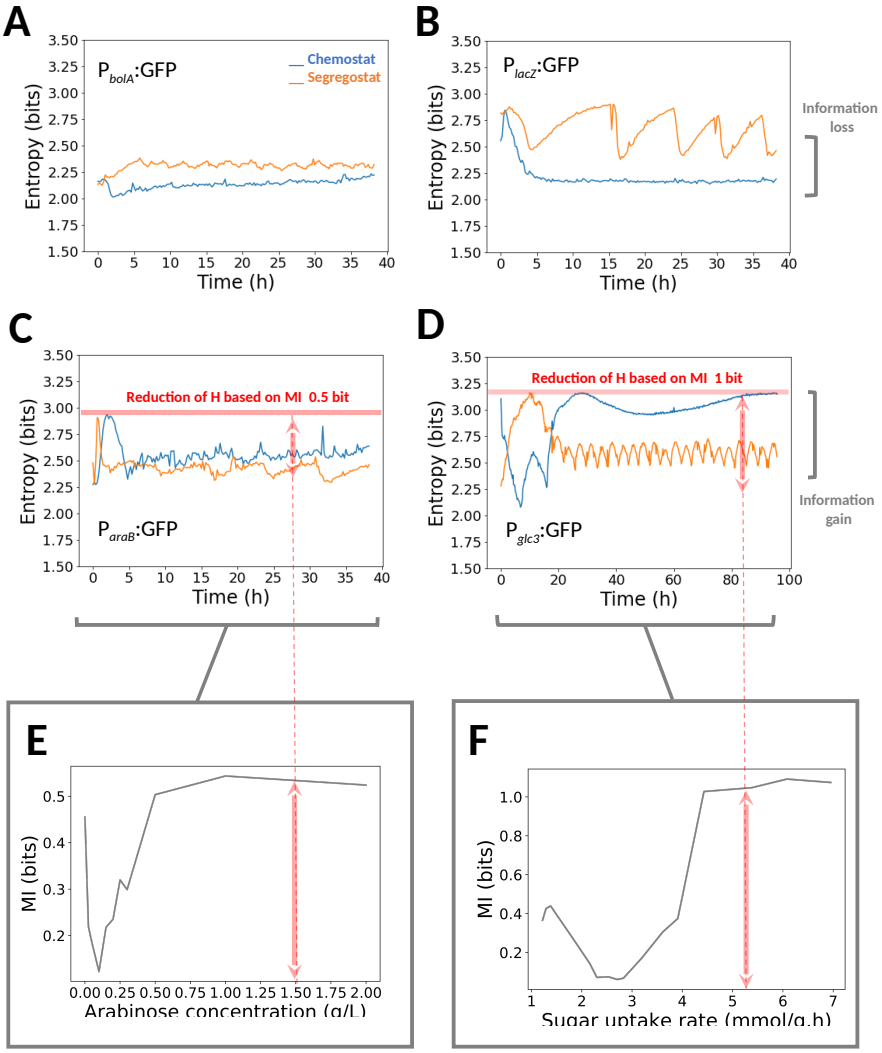

Figure 6.10: Comparative analysis of the entropy profile for the **a** $P_{bolA} : GFP$ system in *E. coli*. **b** $P_{lacZ} : GFP$ system in *E. coli*. **c** $P_{araB} : GFP$ system in *E. coli*. **d** $P_{glc3} : GFP$ system in *S. cerevisiae*. For the $P_{araB} : GFP$ and $P_{glc3} : GFP$ systems, the conditional probabilities, i.e., the GFP distribution of the population exposed at different environmental conditions, have been experimentally determined (Supplementary Note 2), allowing the computation of the mutual information (MI). **e** MI for the $P_{araB} : GFP$ system exposed to different arabinose concentrations. The MI distribution for the $P_{araB} : GFP$ system suggests that, at high arabinose concentration, a gain of information of approximately 0.5 bit has to be expected (the value has been reported by a red line on Figure 3C). **f** MI for the $P_{glc3} : GFP$ system exposed to different sugar uptake rates in a accelerostat cultivation device (Supplementary Figure 4). The MI distribution for the $P_{glc3} : GFP$ system suggests that, at high glucose concentration a gain of information of approximately 1 bit has to be expected (the value has been reported by ared line on Figure 3d)

2. Supplementary information chapter 3

Supplementary Notes

Supplementary Note 1: Model of stress phenotype expression and resolution

A chemostat is usually modelled with the following equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = (S_{in} - S)D - \frac{\mu X}{Y} \tag{6.3}$$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = (\mu - D)X \tag{6.4}$$

However, often, cells have a second state that can be induced when they are stressed, in which case their growth rate decreases. We can represent those two states in a chemostat with the following equations, if we consider that the stressed state is irreversible and cannot go back to non-stressed. As the time required to dilute the stress proteins is quite long, and made longer by the lower growth rate caused by stress, switching to a stress state is much faster and this is a reasonable assumption for a simplified approach. That is without considering the fact that, in a chemostat with non-stressed cells, the stressed-cells would be overtaken slowly and, if not renewed, are likely to simply be washed out before they can go back to their non-stressed state.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = (S_{in} - S)D - \frac{\mu X_1}{Y_{X1}} - \frac{k\mu X_2}{Y_{X2}} \tag{6.5}$$

$$\frac{\mathrm{d}X_1}{\mathrm{d}t} = (\mu - D)X_1 - H_1(S)X_1 \tag{6.6}$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = (k\mu - D)X_2 + H_1(S)X_1 \tag{6.7}$$

$$\mu = \mu_{max}\frac{S}{K_S + S} \tag{6.8}$$

$$\mu_{max,1} = \mu_{max} \tag{6.9}$$

$$\mu_{max,2} = k\mu_{max} \tag{6.10}$$

$$H_1(S) = \begin{cases} H & \text{if } S \leq K_I \\ 0 & \text{if } S > K_I \end{cases} \tag{6.11}$$

Where the variables are defined as:

D	the dilution rate.
k	the ratio of the growth rate of the slower phenotype to the growth rate of the faster phenotype.
K_I	the threshold substrate concentration for the switch.
K_S	the half-saturation constant of the Monod equation.
H	the switching rate of the faster to the slower phenotype.
$H_1(S)$	the switch function from the faster to the slower phenotype.
μ	the instantaneous growth rate of the faster phenotype.
μ_{max}	the maximum growth rate of the faster phenotype.
S	the substrate (usually glucose) concentration in the reactor.
S_{in}	the substrate concentration in the feed.
t	the time.
X_1	the concentration of the faster phenotype in the reactor.
X_2	the concentration of the slower phenotype in the reactor.
Y_{X1}	the yield of the faster phenotype.
Y_{X2}	the yield of the slower phenotype.

We can note the following conditions on the parameters based on the parameters:

- All variables are positive due to the constraints of reality and considering growing phenotypes.
- D, Y_{Xi}, μ_{max}, K_S, K_I and H are strictly positive.

Considering those equations for a chemostat, the question that now occurs is: What happens at steady state?

Due to the switch to a stress conditions, we have two kinds of steady-state to consider: ones where the substrate concentration is above the threshold K_I, and ones where the substrate concentration is below the threshold K_I. Separating those cases let us replace $H_1(S)$ by H or 0 in the equations, greatly simplifying the system. Note that by definition, the steady-state requires the residual substrate concentration S to be constant, therefore all the possible steady-state can be found by splitting the equations this way.

To jump to the solution of the steady-state, go to section 2.

When $S > K_I$

In this case, at steady-state (derivatives are null), there is no switching and we can replace equations 6.5 to 6.7 by the following:

$$0 = (S_{in} - S)D - \frac{\mu X_1}{Y_{X1}} - \frac{k\mu X_2}{Y_{X2}} \tag{6.12}$$

$$0 = (\mu - D)X_1 \tag{6.13}$$

$$0 = (k\mu - D)X_2 \tag{6.14}$$

According to the latter 2 equations, there are 4 possible cases:

1. $X_1 = 0$ and $X_2 = 0$.
2. $X_1 = 0$ and $D = k\mu$.
3. $X_2 = 0$ and $D = \mu$.
4. If $k = 1$, $D = \mu$

These are the expected steady-states and show already known results: in the absence of switching, at steady-state, either the two populations are washed out or not initially present (case 1); and there is an unstable case where if both phenotypes have exactly equal growth rates, they can both survive, in any proportion as they behave as a single population and this whole model becomes irrelevant (case 4).

For the sake of this argument, we can consider $k \neq 1$ as that is a trivial and uninteresting case.

Both phenotypes cannot coexist in this case if $k \neq 1$, as one existing, in the absence of switching, imposes a specific dilution rate that implies, at steady-state, the other no being there.

The last case, a wash-out of both/either phenotype, will always happen for a population if $(\mu - D) = 0$ is impossible. To show when it is forced, finding out when it is possible is the easiest way.

Then, re-writing equations6.12 to6.14 for a steady-state with a single phenotype:

$$(S_{in} - S)D = \frac{X_i}{Y_{Xi}}\,\mu_{max,i}\frac{S}{K_S + S} \tag{6.15}$$

$$\mu_{max,i}\frac{S}{K_S + S} = D \tag{6.16}$$

where i indicates either phenotype, whichever is present since we established that in this case only one can be present.

Since Y_{Xi}, D and K_S are strictly positive, we can deduce:

$$\mu_{max,i}\frac{S}{K_S + S} = D \qquad \Rightarrow \qquad \frac{D}{\mu_{max,i}} = \frac{S}{K_S + S} \tag{6.17}$$

$$S_{in} - S = \frac{X_i}{Y_{Xi}} \qquad \Rightarrow \qquad S = S_{in} - \frac{X_i}{Y_{Xi}} \tag{6.18}$$

Therefore:

$$\begin{aligned}
\frac{D}{\mu_{max,i}} &= \frac{S}{K_S + S} \\
&= 1 - \frac{K_S}{K_S + S} \\
\Leftrightarrow \quad \frac{K_S}{K_S + S} &= 1 - \frac{D}{\mu_{max,i}} \\
&= \frac{\mu_{max,i} - D}{\mu_{max,i}} \\
\Leftrightarrow \quad K_S + S &= \frac{\mu_{max,i} K_S}{\mu_{max,i} - D} \\
\Leftrightarrow \quad S &= \frac{\mu_{max,i} K_S}{\mu_{max,i} - D} - K_S \\
&= \frac{D K_S}{\mu_{max,i} - D}
\end{aligned} \tag{6.19}$$

Plugging equation 6.19 into equation 6.18 and solving for X_i gives:

$$\begin{aligned}
S_{in} - \frac{D K_S}{\mu_{max,i} - D} &= \frac{X_i}{Y_{Xi}} \\
X_i &= Y_{Xi} \left(S_{in} - \frac{D K_S}{\mu_{max,i} - D} \right)
\end{aligned} \tag{6.20}$$

Thus, eventually, from equations 6.17, 6.18, 6.19, and 6.20 we get:

$$\begin{cases}
D &= \mu_{max,i} \frac{S}{K_S + S} \\
S &= \frac{D K_S}{\mu_{max,i} - D} = S_{in} - \frac{X_i}{Y_{Xi}} \\
X_i &= Y_{Xi} \left(S_{in} - \frac{D K_S}{\mu_{max,i} - D} \right) = Y_{Xi}(S_{in} - S)
\end{cases} \tag{6.21}$$

at steady states, for cases 2 and 3. And since $S \geq 0$ and $X_i \geq 0$, for this to be possible, we need:

- $\mu_{max,i} > D$ for $S \geq 0$. When a given phenotype is initially present, a wash-out and no biomass is leftover at steady-state (it becomes case 1) if the maximum growth rate of the phenotype is lower than the dilution rate (as expected).
- $S_{in} > {}^{D K_S}/_{(\mu_{max,i} - D)}$ for $X_i > 0$. This can be rearranged (as per standard chemostat equations) as $D < \mu_{max,i} {}^{S_{in}}/_{(K_S + S_{in})} < \mu_{max,i}$, dilution rate above which there is no glucose consumption at equilibrium as there is not enough substrate to sustain the population.

Those equations and deductions are only valid when $S > K_I$, or:

$$\frac{DK_S}{\mu_{max,i} - D} > K_I \tag{6.22}$$

Firstly, we established that, when there is no wash-out, and there is biomass at steady-state, $\mu_{max,i} > D$ is required. Therefore, the left hand-side of the inequality is always positive.

We can examine when transform the inequality to examine when it is possible or:

$$\begin{aligned}
& \frac{DK_S}{\mu_{max,i} - D} > K_I \\
\Leftrightarrow \quad & DK_S > K_I(\mu_{max,i} - D) && \text{since: } \mu_{max,i} > D \\
\Leftrightarrow \quad & DK_S + DK_I > K_I\mu_{max,i} \\
\Leftrightarrow \quad & D(K_S + K_I) > K_I\mu_{max,i} \\
\Leftrightarrow \quad & D > \frac{K_I\mu_{max,i}}{K_S + K_I} && \text{since: } K_S, K_I > 0
\end{aligned} \tag{6.23}$$

Therefore, steady-state without switching can be reached when $\mu_{max,i}{}^{S_{in}}/_{(S_{in} + K_S)} > D > \mu_{max,i}{}^{K_I}/_{(K_S + K_I)}$ for each phenotype if the three parameters allow (although, for the stressed phenotype, we are assuming that there is no switch back to the first one, which is probably wrong if waiting long enough in this high substrate concentration case).

As a summary of this section, we have for each case described at the start:

1. Wash out of all the cells initially present, when $D > \mu_{max,i}\frac{S_{in}}{(S_{in}+K_S)}$ for all i phenotypes non-zero in the initial conditions.

2. a steady state with only the second phenotype present, which could happen if all cells switched to the second phenotype, and never switched back, maybe in the case of differentiation:

$$\begin{cases}
X_1 = 0 \\
D = k\mu_{max}\ {}^{S}/_{K_S + S} \\
S = {}^{DK_S}/_{k\mu_{max} - D} = S_{in} - {}^{X_2}/_{Y_{X2}} \\
X_2 = Y_{X2}(S_{in} - S) \\
0 \leq S \leq S_{in} \\
k\mu_{max}{}^{S_{in}}/_{(S_{in} + K_S)} > D > k\mu_{max}{}^{K_I}/_{(K_S + K_I)}
\end{cases} \tag{6.24}$$

3. A steady state with only the first phenotype present:

$$\begin{cases} X_2 = 0 \\ D = \mu_{max}\, {}^{S}/_{K_S + S} \\ S = {}^{DK_S}/_{(\mu_{max} - D)} = S_{in} - {}^{X_1}/_{Y_{X1}} \\ X_1 = Y_{X1}(S_{in} - S) \\ 0 \leq S \leq S_{in} \\ \mu_{max}{}^{S_{in}}/_{(S_{in} + K_S)} > D > \mu_{max}{}^{K_I}/_{(K_S + K_I)} \end{cases} \tag{6.25}$$

When $S \leq K_I$

A similar reasoning can be applied to the alternative case, where the switching always occurs.

At steady-state, we get:

$$0 = (S_{in} - S)D - \frac{\mu X_1}{Y_{X1}} - \frac{k\mu X_2}{Y_{X2}} \tag{6.26}$$

$$0 = (\mu - D - H)X_1 \tag{6.27}$$

$$0 = (k\mu - D)X_2 + HX_1 \tag{6.28}$$

The equations can be solved easily when $X_1 = 0$ and $X_2 = 0$, as this means there is no biomass. Notably, this case is incompatible with the assumption that there is switching regardless as $S_{in} = S$ in that case, and in a bioreactor $S_{in} > K_I$, so we come back to the previous section.

In case $X_1 = 0$, the equations become identical to the previous section. The only difference is the condition of existence derived at the end, which becomes $D < \mu_{max,i}{}^{K_I}/_{K_S + K_I}$ instead. This completes all cases where only the second phenotype is present with a region where it is actually likely that all cells stay in the second phenotype, as the substrate concentration would be too low to permit switching back to the first phenotype.

Thus the only case of interest to discuss here is when both X_1 and X_2 are non-zero.

Re-arranging equations 6.26 to 6.28 with those assumptions:

$$(S_{in} - S)D = \frac{\mu X_1}{Y_{X1}} - \frac{k\mu X_2}{Y_{X2}} \tag{6.29}$$

$$\mu = \mu_{max}\frac{S}{K_S + S} = D + H \tag{6.30}$$

$$(D - k\mu)X_2 = HX_1 \tag{6.31}$$

Interestingly, equation 6.30 implies that the first phenotype, through D, H, μ_{max} and K_S, fully determines the substrate concentration, with only H differing from a normal chemostat and no direct influence of the phenotype ratio in the culture.

Based on equation 6.30, we can deduce:

$$\begin{aligned} & S = \frac{(K_S + S)(D + H)}{\mu_{max}} \\ \Leftrightarrow \quad & S\left(1 - \frac{D+H}{\mu_{max}}\right) = K_S \frac{D+H}{\mu_{max}} \\ & S\,\frac{\mu_{max} - D - H}{\mu_{max}} = \\ \Leftrightarrow \quad & S = \frac{K_S(D+H)}{\mu_{max} - D - H} \end{aligned} \tag{6.32}$$

Equation 6.31 can easily be re-arranged into:

$$X_1 = \frac{D - k\mu}{H} X_2 \Leftrightarrow X_2 = \frac{H}{D - k\mu} X_1 \tag{6.33}$$

From there, we can plug into the remaining equation 6.29:

$$\begin{aligned} & (S_{in} - S)\,D = \frac{\mu X_1}{Y_{X1}} + \frac{k\mu}{Y_{X2}}\,\frac{H}{D - k\mu}\,X_2 \\ \Leftrightarrow \quad & X_1 = \frac{(S_{in} - S)\,D}{\frac{\mu}{Y_{X1}} + \frac{k\mu H}{Y_{X2}\,(D - \mu k)}} \end{aligned} \tag{6.34}$$

$$\begin{aligned} \text{and} \quad & (S_{in} - S)\,D = \frac{(D - k\mu)\mu}{H Y_{X1}}\,X_2 + \frac{k\mu X_2}{Y_{X2}} \\ \Leftrightarrow \quad & X_2 = \frac{(S_{in} - S)\,D}{\mu\left(\frac{(D - k\mu)}{H Y_{X1}} + \frac{k}{Y_{X2}}\right)} \end{aligned} \tag{6.35}$$

We can thus write the solution for this case of the steady-state as:

$$\begin{cases} S & = \frac{K_S(D+H)}{\mu_{max} - D - H} \\ \mu & = \mu_{max}\frac{S}{K_S + S} \;=\; D + H \\ X_1 & = \frac{(S_{in} - S)\,D}{\mu/Y_{X1} + k\mu H/Y_{X2}\,(D - \mu k)} \\ X_2 & = \frac{(S_{in} - S)\,D}{\mu\,((D - k\mu)/(H Y_{X1}) + k/Y_{X2})} \\ \frac{X_1}{X_2} & = \frac{D - k\mu}{H} \end{cases} \tag{6.36}$$

For this solution to be possible, we need:

- $S > 0$
- $X_1 > 0$

- $X_2 > 0$

Since $X_1 = X_2{}^{(D\,-\,k\mu)}/_{H}$, we need for both to be positive:

$$D > k\mu \tag{6.37}$$

Since $D + H = \mu$, we actually need:

$$\frac{D}{D+H} > k \tag{6.38}$$

Note that this condition can only be true if $k < 1$, as the left hand side is always smaller than 1. Thus, if the induced phenotype were to grow faster than the initial one, it would be impossible for the two phenotypes to coexist in the reactor. In this case, we are considering a stress phenotype, so $k < 1$ is a reasonable assumption.

In other musings, this condition means at its origin that the dilution rate will be faster than the instantaneous growth rate of the second phenotype.

The condition can be easily re-arranged in terms of D as well:

$$\frac{D}{D+H} > k \qquad \Leftrightarrow \qquad D > \frac{kH}{1-k} \tag{6.39}$$

This ensures that both X_1 and X_2 have the same sign. To ensure they are positive, we need:

$$S_{in} > S \tag{6.40}$$

which is always true if cells grow.

Finally, we need for $S \geq 0$:

$$D < \mu_{max} - H \tag{6.41}$$

The latter also implies that, for both phenotypes to coexist,

$$\mu_{max} > H \tag{6.42}$$

is required.

Those conditions, combined with the one required by this section, $S \leq K_I$, are the only conditions required for the existence of a steady-state with both phenotypes present.

The latter condition can be developed:

$$
\begin{aligned}
& \frac{K_S(D+H)}{\mu_{max}-(D+H)} < K_I \\
\Leftrightarrow \quad & K_S(D+H) < K_I\,(\mu_{max}-(D+H)) \qquad \text{"<" sign unchanged since (6.41)} \\
\Leftrightarrow \quad & DK_S + HK_S < \mu_{max}K_I - DK_I - HK_I \\
\Leftrightarrow \quad & D(K_S+K_I) < \mu_{max}K_I - H(K_S+K_I) \\
\Leftrightarrow \quad & D < \frac{\mu_{max}K_I}{K_S+K_I} - H \qquad (6.43)
\end{aligned}
$$

Similarly to condition 6.42, we can deduce that

$$\mu_{max}\frac{K_I}{K_S+K_I} > H \qquad (6.44)$$

from the previous condition.

Note that condition 6.43 is stronger than condition 6.41, as:

$$
\begin{aligned}
& \mu_{max} - H > \frac{\mu_{max}K_I}{K_S+K_I} - H \\
\Leftrightarrow \quad & \mu_{max} > \mu_{max}\,\frac{K_I}{K_S+K_I} \qquad (6.45)
\end{aligned}
$$

Inequation 6.45 is always true, as $\frac{K_I}{(K_S+K_I)}$ is smaller than one, so condition 6.41 is irrelevant. The deduced condition on the relationship is similiarly stronger in (6.44) than in (6.42).

In summary, the steady-state with both phenotypes present is possible with the non-trivial conditions:

- ${}^{(\mu_{max}K_I)}\!/\!_{(K_S+K_I)} - H > D > {}^{kH}\!/\!_{(1-k)}$
- ${}^{\mu_{max}K_I}\!/\!_{(K_S+K_I)} > H$

Solution of the steady-state

The different states and conditions in which they exist can be gathered as seen on Figure

There is a dilution rate region where no steady state exist. In that region, any attempt to reach steady-state when $S > K_I$ causes the biomass to increase and S to decrease. When $S < K_I$, the cells start switching and substrate consumption decreases, cause S to increase above K_I again. Since cells do not respond instantaneously to change, and in this region where S is close to K_I small changes in concentration can cause large changes in responses, the cells may oscillate between the two states.

$$\frac{\mathrm{d}(SV)}{\mathrm{d}t} = F(t)\ S_{in} - \frac{\mu XV}{Y} \tag{6.46}$$

$$\frac{\mathrm{d}(XV)}{\mathrm{d}t} = \mu XV \tag{6.47}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F(t) \tag{6.48}$$

Limitations of the model and additional notes

There are a few additional things that can be added on this model, in no particular order:

- While it was derived for a chemostat, it can be applied to an exponential fed-batch reactor as well (see section 2). For a fed-batch with less than exponential feeding, the effective growth rate decreases over time. In other words, it would not be a question of whether the region without equilibrium is reached, but when. Then the second question is whether that will cause oscillations or not, which will depend on the exact biological system, and how long the cells stay in the problematic region.

- This model predicts that there is a region where there is no steady-state when there are sharp transitions in cell metabolism, such as with stress induction, or induction of burdensome production with lactose. Oscillations may appear thanks to the delay in response time of the cells, and very small changes in substrate concentration causing large changes in cell response. However, the switch is not a perfect step function, and the delay will be shorter or longer depending on the cells. Further, a constant switching speed may not be a good approximation, at least when close to the switching threshold. All in all, oscillations may appear, but they may be dampened, or the system may stabilize quickly in an intermediate state that this model is incapable of representing.

- In the demonstration, we considered what happens when $K_I < S$ and $K_I > S$. When $K_I = S$, the details of what happen in the model depend on how $H_1(S)$ was defined (here, $H_1(K_I) = H$, therefore the $K_I = S$ would be the limit of the $K_I < S$ case). In practice, however, reality is not completely discontinuous, and this is the edge of an undefined region regardless, so the mathematics of this particular fictional point do not really matter, only the limit region between the cases does.

- We consider here that both phenotypes are growing, no matter how slowly. If the slower phenotype were to have a death rate instead ($k < 0$), several things would happen:

  - No region would exist where only the second phenotype can survive, as even a batch culture would eventually have no cells left. This is expressed by condition 6.39 being always true.

  - Nothing else would change, as in no place in section 2 is there any division by k for any condition. We can therefore expect exactly the same dynamics and region of uncertainty as with $k \geq 0$.

- For a case such as the induction of burden with a second carbon source (such as the induction of BL21 with lactose (co-feeding of glucose and lactose), where oscillations were observed as well [1]), considering co-consumption of carbon sources, the inequations are inverted such that equations 6.21 is valid only when $D < {\mu_{max}}^{K_I}/_{K_S + K_I}$, and equations 6.36 is valid only when $\mu_{max} - H > D > {\mu_{max}}^{K_I}/_{K_S + K_I} - H$ and $D > H^k/_{(1-k)}$. Further, the case of only the second phenotype being present should be stable between $k{\mu_{max}}^{K_I}/_{(K_S + K_I)}$ and $k\mu_{max}$. Therefore, a steady-state (or more than one) always exists at least as long as the induced phenotype is not washed out, and oscillations may only be an intermediate state that will eventually stabilize. Accounting for those oscillations would require investigation of the exact dynamics of the system of interest, and probably require taking into account carbon catabolite repression (When induced, cells stop growing, this increases both glucose and lactose concentrations. Increasing glucose concentration would inhibit induction by carbon catabolite repression, from which similar dynamics to the stress case would follow).

Exponential fed-batch

The traditional equations for a fed-batch reactor are:

$$\frac{\mathrm{d}(SV)}{\mathrm{d}t} = F(t)\ S_{in} - \frac{\mu XV}{Y} \tag{6.49}$$

$$\frac{\mathrm{d}(XV)}{\mathrm{d}t} = \mu XV \tag{6.50}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F(t) \tag{6.51}$$

Where $F(t)$ is the flow rate added to the reactor, and D is the expected growth rate of the cells. In case of exponential feeding, V and ${dv}/{d_t}$ are defined as:

$$V(t) = C_0 + C_1 e^{Dt} \tag{6.52}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = C_1 D e^{Dt} \tag{6.53}$$

Where C_0 is the integration constant, so that $V(0) = V_0$ the initial volume, and C_1 modulates the state of the exponential. Traditionally, C_1 is defined as V_0 (see [2]), which makes $C_0 = 0$ and simplifies the equations significantly. In [3], however, the authors report a common value of $C_1 = {V_0 X_0}/_{S_{in} Y}$ for *B. subtilis*. We can note that, if the initial glucose concentration is equal to S_{in}, then we can approximate $C_1 \approx V_0$. Regardless of the exactitude of that approximation, for the fed-batch to make sense,

we need $V_0 > C_0 > 0$. Therefore, since we are here interested in steady-states only, we can t_0 the original time of the reactor such that $V(t_0) = V_0$, which means that C_0 can arbitrarily set to 0 and the steady-states will be identical.

The equations for the fed-batch can be adapted like for the chemostat:

$$\frac{\mathrm{d}(SV)}{\mathrm{d}t} = F(t)\, S_{in} - \frac{\mu X_1 V}{Y_{X1}} - \frac{k\mu X_2 V}{Y_{X2}} \tag{6.54}$$

$$\frac{\mathrm{d}(X_1 V)}{\mathrm{d}t} = V X_1 \mu - H_1(S) X_1 V \tag{6.55}$$

$$\frac{\mathrm{d}(X_2 V)}{\mathrm{d}t} = V X_2 k\mu + H_1(S) X_1 V \tag{6.56}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F(t) = C_1 D e^{Dt} \tag{6.57}$$

$$V(t) = C_1 e^{Dt} \tag{6.58}$$

Those equations can be set back into concentrations by using the chain rule for $A = S, X_1, X_2$ in the following equation:

$$\frac{\mathrm{d}(AV)}{\mathrm{d}t} = \frac{\mathrm{d}A}{\mathrm{d}t} V + A \frac{\mathrm{d}V}{\mathrm{d}t} = V \frac{\mathrm{d}A}{\mathrm{d}t} + A F(t) \tag{6.59}$$

The equations become:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{F(t)}{V}(S_{in} - S) - \frac{\mu X_1}{Y_{X1}} - \frac{k\mu X_2}{Y_{X2}} \tag{6.60}$$

$$\frac{\mathrm{d}X_1}{\mathrm{d}t} = (\mu - \frac{F(t)}{V} - H_1(S)) X_1 \tag{6.61}$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = (k\mu - \frac{F(t)}{V}) X_2 + H_1(S) X_1 \tag{6.62}$$

Since $^{F(t)}/_{V} = D$, these equation become exactly identical to a chemostat, and the same conclusions can be drawn.

Supplementary Note 2: Extended materials and methods

Strain and growth media

All experiments were performed in minimal mineral media containing (in g/l): K_2HPO_4: 14.6; $NaH_2PO_4 \cdot 2H_2O$: 3.6; Na_2SO_4: 2; $(NH_4)_2SO_4$: 2.47; NH_4Cl: 0.5; $(NH_4)_2H - \text{citrate}$: 1; glucose: 5, thiamine: 0.01, tryptophan: 0.05. The medium is supplemented with a trace element solution totaling 11 ml/l assembled from the following solutions (in g/l), 3/11 of $FeCl_3 \cdot 6H_2O$: 16.7 , 3/11 of EDTA: 20.1, 2/11 of $MgSO_4$: 120 and 3/11 of a metallic trace element solution. The metallic trace element solution contains (in g/L): $CaCl_2 \cdot 2H_2O$: 0.74; $ZnSO_4 \cdot 7H_2O$: 0.18; $MnSO_4 \cdot H_2O$: 0.1; $CuSO_4 \cdot 5H_2O$: 0.1,

$CoSO_4 \cdot 7H_2O$: 0.21. Both the trace element solution and the amino acids were filter sterilized (0.22 μm). To monitor the switching to the sporulating phenotype, *Bacillus subtilis* 168 was modified to express *gfpmut2* under the control of the *SpoIIE* promoter. The strain, *B. subtilis* 168 P_{spoIIE} : $gfpmut2$, was built by *AmyE* chromosomal integration of a casette containing the promoter, $gfpmut2$ and the selection maker, kanamycin.

Cultivation conditions and flow cytometry analysis

Precultures were started from single colonies picked from a LB plate and grown overnight. To avoid starting bioreactor cultivations with sporulating populations, the precultures were done in overfilled shake flasks (20 % of filled volume) without baffles to reach oxygen limitation before glucose. All bioreactor cultivations were performed at 37°c, pH 7 and aeration at 1 VVM. For growth and switching parameters determination, experiments were performed in DASBox mini - bioreactors (Eppendorf) with a stirring speed of 400 rpm and a cultivation volume of 160 ml. Continuous cultivations were performed in Bionet F1 bioreactor with a cultivation volume of 1 l at an agitation speed of 1000 rpm. The transition from batch to continuous was triggered once oxygen rose, signaling the transition from exponential to stationary phase.

Bioreactor cultivations were monitored with a custom made sampling device (the Segregostat) that automatically draws a sample from the reactor and dilutes it with PBS before flow cytometry (FC) analysis (BD Accuri, C6 for the parameter determination and C6+ for the continuous cultures). A total of 40,000 cells were analyzed in each sample where the FL1-A channel is used to visualize the GFP content.

Growth and switching parameter determination

A triplicate batch cultivation in the DASBox was performed to estimate μ_{max} and switching (K_I, H) parameters of our strain. To this end, each bioreactor was inoculated with a OD of roughly 0.08 and samples were collected throughout the batch to assess the OD and the residual glucose concentration. The residual glucose concentration was determined using a D-glucose enzymatic assay kit from Megazyme in 98-well plates with a plate reader. During the batch, FC analysis was performed with our custom platform, the Segregostat, to observe when and at what speed cells were sporulating. The K_S was defined based on the analysis the residual glucose concentration at steady state at a high dilution rates (where cells are not sporulating) during the chemostats with varying dilution rates.

The following parameters were measured or estimated:

- H: 1.1 h^{-1} $\rightarrow$ Average from 0.83, 1.2, and 1.3 with r_2 = 0.98, 0.99 and 0.99 respectively (computed by linear regression over the time there is switching)

- K_S: 0.14 g/L $\rightarrow$ Based on the concentration at D=0.32 h^{-1}, measured as 0.2761 and 0.2432 g/L. From $D = mu_{max}{}^S/_{(K+S)}$, we get $K_S = mu_{max}{}^S/_D - S$

- K_I: [0.01,0.1] g/L → Explained below
- μ_{max}: 0.49 h^{-1} → Average from 0.52, 0.47 and 0.48 with $r_2 > 0.99$ for all three (computed by linear regression over the time of exponential growth)
- k: 0 → Sporulating cells are considered non-growing

To determine the start of sporulation, we considered the time at which the measured optical density dropped. That drop likely corresponds to a change in morphology of the cells. For that stretch of time, samples were taken more often than the automated flow cytometry, providing a better time resolution, in addition to having no absolutely no cross-contamination (using the segregostat with our DasGip in triplicate setup, a few percent leftover from each sample may be measured with the next sample), which is relevant when looking at very low percentages of induction. Based on these OD measurements, we considered the induction threshold was passed during the preceding measured time interval. This corresponds to 0.013 to 0.097 g/L, 0.009 to 0.068 g/L and 0.010 to 0.282 g/L for reactors 1, 2 and 3 respectively. Thus we estimated K_I to be in the range of 0.01 to 0.1 g/L, although considering the cells need time to switch, K_I's order of magnitude is likely the higher end of the range.

Based on these measurements, we could compute the unstable range of dilution rates for *B. subtilis*. The lower end of the range is at $D = 0$ $^{-1}$, as $k = 0$ and $(\mu_{max} K_I)/(K_S + K_I) - H < 0$ ($\mu_{max} < H$). The upper end depends on K_I. Based on the estimated range of K_I, it should be between $(\mu_{max} K_I)/(K_S + K_I) \in [0.03, 0.2]$ h^{-1}. The first replicate of the experiment showcasing a range of dilution rates showed the start of the instability at $D = 0.17$ h^{-1}, which would correspond to a K_I of 0.075 g/L.

3. Supplementary information chapter 4

Supplementary Notes

Supplementary Note 1: Growth production trade-off determination

To compute the growth rate of the strain as a function of the induction strength (fluorescence content) from FC population snapshots, two assumptions are needed:

- GFP is not degraded and thus its concentration in the cells only decreases though growth
- The expression system driving the production of GFP is tight, in other words, during the relaxation phase cells do not produce GFP at all

Both assumptions are reasonable, since: A) GFP is a protein, no degradation tags have been linked to it and the time scale associated to this analysis (only a few hours)

is rather small. B) The T7 expression appears really tight as in the batch phase where no lactose is provided, no fluorescence above the auto-fluorescence of a non GFP containing *E. coli* is observed.

From there, the growth rate associated to different level of GFP can be retrieved from multiple FC analysis gathered during the relaxation phase, i.e., when the population fluorescence is decreasing after the increase that followed induction. With these FC files, the fluorescence of the cells can be sorted and characterized with quartiles. These quartiles are collected over a few FC analysis to grasp how their value decreases over time. If the GFP production did not impact the growth, the value of each quartile would drop at a same rate, the population growth rate. Instead of this, what we observe in the case of *E. coli* BL21 (DE3), is that the greater the quartile is, the slower its value decreases over time. A log scale transformation of the quartile values make a simple linear regression sufficient to retrieve the growth rate (the slope of the line) associated with each quartile (Figure 6.11).

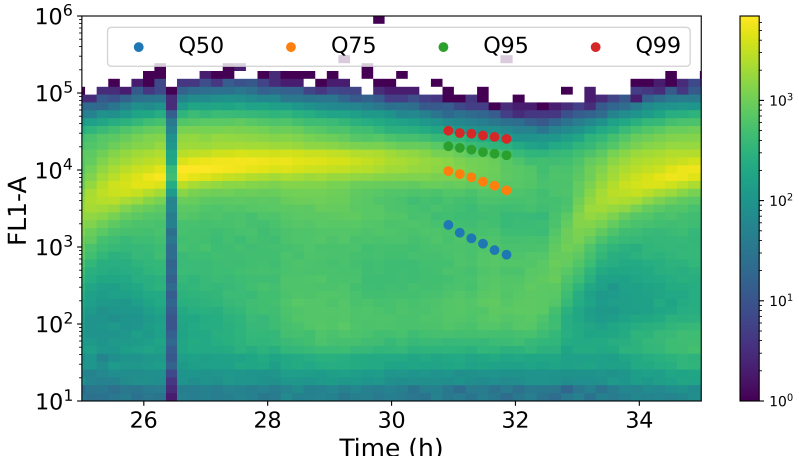

Figure 6.11: Overlapping the quartile values on the density plot during a relaxation phase shows that the slope of the quartile value decrease is inversely proportional to the quartile value.

Supplementary Figures

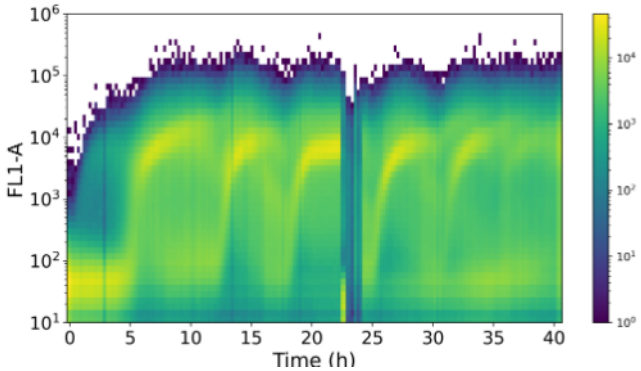

Figure 6.12: Time scatter plot of a replicate Segregostat cultivation of *E. coli* BL21 where lactose is added as pulse (0.5 g) once 50% of the population exhibits a fluorescence below 1000 fluorescence units (in FL1-A channel)

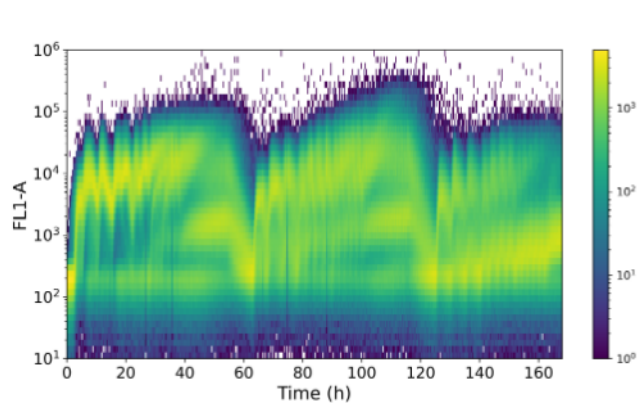

Figure 6.13: Time scatter plot of a replicate of the cultivation where lactose is pulsed at an increasingly high frequency three times in a row with 5 retention times in between.

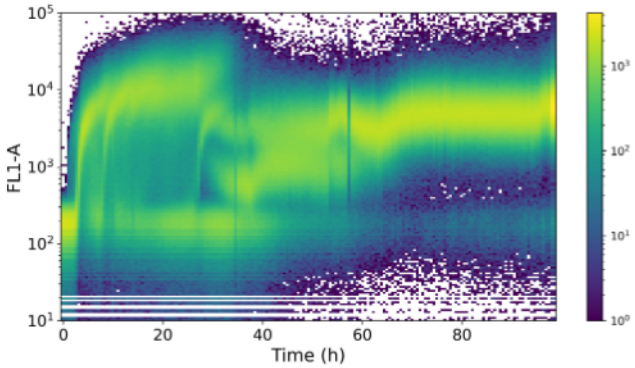

Figure 6.14: Time scatter plot of a replicate of a chemostat where the cultivation starts with glucose as the main carbon source, followed with arabinose and then xylose.

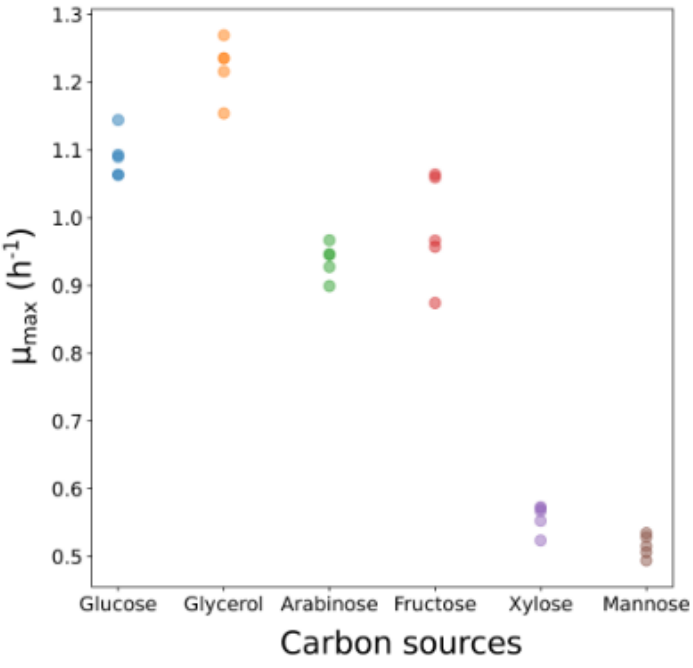

Figure 6.15: Maximum growth rate of *E. coli* BL21 (DE3) on multiple carbon sources (n=5). The mean maximum growth rate on glucose, glycerol, arabinose, fructose, xylose and mannose are respectively 1.09, 1.22, 0.93, 0.98, 0.55 and 0.51 h^{-1}.