Optimal phenotypic switching model for bacterial populations under fluctuating environmental conditions

Auteur : Vandenbroucke, Vincent
Promoteur(s) : Delvigne, Frank
Faculté : Faculté des Sciences appliquées
Diplôme : Master en ingénieur civil biomédical, à finalité spécialisée
Année académique : 2020-2021
URI/URL : http://hdl.handle.net/2268.2/11602

Optimal phenotypic switching model for bacterial populations under fluctuating environmental conditions

Master's thesis carried out to obtain
the degree of Master of Science in Biomedical Engineering
by Vandenbroucke Vincent

Supervisor: Frank Delvigne

University of Liège - School of Engineering and Computer Science
Academic year 2020-2021

Acknowledgements

Before beginning, I would like to thank the people who helped me for this master thesis. This includes everyone on the MIPI team at the laboratory, and in particular: my supervisor, Frank Delvigne, without whom this entire thesis would not have been possible; Juan Andres Martínez Alvarez, who helped me a lot in giving structure to the work, among other things; and Lucas Henrion, who produced the experimental data and with whom I shared ideas on the project. I would also like to thank Dominique Toye, who among other things gave a very good class on bioreactors without which this work would have been much harder, and Abhyudai Singh, with whom I discussed the project a few times. Finally, I thank my friends and family, who supported me in this endeavour.

Abstract

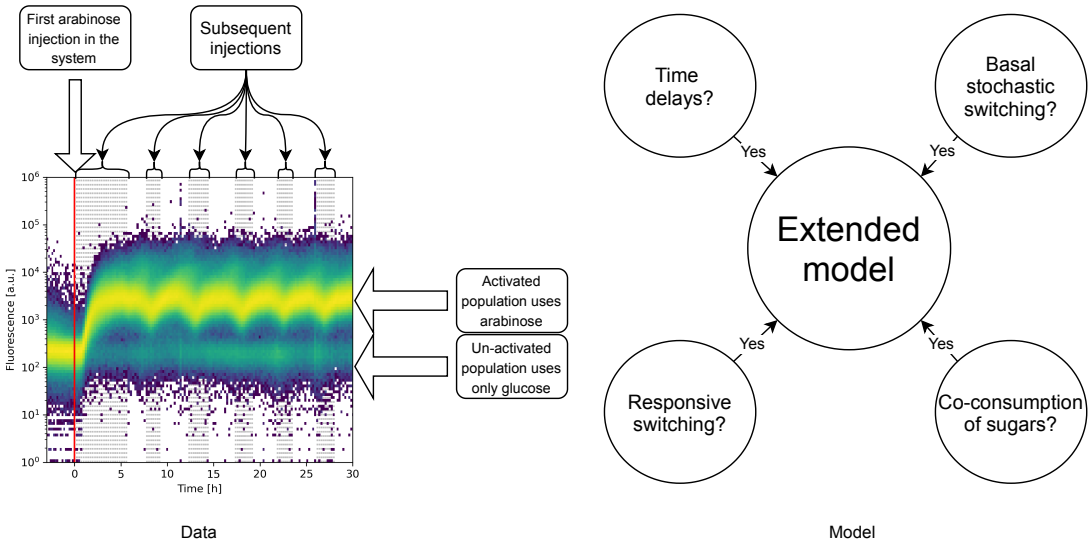
In a fluctuating environment, cell populations need to adapt to survive. There are several ways to achieve it, but in general cells take benefit from biological noise to switch to alternative phenotypes, leading to improved fitness in the new environment. This is a known phenomenon that gives rise to phenotypically heterogeneous populations, but the dynamics of the adaptation are still unclear.

Understanding those dynamics is key if one wants to control gene expression in bio-processes in general, but the literature on ways to reproduce population behaviours quantitatively *in silico* is still nascent; and if a few models are available, they have not been validated based on experimental data.

Therefore, in this work, an existing model was applied and improved to better fit experimental data obtained based on the use of high-throughput, automated, flow cytometry. In particular, a simple model derived from the general model developed by Thattai and Van Oudenaarden [1] was applied to data representing populations of *E. coli* growing under continuous cultivation conditions. The fluctuating environment for the population was based on a constant feed of glucose, but varying amounts of arabinose in the medium.

The existing model used basal stochastic switching to drive the population diversification. It was found insufficient to reproduce the observed data, and several possible new contributions to the model were therefore devised. Each one was considered independently of the others, thanks to a succession of models, each taking into account a different set of contributions, and fitted as well as possible to the data.

Eventually, it was found that several of those contributions could indeed improve the model for this particular case. Moreover, this work also sheds new lights on two important points that are likely applicable in most systems with heterogeneous cell populations: the growth rates of the different phenotypes need to be considered in details if the population is to be modelled correctly, and representing the switching mechanism adequately is likely often easier and more exact by simulating single cells with a stochastic behaviour, rather than using differential equations.



Contents

List of Figures

List of Tables

1 Introduction

Phenotypic heterogeneity in fluctuating environments can be recognized as a key strategy leading to improved fitness among microbial populations. Living cells, when presented with a new environment, may adapt by changing their phenotype to better fit said environment. However, not all the cells in any given population actually adapt to the new conditions. Indeed, cells have evolved to adapt to an environment that may change again, and a part of the population effectively anticipates the next disruption by staying in their old phenotype. Thus a heterogeneous population arises where a part of the cells is adapted to the current environment, and the rest of the population is adapted to the next change; the population is said to hedge its bets. [2, 3]

Understanding the dynamics of populations that hedge their bets, and how individual cells choose to switch from one phenotype to another is key if one wants to control cell population dynamics, gene expression, and bio-processes in general. To study this phenomenon, scientists usually either observe the population dynamics in a controlled, fluctuating environment, or they model *in silico* what could happen to a population hedging its bets. [1, 4–6]

In this work, an attempt at conciliating the two approaches is made, to determine whether the current population models can apply to a case study where *Escherichia coli* adapts to the presence of different sugars in a bioreactor.

2 Experimental data

Like most engineering works, this master thesis is based on observations. In this case, those observations are the experimental results obtained by Lucas Henrion at the laboratory where the thesis was made.

The experiments all followed the same principle: 1 L of medium was first incubated with an *E. coli* culture in a perfectly mixed bioreactor for a few hours to obtain a substantial amount of biomass in the system; then a continuous phase started, where fresh medium was injected and culture medium removed, at a dilution rate of 0.5 h^{-1}.

The device itself is illustrated in FIGURE 1 and allowed two kinds of regulation for the experiments: in "chemostat" mode, the main feed contains both glucose and arabinose, while in "segregostat" mode it contained only glucose, and the arabinose was added independently.

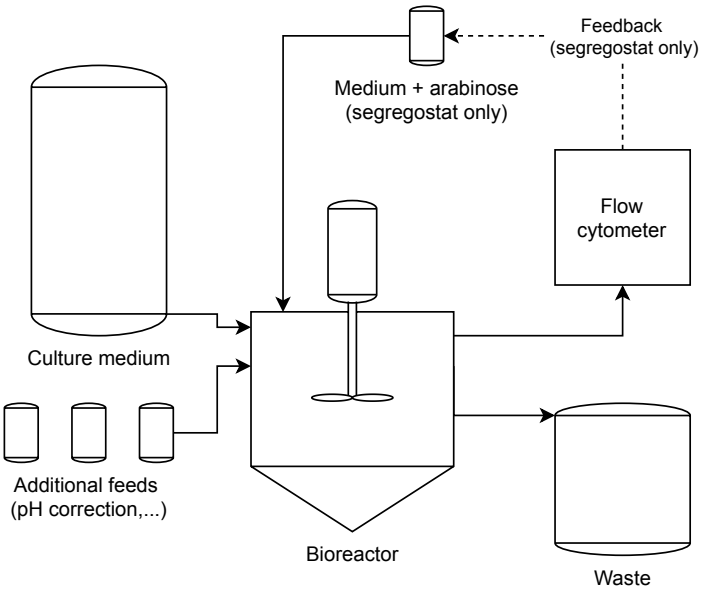


Figure 1: Scheme of the bioreactor setup. In the chemostat, there is no arabinose tank, and the culture medium contains arabinose. In the segregostat, the flow cytometry measurements determine when arabinose is added in the system.

To provide quantitative measurement regarding the bacterial population, an interface in the bioreactor periodically samples some medium for a flow cytometry analysis. In this particular case, the flow cytometry was mainly used to measure the fluorescence of cells that expressed (or not) GFP[1]. The cells in question were *E. coli* cells modified to express GFP when the arabinose operon[2] is induced, therefore the proportion of cells with an active arabinose metabolism can effectively be monitored over time.

In addition to flow cytometry, individual samples could be taken to measure additional parameters in the medium, like the sugar concentrations.

[1] GFP is the green fluorescent protein. When cells express it, and blue light is shown on them, they will emit green light. This lets instruments like the flow cytometer detect GFP content in cells.

[2] An operon is a set of genes under the control of a single regulating region in the DNA (the promoter). In this case the arabinose operon regulates the arabinose metabolism.

2.1 Details on chemostat experiments

In chemostat experiments, glucose and arabinose were present in the feed with concentrations of, respectively, 5 g/L and 1.5 g/L. This allowed to study the way cells used arabinose at a single cell resolution during the transition from a medium containing only glucose, in the batch phase, to a medium with a mix of both sugars.

In one chemostat experiment, the sugar concentrations during the continuous phase were measured through liquid chromatography (HPLC), but no glucose or arabinose was detected in the medium, showing that there was less than 40 mg/L of glucose and less than 20 mg/L of arabinose in the medium, as these were the measurement limits, including right at the start and five hours after the start of the continuous phase.

The flow cytometry results of a representative chemostat experiment are shown in FIGURE 2. The three graphs represent different measurements for a single experiment; with time=0 h being the beginning of the continuous phase. In the first graph (2.a) the evolutions of the cell concentration in the medium, the cell concentration of GFP-positive cells (later called GFP+ cells), which metabolise arabinose, and the concentration of GFP-negative cells (later called GFP- cells), which are considered not to use arabinose, are represented. The delimitation between both type of cells is made according to a threshold of 1000 (a.u.) in the fluorescence measurement (this threshold is the one used by Nguyen et al. [6] on the same equipment). The cell concentration is the one measured by the flow cytometer in "events/µL", and the short term trend is informative. However, the decline of the total concentration over time is not confirmed by OD_{600} measurements, and was observed for other cultures where the biomass is known to increase over time. It can therefore be considered as an artefact of the instruments and cannot be used as a quantitative measurement.

The second graph (2.b) represents the proportions of GFP-positive cells over time. Unlike the first one, this graph is quite reproducible between experiments. The proportion of GFP-positive shows two points of note in the transition between the batch phase and the steady-state: firstly, the GFP+ proportion does not monotonously grow from 0 to the steady-state equilibrium, at around 50–60%. Instead, there is a maximum at ∼80% before stabilisation. And secondly, there is a local minimum after the maximum, which appears after about 20 h in the continuous phase. This local minimum appears in most chemostat experiments, but not all of them.

Finally, in the third graph (2.c), a heat map of the raw fluorescence measurements by the flow cytometer is shown. While this graph cannot be used as-is to compute anything, the way cells vary from one state to another, with an initial massive switch, can give clues as to how the cells behave. It is important to note that the colour represents the number of events detected at a given fluorescence level for a given time, knowing that at each time point 20000 events were analysed, thus the colour effectively represents the proportion of GFP+ cells over time, and is unrelated to the overall cell concentration. Of note on this graph is that nearly all cells switch from GFP- to GFP+ after about one hours, all at the same time. This causes the maximum that appears in the second graph, but the subsequent decrease in proportion is not marked by a significant part of the population visibly switching from GFP+ to GFP-, the transition is much more diffuse.

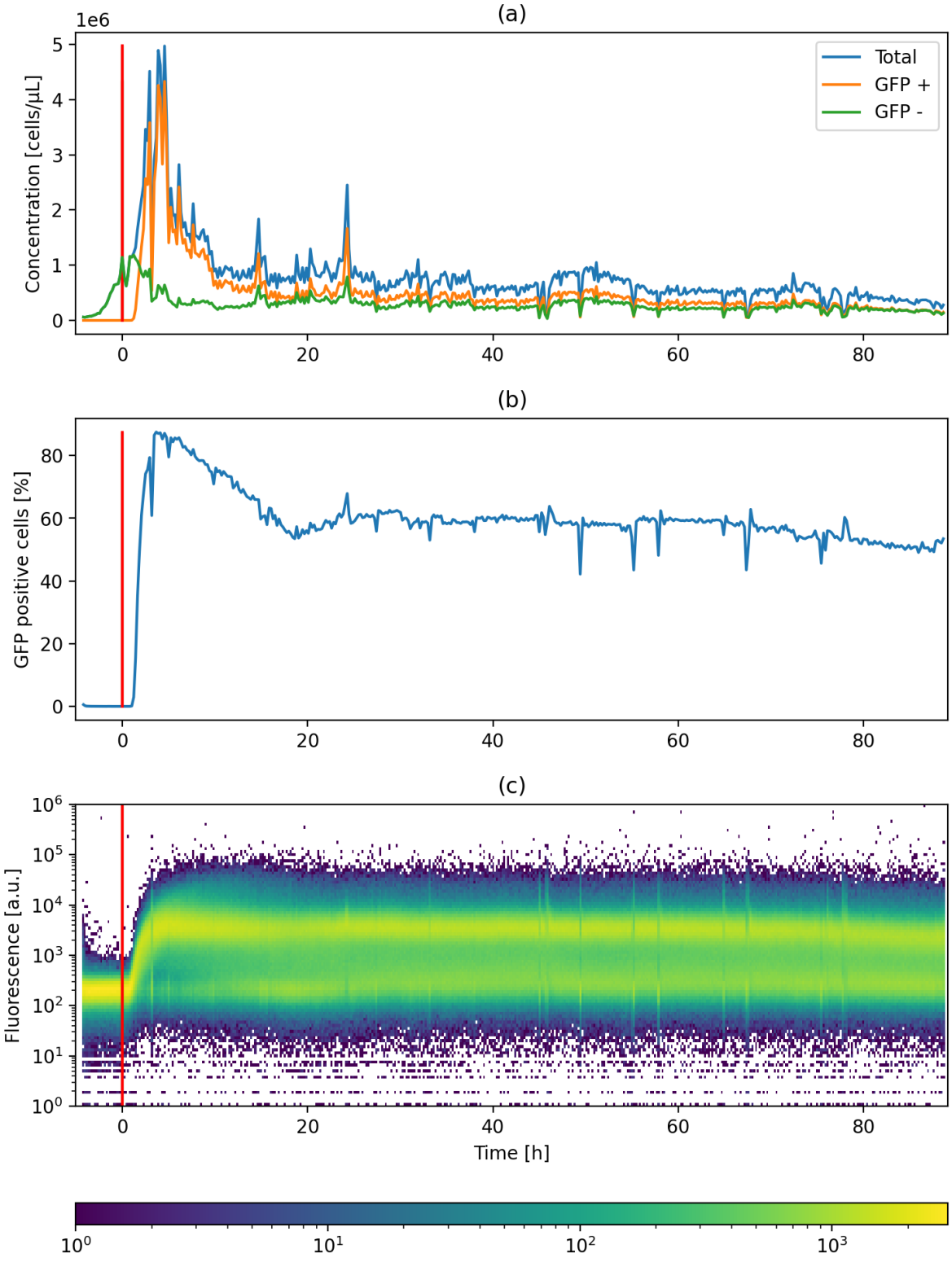


Figure 2: Summary of a representative chemostat experiment. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a fluorescence threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The red line represents the beginning time of the continuous cultivation phase.

2.2 Details on segregostat experiments

The second kind of experimental data comes from segregostat experiments. In the segregostat, the feed was simply growth medium with glucose. In order to induce the arabinose operon, arabinose was periodically injected in the medium, in short pulses of 0.5 g in about 7 mL of medium. That volume is much smaller than the bioreactor and can be considered negligible. In addition, the regulation of the dilution rate is made by controlling both the flow rate of medium coming in and the volume of the reactor: the pump extracting used medium from the reactor removes everything above a certain level of liquid in the reactor, and is a bit faster than the input pump. Therefore, repeated pulses of arabinose spaced several minutes apart do not change the volume of the bioreactor noticeably.

Those pulses of arabinose were not made randomly. In these experiments, a new threshold of GFP fluorescence was defined at 2500 a.u., which is much higher than the GFP+ threshold, and lets one distinguish cells that have the highest GFP contents. The flow cytometer made measurements about every twelve minutes, and, if there was less than 50% of cells more fluorescent than the given threshold, a pulse of arabinose was added to the system. This led to experiments with long series of arabinose pulses, followed by long periods without stimulation.

An example of a representative segregostat experiment can be found in Figure 3. It is presented under the same format as the chemostat experiments, and the grey lines show the periods during which pulses of arabinose were added.

Several things can be observed in those graphs. The main one is that, in Figure 3.b, the proportion of GFP+ cells can be kept close to $\sim$90%, unlike in the chemostat, where the proportion eventually reaches $\sim$ 60%. This proportion can be easily explained by two general observations: on the one hand, on Figure 3.a, the concentration of GFP+ cells spikes whenever arabinose is added, which increases their relative abundance; on the other hand, on Figure 3.c, the heat map shows that GFP+ cells in the segregostat tend to switch back to using only glucose after a while, at which point arabinose is added to the system to compensate. Interestingly, unlike for the chemostat, the tendency to switch back seems much sharper.

Just as with the chemostat, for the concentration measurements (Figure 3.a), the trend of cell concentrations during a single peak can most likely be trusted, but the overall decrease of the peak height over time cannot, and the measurements cannot be considered quantitative.

As for the sugar concentrations, they were followed about once per hour during two segments of a segregostat experiment. The results are shown in Figure 4. When the proportion of cells above the defined threshold is below the 50% line, arabinose is injected in the system, which causes the arabinose concentration to rise. Glucose is not often detected, but it does seem to decrease when cells start to consume the massive amount of arabinose.

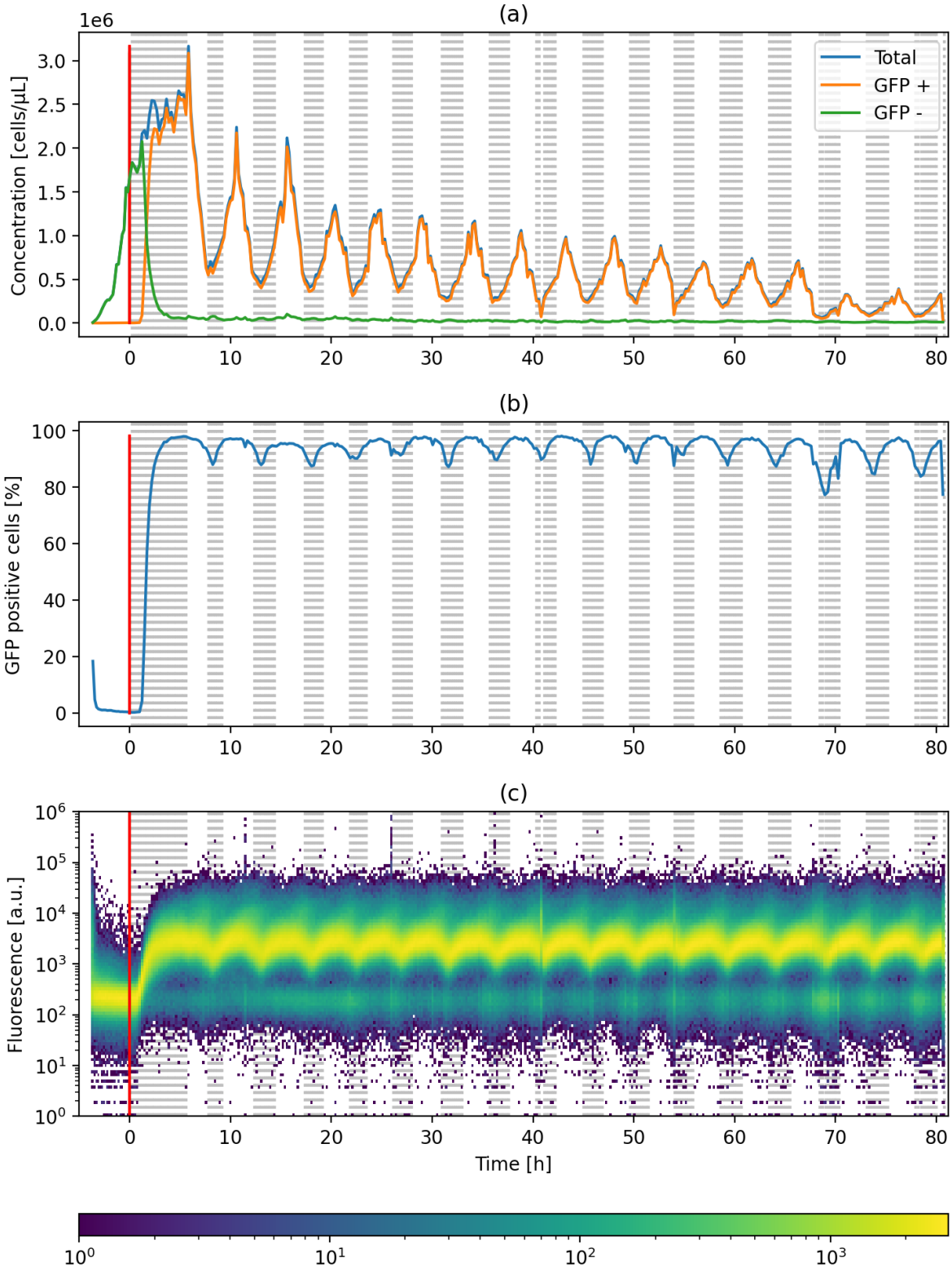


Figure 3: Summary of a representative segregostat experiment. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a fluorescence threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The dashed grey areas show the periods when pulses of arabinose were added into the system. The red line represents the beginning time of the continuous cultivation phase in the experiment.

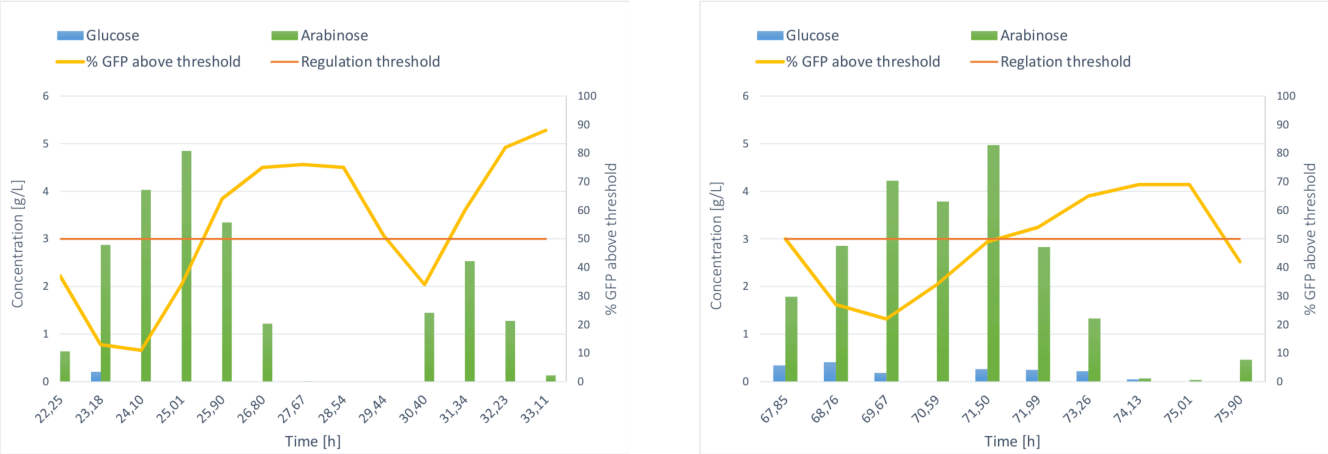


Figure 4: Sugar concentration data for two time series in a single segregostat experiment. The horizontal line indicates the threshold below which the proportion of high fluorescence cells need to be for the system to inject arabinose, and the yellow line indicates the proportion of high fluorescence cells.

2.3 Additional experimental data

In addition to simple chemostat and segregostat experiments, a few experiments were "combined". In those experiments, the first 50 h of culture follow one regulation scheme (i.e. chemostat or segregostat), and it is then switched to the alternative in the following 50 h. The results help show that the cells do not adapt significantly to the regulation used, as behaviours after the switch in regulation are similar to the ones observed directly after the batch phase.

One experiment was made based on the same principle as the segregostat, but the frequency of arabinose pulses was controlled directly rather than based on a GFP threshold. It was mostly unused in the practical part of this work, as it was made much later, but it will be used in the Results and discussion section.

A summary of all the experiments performed is provided in Annex A.2.

3 Problem statement

Those experiments provide relevant data for what happens to a population in a fluctuating environment. Up to now, in the lab, however, the descriptions of the behaviour of the population in these experiments were either qualitative or descriptive (see for example the work by Nguyen et al. [6]).

In order for the data to be useful, however, for a potential use in industrial bioreactors, or simply in order to understand exactly what is happening to the population, a descriptive analysis is not sufficient. Indeed, while a qualitative and descriptive analysis lets one know what happens, in some details, it is not sufficient to discover the underlying complexities of the system and to predict how it will behave in new, untested environments. And predicting what happens in the reactor could provide ways to control it, through a predictive control of the regulation rather than, as in the segregostat, a reactive control.

To predict how the cell population would behave in new environment, a detailed understanding of the kinetics of the cell switching mechanisms, as well as their growth, is required. Considering what is known about said kinetics, as will be shown in the next two subsections, aiming for a model that explains everything perfectly is, at best, a long shot.

Therefore, in this work, a first attempt at a model is made. It aims at proving or disproving the most likely theories regarding the way cells behave in a fluctuating environment. Ideally, the model should quantitatively and qualitatively reproduce the behaviours of the experiments previously discussed. In practice, determining how close the model taking into account the most likely theories can get to reality will provide new insights into how the cells behave, and possibly some useful results for the control of a bioreactor. It will also provide directions as to what information is missing to form a better picture of what is happening.

3.1 State of the art

Before delving into the specifics of how exactly the model could work, it is of interest to take a look at the state of the art regarding models of cell population diversification dynamics, on the one hand, and the behaviour of the arabinose operon, on the other hand.

Firstly, there are different types of models that can be used to represent a microbial population. For the purpose of this work, one division seems particularly interesting to use: on the one hand, there are some mostly empirical models, like the Monod model[7]. On the other hand, there are the mostly mechanistic models, which attempt to use what is known of the chemistry inside the cell to reproduce its behaviour. Obviously, since everything is not known about the cells, but some things are quite obvious, no model can be considered entirely empirical or mechanistic, but the distinction stays quite useful as the approach used to design the model can be quite different.

Empirical models are mostly designed to explain the behaviour of entire populations. Those are called “unstructured” models[8], the best known of which is the Monod model[7, 8]. They work mostly like models designed for chemical equations, in that the growth of the cell is considered exponential (because the rate of cell division is usually independent from cell concentration, the time to double the cell concentration is usually constant as well), and, on top of that exponential growth, the changes in

concentration caused by the sugar consumption and by the reactor in which the cells are placed (e.g. dilution) are added to compute the evolution of the cell and sugar concentrations over time.

The Monod equation, as well as most classical unstructured models, however, only describe a single population with a single phenotype. In order to consider several phenotypes, Thattai and Van Oudenaarden [1] and Kussell and Leibler [4] have added a new component to the simple exponential growth of cells: a switching rate. In particular, in the model from Thattai and Van Oudenaarden [1], two phenotypes and two environments are considered. Each phenotype is considered "optimised" for one of the two environment, meaning it will grow faster in said environment. In addition, for each environment, cells in one phenotypic state have a given switching rate to the other state. The switching rates will, presumably, favour cells switching from their "unadapted" to their "optimised" state. In this way, a bistable population (a population presenting two distinct phenotypes) can be represented over time. Mathematically, Thattai and Van Oudenaarden [1] represent the system, given one environment, as:

$$\frac{\mathrm{d}}{\mathrm{d}t}n_0 = \gamma_0 n_0 - k_1 n_0 + k_0 n_1 \tag{1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}n_1 = \gamma_1 n_1 + k_1 n_0 - k_0 n_1, \tag{2}$$

where n_0 and n_1 are the number of cells in each states, γ_0 and γ_1 are their respective growth rates, and k_0 and k_1 are the basal stochastic switching rates, in the current environment. The basal stochastic switching, later called basal switching, represents the hypothesis that in a given environment cells can switch between states due to biological noise.

This model is very general and can, in that it can technically represent any switching population between two different states depending on the conditions. However, it does not allow for non-binary states of the environment: it is perfect to express something like the absence of one carbon source or the other (and that was a first), but cannot respond to their concentration. It also considers a maximum of two states, although it can be quite easily generalized to having more than. As a matter of fact, Kussell and Leibler [4] consider an arbitrary number of states and the same amount of phenotypes. This can both represent more complex environments and represent, if not exact concentrations, at least variations of sugar concentrations, for example, as three different environments could be a "low", a "medium" and a "high" concentration of sugar. In addition, it may be important to keep in mind that Kussell and Leibler [4] do not consider the kinetics of the transition, only the steady state. In both cases, the purpose was to give a basic framework to work with, and no direct link to experimental was made. To my knowledge, neither these models nor any variation were ever applied to the arabinose operon in *E. coli*.

Mechanistic models, by definition, seek to use what is known of the bacteria to model their behaviour. They are based on the relevant genes present, the concentration of various proteins and other molecules, kinetics of gene translation and transcription, and so on. As a significant portion of the molecules of interest is present in only a few or sometimes a single copy (e.g. DNA), the concentrations cannot be approximated by differential equations in the cells. Instead, what is often used to model the concentrations inside cells is stochastic modelling. [9, 10]

Stochastic modelling includes probabilities in the equations. Since there is only a few molecules of each type, there is only a small chance that the same reaction would happen at the same time in two

different cells. Therefore, by taking into account those probabilities, and simulating a large number of cells, it is possible to reproduce biological noise, and in particular phenotypic heterogeneity. This heterogeneity can arise from something as simple as a positive feedback loop in the gene network: if by chance, a few proteins of one phenotypic state are produced, then they will cause more of those proteins to be produced, thus pushing the cell towards that state, while another cell that did not produce those proteins by chance would stay in its original phenotypic state. [2]

In particular, for the arabinose operon, there are indications that the induction is produced by a single-molecule event. A similar mechanism was evidenced by Choi et al. [11], for lactose, where the DNA is held in a loop by a repressor. It was shown that the complete dissociation of the repressor from the DNA would unravel the DNA loop, allowing for the transcription of the genes that were blocked. Unravelling that DNA loop is much shorter than reforming it, as the repressor in charge of binding two distant sites on the DNA would bind to one site and not be in contact with the other site most of the time. The regulation of the arabinose operon is quite similar ([12]), with a DNA loop bound by the repressor (illustrated in FIGURE 5), and Mäkelä et al. [13] indicate it is probably a single-molecule event that triggers the expression of the arabinose operon.

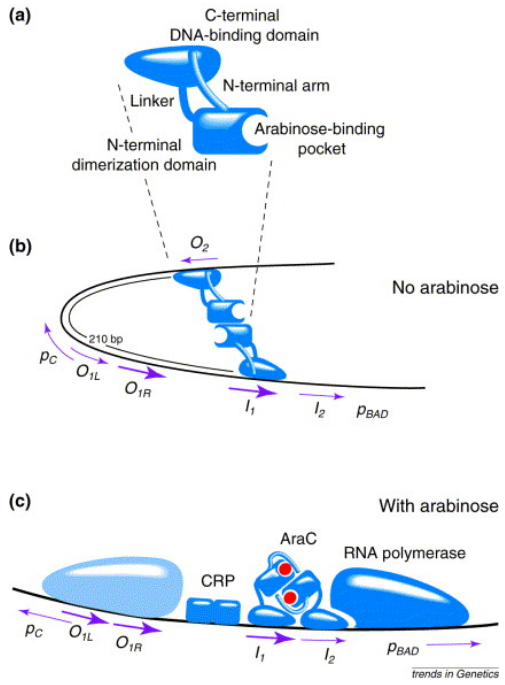


Figure 5: Scheme of the DNA loop repressing the arabinose operon. The figure is reproduced from Schleif [12], with permission from Elsevier. When arabinose is bound to the repressor, the DNA is freed, and takes time to be bound again in case arabinose disappears. The unbound DNA of interest (pBAD and pC in the picture) can then be transcribed.

It is important to note that, while this modelling approach may be much more exact than simple differential equations, it is quite unsuited to directly modelling a cell population in a bioreactor: while it would be possible, the computation time required for the computation would be enormous, as each one of the billions of cells present in the bioreactor would need their own simulations. Therefore an empirical model would be more useful, or, possibly, a semi-empirical model which would take lessons from the mechanistic modelling of reduced amount of cells. Stochastic modelling, is, in addition to being unsuited to large systems, limited by the knowledge of the bacteria: not all kinetic constants may be known for a given gene network, which would make the model inexact.

Another kind of mechanistic model, that may be less dependent of some of the missing knowledge, would be the boolean networks. Boolean networks are, at their core, simplifications of the reality, and

do not seek to simulate the behaviour of cells over time. Therefore they cannot be used for the current problem, but they are nevertheless informative: instead of trying to model the states of cells over time, a boolean model is made to represent the possible states in which cells can be, at equilibrium, and the pathways used to reach equilibrium; this knowledge could then be used in a semi-empirical model. This is made by representing the state of each relevant species in the cell as either "present" or "absent". Then, the interactions between all variables are represented, as a network, which is directly related to the relevant gene network, as can be expected. [14]

In particular, for arabinose, such a model was made by Jenkins and Macauley [14]. The extracellular arabinose levels are represented as two boolean variables: medium and high levels of arabinose. Those two boolean variables lets the authors represent the arabinose levels as "absent", "medium" and "high", three states where usually a boolean models would have only two. However uncommon, those three states allow the authors to determine that, in case of low arabinose concentrations, and in the absence of glucose only part of the bacteria would express the arabinose operon[3]. In case of high level of arabinose in the system instead, all the cells would express the operon, except if there is glucose. The presence of glucose is assumed to inhibit any and all arabinose expression. Those states are summarised briefly in Table 1, although more detailed results are present in reference [14].

Table 1: Table of the possible states of arabinose induction in cells, this data is a part of the data from table 1 in the paper from Jenkins and Macauley [14]. In case there is glucose and medium levels of arabinose, there can be cells in two different states.

Extracellular glucose	Extracellular arabinose	Arabinose operon state
Present	Any level	OFF
Absent	Absent	OFF
Absent	Medium level	ON or OFF
Absent	High level	ON

This boolean model is simplified, compared to reality, but offers a good starting point for the state of the art of the arabinose operon itself. The first point to address is what is called carbon catabolite repression. Namely, the assumption that when glucose, a sugar easier to metabolise than arabinose, is present, then no arabinose will be consumed. This is largely documented [15], including for arabinose [16] and could imply that the cells that consume arabinose do not consume glucose, as glucose consumption would inhibit arabinose consumption. However, recent evidence shows that this is not always the case [17], including for the arabinose operon [18]. Instead, it appears that most cells, in the presence of both arabinose and glucose, consume both sugars at the same time, in various ratios depending on the concentrations of both sugars[18]. Nikolic et al. [18] hypothesize that the uptake of both kinds of sugar at once is due to the global noise in the cell, meaning that even repressed genes can be expressed.

Regarding bistability, the boolean model predicts that it would happen when medium levels of arabinose are present. Medium and high levels of arabinose, however, are arbitrary, and the arabinose concentration can vary with much more subtlety than that. Rather than using a simple 3-states system,

[3]This presupposes that AraC is present in the system, which should be true most of the time, as there should be a basal expression simply due to random noise. At the very least, for any cell, it should be present *eventually*.

a more realistic model could therefore use the whole range of concentrations, with different proportions of cells induced for different concentrations. The exact specifics of how the bistability would arise, however, strongly depend on the kinetics of the switches to and from a phenotype using arabinose: depending on those kinetics, and on cell growth in each phenotype, the overall proportions of cells in each phenotype could vary significantly.

Indeed, the gene network for the arabinose operon, while simple-looking, has various implications on the kinetics of induction of said operon, when combined to the general noise in the system:

- Over a large range of arabinose concentration, the activity of the *araBAD* promoter can vary linearly, rather than just being considered "on" or "off". [19]
- *E. coli* responds to arabinose with a heterogeneous delay, likely dependent on the quantity of transporters at the time of arabinose induction. [20, 21]
- The kinetics of activation of the operon are different from the kinetics of deactivation of the operon. [22]
- The activation of the phenotype switch, as mentioned earlier, is likely caused by a single event: the detachment of the repressor from the DNA loop. This has implications regarding the mean amount of RNA copies in cells at the time scale of an hour. [13]
- As, again, mentioned earlier, not only does the arabinose concentration across time has an effect, but in addition the glucose concentration influences arabinose consumption. [18]

All those factors combined mean that the exact switching kinetics between the two phenotypes, in fluctuating concentrations of glucose and arabinose, are largely unknown.

3.2 Model construction

Given the state of the art on the arabinose operon, building a model for the phenotypic diversification dynamics of the microbial population requires either a semi-empirical or a mostly empirical model, as a mechanistic model would likely involve too many variables for a first attempt at describing the system. The distinction between "empirical" and and "semi-empirical" may be arbitrary and will not be discussed here as it is irrelevant and subjective. Suffice it to say, it seems simpler, in this particular case, to use an existing empirical model and to adapt it to the system than to create a mechanistic model essentially from scratch.

Therefore, the model from Thattai and Van Oudenaarden [1][4] was adapted to the arabinose operon and to the bioreactor. As a naive first approach, it is possible to describe a base model of the reactor through Equations 3 to 6

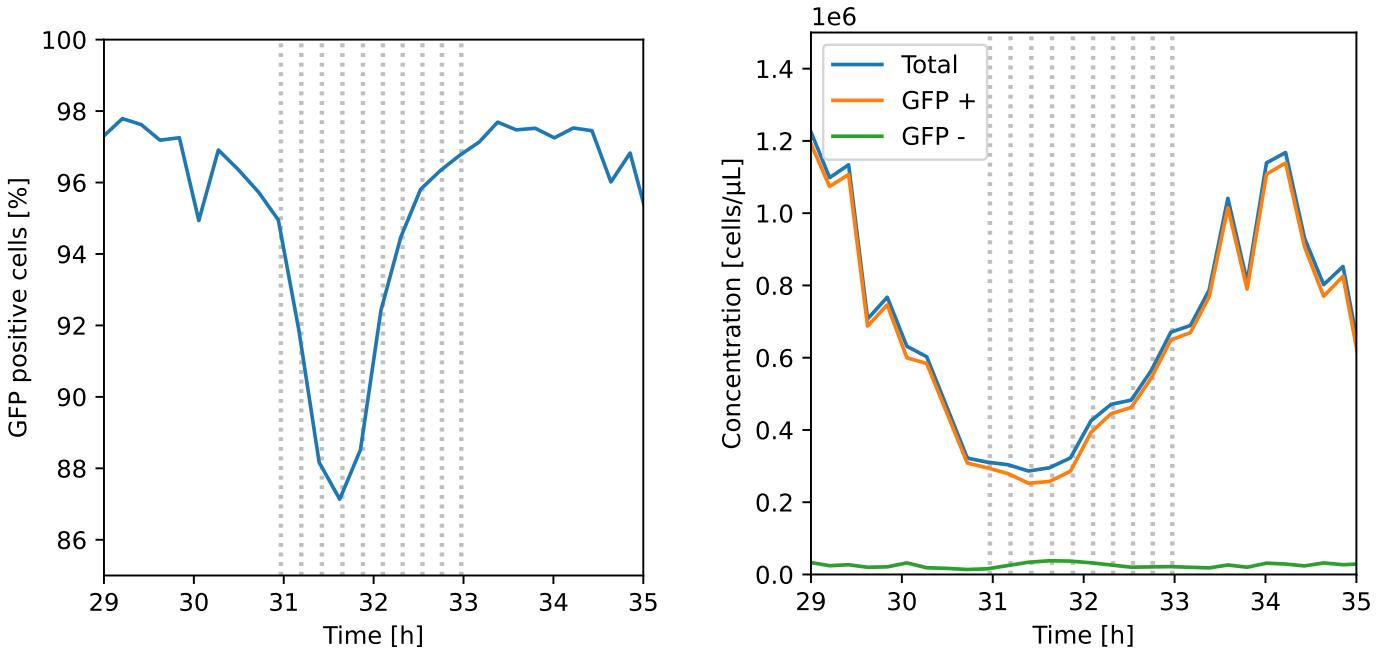
[4]That model is essentially the same as the one from Kussell and Leibler [4], but Thattai and Van Oudenaarden [1] was cited as it is, supposedly, valid out of equilibrium.

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mu_a A - DA - H_1 A + H_2 G \tag{3}$$
$$\frac{\mathrm{d}G}{\mathrm{d}t} = \mu_g G - DG + H_1 A - H_2 G \tag{4}$$
$$\frac{\mathrm{d}C_a}{\mathrm{d}t} = (C_{a,in} - C_a)D - \frac{\mu_a A}{Y_{xa}} \tag{5}$$
$$\frac{\mathrm{d}C_g}{\mathrm{d}t} = (C_{g,in} - C_g)D - \frac{\mu_g G}{Y_{xg}} \tag{6}$$

In those equations:

- A is the (mass) concentration of cells that have an active arabinose operon, later called the activated population.
- G is the (mass) concentration of cells with an inactive arabinose operon, later called the inactive population by opposition to the activated population. It is noted G as those cells use only glucose.
- t is the time.
- D is the dilution rate, as the reactor is a continuous one, all species are diluted over time.
- H_1 and H_2 are the basal stochastic switching rates, supposed constant, from one phenotype to the other. H_1 represents the switching from activated cells to inactive cells.
- C_a and C_g are the concentrations in arabinose and in glucose, respectively, in the medium.
- $C_{a,in}$ and $C_{g,in}$ are the sugar concentrations in the feed of the reactor, they replenish diluted and consume sugar.
- Y_{xa} and Y_{xg} are the yields from arabinose and glucose, respectively, to biomass. They are expressed in g of biomass produced per gram of sugar consumed.
- μ_a and μ_g are the growth rates of both phenotypes. A Monod growth rate is considered, thus $\mu_a = \mu_{a,\max}\frac{C_a}{K_a+C_a}$ and $\mu_g = \mu_{g,\max}\frac{C_g}{K_g+C_g}$; where $\mu_{i,\max}$ and K_i are, respectively, the maximum growth rate and the Monod substrate affinity constant for the relevant phenotypes.

Such a model is really simple and is unlikely to fit the data correctly, as can be expected given the initial peak observed in the chemostat, for example (see FIGURE 2, by contrast the base model would, according to intuition, tend near monotonously towards equilibrium), and as will be shown in the results. Therefore, some additional components must be taken into account.



(a) Proportion of GFP+ cells.

(b) Concentrations of the different sub-populations.

Figure 6: Population behaviour during one series of arabinose pulses, corresponding to a segregostat cycle. In those plots, the grey vertical lines represent the times a pulse of arabinose is added. The proportion of GFP+ cells does not increase immediately after the first pulse is in the system. Instead the GFP+ cells start growing after about half an hour, which increases their relative proportion in the population.

In this work, three such components were added to the model, based on what was observed from the data:

- In the segregostat, there is a time delay between the injection of arabinose and the increase in the proportion of GFP+ cells (see FIGURE 6). It is important to note that the latter delay is considered for the growth of the activated population, thus it is unrelated to the switch and is more likely to be due to intracellular biochemical processes (i.e., time for producing the enzymes and transporters needed for efficient arabinose utilization) than to a lack of arabinose inside the bacteria.

  For consistency, two time delays were added: one for the response to arabinose concentration, and one for the response to glucose, which should be much shorter.

- In the segregostat, peaks in the concentration of GFP+ cells can be observed. Concurrently, no significant peak can be observed in the concentration of GFP- cells (see FIGURE 6.b). This phenomenon has at least one possible simple explanation: firstly the basal switching between phenotypes should be relatively slow, because otherwise any wild fluctuation in the concentration of one phenotype would provoke similar variations, if damped, in the other. In addition, this observation can inform on whether the cells consume arabinose and glucose concurrently. Indeed, the peaks observed are quite repeatable, thus, even if the equipment is not able to provide exact or consistent concentrations, within the peaks, it is at least possible to say that the growth rate of the GFP+ population is much higher than the one from the GFP- population. If GFP+ cells did

not consume any glucose, then it would mean that, while the GFP- population consumes about 5 $gL^{-1}h^{-1}$, the GFP+ population could grow faster using a less efficient carbon source provided at a maximum rate of 3 $gL^{-1}h^{-1}$[5]. This would be quite extraordinary, and the most likely explanation is that GFP+ cells do consume glucose as well. This concurs with the point on co-consumption made in the state of the art.

- As mentioned in the previous point, basal switching should be relatively slow. However, when arabinose is first injected in the system, whether be it in the chemostat or the segregostat, a significant portion of the cells rapidly switch from inactive to activated. Therefore, in addition to to the relatively slow switching, there should be some kind of responsive switching integrated in the model.

This creates the framework for extended phenotypic switching models, depending on which of those points is considered, or how they are implemented. In the rest of the text, they are called the extended models, by contrast with the base model.

3.3 Hypotheses considered for an extended model

Following this reasoning, the hypothesis of this thesis may be written as such: an extended model, considering the co-consumption of the sugars by activated cells, time delays, and responsive switching, is better suited to representing the cell population diversification dynamics than the base model.

Since the model proposed is quite complex, this hypothesis will be divided in four questions, which together define the whole model:

Should the model have a time delay? That time delay is considered to be the time required for the cell to process sugars, meaning that at time t the sugar enters the cell, and at time $t + \Delta t$ the cell will grow by using said sugar. Two time delays will be considered, to stay consistent: one for glucose, and one for arabinose.

Should the model consider co-consumption of sugars? That co-consumption would translate into a higher growth rate for the cells that consume arabinose than for the cells that do not. In particular, for this work, it is considered that, with co-consumption, activated cells grow with a rate μ_{total} that is the sum of the consumption rates of arabinose and glucose, $\mu_{total} = \mu_a + \mu_g$. This is the simplest possibility for co-consumption.

Should the model consider basal switching? Since responsive switching will be considered, is it required for the model to keep the basal switching rates, H_1 and H_2, from Thattai and Van Oudenaarden [1]?

[5] 0.5 g of arabinose are added to 1 L of medium every twelve minutes approximately, but the timing could vary slightly. $3gL^{-1}h^{-1}$ is an over-estimation.

What kind of responsive switching should be considered? In particular, in this work, three possibilities will be considered:

- There may be no responsive switching at all.
- In the literature, Hill functions are considered to describe the activity of the arabinose operon depending on the arabinose concentration. Therefore, Hill functions to describe responsive switching will be considered. In this work, it is considered that the same maximum switching rate can be attained when cells switch from either carbon source to the other.
- From the graphs shown in SECTION 2, it is possible to infer another kind of switching mechanism, that can be approximated, in the model, and for these experiments, by initially forcing a part of the population to switch from inactive to activated.

These four questions are summarised in FIGURE 7. To answer them, all the possible combinations of answers were considered, in different models. These different models will make it possible to discriminate between the useful and useless contributions, by studying the differences between models with and without a given contribution. Finally, when all the questions will be answered, the best of the extended models will emerge as the final model of this work.

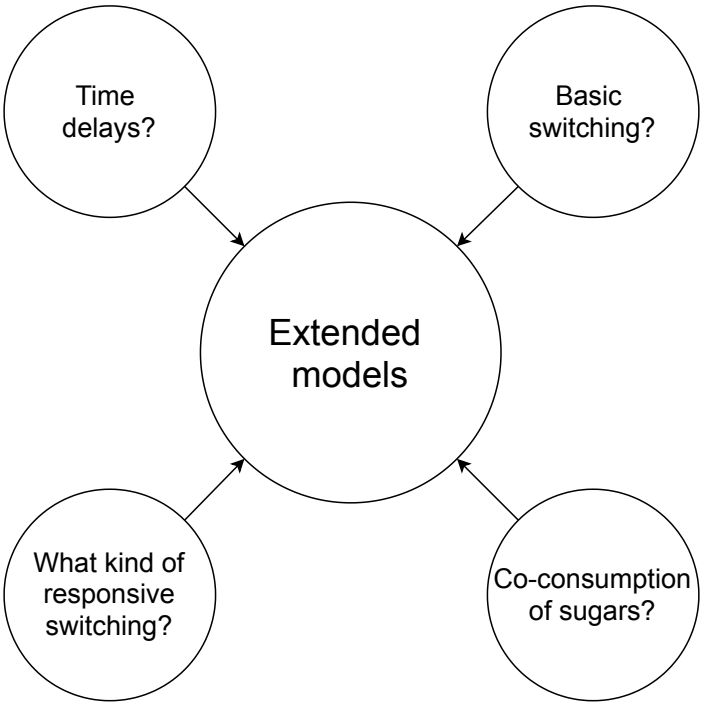


Figure 7: Answering each question separately will allow to form a model that is a combination of the relevant contributions.

4 Materials and methods

Before anything else, the data from the experiments was treated to create the summarising figures shown in Section 2 and have usable data objects.

Then, to determine the usefulness of the extended models, compared to the base one, several main steps were important. Firstly, it was ensured that the base model was well understood. This was made through symbolic analysis and a basic sensitivity analysis of the model.

Afterwards, the mathematical base for the extended model was written. Since the extended models are all based on different combinations of the same base contributions, the latter were coded independently of each other. This method lets one easily iterate over the various models, and made it possible to determine how the effects observed in the extended models appear, which of those contributions is important, and for what reason.

Finally, each model was optimised to fit one or several datasets, through a Metropolis-Hastings algorithm. This lets one determine whether these models are suited to fit the data or not.

The entirety of the code related to the models was made in Python and can be found on gitlab.uliege.be/mipi/standalone/vincent-master-thesis. The repository also contains all the results figures presented later.

4.1 Experiments

All the technical details for the experiments can be found in Annex A.1 and are not fully described here as it is not the object of the thesis.

The data processing was made using the code available at gitlab.uliege.be/mipi/segregostat/segdata ————python[6]. A few transformations were applied to the data for it to fit reality better:

- Between the batch and continuous phase, there is a delay that is not recorded in the data. It is of roughly 0.25 h. Since the continuous phase was considered to begin at a defined $t = 0$ h, and the first measurement was recorded at time 0 h but was in reality taken roughly 0.25 h after the end of the batch phase, all time measurements for the continuous phase were shifted by 0.25 h.
- In the flow cytometry data, measurements sometimes have values of "0" (a.u.). As the flow cytometer should always measure something, all the events (one event is supposedly a measured bacteria) that had a null measurement for any channel were removed from the data.
- Some datasets were much more noisy than others. To remove a big part of the noise, all time points that had more than 15% of their events removed due to null measurements were completely removed from the datasets.

Additional data treatments were considered to decrease the amount of noise, but in the end none was applied as there is always the risk of removing useful data, and the data resulting from the previous transformations was good enough.

[6]This code was made to be used at the laboratory, and further, backward compatibility-breaking revisions may be made. At the time of writing, the last commit was *7dd5ceaca6111a7f8e0767e7db18b247b710c356*, which is also the last commit referenced in the repository containing the models.

4.2 Preliminary analysis

Before even thinking of making a new model, it was necessary to make sure that the base model, derived from Thattai and Van Oudenaarden [1], could not reproduce the observed behaviours. To this end, first, the system was linearised. Since no glucose or arabinose is measured in the medium, the glucose and arabinose were first considered constant to provide a simplified version of EQUATIONS 3 to 6.

Afterwards, a short attempt at linearising the system by considering EQUATIONS 7 and 8 was made.

$$\mu_a = \mu_{a,\max}\frac{C_a}{K_a + C_a} \simeq \mu_{a,0} + \frac{\mathrm{d}\mu_a}{\mathrm{d}C_a}[C_{a,0}] \cdot (C_a - C_{a,0}) + \mathcal{O}\left((C_a - C_{a,0})^2\right) \tag{7}$$

$$\mu_g = \mu_{g,\max}\frac{C_g}{K_g + C_g} \simeq \mu_{g,0} + \frac{\mathrm{d}\mu_g}{\mathrm{d}C_g}[C_{g,0}] \cdot (C_g - C_{g,0}) + \mathcal{O}\left((C_g - C_{g,0})^2\right) \tag{8}$$

In those equations, $\mathcal{O}$ represents an "order of magnitude", the "[ ]" indicate the concentration the derivative was estimated at, the "0" index indicates a point of steady-state, $(C_g - C_{g,0})$ is considered small, as it is a linearisation, and $\mathcal{O}\left((C_g - C_{g,0})^2\right)$ is even smaller, and therefore negligible.

These equations make the system quite complex, compared to simply having constant sugar concentrations, and the results were briefly analysed using the sympy module in Python.

Finally, since the previous analyses were quite limited, the base model was put in equations in Python and solved thanks to the solve_ivp function from the scipy . integrate module. A basic analysis of the impact of the various parameters was then made to see whether the behaviour observed in the chemostat could be reproduced.

The analysis was made by first integrating the set of equations for a given, realistic set of parameters. Then, pairs of analogous parameters for arabinose and glucose (for example, $\mu\max, a$ and $\mu_{g,\max}$ or K_a and K_g) were varied by orders of magnitudes, and the system solved for the new parameters. For example, the starting values used for K_a and K_g were 0.01 g/L, and each took the values 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, and 10, to created 25 possible sets of parameters. The model was then solved for each set of parameters and the results were plotted to analyse the influence of the parameters.

The measured initial biomass of 1.8 g/L was not yet known, at the time, thus an estimated value of 1 g/L was used instead. The estimation comes from the fact that about 2×10^6 cells/µL were measured at the beginning of experiments by the flow cytometer, and a bacteria can be estimated to weigh ~ 500 fg [23].

A few different sets of initial parameters were used, based on which resulting plots looked best, and eventually the ones described in TABLE 2 were settled on.

Table 2: Parameters settled on based on their variation by orders of magnitude. As a reminder, H_1 is the switching rate from the activated to the inactive phenotype. The unrealistically high $\mu_{a,\max}$ was settled on because the sugar concentration profile looked better, but there was no good result anyways, regardless of how realistic the parameters were.

K_g	0.01	K_a	0.01	[h^{-1}]
H_1	10	H_2	0.1	[h^{-1}]
$\mu_{g,\max}$	0.5	$\mu_{a,\max}$	5	[h^{-1}]
$Y_{x,g}$	0.5	$Y_{x,a}$	0.5	[g/g]

4.3 Model design

In order to analyse completely the usefulness of the various contributions, each extended model was studied. In practice this was achieved by coding each of them (i.e. dilution, basal switching, responsive switching, time delays, and co-consumption) independently of the others.In addition, each model was given a set of parameters, but not all of them are useful for all models. For example, the delay would not be a useful parameter for a model where there is no delay considered. The way each model was set, then, was by choosing a set of contributions and the corresponding set of valid parameters.

This way of coding, although it may add some computational burden, lets one easily iterate over the different extended models, and each of them can then be analysed, instead of only comparing the base model and one or two extended models with the most contributions.

4.3.1 Modular contributions

All the contributions to the model are coded as short functions, and each model contains its own list of functions to run. For example, the base model would have a list of contributions with the functions for:

- Dilution
- Growth without co-consumption of sugars
- Basal switching

With this modular system the differential equations for all models, can simply be written, in Python, as:

```
def diff(self, t, x, p):
    # p here is the parameters and memory of the system

    # the concentrations cannot be smaller than 0
    x[x<0]=0
    # x[5] is the proportion of activated cells in the responsive population,
    # see initial forcing section
    if x[5]>1:
```

```
        x[5]=1

    # without contribution, the derivatives are null
    dx=np.zeros_like(x)
    # self.func.values contains the list of contribution functions
    # they are all added successively
    for f in self.func.values():
        f(dx, x, t, p)

    # dx/dt < 0 when x=0 is impossible
    dx[np.all((dx<0,x==0), axis=0)]=0

    # x[5] cannot be bigger than 1:
    if x[5]==1 and dx[5]>0:
        dx[5]=0

    p.store(dx, x, t) # store values to make delay differential equations
                      # (see time delays section)
    return dx
```

This function adds all the contributions together. It is then integrated using the solve_ivp function from the SciPy module (see `https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html`).

4.3.2 Dilution

The one contribution that was present in each model was the dilution, as for all model, the medium is continuously renewed, this is independent from the cells. It was coded as presented below:

```
# Dilution
def dilution(dx, x, t, p):
    dx[0:5]-=p.get_D(t)*x[0:5] # things get diluted
    dx[3:5]+=p.get_D(t)*p.get_Cin(t)
```

In this short piece of code, x is the value of all variables at time t. For reasons that will be made clear, it is a vector of length 6, containing, in order: the concentration of purely inactive cells, the concentration of activated cells, the concentration of inactive cells that may respond to a stimulus (see later), the concentration of arabinose, the concentration of glucose, and the future proportion of activated cells in the responsive population (see later). dx is the vector that will contain the derivative of x, after all the contributions are applied. It is initially initialized as a vector of zeros, and, in this case, the dilution contribution is added to the derivative. Finally, p is an object containing all the required parameters for the model. It provides, in this case, the values of the dilution rate (D) and of the input concentrations in sugars (Cin) at time t. The mention of the time is required as those parameters can be arbitrarily changed depending on the experiment, if required, like in the segregostat where pulses of arabinose are added and Cin changes for a short while.

For ease of use, the next contributions will be mainly expressed in their equivalent mathematical form. The dilution contribution would be written as:

$$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= -DA \tag{9}$$
$$\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= -DG \tag{10}$$
$$\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= (C_{a,in} - C_a)D \tag{11}$$
$$\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= (C_{g,in} - C_g)D \tag{12}$$

Where $+=$ denotes the addition of a contribution to the previous computed value of $\frac{\mathrm{d}A}{\mathrm{d}t}$, and the other symbols are the same as defined in Section 3.2.

4.3.3 Basal stochastic switching

In addition to the dilution, one may add the basal switching, defined in Equations 13 and 14.

$$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= -H_1 A + H_2 G \tag{13}$$
$$\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \quad H_1 A - H_2 G \tag{14}$$

This contribution was not added in every model. Indeed, while it is required in the base model, because the population, grown in the batch phase on glucose only, needs to partly switch to the activated phenotype, the addition of responsive switching to the system may render the contribution obsolete.

It is important to note that, switching, as it is defined there, is considered to appear as soon as traces amount of arabinose are present in the system. In the experiments considered, that is always the case, since arabinose is either injected all the time or injected as soon as not enough arabinose is present to maintain the population of activated cells. In environments where there are long periods of time without arabinose at all, it may be wise to have all cells switch back to an inactive state below a given, very low arabinose threshold.

4.3.4 Co-consumption of sugars

For the co-consumption, two contributions were defined: either cells co-consume, or they do not. This translates into:

For no co-consumption:

$$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= \mu_a A \tag{15}$$
$$\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \mu_g G \tag{16}$$
$$\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_a A}{Y_{xa}} \tag{17}$$
$$\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_g G}{Y_{xg}} \tag{18}$$

For the co-consumption:

$$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= (\mu_g + \mu_a) A \tag{19}$$
$$\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \mu_g G \tag{20}$$
$$\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_a A}{Y_{xa}} \tag{21}$$
$$\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_g (G + A)}{Y_{xg}} \tag{22}$$

In these equations, the consumption rate for glucose is the same for both the activated and inactive populations. Nikolic et al. [18] suggest that the co-consumption is due to global noise in the cell, and this noise may be proportional to the overall growth rate. On the one hand, this means that supposing the two glucose consumption rates are equal is probably wrong, since the growth rates for the two phenotypes are obviously different. On the other hand, however, there is already a lot of parameters in this model, and adding even more would maybe complicate things needlessly. Thus, here, the same glucose consumption rate is hypothesized, and the main idea is that the co-consumption, as it is hypothesized, may bring the model closer to reality than no co-consumption at all, even though it does not make the model exact.

4.3.5 Time delays

The third contribution is time delays. They, unlike the other contributions, cannot be directly added as a set of additional equations. Rather, they need to be integrated within the existing equations, and require the program to have memory as, with time delays, the system becomes a system of delay differential equations (DDEs). Since solve_ivp does not allow for delay by default, two solutions were possible: either use another module, which contains DDEs by default, or code a memory system within the differential equations made for solve_ivp.

Unfortunately, the only Python package that supports DDEs is “diffeqpy” (see pypi.org/project/diffeqpy), which requires Julia as a dependency. Thus, in the spirit of keeping the code easy to use for everyone at the lab, I decided to code the time delays within solve_ivp. This was achieved through the parameter objects which was used earlier: after adding each contribution to the derivatives stored in dx, the current relevant values are stored within the memory. While solve_ivp computes the derivative in some times where the values might be inexact as the algorithm then determines to use shorter time steps (solve_ivp uses the Runge-Kutta algorithm), and those values are stored as well, they should be close enough to reality that it does not matter (this was verified after the fact as well, the error is at most 1%).

Thus, thanks to the memory placed in the object, it is possible to produce time delays, which are directly implemented within the growth contributions. It was considered that bacteria grow with a growth rate dependent on past concentrations, but the sugars are still absorbed as if they were consumed normally. In this manner, the consumption contribution becomes (the co-consumption contribution is similarly affected):

$$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= \mu_a[t-\Delta t_a]A \qquad \text{with} \qquad \mu_a[t] = \mu_{a,\max}\frac{C_a[t]}{K_a + C_a[t]} \tag{23}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \mu_g[t-\Delta t_g]G \qquad \text{with} \qquad \mu_g[t] = \mu_{g,\max}\frac{C_g[t]}{K_g + C_g[t]} \tag{24}$$

$$\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_a[t]A}{Y_{xa}} \tag{25}$$

$$\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_g[t]G}{Y_{xg}}, \tag{26}$$

where $\mu_a[t-\Delta t_a]$ is the growth rate based on past concentrations.

The rationale behind having the delay only for the growth rate of cells is that it is, among other things, due to a delay inside the cells, since the cells need time to produce the enzymes needed for efficient arabinose utilization. Thus the sugar would, under this hypothesis, enter the cells, then translate into a faster growth rate only after Δt_a.

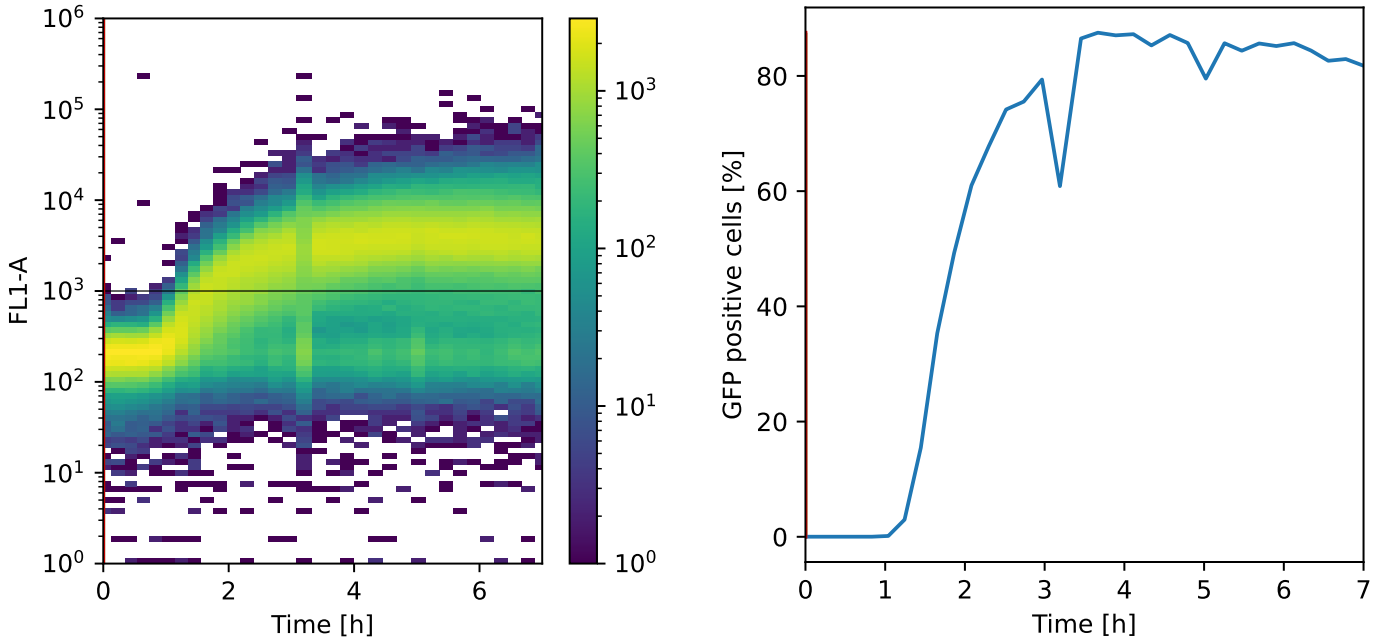
For cells that have just switched to using arabinose, this approximation is probably not valid, as the delay for those would be due to the lack of arabinose transporters [20, 21]. Even then, the reality is probably a combination of both, as there is probably a delay due to diffusion, and a delay due to processing of the arabinose, and both are undistinguishable in a bulk experiment like this one.

4.3.6 Responsive switching

The last contribution to consider is responsive switching. Unlike basal switching, responsive switching is supposed to be the result of the cell sensing its environment. Therefore, it is not simply constant, but it is a function of the arabinose concentration in the medium. It is also the point for which there is the most uncertainty. While there might be some things known about how the operon works, there is not much indication, in the literature or in the experiments, on how fast or when exactly the response comes into play. Among other things, it is unknown whether (or with what probabilities) *E. coli* transmits the state of the arabinose operon to both daughter cells when it divides. For this reason, there were two possibilities considered regarding the phenomenon, each one representing one extreme of what could be happening.

The first possibility is inspired by the behaviour depicted in Figures 8 and 9. On the one hand, when arabinose is first injected in the system, a big part of the cell population switches from not expressing GFP to expressing it (Figure 8). This change happens on the timescale of one hour, thus, for the part of the population that changes, the order of magnitude of the switching rate r_a can be considered to be $\mathcal{O}(1)$ h^{-1}, meaning that about 100% of the population would switch in about an hour. It is also known that such a switching rate is not attained in the the segregostat after the initial switch, because

there would be much wider fluctuations in the inactive cell concentrations otherwise (see FIGURE 6). Moreover (see FIGURE 9), the data on the segregostat shows clearly that, when arabinose is depleted, the cells start switching back to the GFP- phenotype with a similar dynamic, but it is stopped by the addition of arabinose in the system. Those elements contribute to forming a mechanism where the population is separated into two distinct subpopulations: a responsive one, and an "inert" one.



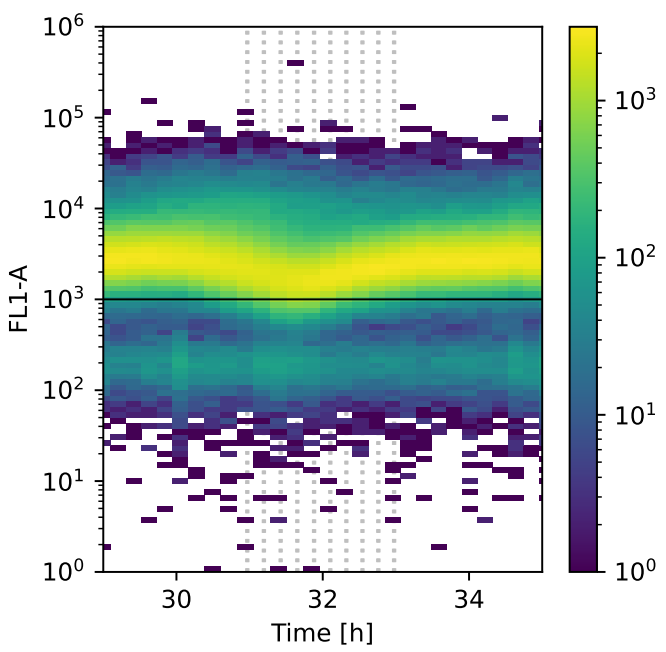
(a) Heat map of the cell fluorescence over time. At the beginning of experiments, the population can clearly be seen switching to arabinose usage; the horizontal line is the threshold above which cells are considered GFP+.

(b) The switch of the population in (a) corresponds to an increase in GFP+ cells here when they cross the fluorescence threshold.

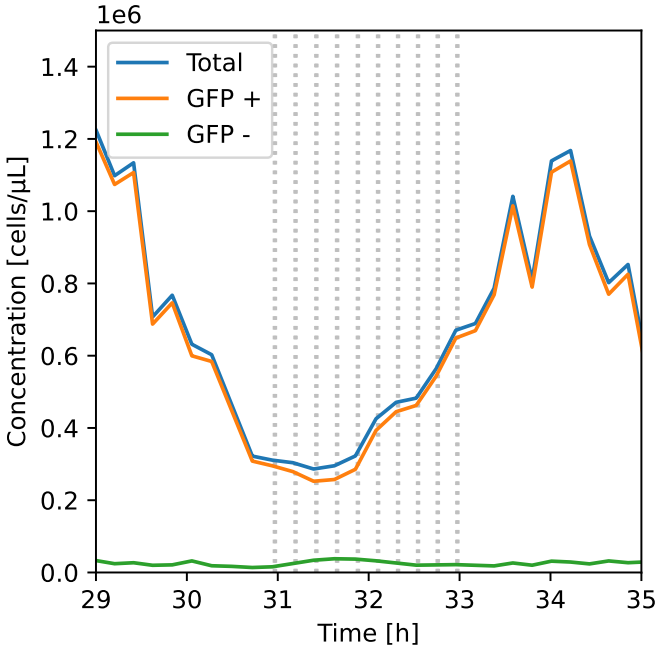
Figure 8: At the beginning of experiments (in this case, the chemostat experiment shown in FIGURE 2), the cell population can be seen switching to using arabinose, after a delay. Said delay is longer than the estimated delay for arabinose consumption, which was about 0.5 h.

Under this hypothesis, the inert population would be present in minority, and would only use glucose throughout the modelled experiment. By contrast, the responsive population would directly respond to the arabinose concentration, and switch rapidly to one phenotype or another when the arabinose concentration passes the required threshold. Such a mechanism is not enough to reproduce the observations, however. Indeed, a large delay can be observed initially, before the hypothesised responsive population starts switching to its activated state (FIGURE 8); and another large delay can be observed between the time the arabinose is depleted (indicated by the decline of the GFP+ cell concentration) and the time the activated cells can be seen expressing less GFP and pass the fluorescence threshold (see FIGURE 9). Reproducing this behaviour for the simulated responsive population can be done by supposing that it will only start changing state when the arabinose concentration is beyond the required threshold for long enough.

A last observation can help complete the behaviour of the responsive population: once the fluorescence of the responsive population starts going down, in the segregostat, it is forced back up before completely



(a) Heat map of the cell fluorescence over time. Cells can be seen expressing less GFP and eventually go below the GFP+ threshold, at time≃32 h.



(b) The drop below the threshold in (a) is visible when looking at the concentrations as well for some of the peaks, but negligible compared to the massive growth of the activated population that happens at the same time.

Figure 9: Population behaviour during one series of arabinose pulses; the same one as shown in FIGURE 6. While the latter figure shows that the main contribution to the change in proportion is the growth, this figure shows that switching does exist. It is simply so delayed compared to growth that it becomes nearly invisible.

switching (see FIGURE 9). This indicates that, while switching, the responsive population is directly sensitive to the arabinose concentration, and it is not forced to finish the transition.

Thus, overall, the responsive population can be summarised as:

- If it is in the inactive state, then it starts switching (with an ascending switching rate r_a) to the activated state if the arabinose concentration was above a threshold for long enough (a time Δt_i) in the system;
- If it is in the activated state, then it starts switching to the inactive state if the arabinose concentration was below the threshold for long enough in the system. That is with a time $\Delta t_f \neq \Delta t_i$, and a descending switching rate $r_d \neq r_a$, as the arabinose operon does not behave the same way when activating or deactivating, as mentioned in SECTION 3.1.
- If the population is partly in both states, then the swithing rate r_a or r_d is chosen based on the arabinose concentration.

The entire implementation of this phenomenon is presented in ANNEX A.3. The most important point is that this implementation explains the quirks in the code that were mentioned earlier. The second population of inactive cells corresponds to the responsive population. When arabinose is added

to the system, the cells rapidly switch from responsive inactive cells to responsive activated cells, while the purely inactive cells are unaffected. The basal switching contribution, defined by H_1 and H_2, is considered to be between the responsive and the non-responsive population. In addition, keeping the "future proportion of activated cells in the responsive population", p, in the derivative vector is a trick to make the combination with time delays possible. Firstly, it is the future proportion because cells can only detect arabinose after the time delay that is considered in the time delays contribution. Therefore, having p depend on the current concentration of arabinose means that it represents the choices that will be made by the population in the future. Secondly, the proportion is tracked rather than having the code remember the past concentrations because the switching rate cannot be implemented as would be intuitively expected. Indeed, if it were, the anticipated equations would be:

$$\frac{\mathrm{d}A_r}{\mathrm{d}t} \mathrel{+}= \quad r_a G_r \tag{27}$$

$$\frac{\mathrm{d}G_r}{\mathrm{d}t} \mathrel{+}= -r_a G_r, \tag{28}$$

with the r denoting the responsive population. With those equations, there would never be 100% of the responsive population that has switched. Instead, the progression of the responsive inactive cell population would follow a decreasing exponential, trending towards its equilibrium value while never quite reaching it — or reaching it very late, when the difference between the real values and the equilibrium values are smaller than the machine epsilon[7]. While this does not usually cause any problem, in this case, one issue arises: it was determined that, when the population is entirely in one state, it cannot switch to the other without waiting for the concentration to be right for long enough. This implies that the responsive population can reach entirely one state or another in a reasonable time. Thus, instead, the equations considered, when switching from the inactive to the activated state, are as follows:

$$\frac{\mathrm{d}A_r}{\mathrm{d}t} \mathrel{+}= \quad r_a[t](G_r + A_r) \tag{29}$$

$$\frac{\mathrm{d}G_r}{\mathrm{d}t} \mathrel{+}= -r_a[t](G_r + A_r) \tag{30}$$

$$\frac{dp[t + \Delta t_a]}{dt} \mathrel{+}= \quad r_a[t + \Delta t_a] \tag{31}$$

Thanks to Equations 29 to 31, p varies linearly over time when switching, which ensures it can reach 0 and 1, at which point the arabinose concentration needs to be consistently beyond the threshold before the population starts switching. r_a for the responsive population simply follows what happened with p a time Δt_a before, to respect the time delay. Some precautions were taken to avoid negative concentrations as well. Some mistakes in the overall concentrations can appear due to the timing, as the responsive population may not be evaluated exactly at a time Δt_a after p started changing, which would cause small differences between the two at any time point, but it is minor and can be corrected by letting the part of the responsive population that should not be there switch towards the one that should when $p = 0$ or $p = 1$.

However, due to the way the memory and the switching were implemented, the integration of this switching mechanism was numerically unstable: firstly, adding to the memory every time the differential

[7]The machine epsilon is the smallest numerical value that can be represented by the computer.

function is called causes some function calls that are then rejected by solve_ivp to be remembered in memory; this means that in some cases there is a $p = 1$ in the memory even though there should not be. In addition to that, it was determined that the cells behaved responsively while switching, meaning that below the threshold arabinose concentration, the proportion of activated cells goes down quickly, and above that threshold, it goes up quickly. This causes instability when the arabinose concentration is close to the threshold: solve_ivp tries to use very small time steps because the derivative is discontinuous. Oscillations around that threshold value also cause the responsive population to stay in a responsive state rather then be locked into one state or another, which makes the predictions wrong.

As will be mentioned in the perspectives, it is probably possible to fix the instability by making the model more complex. However, this is not the purpose of this work; the goal in this case is rather to estimate how close a relatively simple model can get to reality, knowing that it is still an approximation. In this particular case, one additional approximation that can be made concerns the responsive switching after arabinose disappears. Indeed, as previously shown in FIGURE 9, in the segregostat, the GFP+ cells start going down in fluorescence. But the measure that is used to estimate the state of the population is the proportion of cells that have a fluorescence above 1000, and the figure shows that the amount of cells from the descending population that pass below the fluorescence threshold before the cells respond to the added arabinose, compared to the growth of the activated cell population at the same time. This means that even if it were incorporated in the model, there would be very little information on it, and it also means that it can be neglected for all the segregostat and chemostat data (new data with a longer time between arabinose pulses was made available later, and is briefly discussed in the perspectives, but the model is not directly applicable for that particular experiment).

Therefore, for the purpose of this work, the response was simplified: effectively, at the beginning of the experiment, the responsive population is forced to start switching to the activated phenotype after there has been enough arabinose above the threshold for long enough. While it is switching, it is forced to only switch towards the activated state, and no switching back to the inactive state is ever allowed when it has stopped switching as it would be negligible anyways. This contribution is later called “initial forcing”.

This represents one extreme of the switching behaviour: based on a threshold, cells switch, and the decision is held onto for a long time (for the experiments made, the decision is kept for the whole experiment). The other possibility that would make sense is having a switching rate between the phenotypes that directly depends on the arabinose concentration in the bioreactor. The cells would be responsive, directly, to the arabinose concentration, and the slow switching observed in the segregostat would be a consequence of the arabinose concentration being close to the point where there is effectively no switching from the inactive population to the activated population.

To represent such a phenomenon, using a separate responsive population, as for the previous mechanism, is probably not the best option. Rather, if switching rates for the two populations depend on arabinose concentration, and if both sub-populations switch phenotypes with a non-zero rate, then both sub-populations will always be present; hence simply considering inactive and activated cell populations is sufficient to produce the required dynamics.

The question, then, is to know what function might correctly represent the switching rates of the two populations. This function must follow at least two boundary condition: when there is no arabinose at all in the system, all cells should naturally switch to the inactive phenotype. On the other side of the spectrum, if there is a lot of arabinose, most cells should switch to using arabinose. But, regarding the

switching rates, it also stands to reason that there should be a saturation: if there is more than enough arabinose in the system, cells would likely have no way to tell. Hence, cells ought not to switch much faster if there is 15 g/L of arabinose in the system than if there is 10 g/L. They should, however, switch much faster if there is 1 g/L of arabinose than if there is 0.01 g/L.

One family of functions that answer those criteria are Hill's functions. Hill functions appear in a wild range of biological processes involving equilibrium, as shown by Santillán [24]. Briefly[8], a Hill function is a sigmoid function, going from zero in zero to one in ∞, with most of the variation around a determined threshold (see FIGURE 10). Moreover, Hill functions are known to be representative of the promoter activity for arabinose [19, 20], and have a similar shape to the percentage of induced cells reported by Morgan-Kiss, Wadler, and Cronan [25].

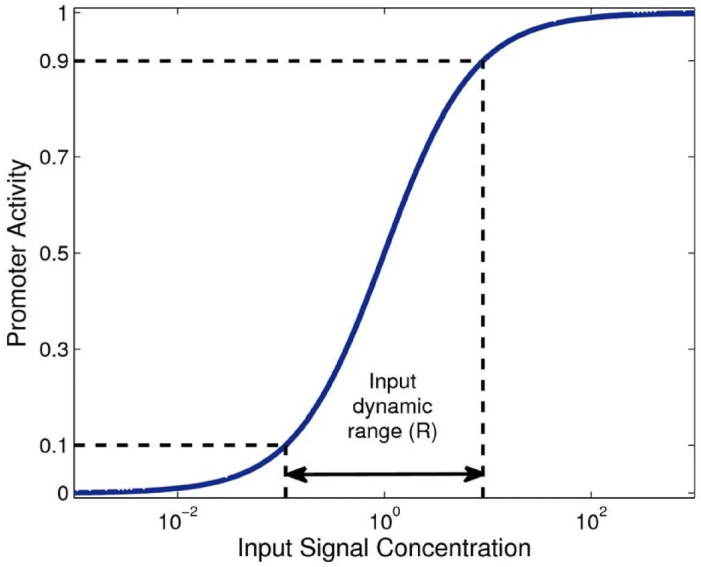


Figure 10: Example of a Hill function, here reproducing the activity of an operon depending on the concentration of its input signal, that could be the arabinose concentration. Before the input dynamic range illustrated, the hill function is close to 0, after it, it is close to one, and when the concentration is within the dynamic range, the promoter activity is nearly linearly correlated with the concentration. The dynamic range is around the concentration threshold of the equations, 1 in the figure. The image is reproduced from Madar et al. [19], which is under Creative Commons Attribution License (`https://creativecommons.org/licenses/by/2.0`).

Hill functions follow the following equations by definition [24]:

$$h^{(1)} = \frac{x^n}{k^n + x^n} \tag{32}$$

$$h^{(2)} = 1 - h^{(1)} \tag{33}$$

$$= \frac{1}{k^n + x^n} \tag{34}$$

Here, if $h^{(1)}$ represents one portion of the population, then $h^{(2)}$ represents the other portion. n is a coefficient that represents how steep the sigmoid is, and k is the threshold concentration at which $h^{(1)} = h^{(2)} = 0.5$. Among others, one of the interesting properties of Hill functions is that:

- In a closed or steady-state system
- With equal growth or dilution of two different phenotypes
- Considering the switching is dependent on the arabinose concentration, so $x = C_a$

[8] Santillán [24] provides a good summary of what a Hill function is, if required.

- Considering switching rates that scale according to the same Hill function

The system can be represented by Equations 35 and 36.

$$\frac{\mathrm{d}A}{\mathrm{d}t} = 0 = r\frac{C_a^n}{k^n + C_a^n}G - r\frac{1}{k^n + C_a^n}A + \alpha \tag{35}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = 0 = -r\frac{C_a^n}{k^n + C_a^n}G + r\frac{1}{k^n + C_a^n}A + \alpha \tag{36}$$

In those equations, α represents the other contributions (equal for both phenotypes at steady state), A and G the two phenotypes, and r a scaling factor, the maximum switching rate that the population can attain. The two lines are equal to each other, thus one can deduce that $h^{(1)} = \frac{A}{G+A}$ thanks to Equations 37 to 41.

$$r\frac{C_a^n}{k^n + C_a^n}G - r\frac{1}{k^n + C_a^n}A + \alpha = -r\frac{C_a^n}{k^n + C_a^n}G + r\frac{1}{k^n + C_a^n}A + \alpha \tag{37}$$

$$\Leftrightarrow \quad \frac{C_a^n}{k^n + C_a^n}G - \frac{1}{k^n + C_a^n}A = -\frac{C_a^n}{k^n + C_a^n}G + \frac{1}{k^n + C_a^n}A \qquad (r \neq 0) \tag{38}$$

$$\Leftrightarrow \quad h^{(1)}G - (1 - h^{(1)})A = -h^{(1)}G + (1 - h^{(1)})A \tag{39}$$

$$\Leftrightarrow \quad 2h^{(1)}(G + A) = 2A \tag{40}$$

$$\Leftrightarrow \quad h^{(1)} = \frac{A}{G + A} \tag{41}$$

This short reasoning shows that, if one neglects the evolutionary advantage of one phenotype over another (i.e. the difference in growth rates), then having Hill functions to represent the switching rates means that, at equilibrium, the proportion of cells having one phenotype or another will follow Hill functions as well. This result is coherent with all the requirements that were stated earlier, and with the observations regarding the amount of induced cells by Morgan-Kiss, Wadler, and Cronan [25]. As can be expected, then, the contribution added is defined by Equations 42 and 43.

$$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= r\frac{C_a^n}{k^n + C_a^n}G - r\frac{1}{k^n + C_a^n}A \tag{42}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= -r\frac{C_a^n}{k^n + C_a^n}G + r\frac{1}{k^n + C_a^n}A \tag{43}$$

Both the initial forcing and this contribution, dubbed the direct responsive switching, are obviously simplifications of what could be really happening, but knowing how the model reacts with the two different possibilities and comparing it to reality can tell us which is closer to reality.

This contribution and the others were summarised in Table 3. Together, they let one form 24 models, presented in Table 4.

Table 3: Summary of all the contributions that can be considered. Dilution and one of the growth contributions are needed in all models. Only one of the responsive switching may be considered in a single model.

Contribution		Necessity	Mathematical expression
Dilution		Necessary	$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= -DA$ $\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= -DG$ $\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= (C_{a,in} - C_a)D$ $\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= (C_{g,in} - C_g)D$
Basal switching		Optional	$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= -H_1A + H_2G$ $\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= H_1A - H_2G$
Growth	No co-consumption	One is needed	$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= \mu_a A$ $\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \mu_g G$ $\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_a A}{Y_{xa}}$ $\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_g G}{Y_{xg}}$
	Co-consumption		$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= (\mu_g + \mu_a)A$ $\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \mu_g G$ $\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_a A}{Y_{xa}}$ $\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_g (G+A)}{Y_{xg}}$
Time delays		Optional	$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= \mu_a[t - \Delta t_a]A$ $\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= \mu_g[t - \Delta t_g]G$ $\frac{\mathrm{d}C_a}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_a[t]A}{Y_{xa}}$ $\frac{\mathrm{d}C_g}{\mathrm{d}t} \mathrel{+}= -\frac{\mu_g[t]G}{Y_{xg}}$
Responsive switching	Initial forcing	Maybe one but not both	$\frac{\mathrm{d}A_r}{\mathrm{d}t} \mathrel{+}= r_a[t](G_r + A_r)$ $\frac{\mathrm{d}G_r}{\mathrm{d}t} \mathrel{+}= -r_a[t](G_r + A_r)$ $\frac{dp[t+\Delta t_a]}{dt} \mathrel{+}= r_a[t + \Delta t_a]$ Only when the responsive population switches
	Direct response		$\frac{\mathrm{d}A}{\mathrm{d}t} \mathrel{+}= r\frac{C_a^n}{k^n + C_a^n}G - r\frac{1}{k^n + C_a^n}A$ $\frac{\mathrm{d}G}{\mathrm{d}t} \mathrel{+}= -r\frac{C_a^n}{k^n + C_a^n}G + r\frac{1}{k^n + C_a^n}A$

Table 4: List of the different models that can be formed, and their associated numbers. Number 0 to 3 do not make sense, as they do not consider any way for the cells to be activated.

	Consumption scheme	Time delays	Basic switching	Responsive switching
0	Arabinose only			
1	Co-consumption			
2	Arabinose only	V		
3	Co-consumption	V		
4	Arabinose only		V	
5	Co-consumption		V	
6	Arabinose only	V	V	
7	Co-consumption	V	V	
8	Arabinose only			Initial forcing
9	Co-consumption			Initial forcing
10	Arabinose only	V		Initial forcing
11	Co-consumption	V		Initial forcing
12	Arabinose only		V	Initial forcing
13	Co-consumption		V	Initial forcing
14	Arabinose only	V	V	Initial forcing
15	Co-consumption	V	V	Initial forcing
16	Arabinose only			Direct response
17	Co-consumption			Direct response
18	Arabinose only	V		Direct response
19	Co-consumption	V		Direct response
20	Arabinose only		V	Direct response
21	Co-consumption		V	Direct response
22	Arabinose only	V	V	Direct response
23	Co-consumption	V	V	Direct response

4.4 Fitting

In order to determine how the models fit with reality, it is important to fit them to the data. To do this, two things are required: a method to evaluate how close to reality the predictions of a model and its associated parameters are, and an algorithm to take advantage of that method and bring the model as close to reality as possible.

4.4.1 Model evaluation

The first thing needed to fit the model to reality is to define exactly what being "close to reality" means. Intuitively, it can be understood quite easily: if the same environment as an experiment is provided to the model, then the results of the model should be close to the measurements taken during the experiments.

The first hurdle that appears, then, is the measurements taken. There is a lot of data, but only part of it is really useful in this context. The heat maps presented in SECTION 2 contain data that is too complex for the extended models, as they only take into account 2 cell populations. The cell concentrations may be used qualitatively, to indicate that there are spikes of activated cells in the segregostat, for example, but they conflict with biomass measurements and the long term trends cannot be trusted. Due to noise, the very short term trend (i.e. between two or three successive measurements) cannot be trusted either. The sugar concentrations were measured in a few experiments, at long intervals, but a big part of the measurements are simply below the detection limit, and there is also much less data available than for the cells. In the end, the only measurement that can really be trusted is the proportion of activated cells in the population.

Even then, there are regularly outliers in the measurements, as can be seen in FIGURE 2. Thus, in order to slightly reduce their impact, the logcosh function from Keras[9] to compare the results of the model to the data. This function computes the $\text{mean}(\log(\cosh(y_{predicted} - y_{measured})))$ for a series of measured y that need to be compared to the computed y. In this case, y is the proportion of activated cells in the system. Effectively, the logcosh behaves as the mean square error when $|y_{predicted} - y_{measured}|$ is small, and as the mean absolute value when it is large. This reduces the impact of the outliers, compared to a simple mean square error.

Finally, it was required to choose a dataset to be compared with the model. The two datasets presented in SECTION 2 were used for this purpose. In order to see whether the model parameters could be generalized, it was adjusted to each set of data separately, and to both sets at once (in which case the logcosh was normalised to the number of measurements overall). The comparison to the chemostat was made for the first 50 h, as that is sufficient to see the main trend of the data. For the segregostat, since the trend has a shorter time scale (due to periodic stimuli following arabinose injection), only 40 h of data were used, which slightly reduces the time required for optimisation. For the comparison to both datasets, 50 h of data from both experiments were considered, for consistency.

4.4.2 Metropolis-Hastings algorithm

Once it is possible to compare the model to the data, the said model can be fitted. However, fitting the data using the SciPy, module, with scipy.optimise, has a lot of drawbacks. Most of the possible methods estimate partial derivatives with respect to the parameters, which, when there are as many as for this model, has a high chance of diverging or to lead to very unlikely values (in my experience at least), may be stuck in local minima, and depends heavily on the initial values provided. There is no guarantee that the algorithm converges toward the desired solution.

[9] From the TensorFlow module, see the documentation at keras.io/api/losses/regression_losses

An alternative to these deterministic algorithms is the Metropolis-Hastings algorithm. This algorithm, in theory, should always converge toward the absolute minimum, if given enough time. In practice, there is no real condition to determine when the best possible solution has been found, but it does have other advantages. Mainly, it is unlikely for the parameters to get too far out of range, and it should continuously improve in precision the longer it is run, regardless of exactly how long that is. Determining the exact parameter values here is not really the goal, simply getting close to the best values and estimating how important they are is enough.

The Metropolis-Hastings algorithm is a stochastic algorithm. What it means is that the algorithm takes initial parameters to the models, changes them a little bit, then checks if the new solution is better, and, depending on the results, it has a certain chance of accepting the new solution. The algorithm can be summarised, in pseudo-code, as such (taken from [26]):

Set starting state $x^{(0)}$, $i = 1$
while convergence not reached **do**
    Generate $y^{(i)} \sim q(y|x^{(i-1)})$
    $\alpha \leftarrow \min\left\{1, \frac{p_X(y^{(i)})}{p_X(x^{(i-1)})} \frac{q(x^{(i-1)}|y^{(i)})}{q(y^{(i)}|x^{(i-1)})}\right\}$
    Generate u uniformly in $[0, 1[$
    if $u < \alpha$ **then**
        $x^{(i)} \leftarrow y^{(i)}$
    else
        $x^{(i)} \leftarrow x^{(i-1)}$
    end if
    $i \leftarrow i + 1$
end while

In this pseudo-code, y and x are sets of parameters. q is a distribution that lets one determine, from a given x, a new y, that is close to x, with a probability distribution $q(y|x)$. Finally, $p_X(x)$ is the probability distribution for a given set of parameters to be right, for the algorithm. i is the iteration number, and α represents the probability to accept y as the new set of parameters. The way q is implemented warrants more details and is discussed in the next section, but p_X is described here.

This probability distribution p is the probability for a given set of parameters to be right, *for the algorithm*. It does not necessarily relate to actual probabilities. As a matter of fact, if one wants to minimize a function $f(x)$ (for example, the logcosh function, as described in the previous section), one may use $p_X(x) \propto \exp(-\beta f(x))$ [26]. This ensures that $\alpha \propto \exp\big(-\beta f(y^{(i)}) + \beta f(x^{(i-1)})\big)$, which is equivalent to saying that if the new set of parameter is close to reality ($f(y^{(i)})$ low), then α is high. In addition, most of the time, in our case (see next section), $\frac{q(x^{(i-1)}|y^{(i)})}{q(y^{(i)}|x^{(i-1)})} = 1$, so that having $f(y^{(i)}) < f(x^{(i-1)})$ means $\alpha = 1$ (because it is replaced by 1 if the computed probability is too high). And α represents the probability to adopt the new set of parameters.

The parameter β in $\exp(-\beta f(x))$ corresponds to how probable it is to accept a solution that is worse than the one currently in use. Indeed, if β decreases, and $\alpha \propto \exp\big(\beta(f(x^{(i-1)}) - f(y^{(i)}))\big)$, where

$f(x^{(i-1)}) - f(y^{(i)}) < 0$ as $y^{(i)}$ is worse than $x^{(i-1)}$, then α increases.

Trying the algorithm on the extended model, with co-consumption, time delays, and initial forcing, showed that typical differences between $f(y^{(i)})$ and $f(x^{(i-1)})$, when comparing to both datasets were of the order of 2×10^{-3}. Thus, to allow for the algorithm to accept a worsening of that order of magnitude with an arbitrary 10% chance, which seems reasonable, a β value of 1000 is estimated in EQUATIONS 44 to 49.

$$\alpha = 0.1 = \frac{p_X\left(y^{(i)}\right)}{p_X\left(x^{(i-1)}\right)} \frac{q\left(x^{(i-1)} \mid y^{(i)}\right)}{q\left(y^{(i)} \mid x^{(i-1)}\right)} \tag{44}$$

$$= \frac{p_X\left(y^{(i)}\right)}{p_X\left(x^{(i-1)}\right)} \qquad \text{(Most of the time)} \tag{45}$$

$$= \exp\big(\beta(f(x^{(i-1)}) - f(y^{(i)}))\big) \tag{46}$$

$$= \exp\big(-\beta \cdot 2 \times 10^{-3}\big) \tag{47}$$

$$\Leftrightarrow \quad -\frac{\log(0.1)}{2 \times 10^{-3}} = \beta \tag{48}$$

$$\Rightarrow \quad \beta \simeq 1000 \tag{49}$$

Thus $\beta = 1000$ was used at the start of the algorithm. However, as the algorithm progresses, or for other models, 1000 might not be the best value. Therefore, periodically, β is adjusted such that after 150 iterations, and every 50 iterations following the 150^{th}:

If the new set of parameter was accepted less than 10% of the time in the last 50 iterations:
$\beta = \beta/1.5$
else if the new set of parameter was accepted more than 90% of the time:
$\beta = 1.5\ \beta$

Decreasing β increases α, as mentioned earlier, thus it increases the proportion of times the new set of parameters is added.

In practice, the algorithm was applied 3 times for each model, for 300 iterations when comparing to both datasets, 600 iterations when comparing to the chemostat, and 300 iterations when comparing to the segregostat. The difference in number of iterations is due to time constrains, as comparing to the segregostat is, on average, slower, but 300 iterations is enough to get a good idea of the optimal parameters already.

4.4.3 Initial parameters

The extended models depend on a lot of different parameters. Before trying to find the exact parameters that should be used, it is interesting to have a look at the order of magnitude they should

have, as these orders of magnitude can then be used as a starting point for the Metropolis-Hastings algorithm.

Firstly, for the growth rates, numbers can vary wildly. They follow the Monod equation ($\mu = \mu_{\max} C/(K+C)$). Thus, for each (one for arabinose, one for glucose), there are two different parameters; $\mu_{\max}$ is the maximum growth rate the cells can reach, and K is the sugar concentration at which half the maximum growth rate is reached. $\mu_{\max}$ is usually of the order of magnitude of 1 h^{-1}, at least for *E. coli* growing on glucose[27, 28]. No data was found for arabinose (using Google Scholar and BioNumbers), but it stands to reason that the order of magnitude should be the same, even though the actual value is expected to be lower for arabinose than for glucose: arabinose requires additional protein synthesis, which would decrease growth rate. Unlike $\mu_{\max}$, K does not have a defined order of magnitude. Sources ([27, 28]) place it anywhere between ∼0.1 g/L and ∼0.01 mg/L. At the very least, there are records of about ∼0.01 g/L for K in a chemostat, thus that was the value used when required. The same order of magnitude was considered for arabinose.

Another point that appears in the growth contribution is the yield, Y_{xs}. It is usually accepted to be 0.5 $g_{biomass}/g_{sugar}$[10], and that was the order of magnitude used for both arabinose and glucose. It can at most vary between 0 and 1 but is unlikely to be below 0.25 or above 0.75, at most.

The basal switching rates, H_1 and H_2, were already determined to be much lower than the growth rates (see FIGURE 9). While their order of magnitude is unknown, taking a tenth of the estimated $\mu_{\max}$ fits the criterion, and no other data is available. Thus, $H_1 = H_2 = 0.01$ h^{-1} were used as first approximations.

The time delay for arabinose use can be estimated from the segregostat as the time needed for the activated cell concentration to start increasing after a series of arabinose pulses was started. That time is around half an hour. For the glucose, there is no real indication, but it is known that glucose is the base food for *E. coli*[15], and requires no metabolic adaptation beyond the production of glucose transporters. Thus, the time delay for glucose should be shorter than the one for arabinose.

For the switching, with initial forcing of the population, orders of magnitude have to be estimated from common sense and from the data. In the case of the responsive population, there are several parameters that come into play:

- There is a basic threshold of arabinose concentration above which cells start switching to using arabinose. This threshold should be below the detection limit of arabinose in the system, as no arabinose was detected in the chemostat, but there were GFP+ cells. That detection limit is at ∼0.02 g/L. In addition, Morgan-Kiss, Wadler, and Cronan [25] report that 0.002 g/L is enough to induce the activation or the arabinose operon. Thus, 0.002 g/L was considered a good first approximation of the parameter.
- There is an additional delay at the beginning of experiments, before the cells start switching, in addition to the normal delay required for arabinose usage. The delay between the beginning of the continuous phase and the time cells start to switch is measured to be about an hour. Thus, subtracting a half hour of normal delay for the use of arabinose, the initial delay can be estimated at about a half hour as well.

[10]Records can be found on BioNumbers: bionumbers.hms.harvard.edu/search.aspx?trm=yield+glucose+coli

- The transition to the activated phenotype happens in about one hour, although the exact time is difficult to measure since there is simultaneous growth and diffusion. Thus, if about 100% of the responsive population switches to using arabinose in about an hour, then the switching rate is about 1 h^{-1}.
- Finally, the last parameter is the proportion of the population that, initially, is responsive, sometimes called p_{init} later. This is difficult to estimate, as there are about 2 hours between the start of the continuous phase in the experiment and the time where most cells have switched; and during part of that time there is growth at different rates. However, in the chemostat, ∼80% of the population, and in the segregostat, ∼90%, are activated cells. Thus the proportion should be relatively high; 90% was used later on.

The alternative to the initial forcing is the direct responsive switching. This contribution depends on 3 different parameters:

- In the Hill function ($\frac{x^n}{k^n+x^n}$), n determines how steep the transition of the sigmoid is. There is no real data to determine this, but it should be relatively steep as, when there is some arabinose but not much, there should be little switching. 3 was therefore chosen as a first approximation.
- The k in the Hill function is equivalent to the threshold used for the initial forcing, and the same value (2×10^{-3}) was used.
- This Hill function is scaled by the maximum rate that the transitions can have. This rate should be approximately the same as the one from the initial forcing, as the responsive switching, ideally, should have a similar effect to the initial forcing, during the first few hours of experiment. The maximum rate chosen, r, was therefore 1 h^{-1} as well.

Finally, a word must be said about the amount of biomass initially in the system. While it does not explicitly appear in the model, it would make sense that changing the amount of biomass would change the way the model reacts, as part of the effects observed probably comes from the growth rates and the dilution. However, a small change in biomass probably would not change the results drastically, while a different order of magnitude would. The initial biomass was measured in one of the experiments to be 1.8 g/L, and this measurement was used for all in silico experiments.

Those numbers are summarised later, in Table 5.

4.4.4 Parameter variation

For the Metropolis-Hastings algorithm, it is necessary to make the set of parameters vary. There are several ways to do this, but the simplest one is to draw a new value for each parameter from an interval centred around the old value. Mathematically, it can be described by Equations 50 and 51.

$$q_i(y_i|x_i) = \begin{cases} \frac{1}{\Delta_{int}} & \text{if } y_i \in \ interval \\ 0 & \text{elsewhere} \end{cases} \tag{50}$$

$$\text{where} \qquad interval = \left[x_i - \frac{\Delta_{int}}{2}; x_i + \frac{\Delta_{int}}{2}\right] \qquad \text{If those values are acceptable for } x_i. \tag{51}$$

In this equation, i denotes one of the parameters to vary, and x and y two sets of parameters. The *interval* can be constant for all values of the parameters, except when said parameter is too close to one of its boundaries (e.g. all parameters in this model need to be greater than 0). In that case, the *interval* has to be cropped to avoid selecting an impossible number, and $1/\Delta_{int}$ is replaced by $1/interval\ length$. In the cases where the *interval* would be identical for y_i and x_i, however, this method allows one to write that $q_i(y_i|x_i) = q_i(x_i|y_i)$, and if it is true for all parameters, $\frac{q(x^{(i-1)}|y^{(i)})}{q(y^{(i)}|x^{(i-1)})} = 1$, meaning that the main factor determining the Metropolis-Hastings progression will be how close the model is to reality, not how the model is varied.

This way to define q is good for parameters for which the order of magnitude is certain, like time delays. However, for a parameter like K_a or K_g, in the Monod equation, the order of magnitude is not known, K_a and K_g can vary between 10^{-1} and 10^{-5} g/L. If the *interval* were defined as 10^{-6}, then it would be possible to reach a good precision if the optimal value is around 10^{-5}, but the algorithm would need at least 10^5 iterations to reach a value of 10^{-1}, if it were required. On the other hand, if the *interval* was chosen to be 10^{-2}, the algorithm would be able to reach all values in a reasonable time, but it would be very unlikely for the algorithm to find anything with a precision of 10^{-5}, and there would be a high chance that $q_i(y_i|x_i) = q_i(x_i|y_i)$ would not be right anymore, as the algorithm may edge very close to 0. Thus, for those parameters whose order of magnitude is unknown, the variation was applied to the logarithm of the parameter, as described in Equations 52 and 53

$$q_i(\log_{10}(y_i)|\log_{10}(x_i)) = \begin{cases} \frac{1}{\Delta_{int}} & \text{if } \log_{10}(y_i) \in \ interval \\ 0 & \text{elsewhere} \end{cases} \tag{52}$$

where

$$interval = \left[\log_{10}(x_i) - \frac{\Delta_{int}}{2}; \log_{10}(x_i) + \frac{\Delta_{int}}{2}\right] \tag{53}$$

If those values are acceptable for x_i.

Effectively, tose equations use $\log_{10}(x_i)$ as a parameter instead of x_i, and maintains all the advantages previously mentioned.

The list of all the parameters that were used, the intervals, and the type of variation applied (linear or logarithmic), is summarised in Table 5.

Table 5: Table of the different parameters, their initial values in the model, and the way they are varied. If the variation type is indicted to be logarithmic instead of linear, then Δ_{int} is an interval applied to the logarithm of the parameter. The acceptable values often contain a larger range than the actual acceptable values, as the algorithm is unlikely to stray too far from the initial values anyways, and it is informative if it does.

Contribution	Parameter	Initial value	Variation type	Δ_{int}	Acceptable values	Unit
Basic switching	H_1	0.1	Log.	0.2	$[0;\infty]$	h^{-1}
	H_2	0.1	Log.	0.2	$[0;\infty]$	h^{-1}
Growth	$\mu_{g,\max}$	1	Log.	0.2	$[0;10]$	h^{-1}
	K_g	0.01	Log.	0.2	$[10^{-5};10^{1}]$	g/L
	Y_{xg}	0.5	Lin.	0.2	$[0;1]$	$g_{cell}/g_{glucose}$
	$\mu_{a,\max}$	1	Log.	0.2	$[0;10]$	h^{-1}
	K_a	0.01	Log.	0.2	$[10^{-5};10^{1}]$	g/L
	Y_{xa}	0.5	Lin.	0.2	$[0;1]$	$g_{cell}/g_{arabinose}$
Time delays	Δt_g	0.1	Lin.	0.02	$[0;\infty]$	h
	Δt_a	0.5	Lin.	0.02	$[0;\infty]$	h
Responsive switching (initial forcing)	Arabinose threshold	0.002	Log.	0.2	$[0;\infty]$	g/L
	Δt_i	0.5	Lin.	0.02	$[0;\infty]$	h
	r_a	1	Lin.	0.2	$[0;\infty]$	h^{-1}
	p_{init}	0.9	Lin.	0.2	$[0;1]$	-
Responsive switching (direct response)	r	1	Lin.	0.2	$[0;\infty]$	h^{-1}
	n	3	Lin.	0.2	$[0;\infty]$	-
	k	0.002	Log.	0.2	$[0;\infty]$	g/L

4.5 Additional notes

There are several small points that have to be noted but do not belong in any other section.

The first one is related to the data communicated for the model. The difference between models and datasets was evaluated using the logcosh function. As already mentioned, this functioned measures the $\text{mean}(\log(\cosh(y_{predicted} - y_{measured})))$. This number is not really meaningful by itself, it only has meaning when compared to other values. Therefore, in the data presented, a "corresponding" error is mentioned. This error is computed as the inverse function of the logcosh, as shown in Equation 54:

$$\text{Corresponding error} = \text{acosh}(\exp(\text{mean}(\log(\cosh(y_{predicted} - y_{measured}))))) \tag{54}$$

When it is expressed in %, it is simply 100 times the value computed here.

Another point regards the evaluation of the fitting. While it is good to have a model that fits the data, a common problem with fitting a lot of parameters is over-fitting, where the model becomes much closer to the data then what should be ideal, because several sets of parameters could fit the data, but the best one is not the one that would fit a slightly different dataset. To control for this, the optimised models were compared to an alternate dataset, shown in Annex A.2.4. It is a combined experiment where the first 55 h were run with the segregostat regulation scheme, and the rest of the experiment was run with the chemostat regulation scheme. Since this dataset is independent from both the datasets used for fitting the models, and since it has both segregostat and chemostat data in it, a good model that was not over-fitted should match with the data for the entire experiment, while an over-fitted model would at best match correctly a single part of the experiment. The logcosh of the optimised models and the equivalent errors were also computed with the new dataset, and are the values described as "new data" in Annexes A.5, A.6 and A.7, which feature the results of the Metropolis-Hastings algorithm.

Finally, it is worth noting that the data for the proportion of GFP+ cells does not change much after the first 20 h, whether it is for the segregostat or the chemostat. Therefore, one of the tool that was used to estimate the over-fitting was the mean value after 20 h. Three segregostat and three chemostat datasets were evaluated in this manner, which corresponds to the beginning of all experiments including the ones from Section 2 and the ones from Annex A.2, excluding the last one (Figure A.2.5) which did not follow the experimental same protocol. In case they were a simple experiment, the mean proportion, going from the first measurement after 20 h to the last measurement, usually around 80 h. In case it was combined experiment, only the first part was used. The mean proportion was therefore computed from the first measurement after 20 h to about 50 h, a bit before the start of the second regulation scheme.

5 Results and discussion

First, a base model, derived from the analysis of Thattai and Van Oudenaarden [1], and shown in Equations 3 to 6, was briefly analysed. It was found unadapted to describe the dynamics of the chemostat experiments, and more complex solutions were therefore considered. Several different contributions, graphically summarised in Figure 7, were combined in different ways to form 24 models.

These models, numbered 0 to 23, are enumerated in Table 4. They include different combinations of 4 types of contributions whose mathematical descriptions are briefly summarised in Table 3. Each model was optimised thanks to a Metropolis-Hastings algorithm, and these optimisations will allow to determine which contributions are important or not to better reproduce the experimental observations. In this manner, in the next sections, it will be possible to determine which extended model is best suited to describe the data, as well as its strong points compared to the base models, and its shortcomings compared to the data.

5.1 Base model

The base model that was derived from Thattai and Van Oudenaarden [1] is described in Equations 3 to 6. If it was to be applied to the chemostat experiment, then one basic assumption that could be made, considering that no arabinose or glucose is ever detected, is to consider the arabinose and glucose concentrations constant. In that case, the equations can be written as Equations 55 to 58:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mu_a A - DA - H_1 A + H_2 G \tag{55}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = \mu_g G - DG + H_1 A - H_2 G \tag{56}$$

$$0 = (C_{a,in} - C_a)D - \frac{\mu_a A}{Y_{xa}} \tag{57}$$

$$0 = (C_{g,in} - C_g)D - \frac{\mu_g G}{Y_{xg}} \tag{58}$$

μ_a and μ_g are constant as well in those equations, as they depend on the sugar concentrations. Equations 57 and 58 can therefore not be considered, as they would imply that the entire system reached steady-state. This means that the sugar concentrations cannot be considered exactly constant. However, as a first rough approximation, one could approximate $\mu_a \simeq \mu_{a,0}$ and $\mu_g \simeq \mu_{g,0}$. It is known that they vary, but at first it is only considered that they do not vary enough for it to matter.

This assumption lets us resolve a simplified version of the equation, shown in Annex A.4. It shows that either the approximation does not stand, or, at best, the cell concentrations can only monotonously tend towards their equilibrium values. This is clearly not what happens in the chemostat, as the concentration of activated cells first increases then decreases, as shown in Figure 2.a.

Thus, a slightly less constraining but much more complicated approximation was tested: if μ_a and μ_g are not constant, then they may be approximated as mentioned by Equations 7 and 8, which are

reproduced below:

$$\mu_a = \mu_{a,\max}\frac{C_a}{K_a + C_a} \simeq \mu_{a,0} + \frac{\mathrm{d}\mu_a}{\mathrm{d}C_a}[C_{a,0}] \cdot (C_a - C_{a,0}) + \mathcal{O}\left((C_a - C_{a,0})^2\right)$$
$$\mu_g = \mu_{g,\max}\frac{C_g}{K_g + C_g} \simeq \mu_{g,0} + \frac{\mathrm{d}\mu_g}{\mathrm{d}C_g}[C_{g,0}] \cdot (C_g - C_{g,0}) + \mathcal{O}\left((C_g - C_{g,0})^2\right)$$

The same approximation must then be made for $A \simeq A_0 + (A - A_0) + \mathcal{O}\left((A - A_0)^2\right)$ and $G \simeq G_0 + (G - G_0) + \mathcal{O}\left((G - G_0)^2\right)$ where $(A - A_0)$ and $(G - G_0)$ are considered small. Without those approximations, the system cannot be linearised, which makes it probably impossible to solve analytically. With those equations, the entire system can be linearised, and it can be checked whether A and G can present oscillations, that would indicate the model may be used to simulate the chemostat.

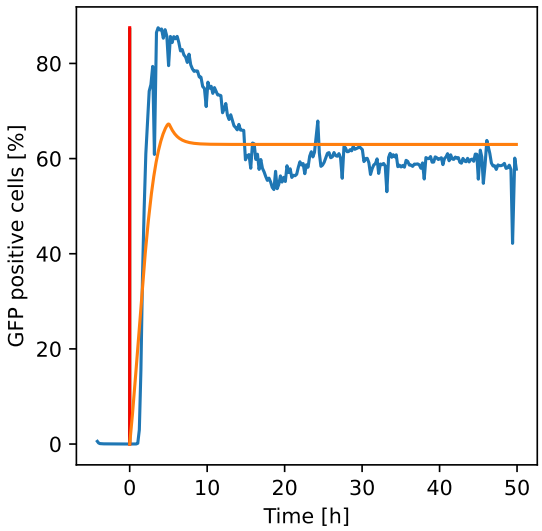
The equations were solved using sympy and showed that the system may indeed present oscillations. However, the solution is very complicated and not much more than that can be said regarding the amplitude or length of said oscillations; the whole solutions are available in the Gitlab repository. It is also good to note that those equations are based on the hypothesis that $(A - A_0)$ is reasonably small, which may not be true as it varies from 0 g/L to its final value, at the very least.

Finally, a basic sensitivity analysis was made, to try and identify which parameters could provide results similar to what is observed in reality. No real good fit was found, however, and while putting graphs here may be illustrative, none of the plots made during the sensitivity analysis came even close to what was found during the optimisation with the Metropolis-Hastings algorithm.

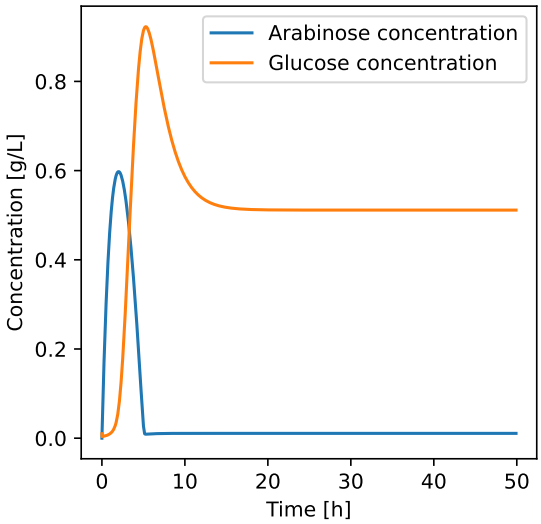
Two possibilities to create a local maximum were made apparent by the optimisation. In one case, it is possible to create a small local maximum, because at the beginning of the chemostat, there are not many activated cells. This lets the arabinose accumulate in the system, and the activated cells can temporarily grow faster. However, this maximum is far from being large enough to even resemble what is seen in reality. This possibility is depicted in Figure 11.

The other possibility it found was the best one for the model, when it comes to fitting the activated cell proportion, and thus it is the one that is mentioned in Annex A.6, where the resulting sets of parameters for the Metropolis-Hastings optimisation for the chemostat dataset are shown. The result is represented in Figure 12. This set of parameters lets the model get much closer to the trend observed in reality. However, the way it does that is by first letting the switching increase the proportion dramatically, then letting dilution and the slower growth of the activated population compared to the inactive one drag the proportion down. This causes the cell concentrations to reach equilibrium close to 0 g/L. Since the behaviour of the cell concentration on the long term is unknown, but should stay reasonably constant after the first few hours (as the OD_{600} does not change much and represents a dense population), having cell concentrations that fall close to zero is not possible.

Thus, as a conclusion to this section, it can simply be said that the base model is not sufficient to describe any of the data meaningfully.



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line indicates the beginning time of the continuous culture in the experiment.



(b) Predicted sugar concentrations by this set of parameters for the model.

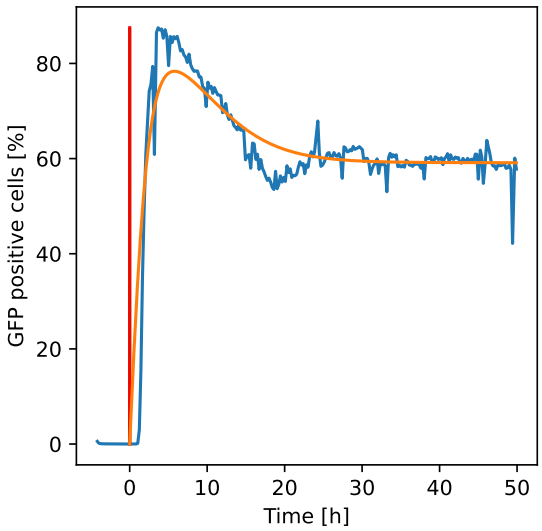
Figure 11: This was the resulting prediction for one of the best sets of parameters found by the Metropolis-Hastings algorithm for the base model. It does present a maximum, thanks to the accumulation of arabinose at the beginning of the experiment, but its dynamics and amplitude are far from reality. The parameters and results for this behaviour were: logcosh= 3.8×10^{-2}, equivalent error of 8.7%, $K_g = 6.0 \times 10^{-3}$ g/L, $K_a = 4.2 \times 10^{-3}$ g/L, $\mu_{g,\max} = 0.11$ h^{-1}, $\mu_{a,\max} = 1.01$ h^{-1}, $Y_{x,g} = 2.9 \times 10^{-2}$ g/g, $Y_{x,a} = 0.98$ g/g, $H_1 = 0.32$ h^{-1} and $H_2 = 0.16$ h^{-1}.

5.2 Convergence of the algorithm

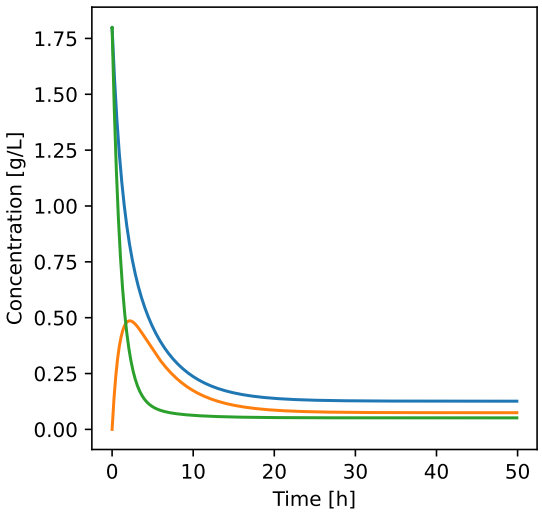
The Metropolis-Hastings algorithm is a stochastic optimisation algorithm assured to converge eventually[26]. However, there is no theoretical assurance that it will converge in any reasonable amount of time. Therefore, here, the convergence will be verified, regardless of the exact model considered.

The first thing that can be verified is that, indeed, the algorithm finds better solutions than the default initial parameters. Plotting the computed logcosh versus the number of iterations made shows several things. Firstly, the algorithm usually converges towards a solution that an equivalent error of less than 20%, or a logcosh lower than about 0.02, within the first 100 iterations at most, as shown for example in FIGURE 13. Depending on the model, the best value found at the beginning of the plateau for the algorithm may be much lower. This plateau is caused by the algorithm having found a set of parameters that is not too bad, then exploring the neighbouring landscape to search for slightly better sets of parameters. In rare cases, when the default values already fit very well the data, the initial decrease that is observed in FIGURE 13 is not visible, and the improvement is much slower, as seen in FIGURE 14.

The initial drop in logcosh, before the plateau means that the algorithm is indeed hindered by the presence of local minima. While it sometimes accepts a model that has a higher logcosh, it is too unlikely for the model to go back towards values as high as for the initial parameters. This leads to



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line indicates the beginning time of the continuous culture in the experiment.



(b) Predicted cell concentrations; in blue, the total concentration; in orange, activated cells; in green, inactive cells.

Figure 12: The base model derived from Thattai and Van Oudenaarden [1], when optimised for the chemostat, shows that it is possible to get the same behaviour as the data for the proportion of GFP+ cells, but at the price of having cell concentrations that tend towards values close to 0.

different minima when running the same algorithm several twice, and means that it is uncertain whether the minimum found is actually close to the absolute minimum.

There are two reasons why this should not be a problem:

1. As will be said later, none of the models can fit correctly both the segregostat and the chemostat data at once, so most of the results that will be found are not quantitative in any case. Rather, the trends shown by the different models will be investigated.

2. Each contribution is represented in several models, and there are only so many ways the different contributions can interact together. While it is possible that not all the possible interactions were found, it is likely possible to deduce the once that are not seen, by contrast to the ones that are.

The fact that no quantitative result is expected also means that having 300 iterations for the algorithm, instead of 600, does not cause any harm: in any case, the algorithm has had time to settle and find likely candidates for the best sets of parameters; those candidates may not be the exact best and the algorithm may be able to find something much better, but in general it does not, and in any case the general shape taken by the optimal model found is unlikely to change too much without increasing the logcosh value too much.

5.3 Extended models

The resulting sets of parameters for the different models are presented in Annexes A.5, A.6 and A.7. All the plots for the best models found can be browsed online at `https://gitlab.uliege.be/mipi/`

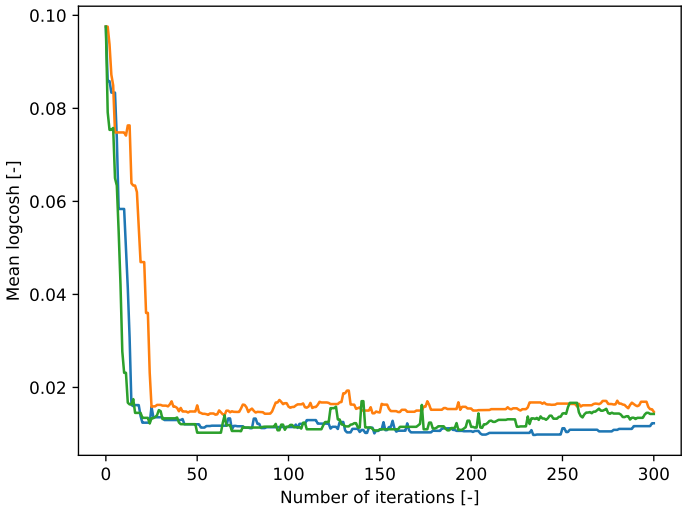


Figure 13: Representative evolution of the mean logcosh over the number of iterations for three different runs of the algorithm. In this case, this is the optimisation of model number 6 for both datasets at once. There is an initial drop of the values followed by a plateau. This is one of the highest curves overall, but the initial logcosh and the plateau are sometimes much lower, depending on the model.

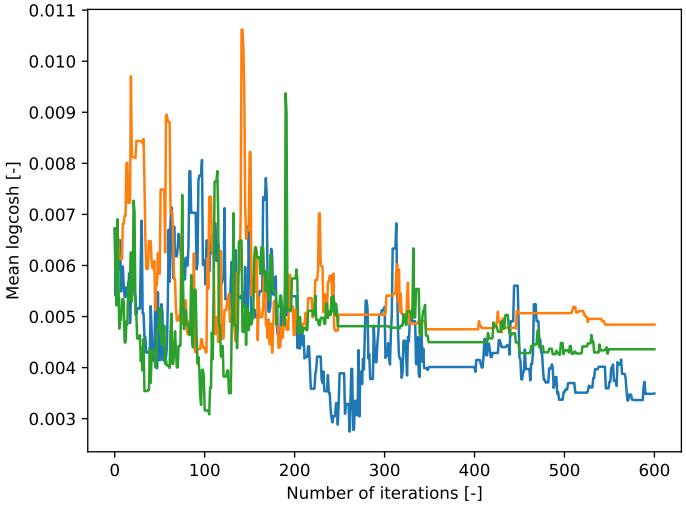


Figure 14: Outlier of the mean logcosh over the number of iterations for three different runs of the algorithm. In cases like this one, there is no initial drop, and only a plateau, as the initial values considered for the optimisation are close to a minimum logcosh.

`standalone/vincent-master-thesis/-/tree/master/results`, and only the relevant ones will be included here. In this section, first, several general observations will be made. Then, each contribution will be analysed. The order of the observations presented here is not the same as in the MATERIALS AND METHODS section. Instead, the most certain observations will be made first, as the least certain ones are linked to each other.

5.3.1 General observations

The first thing that can be noted is that some models do not make sense at all, for various reason. Firstly, models 0 to 3 (see TABLE 4) have neither basal switching nor reactive response. In other words, there is no way, according to those models, for bacteria to form a population of activated cells, and the latter's proportion in the system is always predicted to be 0%. This is obviously false, and the four models were not even optimised.

Similarly, there are two other impossible models, although they are less obvious. Models 9 and 11 use co-consumption, no basal switching, and the initial forcing. The absence of switching, in combination with the initial forcing, mean that once the population has switched, there is no more biomass exchange between the two phenotypes. Since there is co-consumption, and no other contributions after the initial forcing is done, EQUATION 59 can represent the system at that time, and EQUATION 60 can be deduced for both models as the time delay would not change the equations singificantly.

$$\begin{cases} \frac{\mathrm{d}A}{\mathrm{d}t} &= -DA + (\mu_g + \mu_a)A \\ \frac{\mathrm{d}G}{\mathrm{d}t} &= -DG + \mu_g G \end{cases} \tag{59}$$

$$\Rightarrow \frac{\mathrm{d}A}{\mathrm{d}t} = (-D + \mu_g + \mu_a)A \geq (-D + \mu_g)G = \frac{\mathrm{d}G}{\mathrm{d}t} \qquad \begin{array}{l}\text{Because } \mu_g,\ \mu_a \geq 0, \\ \text{And as long as } A > G.\end{array} \tag{60}$$

Since in all experiments there is a point close to the beginning where $A > G$ is desired, and this point is reached thanks to the initial forcing, it means those models can only predict that the proportion $A/(A+G)$ would be growing for the remainder of the experiment. An illustration of the phenomenon is provided in FIGURE 15.

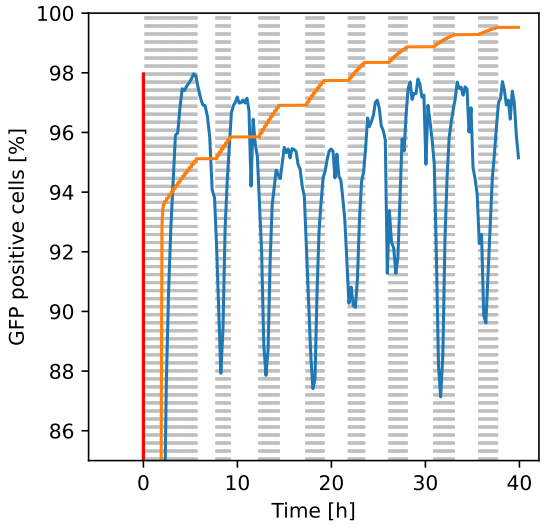
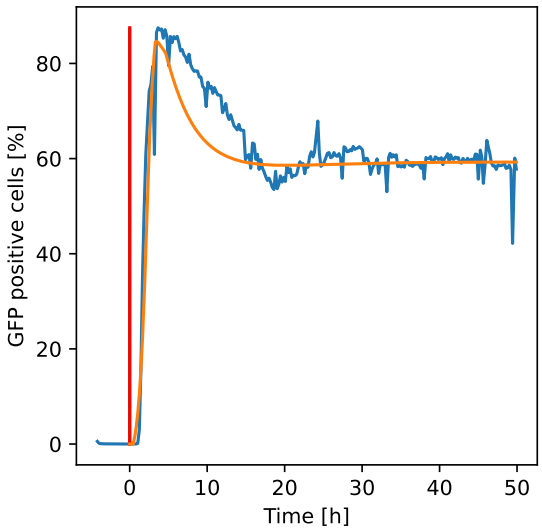


Figure 15: The model number 9 has initial forcing, co-consumption, no basal switching, and no time delay. In this case, the second one is the most important: co-consumption means that activated cells can only grow at least as fast as inactive cells, in all conditions. In this figure, the model number 9 (in orange) was fitted to the segregostat data (in blue). It shows that the best that can be done to fit to the observed oscillations, with this model, is a constant proportion of activated cells when no arabinose is in the system, and a slow increase when there is. The red line represents the beginning time of the continuous phase in the experiment.

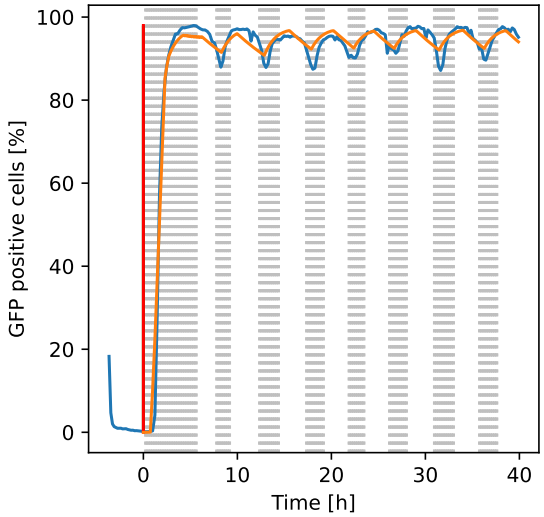
Thus, when talking about "all models" later, models 0 to 3, 9, and 11 are implicitly excluded.

For the remaining models, it is interesting to check whether the equations can explain both the segregostat and the chemostat at once. FIGURES 16 and 17 are representative of what happens for all models and show clearly that no, it is not possible. In the first one, the data is compared to the original datasets used for the optimisation, while in the second an experiment starting as a segregostat and transitioning to a chemostat is used as an independent reference. In the comparison to the original datasets, when optimising for the chemostat, the models consistently show a mean proportion of activated cells, after 20 h, around 60%, which is coherent with chemostat observations[11].

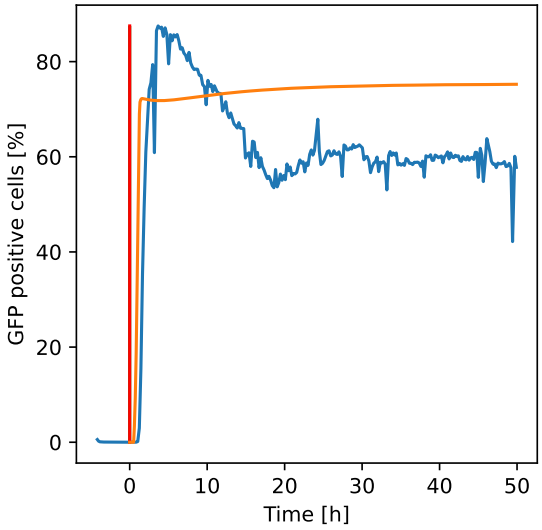
[11] Of 3 experiments starting with the chemostat, one did not reproduce the local minimum after 20 h (see FIGURE A.2.3); it gave a mean of 77%. The other two gave means of 51 and 57%, with 57% corresponding to the dataset for which the models was optimised, further confirming that the model fits its mean value to the data.



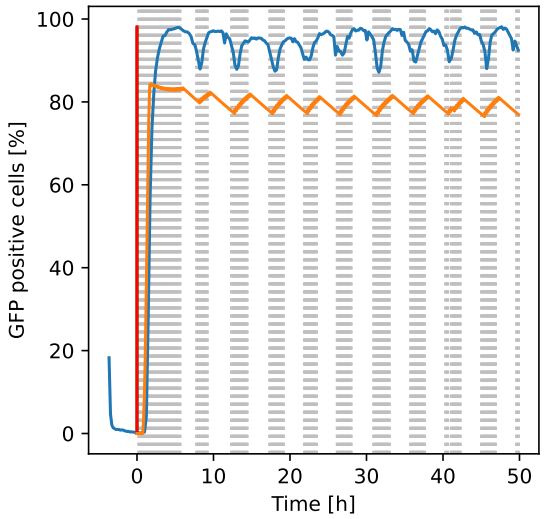
(a) Model fitted to the chemostat.



(b) Model fitted to the segregostat.



(c) Model fitted to both experiments.

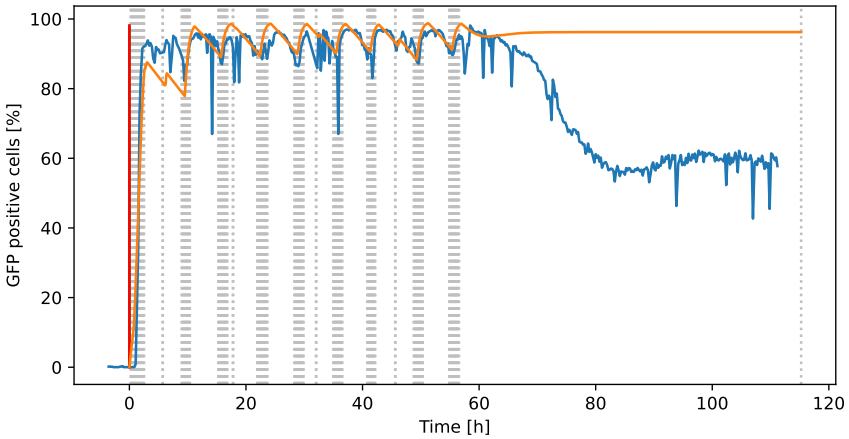


(d) Model fitted to both experiments.

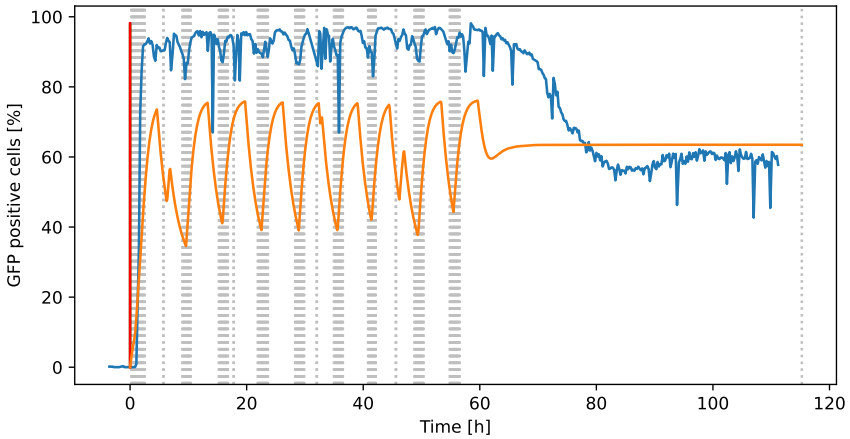
Figure 16: This is a representation of how model number 15 (in orange) can be fitted to either dataset (in blue), in (a) and (b), but cannot be fitted to both at once, in (c) and (d). The model number 15 in particular has co-consumption, basal switching, initial forcing, and time delays, but the same can be said of other models. The red lines are the beginning times of the continuous phases for the experiments.

When optimising for the segregostat, similarly, the mean proportion after a few hours is around 90%, and that is coherent with data showing mean proportions, after 20 h, of $93 \pm 1\%$. However, when optimising for both at once, the mean values obtained by the various models, whether in the chemostat (after 20 h) or in the segregostat, are between 70 and 80%, which is between the chemostat and the segregostat values for the targeted experiments. This is true even when comparing to the independent dataset: in all cases, the mean predicted proportion after the first 20 h is roughly the same for both parts of the experiments, and similar to the ones predicted on the training datasets. Therefore, it can be concluded that no model can be optimised for both datasets at once.

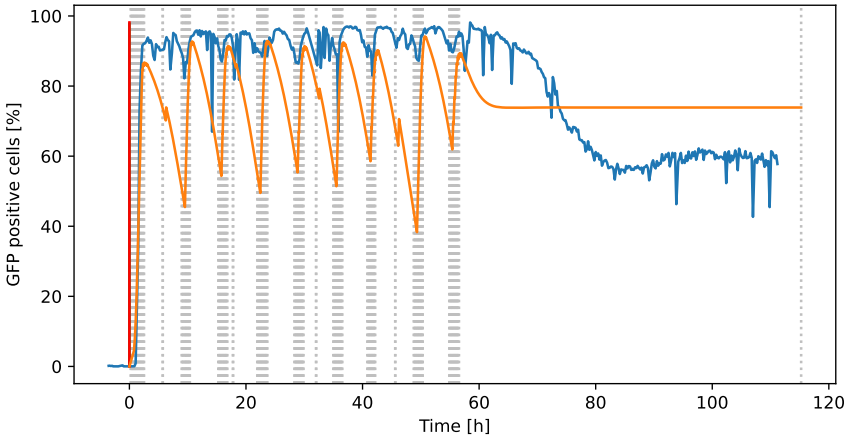
Something else that can be noticed is how realistic, or unrealistic, the various optimised parameters



(a) Model fitted to the segregostat dataset



(b) Model fitted to the chemostat dataset



(c) Model fitted to both experiments at once

Figure 17: Here, the predictions for model number 6 (in orange) are compared to an alternative dataset (in blue), which is a combined experiment with the segregostat coming first. The same phenomenon as on FIGURE 16 can be observed, on the alternate dataset: when the model is fitted to the segregostat, it can only fit segregostat data, and when it is fitted to the chemostat, it can only fit chemostat data. When it is fitted to both, it can fit neither correctly and ends up with mean predicted values between the ones of the segregostat and the chemostat. The red lines represent the beginning time of the continuous cultivation phase in the experiment.

are. It can be observed that four of the six parameters related to growth, $\mu_{g,\max}$, $\mu_{a,\max}$, Y_{xg} and Y_{xa}, are often unrealistic, with $\mu_{\max}$ values of up to nearly 9 h^{-1} when they were supposed to be around 1 h^{-1}, yield values close to 0 or 1 g/g when they should be close to 0.5 g/g. K_g and K_a were within the expected range (10^{-1} to 10^{-5}), even though they were given freedom to go much higher, but with the other parameters out of bound it does not mean much. All the other settings were realistic and relatively constant in comparison.

While one might think that constraining the values for the different parameters may yield more realistic results, the fact that no model can be optimised for both the chemostat and the segregostat at once means that there is no real point to doing so before correcting the source of the incoherence, which will be explored in the rest of the document. The results for the constrained optimisation of the models as they are would at best yield similar results to the ones obtained without much constraints, and at worse would be less informative regarding the contributions that have realistic parameters.

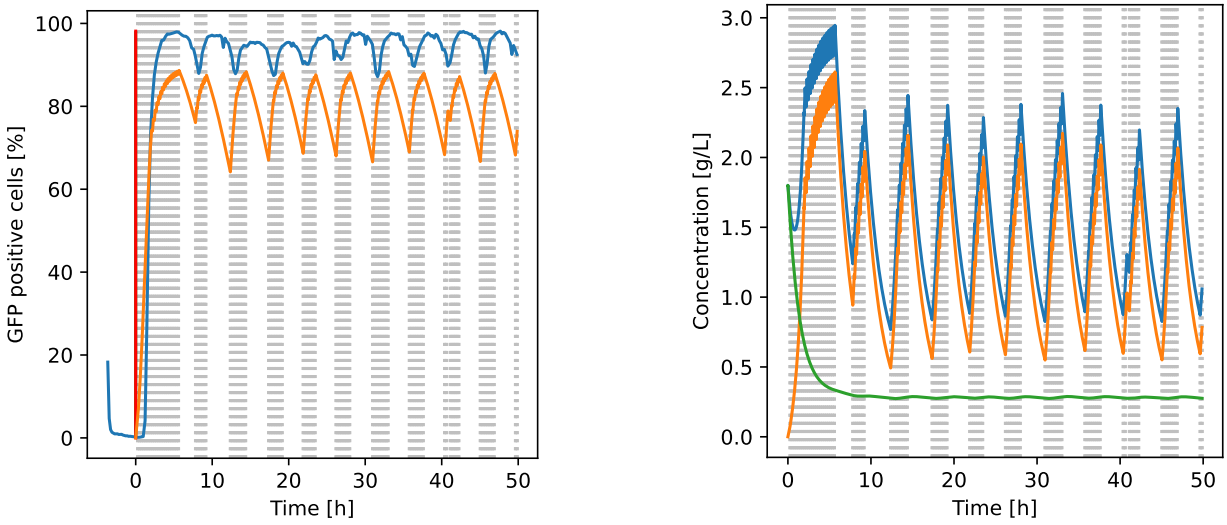
5.3.2 Direct responsive switching

The first contribution that will be discussed is the direct responsive switching. The most informative way to analyse this contribution is to investigate the behaviour of the optimised models in the segregostat, and to compare the models that did not use responsive switching to the ones that did.

All of the models that do not take into account this contribution could produce some kind of oscillations, even if they were optimised only for the chemostat, with a only a few exceptions whose causes will be mentioned later. The oscillations were caused by the variations in activated cell concentration, as illustrated in FIGURE 18.

However, models that do consider responsive switching were found to exhibit several different behaviours, none of which matching with the behaviour previously described. Instead, they present three different behaviours:

- In case there is oscillation observed, the oscillation is mainly caused by the responsive switching, and not by growth, as illustrated in FIGURE 19. This does not correspond to what is observed in segregostat experiments.
- In the previous case, the arabinose was depleted at about half the time without arabinose injected, and the switching started at that point, which is indicated by the maximum cell concentrations. In a few cases, the model resulting from the optimisation towards the segregostat experiment was tuned to have its depletion of arabinose just before the next injection of arabinose. This provokes much sharper drops in activated cell proportions, as shown in FIGURE 20, because the Hill function is very sensitive around the threshold value. This shows the reliance by the model on the frequency of the system, as having small changes in frequency around the time the maximum proportion arrives at would massively change the shape of the curve. This is further illustrated in FIGURE 21, where two different experiments are considered. The phenomenon is a limit of having only two sub-populations considered, with no real delay between the time cells detect a lack of arabinose and the time they start switching, like for the initial forcing (even when a time delay is considered in the system, the detection is delayed, but cells start switching back to their inactive phenotype as soon as the lack of arabinose is sensed). In reality, cells would wait to "see" if the



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line is the beginning time of the experiment.

(b) Predicted cell concentrations; in blue, the total concentration; in orange, the activated population; and in green, the inactive population.

Figure 18: In a typical model without responsive switching, oscillations are due to the activated cell concentration greatly increasing, while the inactive cell concentration barely changes, just like what is observed in experiments. This behaviour is observed in most models without responsive switching. In this case, the base model, optimised for both experiments at once, illustrates the point.

concentration stays down or up, before changing, and there would be a delay dependent on the exact concentrations sensed [21].

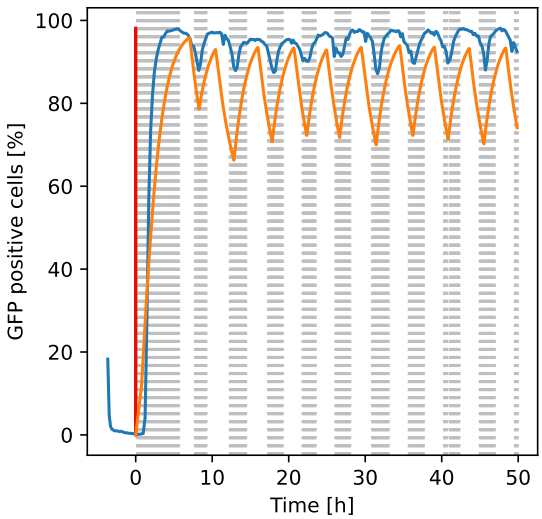
- In case there is no oscillation observed, it is because the algorithm could not find a combination of parameters where the switching was dominant. Rather, it found a "better" solution for the proportion, but the cell concentrations reach steady-state close to 0 g/L as there is too little growth, as shown in FIGURE 22, exactly like for the base model in FIGURE 12.

Overall, those results show that responsive switching, as it was implemented, does not represent reality at all. This may be because the scale of the switching rate is too large, but in addition to that this type of response is too sensitive to arabinose concentration, and something should be implemented to take that into account.

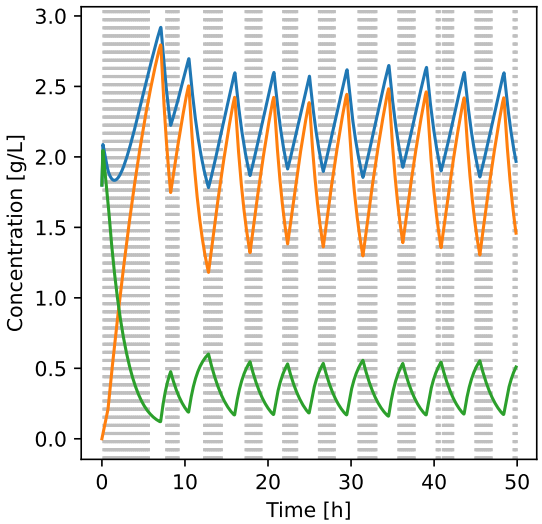
5.3.3 Time delays

Time delays are probably the simplest contribution to analyse, if one does not consider the models with responsive switching.

Firstly, for the time delay on glucose use, whenever one is considered, it is optimised to about 0.1 h or less; in particular it is often found to be of the order of 10^{-2} h when the optimisation is made for the segregostat, and the segregostat was the experiment for which a time delay could have mattered, as there were cyclical changes. Considering that the measurements are made every 12 minutes on average, or about 0.2 h, a shorter delay for glucose use can be considered noise, or at least inconclusive.

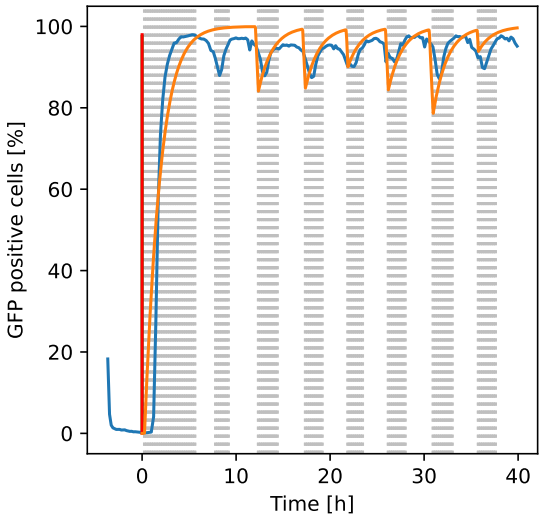


(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line represents the beginning time of the continuous phase in the experiment.

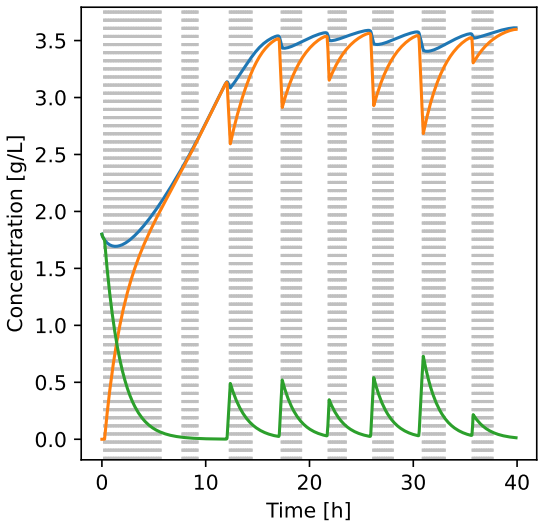


(b) Predicted cell concentrations; in blue, the total concentration; in orange, the activated cell concentration; in green, the inactive cell concentration.

Figure 19: In a model with responsive switching showing oscillations, said oscillations are directly linked to switching between the two subpopulations, as shown in (b). In this case, the model number 23, with responsive switching, basal switching, co-consumption, and time delays, optimised for both experiments at once, illustrates the point.

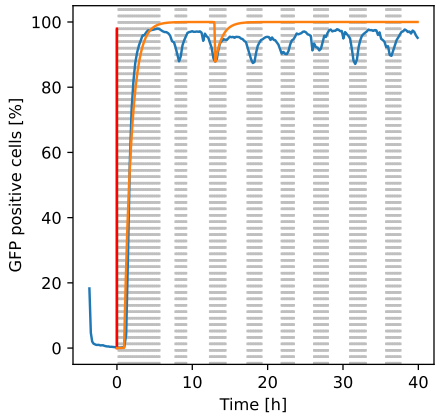


(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line represents the beginning time of the continuous phase in the experiment.

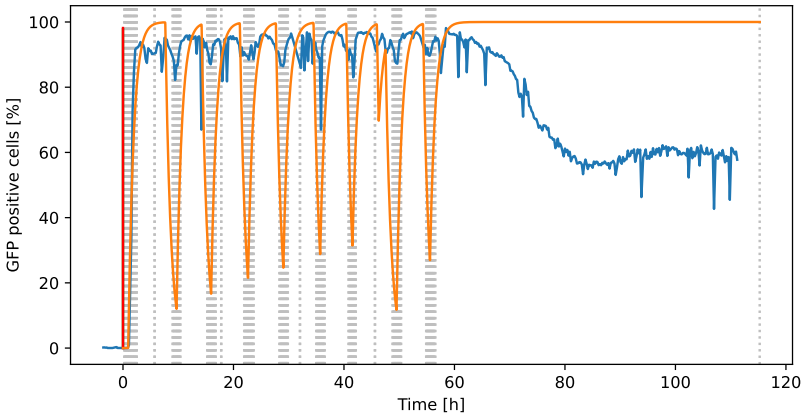


(b) Predicted cell concentrations; in blue, the total concentration; in orange, the activated cell concentration; in green, the inactive cell concentration.

Figure 20: The model number 17 is a model with direct responsive switching and co-consumption only. In this case, it was optimised for the segregostat only. There are sharp drops that appear just before the addition of arabinose, rather than the nice oscillations in some other models or optimisations, as the responsive switching was optimised to be very sensitive to low concentrations of arabinose.

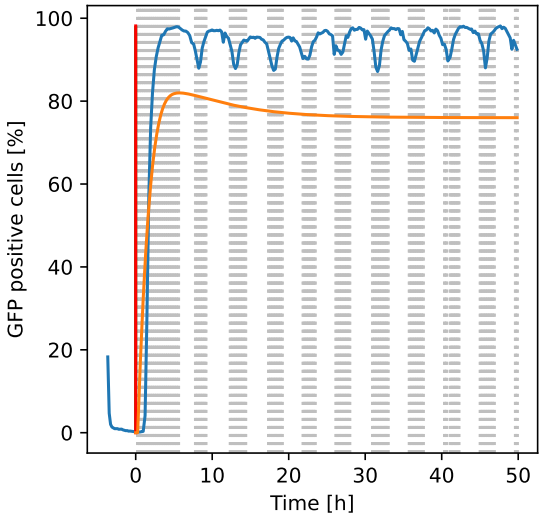


(a) In blue, the measured proportions of GFP+ cells during the segregostat experiment; in orange, the predicted proportions.

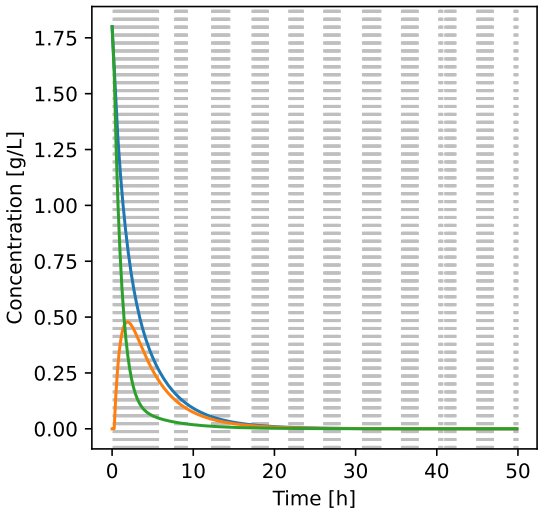


(b) In blue, the measured proportions of GFP+ cells during a combined experiment with the segregostat first; in orange, the predicted proportions

Figure 21: The model number 19 has co-consumption, time delays, no basal switching, and direct responsive switching. In this case, it was optimised for the segregostat only. In the experiment that was the base of the optimisation, only a single drop in proportion appears, because there was slightly less arabinose pulses in the previous series. The drop appears while arabinose is being added because of the time delay. In the second experiment shown, various differences in amounts of arabinose injected and time differences between pulses cause various predicted activated cell proportion drops. The red lines indicate the beginning times of the continuous phases in the experiments.



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line represents the beginning time of the continuous phase in the experiment.



(b) Predicted cell concentrations; in blue, the total concentration; in orange, the activated cell concentration; in green, the inactive cell concentration.

Figure 22: The model number 16 has no co-consumption, no time delay, no basal switching, and direct responsive switching. In this case, it was optimised for both the segregostat and the chemostat. Instead of oscillating, the proportion of activated cells stabilizes close to the mean value between the chemostat and the segregostat; and the cell concentration needs to fall close to zero for that to happen.

The current experiments should therefore be considered without delay for glucose, barring any new information.

Secondly, there is the time delay for arabinose. Since the time delay was put in place to respond to the delay observed after arabinose was added in the segregostat, the segregostat predictions are the ones that should be improved. Thus, here, only the optimisation for the segregostat is considered.

Comparing the optimised models directly is complicated, as some parameters are very different between any pair of models. Thus, instead, the parameters of the models without time delay were applied to the models with time delays. The time delays used were 0.1 h^{-1} and 0.5 h^{-1}, respectively, for the glucose and the arabinose, as those were the orders of magnitude of delays settled on by all the optimisations. Then, comparing the different models to the data showed that the time delay usually brings an increase of about 10% in the precision of the model, as seen in Table 6. Since they were determined as incorrect, the models using direct responsive switching were not considered here. The only model that did not show any improvement is one where no real oscillation was present, as Metropolis-Hastings converged towards a system where all concentrations fell to 0.

Table 6: For this table, the fitted models for the segregostat (first two columns) were compared to the same models (with the same sets of parameters) with added time delays of 0.1 h for glucose and 0.5 h for arabinose (2 following columns). * The average error showed is computed based on the average logcosh of the error on the proportion of GFP+ cells, thus it corresponds to: $100 \cdot \text{acosh}(\exp(\text{mean}(\ln(\cosh(error)))))$. The percent different is computed based on the average error of the model without time delay, as $100(\overline{error}_{no\ delay} - \overline{error}_{delay})/(\overline{error}_{no\ delay})$, where $\overline{error}$ corresponds to the average error previously mentioned.

Models without time delay	Average* error [%]	Models with time delays	Average* error [%]	Percent difference* [%]
4	4.59	6	4.12	10.2
5	4.27	7	3.74	12.3
8	3.35	10	3.29	1.9
12	3.96	14	3.42	13.6
13	3.69	15	3.69	-0.1

This increased accuracy with the time delays shows that, indeed, as can be expected from simple observation, there should be a time delay in the system. Whether said delay should be implemented as it was here is another matter entirely, as there is no real data on whether the sugar concentrations indeed behave as hypothesised, especially considering the strange growth parameters obtained.

5.3.4 Basal stochastic switching

The basal switching contribution was hypothesised to be small, or at least smaller than the growth. With rare exceptions, the optimised parameters for basal switching, H_1 and H_2, have orders of magnitude of 10^{-2} to 10^{-1} h^{-1}. The exceptions were always only one of the parameters being much greater than 10^{-1}.

In the few cases where H_1, the switching rate from activated to inactive cells, was greater than 0.1 h^{-1}, oscillations in the inactive cell concentrations, similar to the ones observed in FIGURE 19, with the responsive switching, could be observed. This proves that the switching rate from the activated phenotype to the inactive one should not be higher than 0.1 h^{-1}. The number itself can be confirmed as the limit by the one model that settled on a switching rate of 0.14 h^{-1}, depicted in FIGURE 23.

On the other hand, for the optimised extended model that has $H_2 \gg 0.1$ h^{-1}, no real incoherence in the cell concentrations can be seen (see FIGURE 24). This raises an interesting point that was thus far unaddressed: there is no real reason, within the model, for the switching rate to be of the same order of magnitude in both directions. In the case of direct responsive switching, the switching rate from the activated phenotype to the inactive one was much greater than 0.1 h^{-1} when there was enough arabinose present in the system, which caused the oscillations in the inactive cell population. In the case of basal switching, having a high switching rate to the inactive state in the models means that whenever the activated cell population increases faster than the other phenotype (which is when there is a lot of arabinose in the system), the predicted inactive population varies more than what is observed. But if there is a constant switching rate from the inactive cell population to the activated cell population, in experiments where the inactive cell concentration is mostly constant, then a high switching rate effectively decreases the observed inactive cell concentration, with no visible impact on the dynamics of the system. The only thing that can be said with the observations made till this point is that, if there were constant switching at the same rate or faster than at the beginning of the experiment, which seems to be about 1 h^{-1}, then a significant part of the cells could always be observed to be "in transition" between their inactive and activated states, in the heat maps of the two experiments (see FIGURE 2.c and FIGURE 3.c). That is not the case, so it can be deduced that H_2 is reasonably low.

Finally, it should be noted that, because biological noise exists for individual cells, the real switching rates are unlikely to be exactly 0 h^{-1}. The real question is whether they are detectable, but here the most important contribution seems to be (and it will be confirmed later) growth, rather than switching, and the models with basal switching are not always better fitted than the ones without. Therefore, while the models with basal switching seem to settle on values between 10^{-2} to 10^{-1} h^{-1}, it cannot really be said at this point that there is any detectable basal switching, only that the basal switching there is is at most of an order of magnitude of 0.1 h^{-1}.

5.3.5 Initial forcing

The direct response, unlike the previous switching contribution, can more easily be analysed using the optimisation for the chemostat, rather than any other. Similar observations to the ones that are described here can be made for the optimisation towards both chemostat and segregostat, but in that case things depend greatly on growth parameters. Since growth parameters are never found to be quite realistic, it is likely they represent the fitting of unadapted growth curves to the model, and fitting for the two experiments at once can make the parameters vary too much to be certain of the significance of the observations.

The reason for the interest towards the chemostat experiment is related to the maximum proportion of GFP+ cells that appears in this kind of experiments, after about 4 hours. The base model, with basal switching only, can reproduce this peak, as seen in FIGURES 11 and 12. However, it comes at a

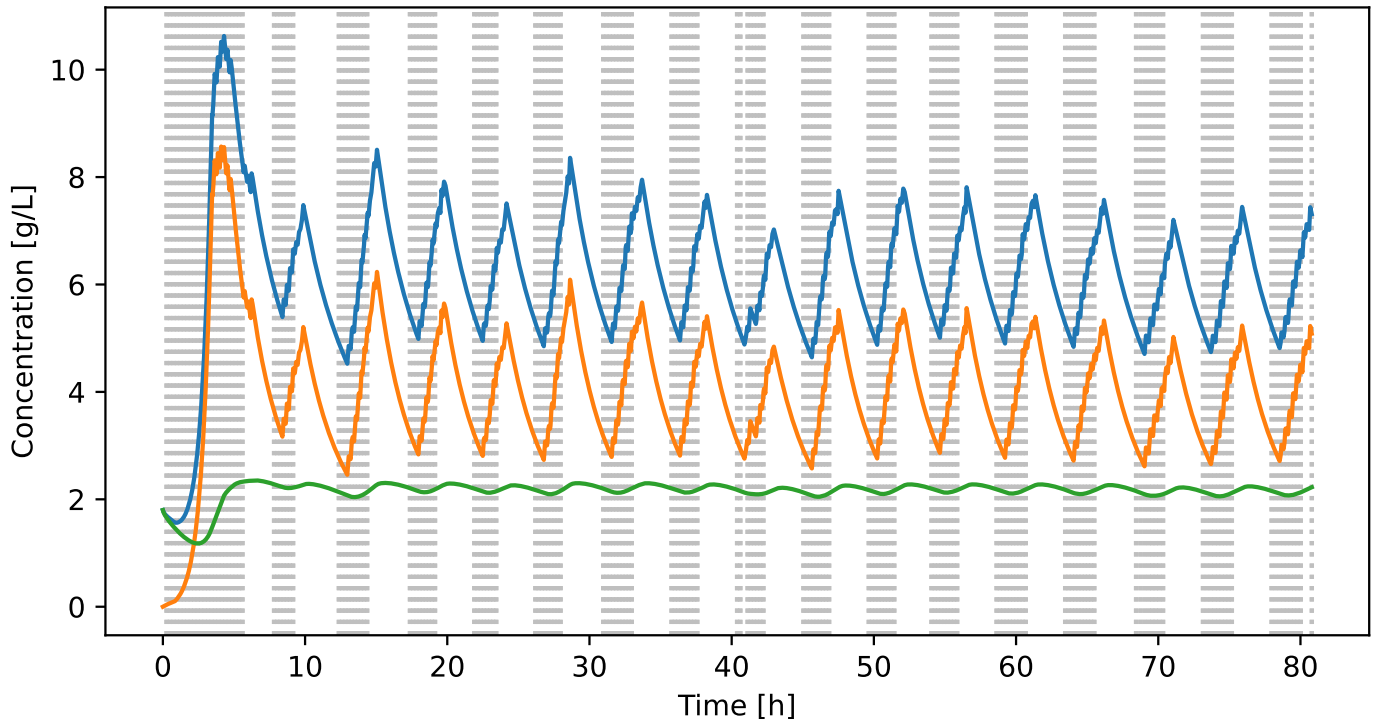


Figure 23: The model number 7, which has co-consumption, time delays, and basal switching, when optimised for the chemostat, had an optimal switching rate H_1 of 0.14 h^{-1}. In this figure, the cells concentration predicted by the model with a segregostat experiment are shown; with the total cell concentration in blue, the activated cells in orange, and the inactive cells in green. The oscillations in the inactive cell concentration, due to switching from the oscillating activated population, start to be visible.

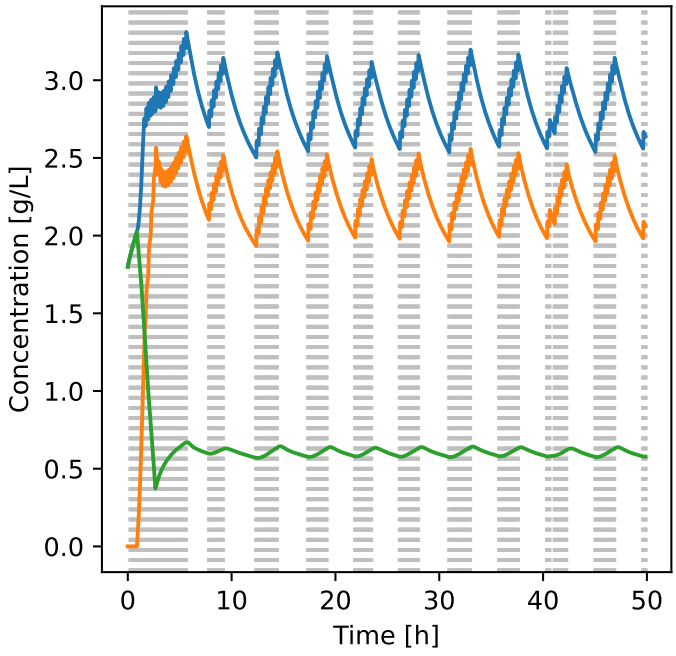
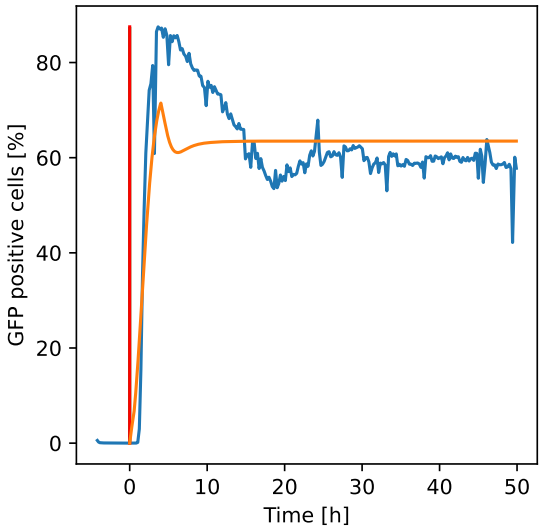


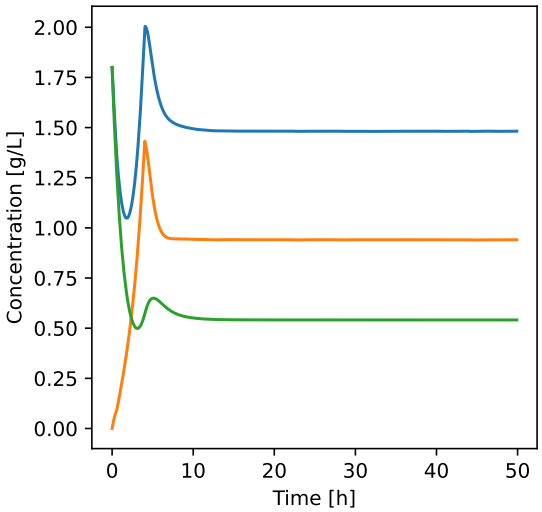
Figure 24: Predicted cell concentrations for the model number 6, optimised to fit both experiments, and shown for the segregostat experiment. In blue, the total cell concentration; in orange, the activated cell concentration, and in green the inactive cell concentration. The optimal switching rates found were $H_1 = 0.22$ and $H_2 = 2.29$ h^{-1}. The relatively high switching rate H_1 from activated to inactive cells causes the small oscillations in inactive cell concentrations, but the very high switching rate H_2 does not cause undue fluctuations as the inactive cell population is nearly constant.

cost: the only way for the model to produce a realistic peak is to have cell concentrations that tend towards a very low value, and otherwise the peak cannot reach the required amplitude.

A time delay does allow a better overshoot of the equilibrium concentration, compared to FIGURE 11, as shown in FIGURE 25. This is because the accumulation of arabinose, which makes the overshoot possible, is amplified by the time delay. Nevertheless, the entire phenomenon is based on the population quickly adapting to an increased arabinose concentration thanks to very different growth rates. Thus a time delay of half an hour cannot produce phenomena at the time scale required to reproduce the observations, as growth rate differences make the system reach equilibrium within a few hours.



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line represents the beginning time of the continuous phase in the experiment.

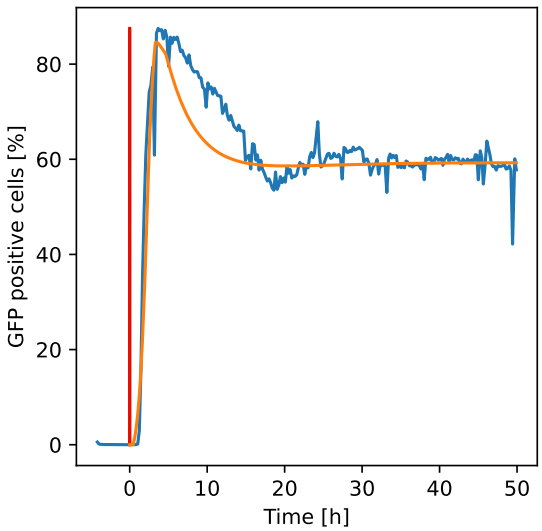


(b) Predicted cell concentrations by model number 6; in blue, the total concentration; in orange, the activated cell concentration; in green, the inactive cell concentration.

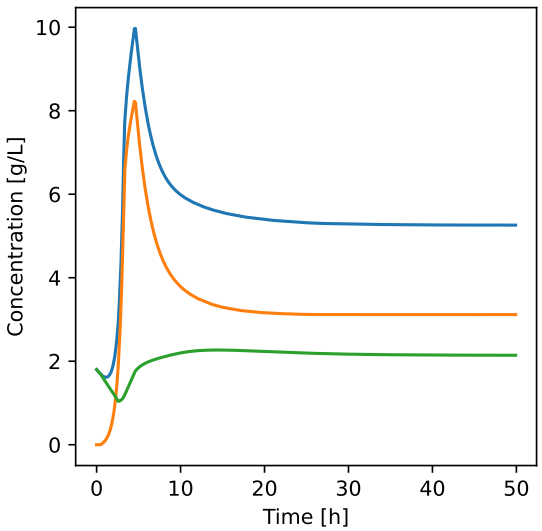
Figure 25: The model number 6 has basal switching, time delays, and no additional contribution. In this case, it was optimised for the chemostat. The time delay allows a bigger overshoot than in FIGURE 11, but the time scale of the phenomenon is measured in hours, not in tens of hours as in the observations.

The experimentally observed decrease in GFP+ proportion, after the overshoot, happens in roughly ten hours. Therefore, it is obvious that the time delay considered in the model is not sufficient to explain this long decrease, and that is where the initial forcing may help. It is composed of two different pieces: there is a delay, attached to the forcing, in addition to the time delay considered for consumption. It causes the GFP+ population to start increasing much faster after an hour instead of the 30 minutes that are required for growth after the normal time delay. This first effect, in theory, should let more arabinose accumulate into the system, but it is difficult to confirm anything as no model has any realistic set of parameters for the growth, and everything is tuned to the data. The second effect has a much higher impact: a part of the population is forced to become activated. This part of the population that switches provokes or magnifies the overshoot. Then, the population decreases more slowly in the model with initial forcing, because without it the growth parameters were made to create a very rapid growth so that the proportion could reach high values very fast. With initial forcing, the growth rate of the activated population that is required to reach the right amplitude at the end of the peak is reduced, as there are much more activated cells present due to switching.

This is the ideal behaviour of the initial forcing, depicted in FIGURE 26. The overall switching rate from inactive to activated cells can be estimated by multiplying the portion of the responsive population by the rate at which said responsive population is forced to switch to the activated phenotype. The results, shown in TABLE 7, highlight that, ideally, for the behaviour described previously, the overall added switching rate is usually between 0.05 and 0.5 h^{-1}. This is of the same order of magnitude as the values usually found for H_1 and H_2, the basal switching rates, which makes sense as growth is still a main contributor to the increasing activated cell proportion within the first hours, the initial forcing simply helps the growth by boosting the population a little bit. 0.05 h^{-1} may seem like very little, but the prediction shown in FIGURE 26 has $r_a p_{init'} = 0.05$ h^{-1}, in this case it is even lower than H_1, and the reason it works is that the initial forcing stops acting after a few hours, when the responsive population has switched. On the other hand, 0.5 h^{-1} may seem like a lot, but the rate at which the proportion changes initially is measured to be around 1 h^{-1}, with a lower limit of about 0.8 h^{-1} in the chemostat, where there are slightly less cells at the peak. That means the difference has to be compensated by growth.



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions by model number 15.



(b) Predicted cell concentrations by model number 15; in blue, the total concentration; in orange, the activated cell concentration; in green, the inactive cell concentration.

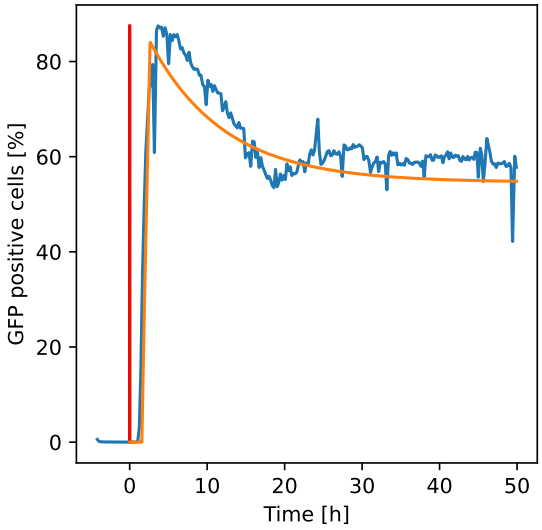
Figure 26: The model number 15 is the most complex model with initial forcing: it has co-consumption, time delays, and basal switching as well. It is illustrative of other models with initial forcing, where the tendencies already shown in FIGURE 25 are amplified. In this particular case, the initial forcing has a global rate $p_{init} \cdot r_a$ of 0.07 h^{-1}, and that it enough for the system to be able to behave much better.

TABLE 7 also contains an outlier, however: for model 13, which has co-consumption, no delay, basal switching, and initial forcing, the estimated switching rate of the population reaches 0.85 h^{-1}. This is because the Metropolis-Hastings algorithm settled, for this model, on another kind of solution which could fit the data. This solution is depicted in FIGURE 27 and fits the data not by boosting the effect of growth, but by forcing most of the cells to switch, then letting them settle to their steady-state concentrations.

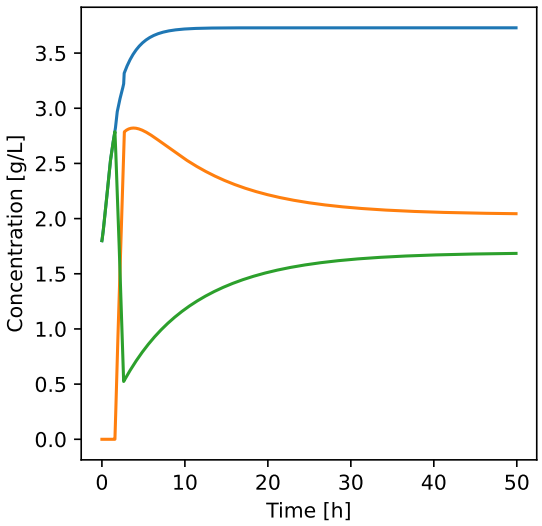
Thus there are two ways the initial forcing can provide the boost needed for the population to present the apparent overshoot. On the one hand, it can improve the effect provided by a faster growth of the

Table 7: Table of the various overall forcing rates for the different models.

Model number	Segregostat optimisation			Chemostat optimisation			Optimisation for both		
	r_a [h^{-1}]	p_{init} [-]	$r_a \cdot p_{init}$ [h^{-1}]	r_a [h^{-1}]	p_{init} [-]	$r_a \cdot p_{init}$ [h^{-1}]	r_a [h^{-1}]	p_{init} [-]	$r_a \cdot p_{init}$ [h^{-1}]
8	0.81	0.60	0.49	0.15	0.21	0.03	0.76	0.65	0.50
10	0.94	0.72	0.68	0.26	0.78	0.20	0.59	0.75	0.44
12	0.09	0.61	0.05	0.13	0.85	0.11	0.54	0.45	0.24
13	0.54	0.74	0.40	0.92	0.93	0.85	0.16	0.33	0.05
14	0.68	0.34	0.23	0.09	0.76	0.07	0.19	0.91	0.17
15	0.51	0.75	0.38	0.24	0.31	0.07	1.24	0.23	0.29



(a) In blue, the measured proportions of GFP+ cells during the experiment; in orange, the predicted proportions. The red line represents the beginning time of the continuous phase in the experiment.



(b) Predicted cell concentrations by model number 13; in blue, the total concentration; in orange, the activated cell concentration; in green, the inactive cell concentration.

Figure 27: The model number 13 has co-consumption, basal switching, and no time delay, in addition to the initial forcing. It is here optimised for the chemostat experiment. With this set of parameters and this model, instead of amplifying an existing phenomenon, the model replaces it by making most of the population switch towards the activated then letting it slowly reach steady-state.

activated population, by temporarily increasing the switching rate to activated cells by a factor 2 or 3 at most. In that case, the total population increases drastically then starts decreasing as there was an overshoot. On the other hand, it is possible for the population to nearly completely switch to the activated state, then stabilise. In the latter case, the overall population does not need to change much, as the switching rate is increased to ten times its normal value, and most of the change observed is due to switching rather than to the population growth. Those are two extreme cases, and one might also consider the intermediate proposition: the boost to switching may be relatively high, higher than what is observed in the optimised models, but the activated cell population may increase very fast as well,

which would create an effect similar to what is seen in FIGURE 26, in terms of cell concentrations, but with different amplitudes depending on exactly how fast cells grow and switch.

The specifics cannot really be determined without knowing the exact growth profiles of the two sub-populations, but one thing can be said for certain: unlike what is observed in the optimised model shown in FIGURE 27, the growth is very likely to play an important role in the initial population distribution. This can be affirmed based on two observations. Firstly, it has already been determined, for the segregostat, that cells do grow much faster when arabinose is added, as too much switching would contradict the observations. While there is no real guarantee regarding the switching, since the entire reasoning behind the initial forcing is that switching could be different at the beginning, there is no reason for the new activated cells created not to grow faster than the inactive cells, much like during the peaks in the segregostat. The second observation is simply this: if one looks at the measured cell concentrations at the beginning of various experiments, one can easily notice, even with the noise present, that the concentration initially increases (see FIGURE 28).

Thus, as a conclusion to this point, it can be said that the initial forcing is a reasonable, and required contribution to add. It helps the population switching from inactive to activated, which temporarily boosts the growth of the entire population as the activated cells grow faster than their inactive counterparts.

5.3.6 Co-consumption of glucose and arabinose

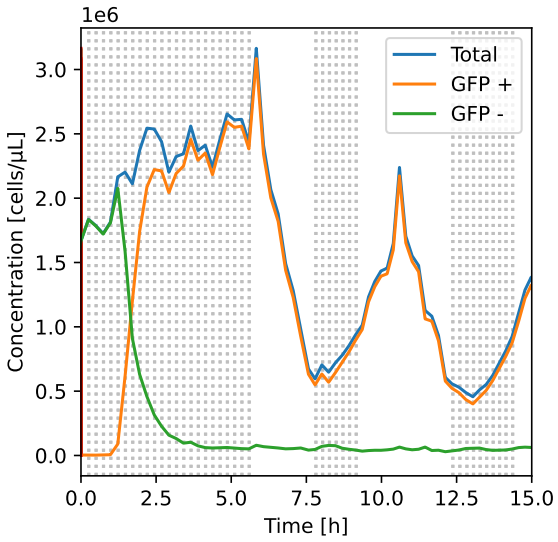
Analysing the contribution of co-consumption is much harder than any of the other contributions because there is no real number that was found. All the optimised models had parameters for growth vastly out of the expected range, like yields close to 0 or 1 g/g, or growth rates of 0.1 or 10 h^{-1}.

This can mean one of two things: either the curves that were fitted to the model were wrong, or the data is over-fitted. The over-fitting possibility seems unlikely, as none of the models could fit correctly both the chemostat and the segregostat at once, and the solution modelled when that was tried always had mean values greater than all chemostat experiment, and lower than all segregostat experiment, indicating that it was not biological variability causing the issue. If the curves are wrong, however, it does not especially mean that the growth contribution is wrong: having a very different, unthought of, switching mechanism, instead of the approximations used, could be enough to throw off any attempt at fitting anything.

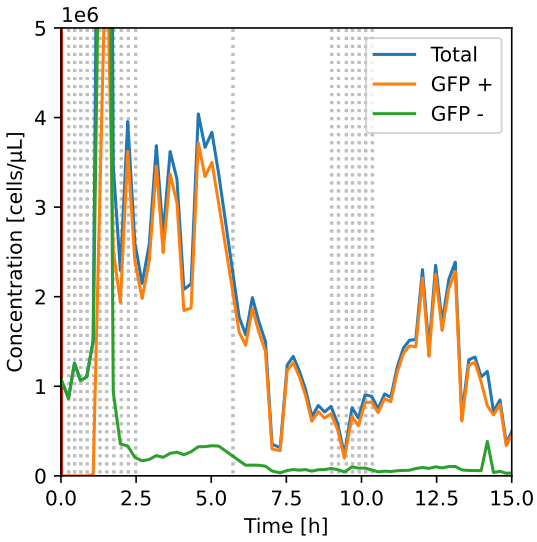
Here, the different arguments for or against co-consumption, or for different kinds of co-consumption, are summarised and discussed to reach as certain a conclusion as possible.

The main argument for co-consumption has already been summarised in the previous part, regarding the initial forcing: cells seem to grow much faster in their activated phenotype than in their inactive phenotype. This observation is valid for both the periodic peaks observed in the segregostat, and the fast growth observed at the beginning of the continuous phase in all experiments.

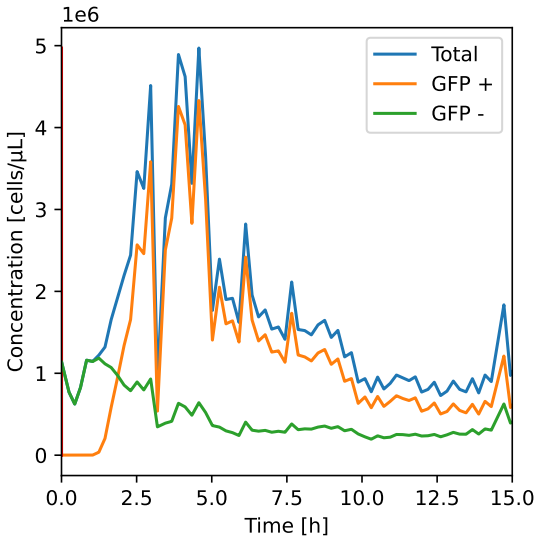
An alternative explanation for these two phenomena is that, somehow, the cells grow very fast on glucose during the observed periods of fast growth when there is arabinose, but their switching rate to using arabinose is nearly perfectly balanced with the increased growth rate, which causes the apparent inactive cell population to stay constant when arabinose is added in the system. While this seems like



(a) Beginning of the segregostat experiment shown in Section 2



(b) Beginning of the combined experiment used in Figure 17



(c) Beginning of the chemostat experiment shown in Section 2

Figure 28: The beginning of all experiments shows that the cell concentration increases in the first few hours, even as the inactive cell concentration decreases. This is visible even through the noise. In (b), the peak that is cut at the very beginning of the experiment was a few measurements of $\mathcal{O}(10^7)$; they are certain to be noise. In (c), the large fluctuation after 3 hours is unlikely to be real.

an overly complex possibility, it is supported by the general information regarding carbon catabolite repression (even though, as mentioned in the state of the art, some studies [17, 18] now show the contrary), and the fact that glucose is more easy to use by the cell.

It is good to note that one possibility that is nigh impossible, would be to have the activated population growing only on arabinose faster than the inactive population growing on glucose. There is 2.5 $gL^{-1}h^{-1}$ of glucose injected in the system, overall, and only 0.75 $gL^{-1}h^{-1}$ of arabinose added in the chemostat, or an average, during the series of pulses, of at most 3 $gL^{-1}h^{-1}$ but normally 2.5 $gL^{-1}h^{-1}$ (the pulse frequency is known after the fact, but may vary a bit, they should be injected every 12 minutes in the system but at worst they would be injected every 10 minutes). Given that bacteria are known to prefer glucose to arabinose, and they do not need to produce additional proteins to degrade it, it is not

possible for the cells to show the peaks they show if one considers little switching and no co-consumption. Indeed, the peaks of activated cell concentrations can be measured to go, in 2.5 h, to five time their initial concentration A_0. Neglecting switching, and without co-consumption, the corresponding growth equation would be:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -DA + \mu_a A \tag{61}$$

$$\Leftrightarrow \quad \frac{dA}{A} = (\mu_a - D)dt \qquad \text{Because variables can be separated} \tag{62}$$

$$\Leftrightarrow \quad \log(A) = (\mu_a - D)t + C_1 \tag{63}$$

$$\Leftrightarrow \quad A = C_2 \exp\left((\mu_a - D)t\right) \tag{64}$$

$$\Leftrightarrow \quad A = A_0 \exp\left((\mu_a - D)t\right) \qquad \text{Considering the beginning of the peak as } t = 0. \tag{65}$$

$$\Leftrightarrow \quad 5A_0 = A_0 \exp\left((\mu_a - 0.5)2.5\right) \qquad \text{Substituting the values} \tag{66}$$

$$\Leftrightarrow \quad \frac{\log(5)}{2.5} + 0.5 = \mu_a \tag{67}$$

$$\simeq 1.1 \text{ h}^{-1} \tag{68}$$

Correspondingly, for the inactive cell concentration, that stays constant over time, the overall growth could be summarised by $\mu_g = D = 0.5$ h^{-1}. Activated cells cannot, without excessive switching, grow twice faster than inactive cells, with about the same concentration of a sugar that is less efficient than glucose.

There is another point that can be made against the previous case: when observing the sugar measurements, for one segregostat experiment (see FIGURE 4), one can note that the glucose concentration decreases when arabinose is injected in the system. If the activated population did not use glucose, and since the inactive cell concentration is constant, there would be no reason for the glucose concentration to vary at all.

There is a last argument, although it may be considered weak, that corroborates the fact that there should be either a fast switching from inactive to activated cells, or the activated cells have to co-consume: for all but one of the optimal models that do not consider co-consumption or direct responsive switching (because that was disproved, so it was not tested), it can be observed that $Y_{xa} > Y_{xg}$. The ratio Y_{xa}/Y_{xg} ranges from 1.8 to more than 60. This would mean that, for those models to be valid, the arabinose should be assimilated better than glucose. This is known to be wrong. Interestingly, the one exception to this observation (model number 12, optimised for both experiments), has $H_2 \simeq 2.3$ h^{-1}, which is much higher than in the rest of the models, confirming that a high switching rate from the inactive to activated phenotypes, if tuned correctly, can reproduce the behaviour observed, at least in part. This argument was said to be weak, however, because most of the models with co-consumption also show $Y_{xa} > Y_{xg}$, although usually the ratio Y_{xa}/Y_{xg} is smaller than for the equivalent model without co-consumption, and the switching rate is never very high. The numbers mentioned in this paragraph are summarised in TABLE 8.

At this point, there are two possibilities that have been considered and not disproved: either the cells co-consume, or they have a very fast switching rate which supplements the activated cell population. It is worth noting, however, that the peaks of activated cell concentration observed in the segregostat

Table 8: Table of the optimised yields of the different models. When no co-consumption is considered, the yield for glucose is always (with one exception where H_2 is very high) much lower than the yield for arabinose, as activated cells need to grow faster. This is a weak argument as what is true for all but one model using arabinose is true for the models considering co-consumption as well, even though the ratio $Y_{x,g}/Y_{x,a}$ is usually lower for the latter.

Model number	Without co-consumption			Model number	With co-consumption		
	Segregostat optimisation				Segregostat optimisation		
	$Y_{x,g}$ [g/g]	$Y_{x,a}$ [g/g]	$Y_{x,g}/Y_{x,a}$ [-]		$Y_{x,g}$ [g/g]	$Y_{x,a}$ [g/g]	$Y_{x,g}/Y_{x,a}$ [-]
4	6×10^{-4}	0.43	677.17	5	0.06	0.95	15.93
6	0.30	0.62	2.06	7	0.10	0.53	5.18
8	0.01	0.66	62.59	/			
10	0.02	0.74	45.76	/			
12	0.16	0.42	2.65	13	0.86	0.12	0.14
14	0.32	0.58	1.82	15	0.04	0.53	14.57
	Chemostat optimisation				Chemostat optimisation		
	$Y_{x,g}$ [g/g]	$Y_{x,a}$ [g/g]	$Y_{x,g}/Y_{x,a}$ [-]		$Y_{x,g}$ [g/g]	$Y_{x,a}$ [g/g]	$Y_{x,g}/Y_{x,a}$ [-]
4	0.02	0.90	56.83	5	0.35	0.98	2.77
6	0.04	0.98	22.08	7	0.89	0.90	1.01
8	0.12	0.65	5.24	/			
10	0.09	0.40	4.63	/			
12	0.35	0.93	2.70	13	0.73	0.07	0.10
14	0.20	0.77	3.93	15	0.90	0.90	1.01
	Optimisation for both				Optimisation for both		
	$Y_{x,g}$ [g/g]	$Y_{x,a}$ [g/g]	$Y_{x,g}/Y_{x,a}$ [-]		$Y_{x,g}$ [g/g]	$Y_{x,a}$ [g/g]	$Y_{x,g}/Y_{x,a}$ [-]
4	0.07	0.64	9.14	5	0.93	0.93	1.00
6	0.04	0.38	9.76	7	0.56	0.63	1.13
8	0.19	0.98	5.09	/			
10	0.32	0.72	2.22	/			
12	0.48	0.24	0.49	13	0.91	0.38	0.42
14	0.10	0.99	9.53	15	0.94	0.44	0.47

have a downward trend as well as an upward trend. This trend can be measured, just as previously, by noting the decrease of the population by a factor 5 in about 2.5 h, as shown in EQUATIONS 69 to 72:

$$\Leftrightarrow \quad A = A_0 \exp\left((\mu_a - D)t\right) \qquad \text{Just as before} \tag{69}$$

$$\Leftrightarrow \quad \frac{A_0}{5} = A_0 \exp\left((\mu_a - 0.5)2.5\right) \qquad \text{Substituting the values} \tag{70}$$

$$\Leftrightarrow \quad \frac{\log(0.2)}{2.5} + 0.5 = \mu_a \tag{71}$$

$$\simeq -0.15 \text{ h}^{-1} \tag{72}$$

The fact that the measured μ_a is < 0 can probably be attributed to switching rates that were not taken into account and measurement error as it is not exactly 5 times in exactly 2.5 hours, but the point stands: the overall population growth, without considering the dilution rate, is very small. Again, this could be explained through switching: the inactive cells would constantly switch massively to their activated state whenever there is some arabinose in the system, and in the meantime, whenever there is not enough arabinose in the system, activated cell would massively switch back to being inactive, which would effectively create an apparent growth rate of ~ 0 h^{-1}. Or, the inactive population could stop switching and grow much slower when there is no arabinose in the system, but having the inactive population growth rate depending on the arabinose concentration seems like a stretch of the imagination. Overall, explaining the segregostat peaks with no co-consumption means explaining them with a very complex and unlikely switching mechanisms. Those can be further disproved as switching between the two populations takes time, as shown by Fritz et al. [20], especially when switching from activated to inactive. This time needed to switch can be easily observed in a different experiment, that was made later, and where the frequency of arabinose induction was much lower, in FIGURE 29. With such a long time needed to switch back from an activated to an inactive population, and with the majority of inactive cells switching to an activated phenotype constantly, the amount of cells in transition would be very high at all time during the segregostat. This is not at all observed in reality (see FIGURE 3.c), and means that the only remaining possible theory is co-consumption.

One may wonder, however, if there were co-consumption, and the switching considered in the other sections were a good approximation of reality, why could none of the models with co-consumption be correctly fitted to the data? The most likely explanation is that there is co-consumption, but not quite in the way it was implemented here. Indeed, Nikolic et al. [18], who showed that co-consumption for glucose and arabinose was possible, analysed sugar consumption by at a single cell resolution. What they found is that cells that consumed more arabinose, also consumed more glucose. And, on the other hand, cells that used little arabinose also used very little glucose. The reason this could happen, according to the authors, is that the faster a cell grows, the more noise there is in gene expression, which could then let all genes, including the ones responsible for arabinose degradation be activated. This explanation would fit nicely with the current data: cells that do not consume arabinose and produce no GFP would be slow-growing cells, that barely consume enough glucose not to be washed away by the dilution rate. On the other side of the spectrum, activated cells would also be using most of the glucose in addition to arabinose, and grow much faster. The only problem with that explanation is that the study did not find any bimodal population in their chemostat experiments (which were at smaller scale then the ones discussed in this work, and with another *E. coli* strain). The conclusions of their work should therefore be confirmed for the experimental conditions used in this work.

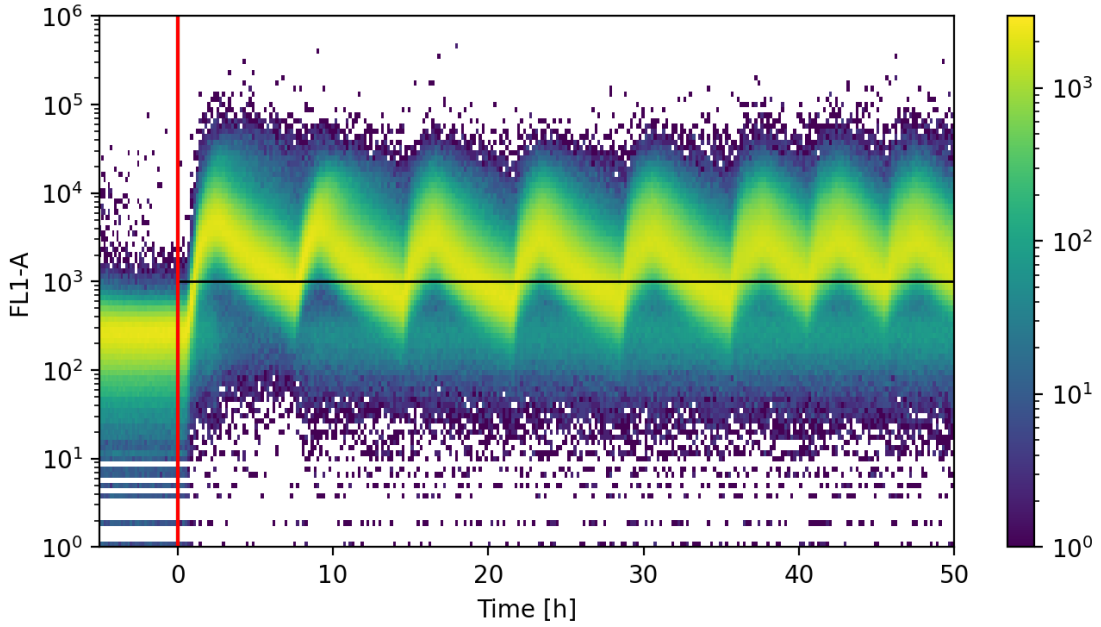


Figure 29: Experiment where 1 g of arabinose was added in the 1 L bioreactor first every 7 h then every 5 h, without segregostat control. The red line represents the beginning time of the continuous cultivation phase, and the horizontal line is the fluorescence threshold. The time needed for the activated cell population to switch to the inactive phenotype is approximately the delay between the time the population first reaches the fluorescence threshold and the time it passes it nearly completely, just before the time arabinose is re-injected in the system after 7 h. That corresponds to 3–4 h.

5.3.7 Summary and further discussion

Before trying to expand on the different ways this work could be continued, a short summary is provided here, putting the work until now in context with what was expected.

The general goal of this work was to design a model that would better the understanding of the population behaviour of *E. coli* in a fluctuating environment where the concentrations of arabinose and glucose vary.

To this end, a list of models, ranging from very simple to much more complex, were put together. The goal was then to identify whether the most complex models were indeed better than the simpler ones, which was done by fitting the models to the data. From the analysis of the results, several conclusions could be drawn.

The two least important and most certain ones regard time delays. While they were initially thought to be able to explain the behaviour of the chemostat, they were found insufficient for the task. Their existence, however, was confirmed: there is a delay between a change in arabinose concentration and the bacteria's change in growth rate. For glucose, no significant delay was found, although a higher importance of such a delay could be made apparent if the glucose concentration were made to vary during the experiment. The exact implementation the delays should have is still unknown, however.

Then, it can be said that the direct responsive switching, as it was envisioned, is not possible: because it had a rate supposed to be around 1 h^{-1}, it caused significant variations in the inactive cell concentrations

which were not observed in reality. It was later shown that any switching mechanism that would have a switching rate from activated to inactive cells greater than ~ 0.1 h^{-1}, during the peaks of the segregostat experiment, would predict oscillations in the inactive cell concentrations unlike what is seen in reality.

The switching mechanism, instead of direct responsive, was determined to be reasonably approximated by a basal switching rate of an order of magnitude smaller or equal to 0.1 h^{-1}, if there is any, and likely an initial forcing such that $r_a \cdot p_{init}$ is close to 0.5 h^{-1}.

The growth mechanism was confirmed to be a kind of co-consumption, but the specifics are unknown. It is likely that the GFP positive cells, instead of using as much glucose as GFP negative ones, instead use more of it.

A point that arises only when all the contributions are considered is that if there is co-consumption, and initial forcing, in the same model, then there must be a basal switching as well, as the two models without it are the models 9 and 11, which were earlier determined as impossible. Therefore, the evidence shows that the best model possible, among the ones considered, is the model 15, which is fully represented by the Equations 73 to 76 when the responsive population is not switching, and otherwise Equations 29 to 31 need to be added in.

$$\frac{\mathrm{d}A}{\mathrm{d}t} = (\mu_a[t - \Delta t_a] + \mu_g)A - DA - H_1 A + H_2 G \tag{73}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = \mu_g G - DG + H_1 A - H_2 G \tag{74}$$

$$\frac{\mathrm{d}C_a}{\mathrm{d}t} = (C_{a,in} - C_a)D - \frac{\mu_a A}{Y_{xa}} \tag{75}$$

$$\frac{\mathrm{d}C_g}{\mathrm{d}t} = (C_{g,in} - C_g)D - \frac{\mu_g (G + A)}{Y_{xg}} \tag{76}$$

(when the responsive population does not switch)

The dynamics of the chemostat and segregostat experiments can be mostly explained thanks to this model, even though the real growth equations are unknown. At the beginning of the continuous phase, the initial forcing amplifies the apparent different concentrations of the two sub-populations, which is caused by different growth rates. After that, most of the effects observed are due to different growth rates of the two populations, rather than switching.

There are, however, two main things that cannot be explained. Firstly, the model cannot be fitted to correspond to both types of experiments at once. Secondly, the time delay cannot explain the local minimum that appears after 20 hours in the chemostat.

Part of the reason is probably due to the fact that the growth curves of the two sub-populations are unknown, but there may very well be more at play. Indeed, the combination of initial forcing and basal switching are a good approximation of the switching phenomena, but they stay an approximation. There is no reason for the part of the population that switches to be independent of the arabinose concentration. In the segregostat, more arabinose is injected overall than in the chemostat, during the series of arabinose pulses. It would make sense for the switching to be more important in that case, which would increase the activated cell proportion from the beginning.

Furthermore, it is important to remember that the initial forcing was an approximation of a phenomenon that was observed in the segregostat, where cells would start switching back to arabinose just before arabinose was injected in the system (see FIGURE 8). This means that in the segregostat, with a time period of the fluctuations of 4 to 5 hours, the initial forcing is a good approximation. However, experiments made later, where the period was defined to be longer (see FIGURE 29), show that more than 5 hours after the start of the decline of the activated cell concentration, the population is still switching. The phenomenon is in this case mainly caused by switching, as can attest the proportion of cells shown in transition between a fluorescence of 3×10^3–10^5, the activated population, and the cells at 2–3×10^2, which is normally the fluorescence of the inactive population. It is not impossible for this switching to take place over 10 to 15 hours. If that is the case, the switching of a part of the population, after the peak of the chemostat, could explain the local minimum.

Finally, the range of validity of the model deserves to be mentioned. It is mostly valid, once fitted, for the chemostat and the segregostat, as the experiments were made, with the caveats mentioned previously. It is known that for lower frequencies of induction, the model would fail, as switching becomes predominant. And, in case of higher frequencies, the environment would, at some point, start resembling a chemostat in terms of arabinose concentrations, since the cells would not have time to consume all the arabinose between two arabinose injections. Thus, at some point, the unexplained local minimum after 20 h is likely to appear.

6 Perspectives

As discussed previously, the model is clearly not the absolute best model that could be used to model the experiments yet. In this section, ways to improve it, to confirm (or infirm) experimentally the basis on which it is established, and whether the model could be used for practical purposes is discussed.

6.1 Better approximation of the phenotypic switching rates

The first point that is to be addressed is the switching mechanism. Initially, as described in the materials and methods, two "forcings" were envisioned: whenever the arabinose concentration would pass a given threshold for long enough, the responsive population would start switching rapidly. This does seem to match the observations, but it comes with a critical problem: instability. Indeed, when the responsive population starts going down, and arabinose is injected, it start going back up. This means that it can change "decision", even while switching. While going up, however, it consumes arabinose, meaning that if the arabinose concentration gets below the threshold before the responsive population switched completely to the activated phenotype, the model would make the population start switching back to its inactive state. In particular, when considering the chemostat, arabinose is constantly injected, and the responsive population of the model can be stuck between switching down and up to keep the arabinose concentration at the threshold. This requires very small time steps, making the integration of the differential equations for the model very slow (in addition to being wrong). When considering the segregostat, the population would never reach a point at which it is entirely in one state, and the delay implemented as a time needed for the cells to commit to the switching decision would be useless. That would effectively imitate the direct responsive switching, and would be wrong for the same reason.

The general idea, however, seems sound. But the mechanism it is supposed to imitate, which is the GFP+ population that slowly passes the fluorescence threshold, means that the responsive population cannot be "stuck" between going up and going down depending on the exact arabinose concentration at the time point considered.

One phenomenon that could help for this problem is hysteresis. Indeed, Jenkins and Macauley [14] found it likely that the arabinose threshold to use in order to switch from an inactive operon to an active one should be higher than the threshold used to decide when to stop using arabinose. This is not a perfect solution, however, as for the right input rates of arabinose and the right cell concentrations, the arabinose concentration could still oscillate above and below, respectively, the maximum and minimum threshold. In such a situation, however, the system would be more similar to a chemostat than anything else and the population would become bistable, there would likely not be a responsive population oscillating anywhere.

Another potential solution would be to have the cells respond not to the current arabinose, but rather to an agglomeration of present and past information. In other words, the behaviour of the responsive population could be dependent on:

$$\int_{present-\Delta t}^{present} C_a(t)\ dt$$

In that case, no time delay needs to be implemented, as the integration already represents the memory of the population. At least in the segregostat, it would also likely stop the population from being "stuck" between two states, although for the chemostat it would still probably be possible.

Another point that needs to be addressed is the probable variable size of the responsive population. Indeed, it is likely that when more arabinose is added, more cells switch to using arabinose. Therefore, instead of having a single population, it would probably be better to to have two states, and, when the conditions are right (i.e. depending on $\int_{present-\Delta t}^{present} C_a(t)\ dt$), a new responsive population would be created, take a percentage of the population that needs to switch, and make it switch before being forgotten by the model as the newly switched cells can be added to the phenotype they switched to. This could explain, potentially, the local minimum after 20 h in the segregostat: a small activated cell population would be made responsive due to the lack of arabinose in the system, and the new responsive population switching back to an inactive state would decrease the proportion of activated cells temporarily. This way of modelling things, however, becomes very complicated, especially considering that nothing stops several responsive population from co-existing in the model.

Such mechanisms would already, probably, be a good approximation of what happens in reality, even without hysteresis. However, there is third point that needs to be addressed, even if it may be less important for the overall exactness of the model: depending on the fluorescence of the cells, they need different times to be marked as "activated cells" again, since time is needed to increase the fluorescence. In addition, it is unknown how much arabinose the cells that are transitioning between the two states actually consume. To address this issue, a third, temporary state could be added to the model, along with a "progress" between the two phenotypes. That way, a responsive population that is very close to having reached the inactive state would take longer to switch back to an activated state.

If one needs to integrate a variable amount of responsive populations, with 3 or more phenotypes considered, then it may reach the limit of what can be modelled with differential equations. The more complexity is added to the model, the more it approaches a mechanistic model. And for mechanistic models, when talking about bacteria, it is often easier to make a stochastic simulation. A given amount of cells — not all of them as that would require tremendous computing power, but in theory, since things are perfectly mixed, one can simply model a part of the volume of the reactor — could be simulated, with the cell behaviour depending on its past and on the environment. Fritz et al. [20] and Megerle et al. [21] note that the behaviour of the cell, at least when no switching has occurred yet, depends on the concentration of arabinose transporters it expresses. One could therefore model the behaviour of any cell as $f(\int_{present-\Delta t}^{present} C_a(t)\ dt, transporter\ concentration)$, supposing the transporter concentration is directly correlated to the fluorescence of the cell. For such a model, one would also need to simulate a stochastic dilution of the cells in the medium, and the behaviour of the two daughter cells that appear after cell division would need to be assessed.

I believe not all of the parameters that would be needed to determine exactly what happens in the bioreactor can be determined solely with similar experiments to what was done until now. Indeed, the experimental setup lets one know the composition of the population at any given time, but it does not follow cells individually. It would be most useful to confirm the switching behaviour through microfluidics experiments tracking single cells over time. While the exact kinetics would be different, the overall principles would be similar.

6.2 Determination of growth parameters

In order to correctly assess switching kinetics, the growth profiles of the various phenotypes are required.

Since the concentration measurements from the flow cytometer cannot be trusted, one thing that should probably be done before anything else is to make follow, hour per hour, the biomass concentration both at the beginning of an experiment, and during one peak of the segregostat. In this manner, the biomass evolution could at least be confirmed; it is very likely already but trusting a measurement known to have instrumental artefacts without a second confirmation can lead to false results.

Another way to simply confirm co-consumption would be to measure the growth parameters of the *E. coli* strain on solely glucose and solely arabinose. Then, trying to fit the rest of the parameters in the model obtained in this work, without co-consumption, either should not work or should provide impossible parameters, if said parameters were not bounded. If it does work, with realistic parameters found, then some conclusion of this work should be put in doubt.

After confirming the evolution of the biomass, the next step would be to experimentally produce direct evidence of the co-consumption of the two phenotypes. This could be achieved in the same manner as Nikolic et al. [18]. Among other techniques, they used *E. coli* populations with reporters for *ptsG* and *araE*. *AraE* was reported with GFP, just like in this work, and *ptsG* was reported through mCherry, a red fluorescent molecule. *PtsG* is the main glucose transporter, indicator of how much glucose the cell is using. Using a population with the two reporters in segregostat and chemostat experiments would allow to determine whether the cells indeed produce a population that uses very little glucose and one that uses a lot of both glucose and arabinose. It would also exclude the presence of other, as of yet undetected, populations of cells: in theory, with those two reporters, four populations could be detected. Cells that use only arabinose, cells that use only glucose, cells that use neither in significant quantities, and cells that use both. In theory, nothing stops all four from being present, but in Section 5.3.6, only two were considered, as there is no known reason to consider more.

In the work of Nikolic et al. [18], the ratio of expression of both *ptsG* and *araE* was dependent on both the concentrations of arabinose and glucose. An important point is that it did not simply depend on the ratio of arabinose and glucose concentrations. When both arabinose and glucose were present at 10 µM, *ptsG* expression was twice as high as *araE* expression. When they were both at 3 mM, the ratio was closer to six or seven. Thus, even the equivalence of the observed sub-populations between the different experiments should probably be verified. One possible to way to verify that is to produce several chemostat experiments, with different arabinose and glucose feeding rates. Additionally, such experiments could provide valuable information on the consumption rates of the different observed phenotypes in different sugar concentrations, at equilibrium, which could provide data on the actual growth parameters of the sub-populations.

If one considers that only two sub-populations are present, speculations can be made on a model that could fit the data. Supposing that the work from Nikolic et al. [18] can be transposed here, then the inactive population should have a maximum growth rate $\mu_{g,\max,GFP-}$ smaller than the maximum growth rate on glucose by the activated population, $\mu_{g,\max,GFP+}$. Supposing $\mu_{g,\max,GFP+}$ to be independent of the arabinose concentration, one could write the growth rates of the two populations as in Equation 77:

$$\mu_g = \mu_{g,\max,GFP-}\frac{C_g}{K_{g,GFP-}+C_g}\mu_a \qquad = \mu_{g,\max,GFP+}\frac{C_g}{K_{g,GFP+}+C_g} + \mu_{a,\max}\frac{C_a}{K_a+C_a} \tag{77}$$

This would probably the simplest expression of the co-consumption, but as of now there is no data to show whether or not it is a good approximation of reality.

6.3 Limitations for the use of the model

As it is now, the model is probably not complete enough to be used in any kind of practical application. It needs to be fitted to an experiment before it can be used to predict it, and then only small ranges of parameter variation would be predicted correctly. Once it is fitted, however, one may use it to replace the feedback loop used in the segregostat: if the time the pulses are needed is predicted in advance, the experiment would not depend on the measurements of the flow cytometer, which can be noisy.

There is at least one thing, however, that may improve the range of parameters over which the model is useful. The reason why the model could not be fitted is probably mostly due to the fact that during the chemostat, the feeding rate of arabinose in the system is at 1.5 $gL^{-1}h^{-1}$, and during the series of pulses of the segregostat it is at an average of ~ 2.5 $gL^{-1}h^{-1}$. It is not out of the question that the same set of parameters could explain both the chemostat and the segregostat experiments at once, if both experiments had an equivalent arabinose feeding rate.

While the model as it is is mostly useless, it is modular. This means that it could easily be used to test different possibilities for the growth and switching contributions, as long as they are based on differential equations and not on stochastic models.

Finally, it can be noted that, once the growth curves and/or a better approximation of the switching mechanism are found, it could be used to predict the result of new experiment. Among other things, it has the potential to be adapted to other bioreactors, or it could be used to study the impact of different arabinose and glucose feeding schemes without needing triplicate: if the model can explain everything, then any experiment explained by the same set of parameters for the model could be considered a duplicate of other experiments that confirm it.

7 Conclusion

In this work, the way *E. coli* changes its gene expression to adapt to different sugars was studied. This work was made possible thanks to flow cytometry data, that allow for quantitative measurements of the relative proportion of the cells that are able to degrade arabinose. Those quantitative measurements were then compared to the application of the model used by Thattai and Van Oudenaarden [1] to describe cell populations, and it was found that said model cannot reproduce the experimental data.

Therefore, several additions were made to the base model. Each of them was independently evaluated to judge whether or not it allowed the new, extended model to better fit the data, and the following results were found:

- A time delay that causes the bacteria to use the arabinose about half an hour after it was injected slightly improves the model.
- The switching behaviour should consider basal stochastic switching. However, that is not sufficient to reproduce the observed population behaviours, and a form of delayed responsive switching ought to be considered.
- Microbial cells are able to use simultaneously glucose and arabinose upon phenotypic adaptation.

However, while the extended model devised in this work can represent the data better than the base model derived from Thattai and Van Oudenaarden's work, it is far from perfect. Indeed, it was established that the model could not fit the data from different kinds of experiments with a single set of parameters. This is due to two main factors:

- The growth of the different sub-populations cannot be approximated by supposing that all cells use glucose at the same rate, and any model that seeks to explain population behaviours requires a good identification of the growth curves of the different phenotypes.
- The switching mechanism considered is better than the absence of responsive switching, but it stays an approximation and is still far from reality. In order to make it much better, however, it would be simpler to use stochastic modelling with single cell resolution rather than keep using differential equations. Indeed, the modifications that would need to be brought to differential equations would likely need to include considerations such as intermediate gene expression states, time delays, and splitting of the population into multiple sub-populations that would behave independently from each other, and all of these would be difficult to integrate in a single deterministic system.

Finally, it needs to be noted that, while those two factors were made obvious for this case study, they are indicative of more general facts regarding the study of heterogeneous cell populations: the growth rates of the various phenotypes of such population need to be investigated thoroughly before any conclusion can be drawn, and switching would probably be easier to model correctly through stochastic modelling than through differential equations in a lot of systems.

References

[1] Mukund Thattai and Alexander Van Oudenaarden. "Stochastic gene expression in fluctuating environments". In: *Genetics* 167.1 (2004), pp. 523–530. DOI: 10.1534/genetics.167.1.523.

[2] Avigdor Eldar and Michael B Elowitz. "Functional roles for noise in genetic circuits". In: *Nature* 467.7312 (2010), pp. 167–173. DOI: 10.1038/nature09326.

[3] Jan-Willem Veening, Wiep Klaas Smits, and Oscar P Kuipers. "Bistability, epigenetics, and bet-hedging in bacteria". In: *Annu. Rev. Microbiol.* 62 (2008), pp. 193–210. DOI: 10.1146/annurev.micro.62.081307.163002.

[4] Edo Kussell and Stanislas Leibler. "Phenotypic diversity, population growth, and information in fluctuating environments". In: *Science* 309.5743 (2005), pp. 2075–2078. DOI: 10.1126/science.1114383.

[5] Hosni Sassi et al. "Segregostat: a novel concept to control phenotypic diversification dynamics on the example of Gram-negative bacteria". In: *Microbial biotechnology* 12.5 (2019), pp. 1064–1075. DOI: 10.1111/1751-7915.13442.

[6] Thai Minh Nguyen et al. "Reducing phenotypic and genotypic instabilities of microbial population during continuous cultivation based on stochastic switching dynamics". In: *bioRxiv* (2021). DOI: 10.1101/2021.01.13.426484.

[7] Jacques Monod. "Recherches sur la croissance des cultures bacteriennes". In: (1942).

[8] Dominique Toye. *Questions avancées en Génie Chimique : BIOTECHNOLOGIE*. French. [Course, not published]. Faculté des Sciences Appliquées (ULiège), 2020. Chap. Réacteurs biologiques.

[9] Darren J Wilkinson. *Stochastic Modelling for Systems Biology, Third Edition*. Hardcover. CRC Press, 2018. ISBN: 9781138549289.

[10] Mads Kaern et al. "Stochasticity in gene expression: from theories to phenotypes". In: *Nature Reviews Genetics* 6.6 (2005), pp. 451–464. DOI: 10.1038/nrg1615.

[11] Paul J Choi et al. "A stochastic single-molecule event triggers phenotype switching of a bacterial cell". In: *Science* 322.5900 (2008), pp. 442–446. DOI: 10.1126/science.1161427.

[12] Robert Schleif. "Regulation of the L-arabinose operon of Escherichia coli". In: *Trends in Genetics* 16.12 (2000), pp. 559–565. DOI: 10.1016/S0168-9525(00)02153-3.

[13] Jarno Mäkelä et al. "In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter". In: *Nucleic acids research* 41.13 (2013), pp. 6544–6552. DOI: 10.1093/nar/gkt350.

[14] Andy Jenkins and Matthew Macauley. "Bistability and Asynchrony in a Boolean Model of the L-arabinose Operon in Escherichia coli". In: *Bulletin of mathematical biology* 79.8 (2017), pp. 1778–1795. DOI: 10.1007/s11538-017-0306-1.

[15] Boris Görke and Jörg Stülke. "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients". In: *Nature Reviews Microbiology* 6.8 (2008), pp. 613–624. DOI: 10.1038/nrmicro1932.

[16] You-Jin Lee and Kyung-Hwan Jung. "Modulation of the tendency towards inclusion body formation of recombinant protein by the addition of glucose in the araBAD promoter system of Escherichia coli". In: *Journal of microbiology and biotechnology* 17.11 (2007), pp. 1898–1903.

[17] Frank Schreiber and Martin Ackermann. "Environmental drivers of metabolic heterogeneity in clonal microbial populations". In: *Current opinion in biotechnology* 62 (2020), pp. 202–211. DOI: 10.1016/j.copbio.2019.11.018.

[18] Nela Nikolic et al. "Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations". In: *PLoS genetics* 13.12 (2017), e1007122. DOI: 10.1371/journal.pgen.1007122.

[19] Daniel Madar et al. "Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli". In: *BMC systems biology* 5.1 (2011), pp. 1–9. DOI: 10.1186/1752-0509-5-111.

[20] Georg Fritz et al. "Single cell kinetics of phenotypic switching in the arabinose utilization system of E. coli". In: *PloS one* 9.2 (2014), e89532. DOI: 10.1371/journal.pone.0089532.

[21] Judith A Megerle et al. "Timing and dynamics of single cell gene expression in the arabinose utilization system". In: *Biophysical journal* 95.4 (2008), pp. 2103–2115. DOI: 10.1529/biophysj.107.127191.

[22] Shmoolik Mangan, Alon Zaslaver, and Uri Alon. "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks". In: *Journal of molecular biology* 334.2 (2003), pp. 197–204. DOI: 10.1016/j.jmb.2003.09.049.

[23] M Loferer-Krößbacher, J Klima, and R Psenner. "Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis". In: *Applied and environmental microbiology* 64.2 (1998), p. 688. DOI: 10.1128/AEM.64.2.688-694.1998.

[24] Moises Santillán. "On the use of the Hill functions in mathematical models of gene regulatory networks". In: *Mathematical Modelling of Natural Phenomena* 3.2 (2008), pp. 85–97. DOI: 10.1051/mmnp:2008056.

[25] Rachael M Morgan-Kiss, Caryn Wadler, and John E Cronan. "Long-term and homogeneous regulation of the Escherichia coli araBAD promoter by use of a lactose transporter of relaxed specificity". In: *Proceedings of the National Academy of Sciences* 99.11 (2002), pp. 7373–7377. DOI: 10.1073/pnas.122227599.

[26] Denoël Vincent, Geurts Pierre, and Duchesne Laurine. *Éléments de processus stochastiques – Cours et projet d'informatique*. French. [Course, not published]. Faculté des Sciences Appliquées (ULiège), 2019.

[27] Urs Lendenmann, Mario Snozzi, and Thomas Egli. "Growth kinetics of Escherichia coli with galactose and several other sugars in carbon-limited chemostat culture". In: *Canadian journal of microbiology* 46.1 (1999), pp. 72–80. DOI: 10.1139/w99-113.

[28] Heinrich Senn et al. "The growth of Escherichia coli in glucose-limited chemostat cultures: a re-examination of the kinetics". In: *Biochimica et Biophysica Acta (BBA)-General Subjects* 1201.3 (1994), pp. 424–436. DOI: 10.1016/0304-4165(94)90072-8.

Annexes

A.1 Experimental setup

The experimental setup is described completely by Nguyen et al. [6]; the relevant part of their MATERIALS AND METHODS section is reproduced here, with permission from Frank Delvigne (last author), and knowing that the paper is under a Creative Commons license (CC-BY-NC-ND 4.0 International license).

A.1.1 Strains and media

E. coli K12 W3110 wildtype was transformed with pBbB8k(4480bp) plasmid with the P_{araBAD}::GFP, which is an arabinose inducible promoter system. This plasmid belongs to a library of expression vectors compatible with the BglBrick standard[12]. Pre-cultures and cultures were performed on defined mineral salt medium containing (in $g.L^{-1}$): K_2HPO_4 14.6, $NaH_2PO_4 \cdot 2\,H_2O$ 3.6; Na_2SO_4 2; $(NH_4)_2SO_4$ 2.47, NH_4Cl 0.5, $(NH_4)_2$-H-citrate 1, glucose 5, thiamine 0.01. Thiamine was sterilized by filtration (0.2 µm). The medium was supplemented with 3 $mL.L^{-1}$ of trace element solution, 3 $mL.L^{-1}$ of a $FeCl_3 \cdot 6\,H_2O$ solution (16.7 $g.L^{-1}$), 3 $mL.L^{-1}$ of an EDTA solution (20.1 $g.L^{-1}$) and 2 $mL.L^{-1}$ of a $MgSO_4$ solution (120 $g.L^{-1}$). The trace element solution contained(in $g.L^{-1}$): $CaCl_2 \cdot H_2O$ 0.74, $ZnSO_4 \cdot 7\,H_2O$ 0.18, $MnSO_4 \cdot H_2O$ 0.1, $CuSO_4 \cdot 5\,H_2O$ 0.1, $CoSO_4 \cdot 7\,H_2O$ 0.21. Kanamycin (50 µg/mL)was used as a selective marker.

A.1.2 Segregostat cultivations

Cultures in segregostat mode were performed in lab-scale stirred bioreactor (Biostat B-Twin, Sartorius) with 1L working volume. The batch phase was started at an OD_{600} value of 0.5. The pH was maintained at 7.0 by automatic addition of ammonia solution 25%(w/v) or phosphoric acid 25% (w/v). The temperature was maintained at 37°C under a continuous stirring rate of 800 rpm and an aeration rate of 1 vvm. The [dissolved oxygen](DO) signal was used as an indicator for switching to the continuous operation modes (a rapid increase of the DO signals after depletion of glucose in the initial batch medium, observed typically after 3–5 h). The feeding medium, containing 5 g/L glucose, was continuously fed with a dilution rate of 0.5 h−1. According to the sequences controlled by the online [Flow cytometry] (FC) software, pulse of arabinose was injected in order to quickly increase ([in] approx. 30 seconds) the global arabinose concentration to 0.5 g/L[13]. On-line sampling was performed every 12 minutes and processed according to the dilution sequence set through the following series of steps. First, the sample is automatically transferred to [the flow cytometer] C6 FC (BD Accuri C6, BD Biosciences) and analyzed at a medium flow rate (33 $\mu l.min^{-1}$). All the data related to the different parameters (mean, median, CV) are displayed in real-time during the cultivation. Then a tailor-made feedback control loop MATLAB script based on the FC data regulated the actuator profile for the arabinose

[12]Brognaux A, Han S, Sorensen SJ, Lebeau F, Thonart P, Delvigne F. Alow-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors. Microb Cell Factories. 2013;12.

[13]Sassi H, Nguyen TM, Telek S, Gosset G, GrunbergerA, Delvigne F. Segregostat: a novel concept to control phenotypic diversification dynamics on the example of Gram-negative bacteria. Microb Biotechnol. United States; 2019;12:1064–75.

addition. Within this script, FC events were gated based on forward scatter (FSC)[14] to distinguish to single cells and clumps. Then GFP positive events were gated based on a fluorescence threshold value of 2500[15] (arbitrary units, FL1-A; excitation 488 nm, emission 533 nm) (...). Control policy was set for actuator activation when the fraction of GFP positive cells was measured to be below 50%. Actuator activation then exerted arabinose pulsing by the use of a digital control system comprising a peristaltic pump (Watson Marlow, 101 UR). (...) Data were exported as .fcs files (…).

A.1.3 Online flow cytometry (FC) analyses

The online flow cytometer platform comprises three modules (…): (i) a conventional culture device, (ii) a physical interface for sampling and dilution containing peristaltic pumps and mixing chamber, and (iii) a detection device, in this work an Accuri C6 flow cytometer (BD Accuri, San Jose CA, USA) was used. Briefly, sample processing comprises the following steps: (i) sample acquisition, (ii) online FC analysis, (iii) dilution threshold and (iv) feedback control loop. The sample is entered and removed from the mixing chamber based on silicone tubing (internal diameter: 0.5 mm; external diameter: 1.6 mm, VWR, Belgium) and peristaltic pumps (400FD/A1 OEM-pump 13 rpm and 290 rpm, Watson Marlow). Before and after each experiment, all the connection parts (tubing, pumps and mixed chamber) were cleaned with ethanol and rinsed with filtered PBS as set-up in a previous protocol[16].

A.1.4 Additional precision

In addition to what was described by Nguyen et al. [6] previously, it can be noted that for the chemostat experiment a feed containing 1.5 g/L of arabinose, in addition to the glucose and medium, was used.

[14] In the experiments used in this work, the minimum threshold was 20000

[15] this value was modified from the original text to fit the data of this work

[16] Brognaux A, Han S, Sorensen SJ, Lebeau F, Thonart P, Delvigne F. Alow-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors. Microb Cell Factories. 2013;12.

A.2 Experimental results all the available experiments

The following Figures A.2.1 to A.2.5 represent the results of all the experiments that were made. Briefly:

- Figure A.2.1 is a simple chemostat experiment, although the results shown there could not be replicated.
- Figure A.2.2 is a segregostat experiment, with more noise than the experiment shown in the text.
- Figure A.2.3 is an experiment where the first 50 hours of the continuous phase are a chemostat experiment, then the rest of the experiment is made as a segregostat.
- Figure A.2.4 is similar to the previous figure, but the segregostat part was made before the chemostat.
- Figure A.2.5 was only made towards the end of the master thesis, it represents an experiment where, every 7 h, then every 5 h, 1 g of arabinose was added.

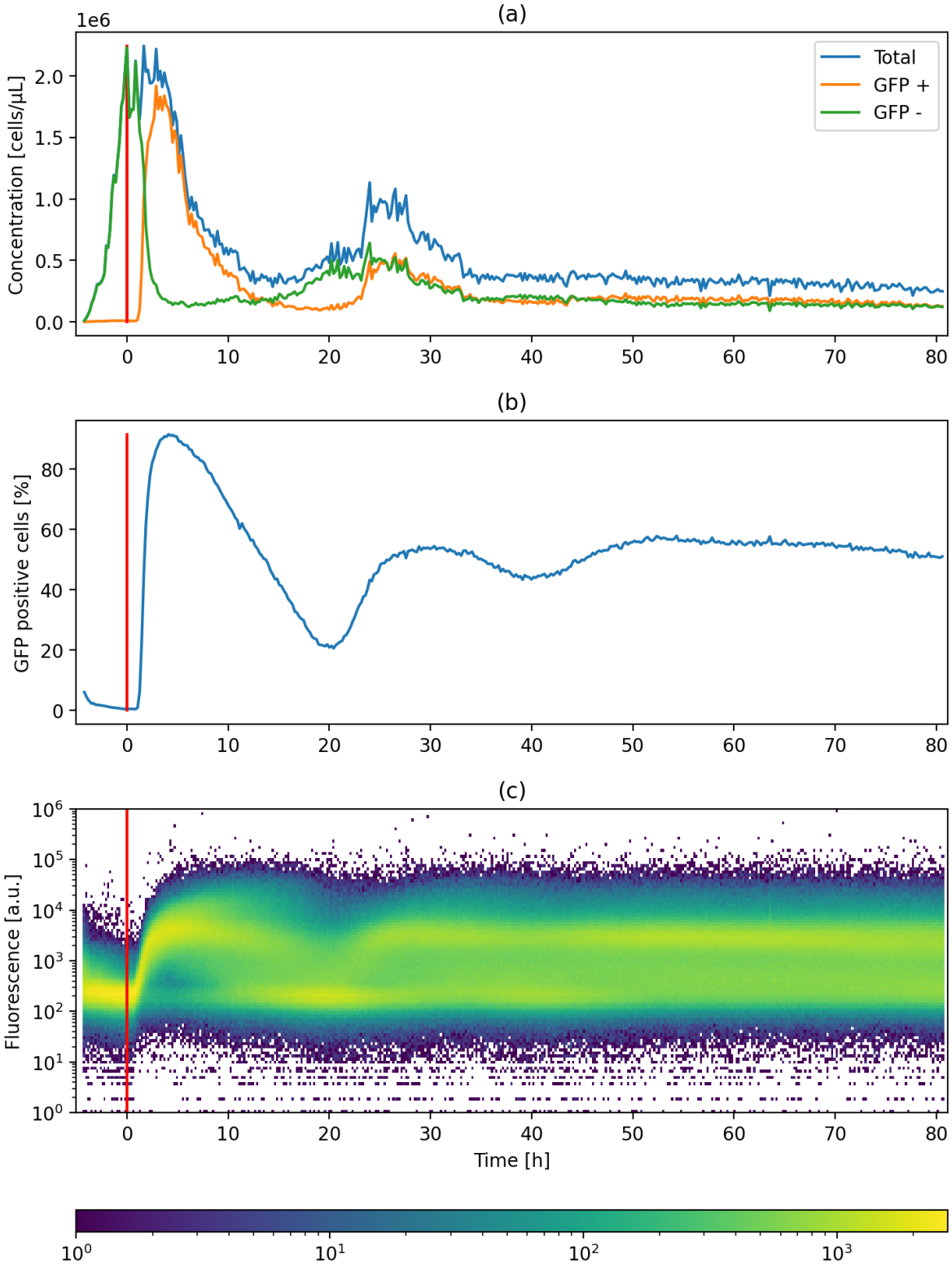


Figure A.2.1: Summary of a chemostat experiment. The oscillating proportion of GFP+ cells could not be reproduced in other experiments. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The red line represents the beginning time of the continuous cultivation phase.

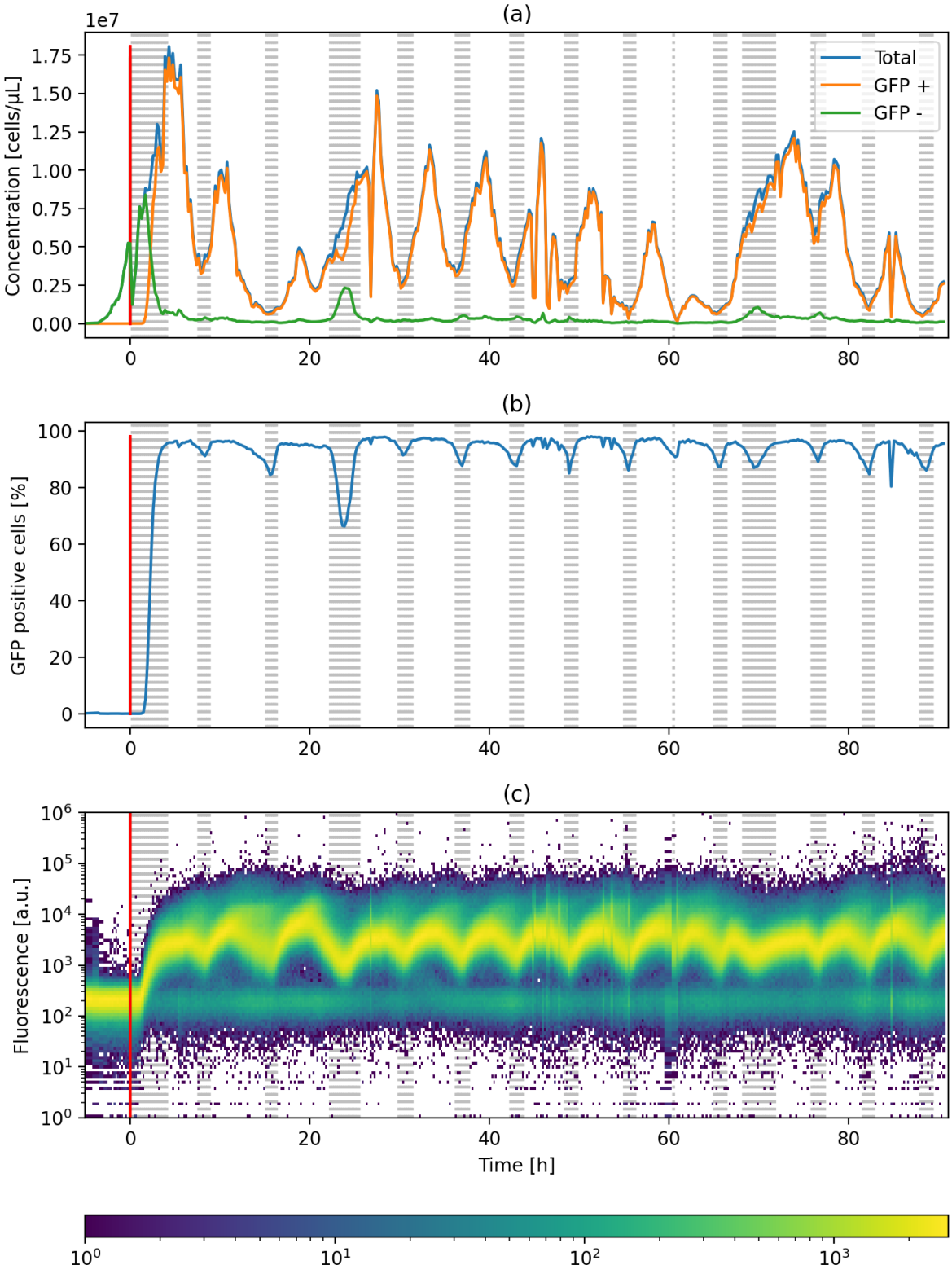


Figure A.2.2: Summary of a segregostat experiment. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The dashed grey areas show the periods when pulses of arabinose were added into the system. The red line represents the beginning time of the continuous cultivation phase.

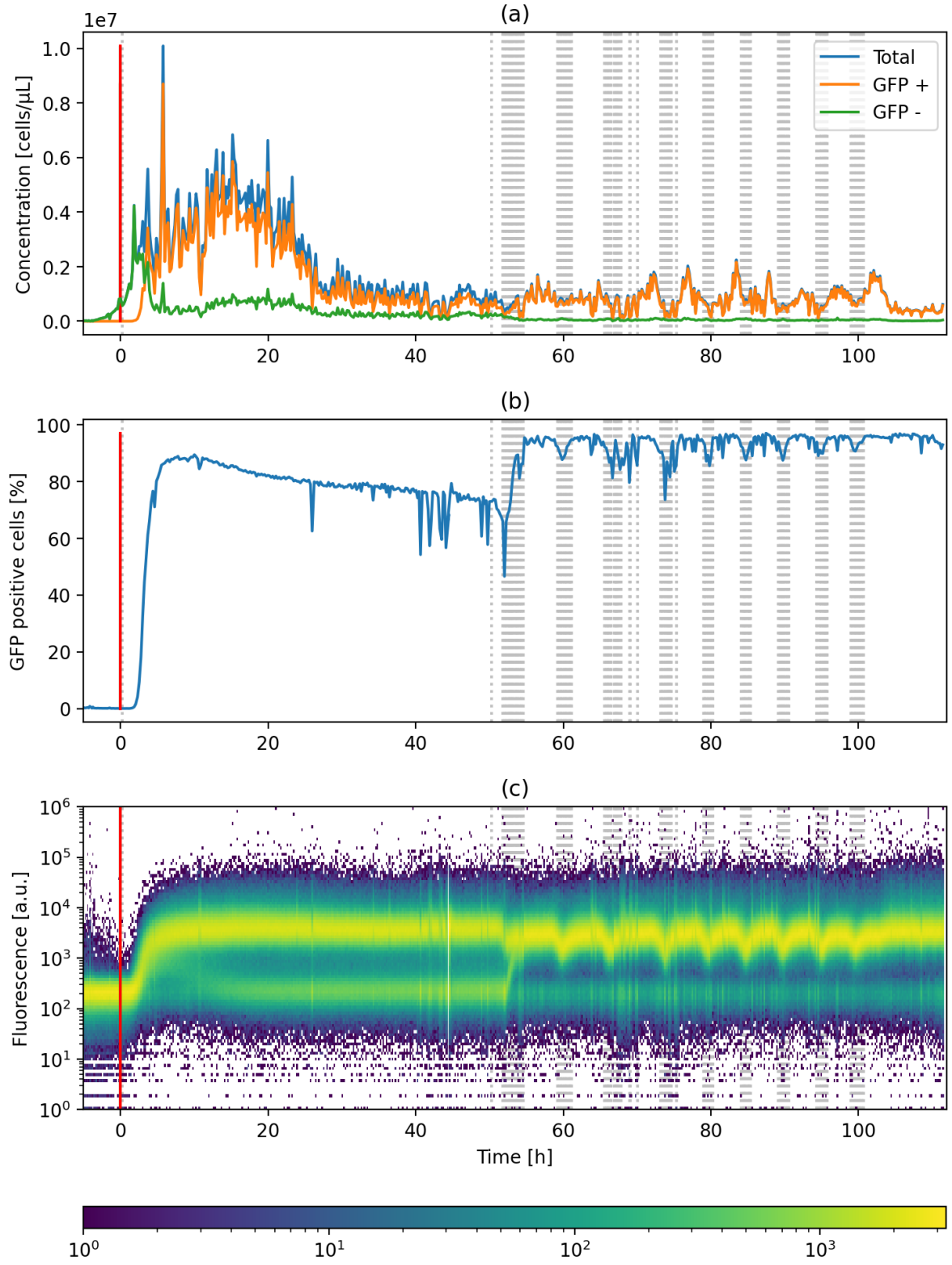


Figure A.2.3: Summary of an experiment where a segregostat phase follows a chemostat phase. Notably, no local minimum is observed after 20 h in the GFP+ proportion here. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The dashed grey areas show the periods when pulses of arabinose were added into the system. The red line represents the beginning time of the continuous cultivation phase.

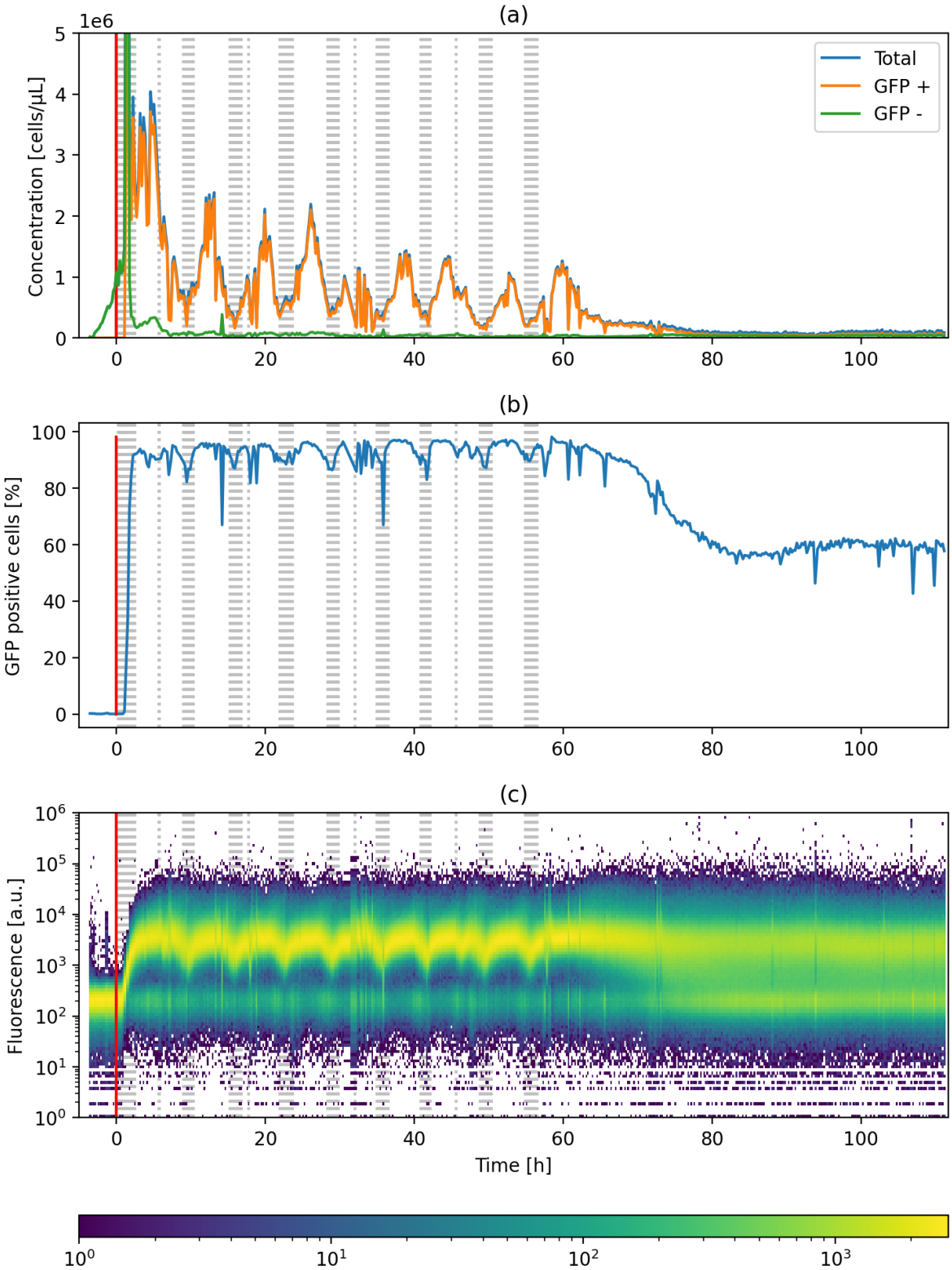


Figure A.2.4: Summary of a combined experiment with the segregostat phase first. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The dashed grey areas show the periods when pulses of arabinose were added into the system. The red line represents the beginning time of the continuous cultivation phase. In (a), the first few cropped measurements were of the order of magnitude of 10^7, which is obviously a measurement error and further shows that the concentration measurements cannot be trusted.

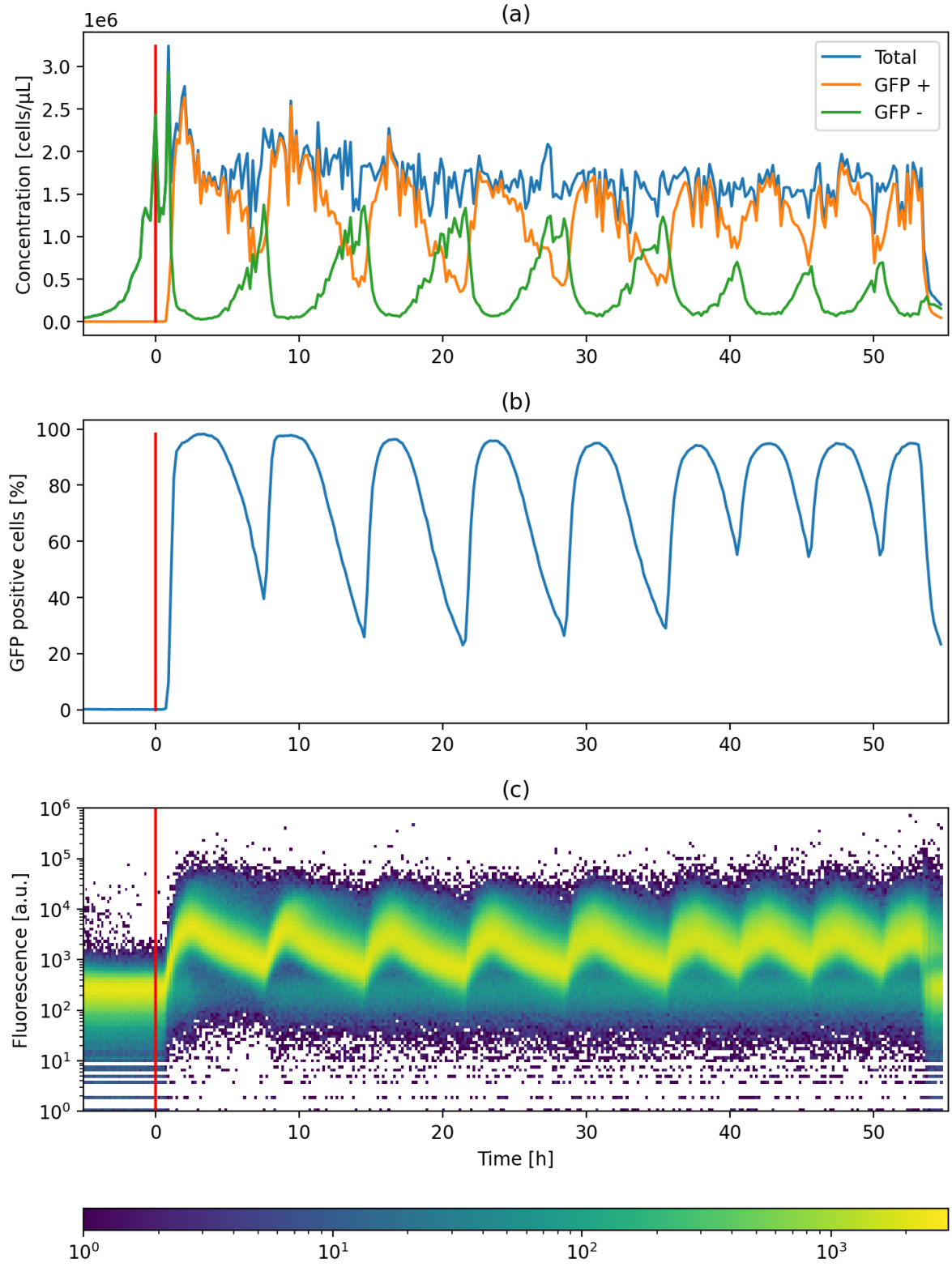


Figure A.2.5: Summary of an experiment with regulation at variable intervals. Every increase in GFP+ population corresponds to an injection of 1 g of arabinose in the system, first every 7 h, then every 5 h. In (a), the cell concentrations as measured by the flow cytometer; in (b), the proportion of GFP+ cells (with a threshold of 1000), and in (c) the heat map of how many events were counted for different fluorescence values — for each time point, the same amount of counted events was measured, the colour is therefore unrelated to the global concentration. The red line represents the beginning time of the continuous cultivation phase.

A.3 Full initial implementation of the observed switching mechanism

Before the initial forcing was decided upon, the full hypothesized mechanism was implemented. The idea was that when the responsive population is entirely composed of a single phenotype, it can only start switching if it detects arabinose is present for long enough. The variable used to detect this is p.t_o. It represents the time at which the concentration required for switching was detected, if it was, or is None if the past arabinose concentration did not allow for switching. p.get_ra(), p.get_rd(), p.get_dti(), p.get_dtf() are the functions that allow to retrieve the various parameters from the parameter and memory object p. dti and dtf are, respectively, the times during which the arabinose needs to be high enough or low enough for the population to start switching. ra and rd are the rates at which the population can switch, respectively, to inactive and activated states.

Another layer of complexity is added by the fact that a time delay may be considered between the time the cells detect enough (or not enough) arabinose and the time they act on it. To remedy this, the "future" proportion of cells is kept in memory instead of the present one. The future proportion is part of the variables integrated by solve_ivp. It depends on the present concentrations of glucose and arabinose, but the cells will only be able to access it after a time delay Δt_a, which is also the time delay before cells can grow on arabinose.

Overall, those different constraints can be put together in 4 different contributions:

- react_down(dx, x, t, p) and react_up(dx, x, t, p) detect, respectively, when the future population needs to start switching phenotypes.
- react_mid(dx, x, t, p) is responsible of what happens when the future proportion of the population is neither fully made up of activated cells, nor fully made up of inactive cells: then, it follows the arabinose concentration in the bioreactor.
- follow_p(dx, x, t, p) accesses the present value of the proportion of the responsive population that should be in each phenotypes, and modifies the derivatives of the cell concentration accordingly.

The code for this implementation is presented here and available at `https://gitlab.uliege.be/mipi/standalone/vincent-master-thesis/-/blob/8596593ba3caa85bc2a55c11a8711da10ec18625/segmodels/contributions.py`, at commit 8596593ba3caa85bc2a55c11a8711da10ec18625.

```
# x:
# 0 - G,   (non-responsive inactive population)
# 1 - A_r, (responsive activated population)
# 2 - G_r, (responsive inactive population)
# 3 - Ca,  (concentration of glucose)
# 4 - Cg   ( "             of arabinose)
# 5 - future A_r/(A_r+G_r)
# dx: derivative of x

def follow_p(dx, x, t, p):
    prop=p.get_present_p(t)
```

```
    tot=np.sum(x[1:3])
    da=p.get_present_dp(t)*tot

    #if da!=0: # just need to follow the rate from p
              # should never be 0 if prop!=0 and prop!=1
    # else
    if da==0 and prop!=x[1]/np.sum(x[1:3]):
        # p is constant at 0 or 1 and ccttions need to be brought in sync
        if prop==1: # small adjustment needed because growth and
                    # numeric error caused a shift
            da=p.get_ra()*tot
        elif prop==0:   # small adjustment needed because growth and
                        # numeric error caused a shift
            da=-p.get_rd()*tot

    dx[1]+=da
    dx[2]-=da


def react_down(dx, x, t, p):
    if x[5]==1:
        if x[3]>p.get_thresh(): # Ca > thresh
            if p.t_o is not None:
                p.t_o = None
        else: # Ca <= thresh
            if p.t_o is None:
                p.t_o=t
            else:
                # not enough arabinose
                # countdown to reaction was started beforehand
                # if countdown finished, p decreases
                if t>p.t_o+p.get_dtf():
                    dx[5]-=p.get_rd()

def react_up(dx, x, t, p):
    if x[5]==0:
        if x[3]<=p.get_thresh(): # Ca <= thresh
            if p.t_o is not None:
                p.t_o = None
        else: # Ca > thresh
            if p.t_o is None:
                p.t_o=t
            else:
                # not enough arabinose
                # countdown to reaction was started beforehand
                # if countdown finished, p decreases
                if t>p.t_o+p.get_dti():
                    dx[5]+=p.get_ra()
```

```
def react_mid(dx, x, t, p):
    if x[5]>0 and x[5]<1:
        if x[3]>p.get_thresh(): # Ca > thresh
            dx[5]+=p.get_ra()
        else:
            dx[5]-=p.get_rd()
```

Since this implementation was unstable, as mentioned in the main text, the final implementation simply did not use react_down and had added code to force the responsive population, once it started switching for the first time, to completely switch to using arabinose regardless of arabinose concentrations.

A.4 Solution of the simplified model derived from Thattai and Van Oudenaarden [1]

The assumption that μ_a and μ_g are constant let us with the set of equations:

$$\begin{cases} \frac{\mathrm{d}A}{\mathrm{d}t} &= \mu_{a,0}A - DA - H_1A + H_2G \\ \frac{\mathrm{d}A}{\mathrm{d}t} &= \mu_{g,0}G - DG + H_1A - H_2G \end{cases} \tag{A.4.1}$$

For ease of use, here $\mu_{a,0} - D - H_1$ is renamed a_1 and $\mu_{g,0} - D - H_2$ is renamed g_1 Those equations can be solved by first noting that:

$$G = \frac{\frac{\mathrm{d}A}{\mathrm{d}t} - a_1A}{H_2} \tag{A.4.2}$$

and

$$\frac{\mathrm{d}A}{\mathrm{d}t} = g_1G + H_1A \tag{A.4.3}$$

$$= \frac{g_1}{H_2}\left(\frac{\mathrm{d}A}{\mathrm{d}t} - a_1A\right) + H_1A \tag{A.4.4}$$

Then, taking the derivative of Equation A.4.1, one gets:

$$\frac{\mathrm{d}^2A}{\mathrm{d}t^2} = a_1\frac{\mathrm{d}A}{\mathrm{d}t} + H_2\frac{\mathrm{d}A}{\mathrm{d}t} \tag{A.4.5}$$

$$= a_1\frac{\mathrm{d}A}{\mathrm{d}t} + g_1\left(\frac{\mathrm{d}A}{\mathrm{d}t} - a_1A\right) + H_1H_2A \tag{A.4.6}$$

$$\Leftrightarrow \qquad 0 = -\frac{\mathrm{d}^2A}{\mathrm{d}t^2} + \alpha\frac{\mathrm{d}A}{\mathrm{d}t} + \beta A \qquad \text{(for ease of use)} \tag{A.4.7}$$

Such a differential equation homogeneous, meaning that there is no term independent from A. Such equations are known to have a solution of the form:

$$A = C_1\exp(-\lambda_1 t) + C_2\exp(-\lambda_2 t) \tag{A.4.8}$$

Where λ_1 and λ_2 are solutions of the characteristic polynomial $0 = -\lambda^2 + \alpha\lambda + \beta$. A solution such as Equation A.4.8 is valid only if the characteristic polynomial does not have twice the same root; if it did the solution would not be able to stabilise at a constant value other than 0 anyways. As it is, the equation can only tend towards something else than zero if either λ_1 or λ_2 is 0.

While this is not entirely impossible, one does not need to solve the equations to realise that if either characteristic value (let's say λ_2 for the sake of the argument) is null, then Equation A.4.8 becomes Equation A.4.9:

$$A = C_1\exp(-\lambda_1 t) + C_2 \tag{A.4.9}$$

Depending on the value of λ_1, this equation may either tend towards infinity in $t = \infty$, or tend monotonously towards C_2 (if $\lambda_2 = 0$, then it is not possible for λ_1 to have an imaginary part).

A.5 Results of the various optimisations for the segregostat

Models 0 to 3 do not make sense, and no optimisation was made for them. Models 9 and 11 were optimised, and the results are put here, but they do not make sense either and were not compared to another dataset. The resulting fit are presented below, while the parameter values are given on the next 2 pages.

Table A.5.1: Fitting value adjustment to the segregostat data for the various models, with the parameters shown on the next pages.

Model number	logcosh of fitted data [-]	logcosh of new data [-]	Corresponding error of fitted data [-]	Corresponding error of new data [-]
4	1.23×10^{-3}	2.19×10^{-2}	4.95×10^{-2}	2.10×10^{-1}
5	7.73×10^{-4}	2.28×10^{-2}	3.93×10^{-2}	2.14×10^{-1}
6	9.09×10^{-4}	2.30×10^{-2}	4.26×10^{-2}	2.15×10^{-1}
7	4.13×10^{-4}	2.10×10^{-2}	2.87×10^{-2}	2.06×10^{-1}
8	3.60×10^{-4}	2.20×10^{-2}	2.68×10^{-2}	2.10×10^{-1}
9	8.97×10^{-4}			
10	9.51×10^{-4}	2.11×10^{-2}	4.36×10^{-2}	2.06×10^{-1}
11	4.70×10^{-4}			
12	4.97×10^{-4}	2.27×10^{-2}	3.15×10^{-2}	2.14×10^{-1}
13	5.24×10^{-4}	2.01×10^{-2}	3.24×10^{-2}	2.01×10^{-1}
14	3.81×10^{-4}	2.20×10^{-2}	2.76×10^{-2}	2.10×10^{-1}
15	3.04×10^{-4}	2.12×10^{-2}	2.47×10^{-2}	2.07×10^{-1}
16	1.53×10^{-3}	1.85×10^{-2}	5.53×10^{-2}	1.93×10^{-1}
17	1.83×10^{-3}	4.09×10^{-2}	6.06×10^{-2}	2.88×10^{-1}
18	8.83×10^{-4}	2.00×10^{-2}	4.20×10^{-2}	2.01×10^{-1}
19	1.55×10^{-3}	4.54×10^{-2}	5.57×10^{-2}	3.04×10^{-1}
20	1.91×10^{-3}	2.18×10^{-2}	6.19×10^{-2}	2.10×10^{-1}
21	1.54×10^{-3}	2.47×10^{-2}	5.55×10^{-2}	2.23×10^{-1}
22	7.71×10^{-4}	2.14×10^{-2}	3.93×10^{-2}	2.07×10^{-1}
23	7.75×10^{-4}	2.65×10^{-2}	3.94×10^{-2}	2.31×10^{-1}

Table A.5.2: Growth parameter values resulting from fitting the models to the segregostat data.

Model number	$\mu_{g,\max}$	$\mu_{a,\max}$	K_g	K_a	$Y_{x,g}$	$Y_{x,a}$
	[h^{-1}]	[h^{-1}]	[g/L]	[g/L]	[g/g]	[g/g]
4	1.31	9.99×10^{-1}	1.54×10^{-2}	9.64×10^{-3}	6.30×10^{-4}	4.26×10^{-1}
5	3.76×10^{-1}	3.88	5.09×10^{-2}	1.03	5.98×10^{-2}	9.53×10^{-1}
6	2.72×10^{-1}	1.97	2.94×10^{-3}	1.01×10^{-2}	3.00×10^{-1}	6.17×10^{-1}
7	7.28×10^{-1}	3.27	2.00×10^{-2}	1.08×10^{-3}	1.03×10^{-1}	5.31×10^{-1}
8	5.48×10^{-1}	1.18	5.05×10^{-3}	1.73×10^{-2}	1.06×10^{-2}	6.65×10^{-1}
9	4.03×10^{-1}	3.20	2.40×10^{-3}	2.37×10^{-2}	3.43×10^{-1}	6.99×10^{-2}
10	7.58×10^{-1}	9.04×10^{-1}	1.80×10^{-2}	2.65×10^{-3}	1.61×10^{-2}	7.35×10^{-1}
11	3.15	4.45	1.37×10^{-3}	1.04×10^{-3}	8.51×10^{-1}	9.49×10^{-2}
12	5.31×10^{-1}	3.90	3.95×10^{-2}	2.74×10^{-2}	1.59×10^{-1}	4.20×10^{-1}
13	1.47×10^{-1}	3.71×10^{-1}	2.30×10^{-3}	1.42	8.63×10^{-1}	1.19×10^{-1}
14	2.34×10^{-1}	6.80×10^{-1}	2.17×10^{-3}	1.65×10^{-2}	3.21×10^{-1}	5.84×10^{-1}
15	4.55	8.78×10^{-1}	5.29×10^{-2}	5.88×10^{-3}	3.61×10^{-2}	5.26×10^{-1}
16	2.33	4.92×10^{-1}	1.04×10^{-2}	7.84×10^{-2}	4.15×10^{-3}	6.70×10^{-1}
17	4.50×10^{-1}	1.24×10^{-1}	1.89×10^{-2}	7.05×10^{-3}	5.64×10^{-1}	8.90×10^{-1}
18	1.19	4.50×10^{-1}	4.13×10^{-3}	1.33×10^{-2}	1.51×10^{-2}	9.74×10^{-1}
19	8.70	2.84×10^{-2}	1.41×10^{-3}	1.02×10^{-4}	9.61×10^{-1}	4.34×10^{-1}
20	1.67×10^{-1}	3.19×10^{-1}	5.29×10^{-3}	2.63×10^{-3}	7.17×10^{-1}	7.61×10^{-1}
21	8.47	3.59×10^{-1}	4.30×10^{-3}	6.45×10^{-2}	5.71×10^{-3}	2.06×10^{-1}
22	5.95×10^{-2}	8.89×10^{-3}	1.77×10^{-3}	1.73×10^{-2}	3.28×10^{-1}	7.03×10^{-1}
23	2.19	3.96×10^{-1}	8.49×10^{-3}	5.46×10^{-3}	3.79×10^{-2}	9.82×10^{-1}

Table A.5.3: Contribution parameter values resulting from fitting the models to the segregostat data.

Model number	H_1 [h^{-1}]	H_2 [h^{-1}]	Δt_g [h]	Δt_a [h]	p_{init} [-]	Δt_i [h]	r_a or r [h^{-1}]	threshold or k [g/L]	n [-]
4	3.25×10^{-2}	2.15×10^{-1}							
5	1.98×10^{-2}	8.38×10^{-2}							
6	1.57×10^{-2}	1.74×10^{-1}	6.23×10^{-2}	4.95×10^{-1}					
7	1.94×10^{-2}	5.11×10^{-2}	1.39×10^{-2}	5.34×10^{-1}					
8					6.04×10^{-1}	6.31×10^{-1}	8.09×10^{-1}	1.26×10^{-3}	
9					9.13×10^{-1}	5.14×10^{-1}	1.34	4.87×10^{-3}	
10			9.46×10^{-2}	4.80×10^{-1}	7.25×10^{-1}	4.91×10^{-1}	9.40×10^{-1}	1.04×10^{-3}	
11			1.05×10^{-1}	4.61×10^{-1}	5.76×10^{-1}	4.24×10^{-1}	3.61×10^{-1}	6.12×10^{-1}	
12	1.85×10^{-2}	4.72×10^{-1}			6.11×10^{-1}	4.69×10^{-1}	8.60×10^{-2}	5.01×10^{-3}	
13	4.37×10^{-2}	5.13×10^{-1}			7.44×10^{-1}	4.57×10^{-1}	5.42×10^{-1}	3.31×10^{-3}	
14	4.00×10^{-2}	6.10×10^{-1}	1.32×10^{-1}	4.29×10^{-1}	3.42×10^{-1}	5.18×10^{-1}	6.83×10^{-1}	1.31×10^{-3}	
15	2.49×10^{-2}	7.61×10^{-2}	8.10×10^{-2}	4.92×10^{-1}	7.49×10^{-1}	4.85×10^{-1}	5.09×10^{-1}	9.61×10^{-4}	
16							4.32×10^{-1}	3.30×10^{-3}	2.35
17							5.62×10^{-1}	7.26×10^{-3}	3.32
18			7.88×10^{-2}	4.87×10^{-1}			6.31×10^{-1}	1.15×10^{-3}	3.20
19			9.97×10^{-2}	6.89×10^{-1}			1.04	4.30×10^{-3}	2.83
20	3.75×10^{-2}	2.41×10^{-1}					3.32×10^{-1}	1.20×10^{-2}	2.73
21	3.94×10^{-2}	2.68×10^{-2}					4.48×10^{-1}	5.35×10^{-2}	3.07
22	5.89×10^{-2}	6.26×10^{-2}	7.94×10^{-2}	6.38×10^{-1}			1.11	1.86×10^{-3}	2.29
23	6.43×10^{-2}	4.46×10^{-2}	8.55×10^{-2}	5.50×10^{-1}			8.10×10^{-1}	3.41×10^{-3}	3.07

A.6 Results of the various optimisations for the chemostat

Models 0 to 3 do not make sense, and no optimisation was made for them. Models 9 and 11 were optimised, and the results are put here, but they do not make sense either and were not compared to another dataset. The resulting fit are presented below, while the parameter values are given on the next 2 pages.

Table A.6.1: Fitting value adjustment to the chemostat data for the various models, with the parameters shown on the next pages.

Model number	logcosh of fitted data [-]	logcosh of new data [-]	Corresponding error of fitted data [-]	Corresponding error of new data [-]
4	1.76×10^{-3}	3.00×10^{-2}	5.94×10^{-2}	2.46×10^{-1}
5	3.70×10^{-3}	4.08×10^{-2}	8.60×10^{-2}	2.88×10^{-1}
6	3.68×10^{-3}	3.43×10^{-2}	8.59×10^{-2}	2.64×10^{-1}
7	2.61×10^{-3}	3.00×10^{-2}	7.23×10^{-2}	2.46×10^{-1}
8	8.53×10^{-4}	4.59×10^{-2}	4.13×10^{-2}	3.05×10^{-1}
9	4.62×10^{-3}			
10	7.10×10^{-4}	4.91×10^{-2}	3.77×10^{-2}	3.16×10^{-1}
11	3.85×10^{-3}			
12	8.02×10^{-4}	4.08×10^{-2}	4.01×10^{-2}	2.88×10^{-1}
13	1.54×10^{-3}	3.82×10^{-2}	5.54×10^{-2}	2.78×10^{-1}
14	7.54×10^{-4}	3.25×10^{-2}	3.88×10^{-2}	2.56×10^{-1}
15	1.44×10^{-3}	3.36×10^{-2}	5.37×10^{-2}	2.61×10^{-1}
16	1.72×10^{-3}	3.89×10^{-2}	5.86×10^{-2}	2.81×10^{-1}
17	3.63×10^{-3}	5.89×10^{-2}	8.53×10^{-2}	3.47×10^{-1}
18	8.28×10^{-4}	3.04×10^{-2}	4.07×10^{-2}	2.48×10^{-1}
19	1.29×10^{-3}	9.21×10^{-2}	5.07×10^{-2}	4.36×10^{-1}
20	2.60×10^{-3}	3.75×10^{-2}	7.22×10^{-2}	2.75×10^{-1}
21	4.67×10^{-3}	1.11×10^{-1}	9.67×10^{-2}	4.80×10^{-1}
22	1.30×10^{-3}	3.30×10^{-2}	5.09×10^{-2}	2.58×10^{-1}
23	2.70×10^{-3}	1.12×10^{-1}	7.36×10^{-2}	4.82×10^{-1}

Table A.6.2: Growth parameter values resulting from fitting the models to the chemostat data.

Model number	$\mu_{g,\max}$	$\mu_{a,\max}$	K_g	K_a	$Y_{x,g}$	$Y_{x,a}$
	[h^{-1}]	[h^{-1}]	[g/L]	[g/L]	[g/g]	[g/g]
4	3.73	3.15×10^{-1}	6.13×10^{-3}	1.77×10^{-3}	1.59×10^{-2}	9.01×10^{-1}
5	3.67×10^{-2}	1.76	1.23×10^{-2}	2.71×10^{-3}	3.54×10^{-1}	9.80×10^{-1}
6	1.40×10^{-2}	1.35	1.96×10^{-3}	7.23×10^{-4}	4.44×10^{-2}	9.80×10^{-1}
7	3.89×10^{-1}	1.73	4.20×10^{-2}	1.81×10^{-2}	8.95×10^{-1}	9.05×10^{-1}
8	1.29	3.39	4.33×10^{-3}	1.06×10^{-3}	1.24×10^{-1}	6.52×10^{-1}
9	1.94	6.94×10^{-1}	1.12×10^{-2}	5.29×10^{-3}	8.42×10^{-2}	1.86×10^{-3}
10	6.24×10^{-1}	1.22	8.03×10^{-2}	5.99×10^{-3}	8.69×10^{-2}	4.03×10^{-1}
11	1.71	1.13	1.05×10^{-2}	4.68×10^{-2}	9.02×10^{-1}	9.83×10^{-3}
12	4.74×10^{-1}	2.01	2.62×10^{-2}	1.26×10^{-2}	3.45×10^{-1}	9.32×10^{-1}
13	8.89×10^{-1}	1.80	4.24×10^{-3}	1.11×10^{-2}	7.25×10^{-1}	7.04×10^{-2}
14	1.02	2.68	4.80×10^{-2}	4.52×10^{-3}	1.97×10^{-1}	7.74×10^{-1}
15	3.73×10^{-1}	1.67	7.92×10^{-4}	2.44×10^{-3}	8.95×10^{-1}	9.02×10^{-1}
16	6.71	2.14×10^{-1}	1.55×10^{-1}	3.57×10^{-3}	2.06×10^{-2}	1.31×10^{-1}
17	1.26	7.86×10^{-1}	1.71×10^{-3}	1.95×10^{-2}	9.25×10^{-1}	9.42×10^{-1}
18	6.57	2.08×10^{-1}	2.66×10^{-3}	4.81×10^{-2}	2.22×10^{-2}	6.27×10^{-1}
19	4.77×10^{-1}	1.15	1.53×10^{-2}	2.53×10^{-2}	9.13×10^{-1}	2.84×10^{-1}
20	1.91	4.38×10^{-1}	3.07×10^{-2}	5.53×10^{-2}	2.70×10^{-2}	5.80×10^{-1}
21	4.36×10^{-1}	1.28	2.40×10^{-2}	1.87×10^{-2}	5.59×10^{-1}	2.44×10^{-1}
22	1.27	1.40×10^{-1}	3.92×10^{-3}	4.27×10^{-3}	1.33×10^{-2}	1.64×10^{-1}
23	5.11×10^{-1}	5.46×10^{-1}	6.61×10^{-3}	7.29×10^{-3}	8.24×10^{-1}	1.13×10^{-1}

Table A.6.3: Contribution parameter values resulting from fitting the models to the chemostat data.

Model number	H_1 [h^{-1}]	H_2 [h^{-1}]	Δt_g [h]	Δt_a [h]	p_{init} [-]	Δt_i [h]	r_a or r [h^{-1}]	threshold or k [g/L]	n [-]
4	7.34×10^{-2}	3.74×10^{-1}							
5	3.04×10^{-1}	5.49×10^{-2}							
6	3.57×10^{-1}	1.35×10^{-1}	4.28×10^{-2}	6.00×10^{-1}					
7	1.36×10^{-1}	8.62×10^{-2}	4.70×10^{-2}	6.00×10^{-1}					
8					2.10×10^{-1}	4.56×10^{-1}	1.52×10^{-1}	7.49×10^{-4}	
9					5.95×10^{-1}	4.83×10^{-1}	5.40×10^{-1}	2.21×10^{-3}	
10			1.01×10^{-1}	4.27×10^{-1}	7.79×10^{-1}	5.68×10^{-1}	2.55×10^{-1}	9.24×10^{-4}	
11			9.88×10^{-2}	5.08×10^{-1}	5.47×10^{-1}	4.54×10^{-1}	7.26×10^{-1}	1.13×10^{-3}	
12	3.46×10^{-2}	2.28×10^{-2}			8.49×10^{-1}	5.24×10^{-1}	1.30×10^{-1}	3.30×10^{-3}	
13	4.75×10^{-2}	4.29×10^{-2}			9.27×10^{-1}	5.20×10^{-1}	9.21×10^{-1}	1.51×10^{-3}	
14	7.05×10^{-2}	1.54×10^{-1}	8.24×10^{-2}	4.74×10^{-1}	7.58×10^{-1}	5.85×10^{-1}	9.20×10^{-2}	7.74×10^{-3}	
15	8.84×10^{-2}	1.05×10^{-3}	2.08×10^{-2}	5.09×10^{-1}	3.10×10^{-1}	4.54×10^{-1}	2.40×10^{-1}	4.21×10^{-3}	
16							4.14×10^{-1}	4.62×10^{-3}	3.37
17							2.10×10^{-1}	6.68×10^{-3}	3.11
18			9.19×10^{-2}	6.01×10^{-1}			4.47×10^{-1}	2.96×10^{-2}	2.02
19			9.76×10^{-2}	5.29×10^{-1}			2.85×10^{-1}	1.65×10^{-3}	3.15
20	1.33×10^{-1}	2.04×10^{-2}					2.92×10^{-1}	3.72×10^{-3}	4.57
21	6.19×10^{-2}	4.61×10^{-2}					6.68×10^{-1}	1.70×10^{-3}	3.06
22	3.70×10^{-2}	4.00×10^{-2}	6.57×10^{-2}	4.22×10^{-1}			5.21×10^{-1}	1.28×10^{-2}	2.92
23	8.49×10^{-2}	9.33×10^{-2}	4.21×10^{-2}	5.47×10^{-1}			8.26×10^{-1}	3.91×10^{-4}	2.84

A.7 Results of the various optimisations for both experiments at once

Models 0 to 3 do not make sense, and no optimisation was made for them. Models 9 and 11 were optimised, and the results are put here, but they do not make sense either and were not compared to another dataset. The resulting fit are presented below, while the parameter values are given on the next 2 pages.

Table A.7.1: Fitting value adjustment to the data for the various models, with the parameters shown on the next pages.

Model number	logcosh of fitted data [-]	logcosh of new data [-]	Corresponding error of fitted data [-]	Corresponding error of new data [-]
4	1.32×10^{-2}	2.26×10^{-2}	1.63×10^{-1}	2.13×10^{-1}
5	1.28×10^{-2}	1.49×10^{-2}	1.60×10^{-1}	1.73×10^{-1}
6	9.79×10^{-3}	1.53×10^{-2}	1.40×10^{-1}	1.76×10^{-1}
7	1.08×10^{-2}	1.39×10^{-2}	1.47×10^{-1}	1.67×10^{-1}
8	1.35×10^{-2}	3.16×10^{-2}	1.65×10^{-1}	2.53×10^{-1}
9	1.14×10^{-2}			
10	9.28×10^{-3}	2.16×10^{-2}	1.36×10^{-1}	2.08×10^{-1}
11	1.02×10^{-2}			
12	1.26×10^{-2}	1.28×10^{-2}	1.59×10^{-1}	1.60×10^{-1}
13	1.09×10^{-2}	1.29×10^{-2}	1.48×10^{-1}	1.61×10^{-1}
14	1.01×10^{-2}	1.96×10^{-2}	1.43×10^{-1}	1.99×10^{-1}
15	1.10×10^{-2}	1.44×10^{-2}	1.49×10^{-1}	1.70×10^{-1}
16	1.35×10^{-2}	1.30×10^{-2}	1.65×10^{-1}	1.61×10^{-1}
17	6.18×10^{-3}	1.67×10^{-2}	1.11×10^{-1}	1.83×10^{-1}
18	1.24×10^{-2}	1.22×10^{-2}	1.58×10^{-1}	1.57×10^{-1}
19	5.88×10^{-3}	9.16×10^{-3}	1.09×10^{-1}	1.36×10^{-1}
20	1.38×10^{-2}	2.02×10^{-2}	1.66×10^{-1}	2.02×10^{-1}
21	1.40×10^{-2}	1.29×10^{-2}	1.68×10^{-1}	1.61×10^{-1}
22	1.26×10^{-2}	1.60×10^{-2}	1.59×10^{-1}	1.79×10^{-1}
23	5.60×10^{-3}	1.33×10^{-2}	1.06×10^{-1}	1.63×10^{-1}

Table A.7.2: Growth parameter values resulting from fitting the models to both experiments at once.

Model number	$\mu_{g,\max}$	$\mu_{a,\max}$	K_g	K_a	$Y_{x,g}$	$Y_{x,a}$
	[h^{-1}]	[h^{-1}]	[g/L]	[g/L]	[g/g]	[g/g]
4	3.36	1.69	1.51×10^{-2}	2.87×10^{-3}	6.99×10^{-2}	6.38×10^{-1}
5	4.68×10^{-1}	1.40	9.00×10^{-2}	1.35×10^{-2}	9.27×10^{-1}	9.28×10^{-1}
6	4.06	3.19	3.66×10^{-2}	5.13×10^{-3}	3.92×10^{-2}	3.83×10^{-1}
7	7.36×10^{-1}	3.35	1.10×10^{-2}	4.91×10^{-3}	5.57×10^{-1}	6.30×10^{-1}
8	5.14×10^{-1}	1.84	4.92×10^{-2}	7.13×10^{-3}	1.92×10^{-1}	9.79×10^{-1}
9	5.81×10^{-1}	1.43	2.24×10^{-2}	1.19×10^{-2}	7.71×10^{-1}	5.03×10^{-2}
10	5.02×10^{-1}	1.68	1.06×10^{-2}	7.92×10^{-3}	3.25×10^{-1}	7.22×10^{-1}
11	7.26×10^{-1}	3.51	5.21×10^{-3}	4.64×10^{-3}	4.87×10^{-1}	2.72×10^{-2}
12	2.02	2.50	1.48×10^{-3}	1.17×10^{-1}	4.84×10^{-1}	2.37×10^{-1}
13	7.07	5.41	1.19×10^{-1}	4.38×10^{-3}	9.07×10^{-1}	3.81×10^{-1}
14	6.05×10^{-1}	2.23	1.08×10^{-2}	2.95×10^{-3}	1.04×10^{-1}	9.92×10^{-1}
15	5.26	4.10	1.50×10^{-2}	1.66×10^{-2}	9.40×10^{-1}	4.38×10^{-1}
16	8.29×10^{-1}	1.26×10^{-1}	5.02×10^{-2}	4.39×10^{-3}	2.15×10^{-2}	7.02×10^{-1}
17	7.91	4.76×10^{-1}	1.18×10^{-2}	2.32×10^{-3}	2.05×10^{-1}	8.86×10^{-1}
18	1.19	2.15×10^{-1}	1.84×10^{-2}	4.85×10^{-3}	1.59×10^{-2}	4.29×10^{-1}
19	1.79×10^{-1}	5.88×10^{-1}	1.32×10^{-2}	2.28×10^{-3}	4.58×10^{-1}	8.90×10^{-1}
20	2.59	3.97×10^{-1}	2.64×10^{-2}	1.31×10^{-2}	3.47×10^{-2}	6.51×10^{-1}
21	5.94×10^{-1}	2.97×10^{-1}	1.02×10^{-2}	3.38×10^{-3}	6.46×10^{-3}	8.00×10^{-1}
22	7.44	3.46×10^{-1}	1.81×10^{-2}	8.45×10^{-2}	3.03×10^{-2}	5.79×10^{-1}
23	2.97	3.49×10^{-1}	7.47×10^{-3}	6.06×10^{-3}	2.78×10^{-1}	5.74×10^{-1}

Table A.7.3: Contribution parameter values resulting from fitting the models to to both experiments at once.

Model number	H_1 [h^{-1}]	H_2 [h^{-1}]	Δt_g [h]	Δt_a [h]	p_{init} [-]	Δt_i [h]	r_a or r [h^{-1}]	threshold or k [g/L]	n [-]
4	1.44×10^{-2}	1.84×10^{-1}							
5	6.42×10^{-2}	1.04×10^{-1}							
6	2.65×10^{-2}	5.97×10^{-2}	2.50×10^{-2}	5.21×10^{-1}					
7	5.16×10^{-2}	4.04×10^{-2}	2.92×10^{-2}	6.73×10^{-1}					
8					6.52×10^{-1}	5.08×10^{-1}	7.60×10^{-1}	1.95×10^{-3}	
9					7.01×10^{-1}	5.30×10^{-1}	9.70×10^{-1}	9.44×10^{-4}	
10			1.00×10^{-1}	5.00×10^{-1}	7.49×10^{-1}	4.95×10^{-1}	5.92×10^{-1}	1.74×10^{-3}	
11			1.04×10^{-1}	5.16×10^{-1}	4.59×10^{-1}	4.83×10^{-1}	4.25×10^{-1}	2.94×10^{-3}	
12	2.23×10^{-1}	2.29			4.50×10^{-1}	6.26×10^{-1}	5.44×10^{-1}	3.82×10^{-5}	
13	3.21×10^{-2}	5.11×10^{-2}			3.32×10^{-1}	4.72×10^{-1}	1.56×10^{-1}	1.78×10^{-3}	
14	2.45×10^{-2}	2.81×10^{-2}	8.88×10^{-2}	3.88×10^{-1}	9.06×10^{-1}	5.26×10^{-1}	1.88×10^{-1}	4.68×10^{-3}	
15	2.15×10^{-2}	5.27×10^{-3}	1.45×10^{-1}	3.85×10^{-1}	2.35×10^{-1}	5.09×10^{-1}	1.24	2.72×10^{-4}	
16							5.28×10^{-1}	1.42×10^{-3}	3.24
17							1.69×10^{-1}	3.32×10^{-2}	2.89
18			2.05×10^{-1}	5.79×10^{-1}			7.33×10^{-1}	6.77×10^{-4}	2.51
19			7.16×10^{-2}	4.98×10^{-1}			1.84×10^{-1}	2.72×10^{-2}	3.05
20	7.96×10^{-2}	1.18×10^{-1}					4.63×10^{-1}	1.67×10^{-3}	3.05
21	1.96×10^{-1}	4.20×10^{-2}					4.72×10^{-1}	1.34×10^{-3}	3.06
22	8.99×10^{-2}	4.00×10^{-2}	9.53×10^{-2}	4.90×10^{-1}			8.19×10^{-1}	1.38×10^{-2}	2.42
23	1.48×10^{-2}	1.66×10^{-1}	1.23×10^{-1}	4.64×10^{-1}			1.77×10^{-1}	4.74×10^{-2}	3.01