Deterministic models to decipher the lag phase duration during diauxie

Florian Dupeuble, Alain Rapaport, Thomas Guilmeau, Josué Tchouanti, Brice Enjalbert, Carine Bideaux, Jean-Philippe Steyer, Aida Feddaoui, Jérome Harmand

To cite this version:
Florian Dupeuble, Alain Rapaport, Thomas Guilmeau, Josué Tchouanti, Brice Enjalbert, et al.. Deterministic models to decipher the lag phase duration during diauxie. MATHMOD 2022, Jul 2022, Vienna, Austria. hal-03610317
Deterministic models to decipher the lag phase duration during diauxie


* INRAE, Univ. Montpellier, LBE, Narbonne, France (e-mail: florian.dupeuble@insa-lyon.fr, thomas.guilmeau@ensta-paris.fr, jean-philippe.steyer@inrae.fr, jerome.harmand@inrae.fr).
** INRAE, Univ. Montpellier, MISTEA, Montpellier, France (e-mail: alain.rapaport@inrae.fr, daliyoucef.manel@gmail.com)
*** CMAP, CNRS, Ecole Polytechnique, IP Paris, 91128 Palaiseau, France (e-mail: josuetchouanti@yahoo.fr)
**** TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France (e-mail: brice.enjalbert@insa-toulouse.fr)

Abstract: The deterministic model developed by Graham et al. [2020], which is the approximation in large population of a stochastic model, allowed the authors to propose a ‘macroscopic description’ of metabolic heterogeneity of Escherichia coli growing on glucose and xylose. However, these models did not include any mechanistic model to explain the variations of the duration of the ‘lag-phase’ observed when the glucose is exhausted and before the xylose is being consumed. Here, we propose a deterministic mechanistic model to explain how E. coli switches its consumption of a sugar to another one depending on the dynamic of intracellular XylR molecules. The model is developed and investigated numerically. It reveals some important observability issues.

Keywords: Diauxic growth, deterministic modeling, ODE, mass-balance models, simulation

1. INTRODUCTION

Described for the first time by Monod, the diauxic growth consists in a biphasic growth in a bacterial population consuming two different sugars in a closed medium, Monod [1942]. To study this phenomenon, a number of models have been proposed in the literature. However, a specific mechanism able to modulate the length of the lag phase as a function of its state was not proposed neither modeled and in all models the population was supposed to behave uniformly with respect to the two sugars, Turon [2015], Dedem and Moo-Young [1975], Liquori et al. [1981]. In Barthe et al. [2020], it was proposed for a pure culture of Escherichia coli growing on glucose and xylose that ‘metabolic heterogeneity’, i.e. the fact that all cells do not readily switch from glucose to xylose when glucose is exhausted, could be explained by a specific internal process related to the ‘XylR transcription factor’, Song and Park [1997], Laikova et al. [2001], Schmidt et al. [2016]. This molecule is known to play a central role in the activation of the xylose pathway. In particular, depending on its history, a given cell would not contain the same number of XylR molecules as the others. The dynamic at which such a molecule would attach on a specific genomic site could explain the modulation of the duration of the lag-phases sometimes observed in the diauxic growth of this microorganism. The deterministic model developed by Graham et al. [2020], which is the approximation in large population of a stochastic model, allowed the authors to propose a ‘macroscopic description’ of metabolic heterogeneity. However, no mechanistic modeling was proposed.

Here, we propose a deterministic mechanistic model to explain how E. coli switches its consumption of one sugar to another depending on the dynamic of intracellular XylR molecules. The paper is organized as follows: first, the model proposed by Graham et al. [2020], hereafter called ‘macroscopic model’, is recalled together with the main modeling assumptions. Second, a deterministic mechanistic model (called the ‘compartmental model’) is presented. In a third part, simulations allowing us to provide a first analysis of the dynamical behaviours of both models are presented and used to compare their predictions. Finally, a number of modeling recommendations are suggested and discussed.

2. A MACROSCOPIC DETERMINISTIC MODEL

The model proposed by Graham et al. [2020] considered a pure culture of E. coli growing on two sugars, glucose (which concentration is noted $s_1$) and xylose ($s_2$). The population of bacteria is divided in two sub-populations denoted $n_1$ and $n_2$, depending on the sugars on which they are able to grow. The sub-population $n_1$ can only consume glucose while $n_2$ preferably consumes $s_1$ but is also able to consume $s_2$. Both sub-populations produce and consume acetate noted $a$. Populations $n_1$ and $n_2$ grow at rates $b_1(s_1, s_2, a)$ and $b_2(s_1, s_2, a)$, respectively. Each cell has an internal and continuous production of XylR which
accelerates in the presence of $s_2$ but which is inhibited by $s - 1$ due to the catabolic repression. When a XylR molecule attaches onto a specific genomic site, the xylose pathway may be ‘activated’: at this instant, the cell is no longer of type $n_1$ and becomes part of $n_2$. The sub-population $n_2$ is then able to consume $s_2$ but under the repression of $s_1$. It means that $n_2$ consumes $s_2$ under the condition that $s_1$ remains under a given threshold. This mechanism is modeled using a function $\eta_1(s_1, s_2)$ which is the specific rate at which $n_2$ converts into $n_1$. Similarly, if glucose increases again in the medium, $\eta_2(s_1)$ is a function characterizing the switching rate of $n_2$ bacterial fraction into $n_1$.

These dynamics are characterized as follows:

$$\begin{align*}
\frac{dn_1}{dt} &= (b_1(s_1, s_2, a) - \eta_1(s_1, s_2))n_1 + \eta_2(s_1)n_2 \\
\frac{dn_2}{dt} &= (b_2(s_1, s_2, a) - \eta_2(s_1))n_2 + \eta_1(s_1, s_2, a)n_1
\end{align*}$$

with $b_1(s_1, s_2, a) = \mu_1(s_1, a) + \mu_3(a)$, $b_2(s_1, s_2, a) = \mu_2(s_2, a) + \mu_3(a)$ and where $\mu_1(s_1, a) = \frac{\mu_1 s_1}{s_1 + s_1 + s_1}$, $\mu_2(s_2, a) = \frac{\mu_2 s_2}{s_2 + s_2 + s_2}$, $\mu_3(a) = \frac{\mu_3 a}{a + a + a + a}$, $\eta_1(s_1, s_2) = \frac{\eta_1 s_1}{s_1 + s_1 + s_1}$ and $\eta_2(s_1) = \frac{\eta_2 s_2}{s_2}$.

Acetate being produced by both sub-populations, and noting $\theta_1$ and $\theta_2$ production yields of acetate by $n_1$ and $n_2$, and $q_1$, $q_2$ and $q_3$ the yield coefficients of the biomass fractions one has:

$$\begin{align*}
\frac{dn_1}{dt} &= (\mu_1 + \mu_3 - \eta_1)n_1 + \eta_2n_2 \\
\frac{dn_2}{dt} &= (\mu_2 + \mu_3 - \eta_2)n_2 + \eta_1n_1 \\
\frac{ds_1}{dt} &= \frac{\mu_1}{\theta_1}n_1 \\
\frac{ds_2}{dt} &= \frac{\mu_2}{\theta_2}n_2 \\
\frac{ds_3}{dt} &= \frac{\mu_3}{q_3} (n_1 + n_2) + \theta_1\mu_1n_1 + \theta_2\mu_2n_2
\end{align*}$$

Model parameters have been identified using real data to come up with this first candidate model, cf. Graham et al. [2020]. Parameters values are taken from Barthe et al. [2020]. This model allows us to simulate variations of the time lag, either in changing initial conditions or model parameters. However, it remains a macroscopic model in the sense that no mechanism explicitly explains the emergence of sub-populations from an initial homogeneous species. Following the mechanistic hypothesis proposed in Barthe et al. [2020], we propose in the next section a more detailed model called the ‘Compartmental model’.

3. COMPARTMENTAL MODEL

In this section, we develop a mechanistic model of E. coli growing on glucose and xylose in describing specifically the dynamic of the activation of the xylose pathway via the fixation of the XylR molecule on a specific genomic site.

In terms of biomass, the model is based on the same important assumption than the macroscopic model but both populations are further subdivided into compartments depending on the number of XylR they contain. As already mentioned, it comes from the hypothesis that the ‘transcription factor’, i.e. the XylR molecule, can activate the xylose pathway as soon as it is fixed on a specific genomic site allowing bacteria to consume xylose. However, it exists ‘trap sites’ (which are xylA promoters on plasmids) on which XylR can fix too but without giving E. coli the ability to consume xylose. In the following, we denote by $N$ the number of trap sites. Accounting for the genomic site allowing the bacteria to be activated if the XylR attaches to it, the total number of sites on which XylR can fix is $n = N + 1$.

Let us note $X_i$ the non-activated and $Y_i$ the activated bacteria of the compartmental model. $X_i$ contain $i$ XylR molecules ($i = 0...N$). On the opposite, $Y_i$ contain $i + 1$ XylR molecules : $i$ are fixed on trap sites and $1$ is fixed on the promoter. For example, $X_5$ means that all cells within this variable contains $5$ XylR fixed onto $5$ trap sites, while the activated cells in the class $Y_5$ contain $6$ XylR of which $5$ are fixed onto $5$ trap sites and one onto the genomic site. Likewise, $X_0$ is used to describe a cell containing no XylR at all, whereas $Y_0$ is a cell with one XylR fixed onto the ‘promoter site’.

In the macroscopic model, ‘biomass activation’ is modelled by the switching of a fraction of cells from $n_1$ to the compartment $n_2$ at a rate that depends on glucose and xylose concentrations. In the compartmental model, this passage is modeled with more details as the result of the intracellular XylR dynamic. In other words, the underlying idea is that while the lag duration is mostly the result of model parameters in the macroscopic model, it will be the result of the initial distribution of the biomass in the different compartments in the compartmental model, once its parameters will have been set once and for all.

The XylR dynamic depends on three distinct processes that are now being described.

3.1 Modelling XylR production

During a time $dt$, some XylR molecules are produced inside each cell. When a XylR molecule is produced, it is assumed it attaches to a site. If this site is a trap site, a cell of the class $X_i$ quits this class and becomes a cell within the class $X_{i+1}$. Following the same reasoning, a cell of the class $Y_i$ quits this class and becomes a cell within the class $Y_{i+1}$.

Now, if the XylR attaches to the promoter site, the cell becomes activated: a cell of the class $X_i$ quits this class and becomes a cell within the class $Y_i$.

At which rate does this process occur? The intracellular XylR production rate is continuous but modulated by the concentrations in glucose and xylose in the environment: it is repressed by glucose (catabolic repression occurring even at low glucose concentration) but promoted by xylose. We define $\rho_0(s_1, s_2)$ the specific production rate of XylR by one compartment as:

$$\rho_0(s_1, s_2) = \frac{k_1^4}{k_1^4 + s_1} (p_0 + \frac{\rho_0 s_2}{k_2 + s_2})$$

This function has the specific property of acting as a switching function being ‘activated’ depending on the
Once attached to a site, it is hypothesized that a XylR molecule can degrade and detach. We define the specific degradation rate $\alpha(s_1)$ similar to $\eta_2(s_1)$.

$$\alpha(s_1) = \frac{s_1}{k_3 + s_1}$$

Following the same reasoning as in the previous section, this rate is modulated by the probability of such process to happen. And thus, we come up with the following dynamics:

$$\begin{cases}
\dot{X}_0 = -\rho_0 X_0 \\
\dot{X}_i = -\rho_0 X_i + \frac{N - i + 1}{N - i + 2} \rho_0 X_{i-1} \\
\dot{X}_N = -\rho_0 X_N + \frac{1}{N+1} \rho_0 X_{N-1} \\
\dot{Y}_0 = -\rho_0 Y_0 + \frac{1}{N+1} \rho_0 Y_{0} \\
\dot{Y}_i = -\rho_0 Y_i + \frac{1}{N - i + 1} \rho_0 Y_{i+1} + \rho_0 Y_{i-1} \\
\dot{Y}_N = \rho_0 X_N + \rho_0 Y_{N-1} \\
\end{cases}$$

### 3.3 Cell division

The last process considered is the fate of the XylR molecules during cell division. In such a case, we distinguish between two different situations.

- If the cell is in a state where the production of XylR is sufficiently high (thus for cells growing on xylose), postulating that a cell divides into two exactly identical cells, it is assumed that it gives birth to two cells containing the same number of XylR. To sum up, $Y_i$, when growing on xylose (but not on acetate) will always divide into two new $Y_i$.

- If the cell is growing on glucose or acetate, the production of XylR molecules is slowed down: it does not allow the production of two identical daughters and it is assumed that the two cells produced share the number of XylR of the mother. However, how modeling the way the number of XylR are shared between the two daughters? Several modeling possibilities were hypothesized and it was finally chosen that the daughter cells shared half-half of the XylR. This dynamic is making more complex with odd $i$: in this case, one daughter cell will receive $\frac{i+1}{2}$ XylR proteins and the other one $\frac{i}{2}$.

In addition, we assume that during division, $X_i$ (resp. $Y_{i}$) can only produce $X_i$ (resp. $Y_{i}$).

To model the transition from a compartment to another one for the cell division process, we use a matrix $M$ of dimension $n \times n$ that we multiply by the vector $X = (X_0, ..., X_i, ..., X_N)$ or the vector $Y = (Y_0, ..., Y_i, ..., Y_N)$. Note that the dimensions of the matrix $M$ depends on $N$. Let us first define a $M_{\text{div}}$ matrix (of the same size as $M$), which is the new population of cells after a division without taking into account the mother cells disappearance. For $j \in [1, n]$ : $M_{\text{div}}^{i\frac{1}{2}} = 1$ and $M_{\text{div}}^{i\frac{1}{2}+1, j} = 1$ for $j$ even and $M_{\text{div}}^{i\frac{1}{2}+1, j} = 2$ for $j$ odd and $M_{\text{div}}$ equals 0 everywhere else. The terms $M_{\text{div}}^{i\frac{1}{2}+1, i} = 1$ and $M_{\text{div}}^{i\frac{1}{2}+1, i} = 1$ correspond to the $(i+1)$ compartment with odd number of XylR dividing half into the $\frac{i}{2}$ and half the $\frac{i}{2} + 1$ compartments. The $M_{\text{div}}^{i\frac{1}{2}+1, i}$ correspond to the $(i+1)$ compartment with integer number of XylR dividing into the $\frac{i+1}{2}$ compartment.

$M$ is used to characterize the total variation of the population due to cell division. To obtain it, we have to subtract the population after division to the population before division in computing $M = M_{\text{div}} - I_n$. 

![Fig. 1. $\rho_0$ function of $s_1$ and $s_2$](image-url)
Supposing the growth rates are identical to those of the macroscopic model, we have:

\[
\begin{aligned}
\dot{X}_i &= (\mu_1 + \mu_3)(MX)_i \\
\dot{Y}_i &= \mu_3(MY)_i + \mu_2 Y_i \\
\end{aligned}
\]  

### 3.4 Sugar dynamics

The evolution of glucose and xylose, noted \( s_1 \) and \( s_2 \) as in the previous model, and the evolution of acetate are the same as in the previous model. With \( \tilde{X} \) and \( \tilde{Y} \) for the different simulations are taken from Barthe et al. (2020). In this section, we perform a number of simulations in order to compare predictions of both models. Parameters used for the different simulations are taken from Barthe et al. [2020].

\[
\begin{aligned}
\dot{s}_1 &= -\frac{1}{q_1} \mu_1 \tilde{X} \\
\dot{s}_2 &= -\frac{1}{q_2} \mu_2 \tilde{Y} \\
\dot{a} &= -\frac{1}{q_3} \mu_3(\tilde{X} + \tilde{Y}) + \theta_1 \mu_1 \tilde{X}s_1 + \theta_2 \mu_2 s_2 \tilde{Y} \\
\end{aligned}
\]  

### 3.5 Model dynamics

Adding the different components of the dynamic, we finally have:

\[
\begin{aligned}
\dot{X}_0 &= (\mu_1 + \mu_3)(X_0 + X_1) + \alpha Y_0 + \alpha X_1 \\
&\quad -\rho_0 X_0 \\
\dot{X}_i &= (\mu_1 + \mu_3)(MX)_i + \frac{1}{i + 1} \alpha Y_i + \alpha X_{i+1} \\
&\quad -\alpha X_i - \rho_0 X_i + \frac{N - i + 1}{N - i + 2} \rho_0 X_{i-1} \\
\dot{X}_N &= (\mu_1 + \mu_3)(MX)_N + \frac{1}{N} \alpha Y_N - \alpha X_N \\
&\quad -\rho_0 X_N + \frac{1}{N + 1} \rho_0 X_{N-1} \\
\dot{Y}_0 &= \mu_3(MY)_0 + \mu_2 Y_0 + \frac{1}{2} \alpha Y_1 - \alpha Y_0 \\
&\quad -\rho_0 Y_0 + \frac{1}{N + 1} \rho_0 X_i \\
\dot{Y}_i &= \mu_3(MY)_i + \mu_2 Y_i + \frac{i + 1}{i + 2} \alpha Y_{i+1} - \alpha Y_i \\
&\quad -\rho_0 Y_i + \frac{1}{N - i + 1} \rho_0 X_i + \rho_0 Y_{i-1} \\
\dot{Y}_N &= \mu_3(MY)_N + \mu_2 Y_N - \alpha Y_N + \rho_0 X_N + \rho_0 Y_{N-1} \\
\dot{s}_1 &= -\frac{1}{q_1} \mu_1 \tilde{X} \\
\dot{s}_2 &= -\frac{1}{q_2} \mu_2 \tilde{Y} \\
\dot{a} &= -\frac{1}{q_3} \mu_3(\tilde{X} + \tilde{Y}) + \theta_2 \mu_2 s_2 \tilde{Y} + \theta_1 \mu_1 \tilde{X}s_1 \\
\end{aligned}
\]  

### 4. SIMULATIONS

In this section, we perform a number of simulations in order to compare predictions of both models. Parameters used for the different simulations are taken from Barthe et al. [2020].

As already mentioned, changing the distribution of the initial biomass in the compartmental model is related to control the lag duration. In other words, we would like to explain and predict the differences of lag durations using the compartmental model by the initial distribution of XylR in the population while it directly depends on model parameters in the macroscopic model.

First, we compare both models using numerical simulations for ‘equivalent’ initial conditions (i.e. the same proportion of active/non-active bacteria and quantities of sugar). The initial conditions in sugar concentrations and in biomass are thus equal: in the compartmental model, uniform initial conditions are chosen. For \( n_1(0) = X(0) = 0.28 \times 0.75, n_2(0) = Y(0) = 0.28 \times 0.25, s_1 = 10(g.L^{-1}), s_2 = 10(g.L^{-1}), a = 0 \). Time in hours.

As we can see in Figure (2), the lag duration is longer in the compartmental model compared to the macroscopic model. For \( n_1(0) = X(0) = 0.28 \times 0.75, n_2(0) = Y(0) = 0.28 \times 0.25 \), we observed time lags of 1.51 h and 1.16 h, respectively.

Now, another objective is to be able to better understand how the initial conditions of the compartmental model allows to ‘control the lag’. Said otherwise, the question is to ‘link’ initial conditions of the compartmental model (initial biomass distribution) with the fraction of activated/inactivated biomasses of the macroscopic model. To do so, we proceed as follows:

![Image](image_url)
For a given fraction of initial biomass \( n_1(0) \) and \( n_2(0) \), we generate biomass and sugar profiles using the macroscopic model. We call these profiles ‘reference profile’ in the following; we identify some initial conditions for the compartmental model using an optimization approach that allows us to reproduce as well as possible the reference profile. To do so we used a least squared method. We did this work for the initial conditions given by \( n_1(0) = X(0) = 0.28 \times 0.75, n_2(0) = Y(0) = 0.28 \times 0.25, s_1 = 10(g.L^{-1}), s_2 = 10(g.L^{-1}), a = 0 \). Time in hours.

- For a given fraction of initial biomass \( n_1(0) \) and \( n_2(0) \), we generate biomass and sugar profiles using the macroscopic model. We call these profiles ‘reference profile’ in the following;
- We identify some initial conditions for the compartmental model using an optimization approach that allows us to reproduce as well as possible the reference profile. To do so we used a least squared method.

We did this work for the initial conditions given by \( n_1(0) = X(0) = 0.28 \times 0.75, n_2(0) = Y(0) = 0.28 \times 0.25, s_1 = 10(g.L^{-1}), s_2 = 10(g.L^{-1}), a = 0 \). Repeating the previous described optimization procedure from initial conditions of the biomass distribution of the compartmental model (the degrees of freedom of the optimization) allowed us to highlight a very important property of the compartmental model, that is the fact that it is not observable. Indeed, in Figure 4, we plotted 20 optimized biomass distributions allowing the compartmental model to reproduce with a comparable and acceptable accuracy the trajectories of the macroscopic model departing from different sets of initial conditions. Since many sets of initial conditions predict the same trajectories of the sugar and of the total biomass, the system is probably not observable.

A second idea is to ‘think practically’: in practice, prior to any experimental work, the biomass is prepared in specific conditions (cf. for instance Barthe et al. [2020]). The followed idea is to use the compartmental model as a virtual process to run simulations - several times - with arbitrary initial conditions on \( 10 \ g/L \) of glucose only. The final - very similar - distributions of biomass obtained under these conditions (starting from several random initial conditions) are plotted in Figure 6. Said otherwise, when the biomass is growing on glucose only, the number of degrees of freedom to the number of parameters of this distribution. The problem then is to know in which distribution we should restrict the search for the solution...

To investigate this question, we run again the optimization procedure described previously but in limiting the number of freedom degrees for the parameters of a normal distribution. The results of 20 optimizations are plotted in Figure 5. Unfortunately, although the number of freedom degree is drastically reduced, the results indicate that the system might not be observable.
Fig. 6. Final distributions of biomass after growing on 10(g.L⁻¹) glucose, starting with random initial distributions (50 repetitions)

Fig. 7. Final distributions of biomass after growing on 10(g.L⁻¹) of xylose starting with random initial distributions (50 repetitions)

the biomass compartments with a low number of XylR are attracting whatever the initial conditions are. Now, let us repeat this procedure but assuming there is only xylose in the medium. The results are now plotted in Figure 7. The compartments of both activated and inactivated biomasses with a low number of XylR are now attracting. In both cases, the final distributions obtained can obviously be approximated with a normal distribution in low XylR compartments of X and/or in Y.

Taking these results into account, we can further constraint the optimization algorithm and only consider that the compartments with low-XylR can be non-zero. If we retain only 3 or 4 first compartments, the number of parameters is not really more important than when considering parameters of the distributions. However, after several tests, it has not been possible to find only one set of initial conditions to explain the generated data. Thus, at present time, the observability problem of the compartmental model remains a key issue, to be investigated deeper in a coming work.

5. CONCLUSION

In this paper, we presented a new deterministic model of the diauxic growth observed in E. coli growing on glucose and xylose to find a possible mechanistic modeling of the lag phase duration. We compared its predictions to those obtained with another model that has been confronted to experimental data, and showed that they could give similar predictions of the lag phase duration. However, the lag time was sometimes different, highlighting a problem of unobservability. We have pursued several avenues, including the modifications of the model or the reduction of identified parameters, which unfortunately did not lead to the resolution of this problem which remains open at present time. Different techniques could be tested to free ourselves of the uniqueness of the solutions (set membership, multi-valued observers, change of coordinates...) as an alternative to reducing the model.

ACKNOWLEDGEMENTS

This work was supported by the 3BCAR Carnot (project HME) and the ANR (project JANUS; ANR-19-CE43-0004-01).

REFERENCES


