Synthetic Biobots

Math Model

Synthetic
Biobots

Author: José Santiago Jara-Sarracino

iGEM Design League

Season 2022

Introduction

Metabolite synthesis by enzymatic reactions is a complex process for every life being on Earth. In nature, the synthesis machinery is a system with a big amount of components and variables that affect the final result. One of the most important points to take into account about such mechanisms is the time taken to carry out each of those reactions. If during a metabolic pathway an enzyme is expressed a lot, but its substrate is not present in the required amount, the protein won't be able to work. Therefore, the organism is gonna suffer the consequences of wasting amino acids that could have been used in a more urgent task. Because of that, evolution has favored the apparition of many tools that allow cells to optimize what is translated and the moment it is done in order to keep homeostasis.

One of them can be reflected in some transcription networks configuration. A very common arrangement in metabolic transcription networks to solve this problem is the creation of temporal programs, which allow proteins to be translated more or less at the same time that its metabolite is accumulated. This is often achieved by the sequential activation of the promoters implied in the pathway, in which each of them uses the same transcription factor, but with differential affinity for it. That means that some of them will be more active with less concentration of the transcription factor than the others, consequently setting up a serial expression of the genes according to the increasing activator concentration. Those configurations are called Single Input Modules (SIM), since a whole set of genes are regulated by a single factor.[1]

During synthetic biology assays, an expression system is obligated to produce proteins that usually don't represent any utility for its own survival. In cases when it is needed to express many genes, as it happens with metabolic pathways, the resources saving mechanisms are ignored. We then hypothesised that, as our project involves a metabolic pathway, it could be useful to implement an artificial temporal program in order to optimize production.

1 A design that mimics nature to optimize productivity

We found out that three of the enzymes involved in the piperine synthesis are overexpressed in the *P. nigrum* mature fruit (our biobricks IDLBB_023105, IDLBB_023108 and IDLBB_023110, which can be found at the end of the Assembly document), but the other 8 enzymes are not as abundant in the same sample. These three proteins participate in the last part of the pathway, with one of them being Piperamide Synthase, the one which carries out the final step in piperamide synthesis. Since it seems that evolution itself developed a mechanism to separate expression in time for them, this was a hint for us about the importance of timing control in this particular molecular factory. Furthermore, since we were aware of one of the known mechanisms to acomplish that (SIMs) and its utility, we had a starting point to base on.

However, getting a set of 11 promoters activated by the same transcription factor but with different affinities for it sounds very difficult to accomplish. Even more, because using 11 transcriptional units would imply to raise up enormously the amount of DNA parts that would need to be inserted in the expression system, in addition to the extra effort to pull it off and the higher chances of failure.

Here we came up with the idea of creating a completely artificial temporal program by using two promoters: a constitutive one that would be on charge of the first block of proteins, and an inducible one that would produce the last proteins -those that we identified were overexpressed in the mature fruit- once their substrate demand could be sufficed.

Nevertheless, that solves just part of the problem, because the big amount of promoters, terminators and Ribosome Binding Sites needed would be exceedingly large. That could have been completely solved if we could use bacterias as an expression system thanks to the use of poli-cistronic constructs, but as they're not completely reliable to express correctly eucariotic proteins (and having 11 of them increases dangerously the probability of struggling), we turned them down in favor of yeast. In this point we came across with a very powerful tool: 2A peptides. They're peptide sequences that can be interpolated in between two genes, linking them together and allowing their co-expression by a single promoter. They can be used as a poli-cistrionic-like unit, and thus we decided to include them.[4]

It was found that the PTE2A construct is very efficient for quad-cistronic constructs in mice, since it was able to express a good amount of proteins. The PTE2A construct showed some difficulties when translating the last fragments of the transcriptional unit though[4]. However, this was actually something useful to polish the temporal program timing, as it will be discused in the Model section.

Finally, we decided that the protein expression order needed to be something similar to what we show in Figure 1, where each number inside a yellow circle represents the expression time position proposed. As the piperamide synthesis pathway involves two sub-pathways, each of them beginning with an amino acid as the first substrate (Phenilalanine and Lysine) and coming together with piperamide synthase, we labeled the enzyme names according to the sub-pathway it belongs to and the reaction position number in the sub-pathway. The timing must be adjusted to the reaction step in which the pathway is in. Combining a constutive and an inducible promoter plus the smaller adjustments provided by the positioning in the poli-cistronic constructs, we have a powerul mechanism to control the time in which every protein is going to be produced. We established the position of each gene in three constructs (Figure 2) two of them managed by a constitutive promoter and an inducible one. The positioning was made taking into account that the expression level decreases as you move downstream from the promoter.

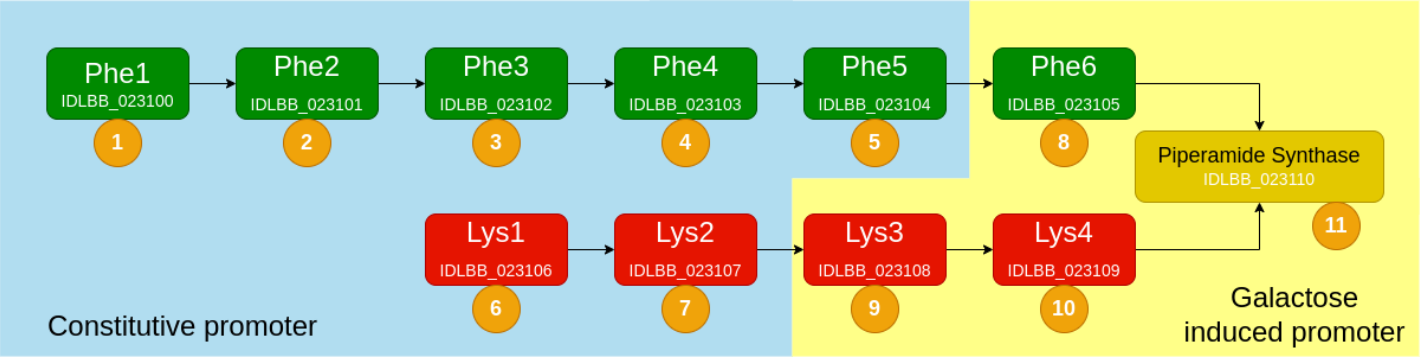

Figure 1: Temporal program proposal. Each number in a yellow circle represents the order the gene is going to be expressed. To accomplish that, we combined the use of a constitutive promoter for the proteins in the blue block and a galactose induced promoter in the yellow block. This last block contains the three proteins that we found out are overexpressed in mature black pepper fruit, except for Lys4, which was not found in the same amount. To polish even more the timing inside each block, we took advantage of the production differences that the 2A peptides imply in artificial poli-cistronic constructions (read the Model section).

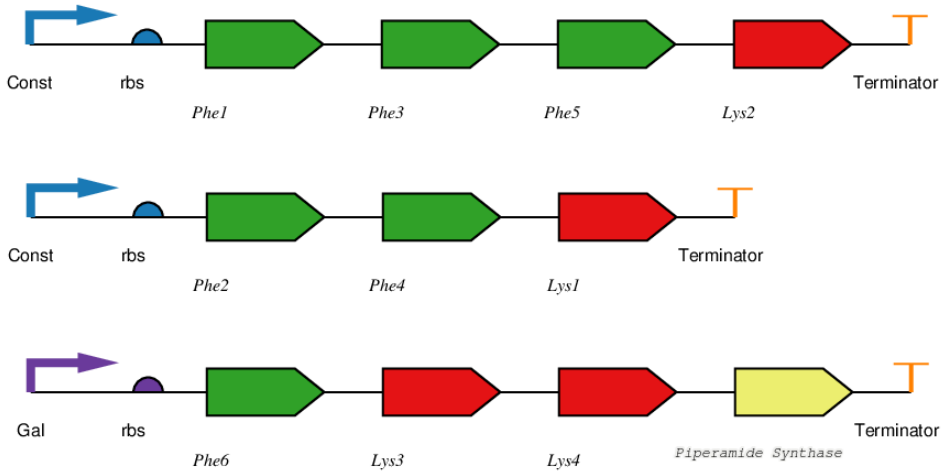

Figure 2: Final design of the artificial poli-cistronic constructs. Green genes carry out the phenilalanine subpathway, whereas red ones are implied in the lysine subpathway. Both of them come together when piperamide synthase catalyzes the final reaction to form piperamide. Between each gene exists a 2A Peptide that separates two genes during translation (not in the figure). The decision of using a constitutive and an inducible promoter was taken because the inducible one can be expressed at any desired time, allowing to create a temporal program.

2 Model

2.1 Differential equations system

The software Simbiology was used to simulate a mathematical model of the expression system. It was developed as a mix between a transcription network and an enzyme kinetics simulation, having Michaelis-Menten and Hill type reactions to produce each compound in the pathway. At the same time, the enzyme production is simulated for both constitutive and inducible promoters, as well as the consequent amino acid consumption. The galactose uptake was also included.

2.2 Enzyme production

To begin, the expression of proteins was different depending on the promoter used. This is very important, because the inducible promoter activity must increase as its activator (galactose) does, whereas the constitutive one has a constant transcription level. Thus, we proposed two similar but distinct differential equations as a general base for enzyme synthesis. First, we present the equation used for the constitutive promoter.

$$\frac{dY}{dt} = \beta P_n H_1 H_2 - \alpha Y \tag{1}$$

Where Y is the enzyme concentration, whose level depends on time (t). Here, β represents the promoter strength; it is the theoretical maximum protein production that it can achieve in molarity per second, taking into account transcription and translation. The mean time taken in yeast to translate a

protein is about 1 minute as well as the mean time for transcription, adding up to a rate of one protein made every two minutes.[1] We used this value ($0.0083M/second$) for β.

P_n is related to the fact that 2A Peptides have shown to change each individual gene translation level according to its position in the transcriptional unity. Results indicate that the first gene is always the most translated, the second one is produced about 80% compared to the first one, the third one is near 50% and the last one is around 45%. These data are based on the PTE2A quad-cistronic construct, which is the one we propose for our project.[4] Therefore we use the subindex n in P_n to indicate the gene position in the transcriptional unity, and the constants' values are $P_1 = 1, P_2 = 0.8, P_3 = 0.5$ and $P_4 = 0.45$.

H_1 and H_2 are increasing functions whose ranges span between 0 and 1. They represent the effect of amino acid availability on the cell as a limiting factor for protein production, considering Phenilalanine (H_1, equation 2) and Lysine (H_2, equation 3) as the most important ones, since they're highly demanded by this metabolic pathway. We used the Hill function for H_1 and H_2, which reaches half of the maximum speed when the amino acid concentration ($[Phe]$) and $[Lys]$) equals a constant (K_{Phe} and K_{Lys}) with concentration units in an analogous way to the Michaelis constant K_m for enzyme kinetics.

$$H_1 = \left(\frac{[Phe]}{K_{Phe} + [Phe]}\right) \tag{2}$$

$$H_2 = \left(\frac{[Lys]}{K_{Lys} + [Lys]}\right) \tag{3}$$

Finally, α is the protein removal rate. It tells the amount of protein that is degraded or diluted in units of $1/time$. Hence, αY can be interpreted as the probability of a given amount of protein to be lost at a given time. The final equation is shown in (4). It is important to notice that the maximum production rate β will be reached only if the amino acids concentrations are high enough and the position in the poli-cistronic construct is the first one. This last point is very relevant for the design of the model, since the gene position has a strong effect on the rate its product is accumulated. Hence, the first gene will be accumulated faster than the next, the second one faster than the third one, and so on. This means that the time when each enzyme reaches its activity threshold will strongly depend on its position in the construct.

$$\frac{dY}{dt} = \beta P_n \left(\frac{[Phe]}{K_{Phe} + [Phe]}\right)\left(\frac{[Lys]}{K_{Lys} + [Lys]}\right) - \alpha Y \tag{4}$$

On the other hand, the inducible promoter used a very similar equation (Equation 5), but with the difference that it uses another Hill function that depends on the galactose concentration. In this way, as the galactose level increases, also the promoter activity does.

$$\frac{dY}{dt} = \beta P_n \left(\frac{[Phe]}{K_{Phe} + [Phe]}\right)\left(\frac{[Lys]}{K_{Lys} + [Lys]}\right)\left(\frac{[Gal]}{K_{Gal} + [Gal]}\right) - \alpha Y \tag{5}$$

The values of K_{Phe} and K_{Lys} were obtained from their basal intracellular concentrations found in *E. coli* during exponential growth phase[5]. Even though the values must differ to the real ones found in yeast, they have been more studied and work as an acceptable base for this model. Similarly, the K_{Gal}

constant was obtained from the same source, but using the amount given for hexoses-P, because one of the main sources of carbon for yeast in this project will be galactose and it is expected to reach a similar concentration.

2.3 Protein amino acid usage

To model the amino acid usage by protein production, specifically Lysine and Phenilalanine since they're the most used ones in piperamide synthesis, we used Equation 6.

$$\frac{da}{dt} = -a\beta\left(K_{constitutive} + K_{inducible}\left(\frac{[Gal]}{K_{Gal} + [Gal]}\right)\right) \tag{6}$$

Where a represents the amino acid concentration and β is the promoter strength (we assume it is the same for both promoters). $K_{constitutive}$ and $K_{inducible}$ are constants related to the amount of the amino acid that the proteins need depending on its amount in the sum of all the primary structures. These quantities are given in 1/M, and can be interpreted as the concentration in molarity of the given amino acid required to produce 1M of total protein per promoter; this is, how many amino acids are required to fully translate a transcriptional unit (if it is the inducible one) or two (the sum of the two constitutive units). Finally, the Hill function is used again to indicate that the amino acid usage of the inducible promoter proteins increases as the amount of galactose in the cell does.

2.4 Enzyme kinetics

To model the enzyme activity we used Hill kinetics, which is based on Michaelis-Menten kinetics but taking into account the effect of cooperative polymeric enzymes by adding a Hill coefficient (Equation 7).

$$\frac{dP}{dt} = K_{cat}[E]\left(\frac{[S]^n}{K_m^n + [S]^n}\right) \tag{7}$$

In this equation, the change of product concentration (P) over time (t) increases as the enzyme concentration [E] does. The catalytic constant K_{cat} is the amount of product synthesised per second. Equation (7) is the Michaelis-Menten equation when the used enzyme works as a monomer, when the Hill coefficient $n = 1$. It can be seen that the velocity reaches half of the maximum speed once the substrate concentration [S] equals the Michaelis constant (K_m).

2.5 Galactose uptake

The galactose consumption was modeled again by Michaelis-Menten kinetics. The V_{max} and K_m were proposed based on experimental results that show that a whole yeast population can take all of the galactose in a culture medium in about 6 hours, so we established the values of 0.0047 M/s and 0.004 M, respectively, that fit this result[9]. As this model assumes that galactose is unlimited and it works just as a transcription inductor, an exact accuracy about the galactose uptake kinetics is not a major problem.

2.6 Enzyme constants

In order to increase accuracy and predictability in the simulations, we made an extensive research on databases and papers that reported values for the constants K_{cat} and K_m for each protein. However, since this pathway has not been intensively studied, experimental data for every single enzyme is not available. Also, considering that some of the proteins are a proposal based on genome annotation, there is even less information for them. Thus, most of the data were retrieved from distinct enzymes with the same function or similar enzymes with different function (Table 1). Still we had to propose arbitrary constants for IDLBB_023104.

Also, it is important to point out that some of the consulted papers didn't provide directly the information, so we had to calculate the variables with the supplied data. Even so, some of the information found didn't have the correct units; for example, the paper we used for IDLBB_023101 gave a V_{max} in units of $1/time$ instead of $concentration/time$. Therefore, in such cases we opted to use V_{max} as if it was K_{cat} since it does use $1/time$ as unit.

3 Results and discussion

3.1 Amino acid consumption

Results showed a much bigger demand of Phenilalanine than on Lysine (Figure 3). This must be due to the big amount of enzymes involved in the phenilalanine sub-pathway, and specially because of the high K_{cat} and low K_m of the first enzyme. Also, as t This means that the limiting reactant of the last reaction is the last one produced by the phenilalanine sub-pathway, piperic acid.

Table 1. Enzyme constants

Enzyme	**Km** (μM)	**Kcat** ($1/s$)	**Phe amount**	**Lys amount**	**Reference**
IDLBB_023100	698.63	28.533	27	68	Different enzyme, same function [10]
IDLBB_023101	0.61	0.2033	27	62	Different enzyme, same function [11]
IDLBB_023102	143.03	0.0347	24	42	Exact enzyme [2]
IDLBB_023103	47	0.013	12	39	Different enzyme, same function [10]
IDLBB_023104	4300	30	23	50	Arbitrary data
IDLBB_023105	43.3	196	25	61	Similar enzyme, different function [8]
IDLBB_023106	640	0.216	34	64	Exact enzyme [12]
IDLBB_023107	2165.5	159	31	60	Different enzyme, same function [10]
IDLBB_023108	25	0.346666	17	53	Similar enzyme, different function [3]
IDLBB_023109	15	0.1329	14	34	Different enzyme, same function [6]
IDLBB_023110	342 (piperoil CoA) 7600 (piperidine)	1.01	23	40	Exact enzyme [7]

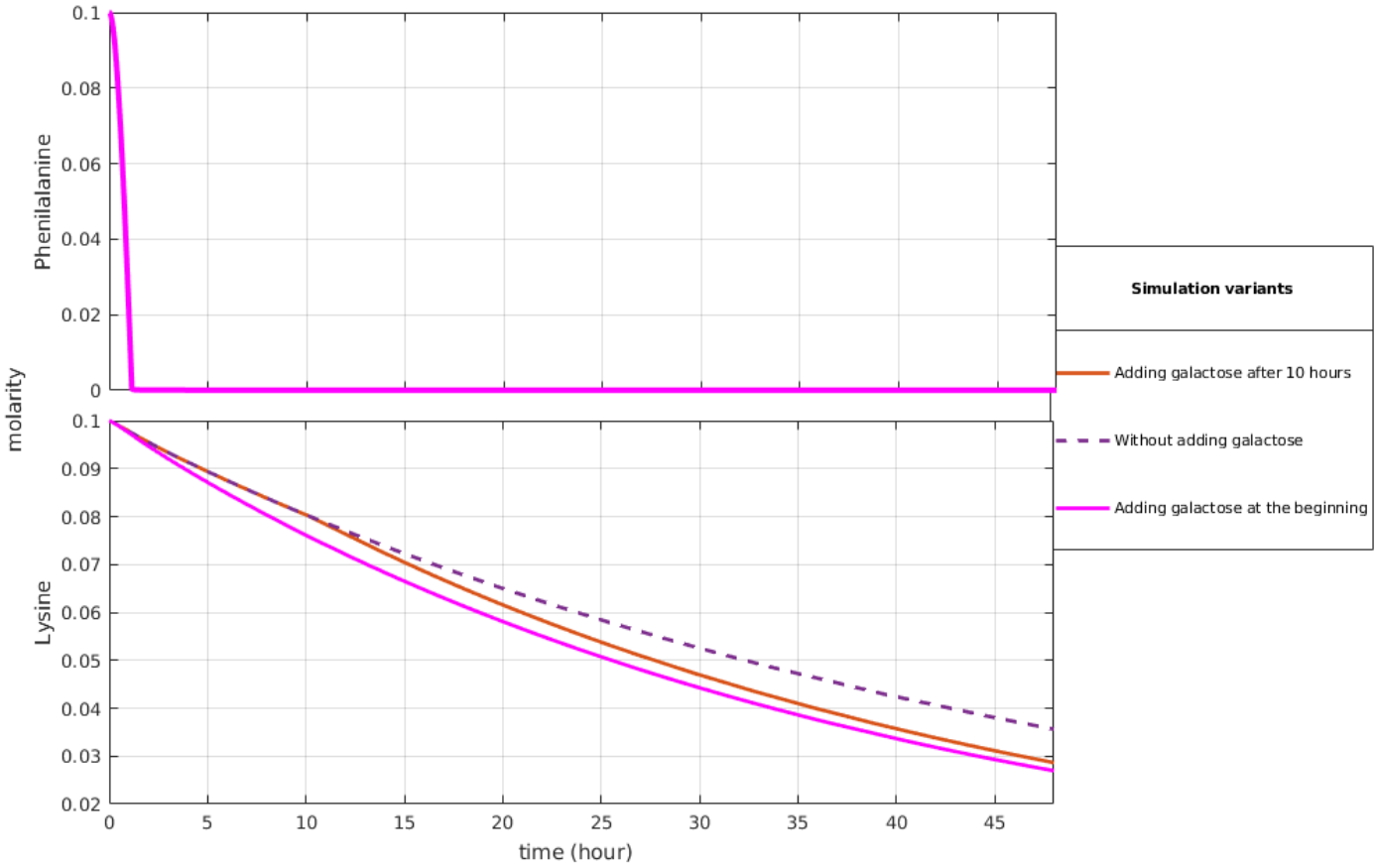

Figure 3: Simulations of the influence of inducting the GAL promoter on amino acid consumption at different times.

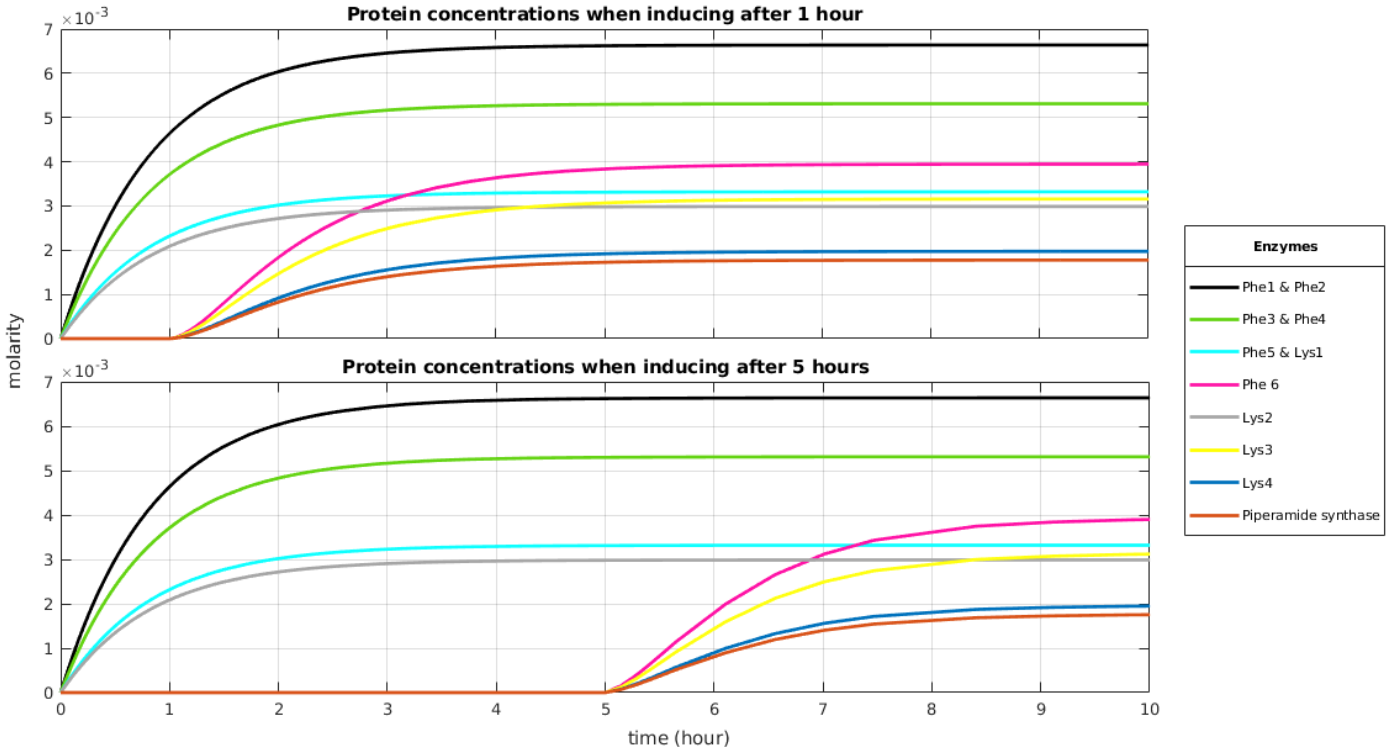

Figure 4: Simulations of the influence of inducting the GAL promoter on amino acid consumption at different times.

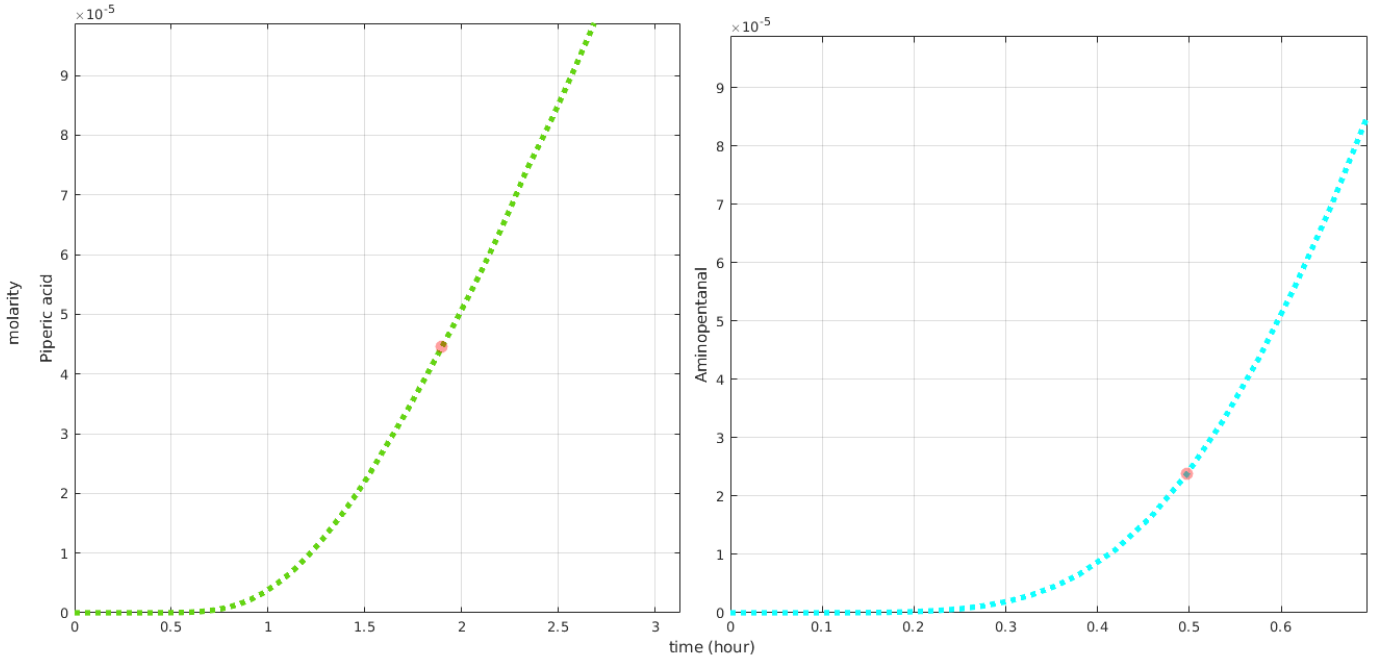

Figure 5: Moments when the last substrates produced by the constitutive promoters reach their respective K_m. Piperic acid reaches its K_m $(4.33x10^{-5}M)$ after 1 hour and 53 minutes, whereas aminopentanal reaches its K_m $(2.5x10^{-5}M)$ after half an hour.

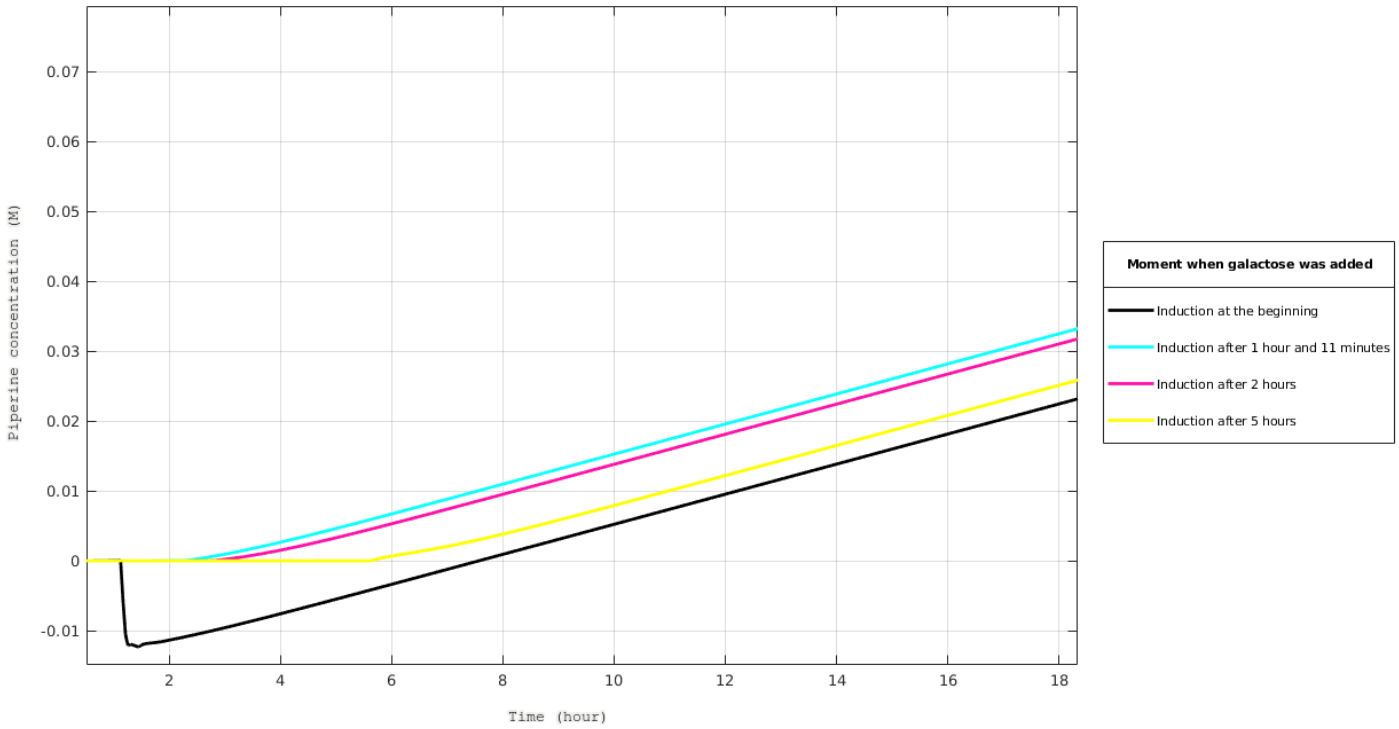

Figure 6: Piperine Production

References

[1] Alon, U. (2020). *An introduction to systems biology: design principles of biological circuits*. 2nd Edition. Chapman and Hall/CRC.

[2] Heo, K. T., Kang, S.-Y., Jang, J.-H., and Hong, Y.-S. (2017). Sam5, a coumarate 3-hydroxylase from saccharothrix espanaensis: New insight into the piceatannol production as a resveratrol 3'-hydroxylase. *ChemistrySelect*, 2(28):8785–8789.

[3] Koo, H. J., Vickery, C. R., Xu, Y., Louie, G. V., O'maille, P. E., Bowman, M., Nartey, C. M., Burkart, M. D., and Noel, J. P. (2016). Biosynthetic potential of sesquiterpene synthases: product profiles of egyptian henbane premnaspirodiene synthase and related mutants. *The Journal of Antibiotics*, 69(7):524–533.

[4] Liu, Z., Chen, O., Wall, J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., and Liu, J. (2017). Systematic comparison of 2a peptides for cloning multi-genes in a polycistronic vector. *Scientific reports*, 7(1):1–9.

[5] Milo, R. and Phillips, R. (2015). *Cell biology by the numbers*. Garland Science.

[6] Muramatsu, H., Mihara, H., Kakutani, R., Yasuda, M., Ueda, M., Kurihara, T., and Esaki, N. (2005). The putative malate/lactate dehydrogenase from pseudomonas putida is an nadph-dependent δ1-piperideine-2-carboxylate/δ1-pyrroline-2-carboxylate reductase involved in the catabolism of d-lysine and d-proline. *Journal of Biological Chemistry*, 280(7):5329–5335.

[7] Schnabel, A., Athmer, B., Manke, K., Schumacher, F., Cotinguiba, F., and Vogt, T. (2021). Identification and characterization of piperine synthase from black pepper, piper nigrum l. *Communications biology*, 4(1):1–10.

[8] Schnabel, A., Cotinguiba, F., Athmer, B., Yang, C., Westermann, B., Schaks, A., Porzel, A., Brandt, W., Schumacher, F., and Vogt, T. (2020). A piperic acid coa ligase produces a putative precursor of piperine, the pungent principle from black pepper fruits. *The Plant Journal*, 102(3):569–581.

[9] Shafaghat, H., Najafpour, G. D., Rezaei, S. P., and Sharifzadeh, M. (2010). Optimal growth of saccharomyces cerevisiae (ptcc 24860) on pretreated molasses for ethanol production: Application of response surface methodology. *Chemical Industry and Chemical Engineering Quarterly*, 16(2):199–206.

[10] Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., et al. (2012). Sabio-rk—database for biochemical reaction kinetics. *Nucleic acids research*, 40(D1):D790–D796.

[11] Zhang, B., Lewis, K. M., Abril, A., Davydov, D. R., Vermerris, W., Sattler, S. E., and Kang, C. (2020). Structure and function of the cytochrome p450 monooxygenase cinnamate 4-hydroxylase from sorghum bicolor. *Plant physiology*, 183(3):957–973.

[12] Zhou, N., Zhang, A., Wei, G., Yang, S., Xu, S., Chen, K., and Ouyang, P. (2020). Cadaverine production from l-lysine with chitin-binding protein-mediated lysine decarboxylase immobilization. *Frontiers in bioengineering and biotechnology*, 8:103.

Contents

List of Figures

List of Tables