

Gene variant design space for biosensor-based pathway dynamic regulation

María Camarena^a, Yadira Boada^{a,b}, Jesús Picó^a, Pablo Carbonell^a

^a Synthetic Biology and Biosystems Control Lab, I.U. de Automática e Informática Industrial (ai2).
Universitat Politècnica de València, Camí de Vera S/N, 46022, Valencia, Spain.

^b Centro Universitario EDEM, Escuela de Empresarios, Muelle de la Aduana s/n, 46024, Valencia, Spain.

E-mails: macape7@etsii.upv.es, yaboa@upv.es, jpico@ai2.upv.es, pjcarbon@isa.upv.es

Abstract

Inside a cell, protein, production and biosensor pathways can be genetically engineered within a dynamic regulation architecture that provides robustness to cell factories. Here we investigated how the selection of gene variants and their associated expression efficiency and kinetic parameters can lead to a wide diversity of dynamic responses in terms of protein or metabolite production. Results show that there is a trade-off between gene expression efficiency and pathway performance, and it can be eventually related to the evolutionary fingerprint of each gene variant. Therefore, the organism source of gene variants is a factor that needs to be considered in the design of dynamic regulation for genetic circuits.

Keywords: dynamic regulation, biosensors, cell factories, biomanufacturing.

1 INTRODUCTION

The current changing economic and structural context promotes constant evolution in all industrial sectors. Existing industries must face these changes, and new industries are structured on the basis of the specific geopolitical scenario. Biomanufacturing aims to be an industrial model of biochemical production that adapts to market needs while respecting established institutional frameworks. Industrial processes based on biomanufacturing by means of biological processes not only provide viable biological models at mass production levels, but also offer cheaper and less risky results at a smaller-scale, as these processes are intrinsic to biological production [13]. This type of industry based on biomanufacturing boosts the bioeconomy, which acts as a nexus between economy and ecology, ensuring economic and technological development while preserving and maintaining the planet's natural conditions. The development of the bioeconomy is underway in more than 50 countries [5], and the industrial processes that promote it are expanding, enhancing the current market opportunities.

Metabolic engineering is based on the modification of organisms to engineer metabolic pathways to produce substances that would not be produced biologically in

that organism (such as drugs or substances of pharmacological and commercial interest). This is why metabolic engineering is usually linked to genetic engineering, whereby the DNA of organisms is uniquely altered to express the genes of interest in the most optimal way possible. By using genetic modification and metabolic engineering [8], production of metabolites of interest can be achieved by dynamically adjusting cellular metabolic parameters, such as, the kinetic constant or affinity rate of target metabolite [7]. However, maintaining a homogeneous and controllable production of a metabolite in a biofactory is not so simple. This is due to the time-varying and non-homogeneous conditions of these production pathways inside the cell, which are difficult to control by dynamic adjustments. In [2], a biosensor based model was proposed to measure the naringenin production as a target metabolite using bacteria as a host microorganism. The biosensor is based on kaempferol, which is another metabolite used as a proxy of the naringenin level. To model a metabolic pathway dynamics, the kinetic parameters of a single enzyme variant for each enzyme from the naringenin synthesis pathway are used.

Here we describe a new approach that takes into account the kinetic parameters of different enzyme variants that catalyses kaempferol from naringenin. Considering both enzyme parameters and phylogenetic distances between host organisms (bio-producers), we obtained the design space for tuning the biosensor dynamic.

2 OBJECTIVE OF THE STUDY

Naringenin is a flavonoid with broad pharmaceutical interest due to its antioxidant, anti-inflammatory, cardioprotective, antitumor properties, among others [10], and herein lies its pharmaceutical and commercial interest. The construction of viable genetic circuits capable of producing significant amounts of naringenin must consider several factors. In [2], the circuit itself considers the possibility of monitoring naringenin production by synthesizing a biosensor that consumes naringenin to provide a control signal. Therefore, there must be a trade-off between the amount of naringenin consumed and the kaempferol produced.

The objective of this study is to assess how the variation of three fundamental parameters on which the production of the biosensor will depend can impact naringenin production. The aim is to identify the proper conditions between kaempferol production and naringenin degradation. In this regard, the two enzymes involved in the synthesis of kaempferol from naringenin will be considered. For both F3H and FLS, enzymes of study as shown in Figure 1, the parameters to be analyzed are: K_m , k_{cat} and the F3H and FLS enzyme concentration respectively. The enzyme concentration depends on the efficiency of gene expression in the host organism. This expression efficiency can depend on the phylogenetic distance from the organism from which the gene originates to the host organism, into which the gene is to be inserted [6]. This point will be discussed in more detail in section 3.2 below.

For this purpose, we assumed a scenario where the two genes that synthesize both enzymes are varying, while all other genetic circuits remain constant. Since experimental of enzyme parameters is not complete for all known enzyme variants, in this study we intend to sample the design space defined by the parameter ranges of each enzyme and evaluate their dynamic responses. In summary, the object of this study is to assess the effect produced on a regulated system described in [2] by varying the three parameters mentioned above, in order to establish a relationship between enzymatic efficiency (independent of phylogeny) and gene expression efficiency (directly dependent on taxonomic distance).

3 MATERIALS AND METHODS

3.1 MODELING THE PATHWAY DYNAMIC REGULATION ARCHITECTURE

Figure 1 shows the naringenin pathway, the biosensor and the antithetic controller to dynamic naringenin regulation. This genetic is considered our cell factory inside the host organism. Our approach assumes that for every k -cell, the kinetics of the enzyme-catalyzed reactions involved in the naringenin metabolic pathway from L -tyrosine to naringenin can be modeled as the set of rate equations 1 obtained from mass balance equations and considering dilution due to cell growth rate μ .

$$\frac{d[S_0^k]}{dt} = V_0 - V_{S_0} - \mu[S_0] \quad (1)$$

$$\frac{d[P_i^k]}{dt} = V_{S_i} - V_{P_i} - \mu[P_i] \quad (2)$$

For each reaction, V_j are the fluxes (molecules·min⁻¹). S_i and P_i is the number of molecules of the substrates and products in the i

step of the pathway, respectively, i.e., L -tyrosine, p -coumaric acid, p -coumaroyl-CoA, naringenin chalcone, and naringenin, which is the product of interest.

We assume that the flux V_0 corresponding to the precursor L -tyrosine remains constant, and all fluxes V_j obey Michaelis-Menten kinetics [2] as follows:

$$V_j = k_{cat_j} [E_j] \frac{[S_j]}{K_{m_j} + [S_j]} \quad (3)$$

where k_{cat_j} is the catalytic rate of each enzyme (min⁻¹), and K_{m_j} the Michaelis-Menten constant for each substrate. The enzyme kinetic parameters were obtained from Brenda [11].

The amount of the enzymes corresponding to each step of the production pathway, i.e. TAL, 4CL, CHI and CHS [2], were selected for a nominal case so that the precursor flux V_0 yields 1 g L⁻¹ of naringenin as in [2].

Similarly, for the biosensor, and using the same assumptions as for the metabolic pathway, the kinetics of the enzyme-catalyzed reactions involved in the metabolic pathway from naringenin to kaempferol, i.e. FLS and F3H, were modeled as:

$$\frac{d[Di^k]}{dt} = V_N - V_{Di} - \mu[Di] \quad (4)$$

$$\frac{d[Ka^k]}{dt} = V_{Di} - \mu[Ka] \quad (5)$$

where V_j are the fluxes of each reaction (molecules·min⁻¹), Di is Dihydrokaempferol, and Ka is kaempferol, which is the effector flavonoid measured by a biosensor promoter. As for the pathway model 1, the flux V_{Di} obeys Michaelis-Menten kinetics:

$$V_N = k_{catF3H} [F3H] \frac{[N]}{K_{mN} + [N]} \quad (6)$$

$$V_{Di} = k_{catFLS} [FLS] \frac{[Di]}{K_{mDi} + [Di]} \quad (7)$$

The kinetic parameters were also obtained from the Brenda database.

In Figure 1, the biosensor uses a Qdor protein and a qdoR-PqdoI transcription factor that inhibit the production of the anti- σ molecules [12]. Kaempferol captures the QdoR transcription factor and inactivates it. Further details of the transcription/translation (TX/TL) model are given in [2].

Finally, for the controller in Figure 1, we consider a model based on ordinary differential equations for the antithetic feedback integral controller proposed in

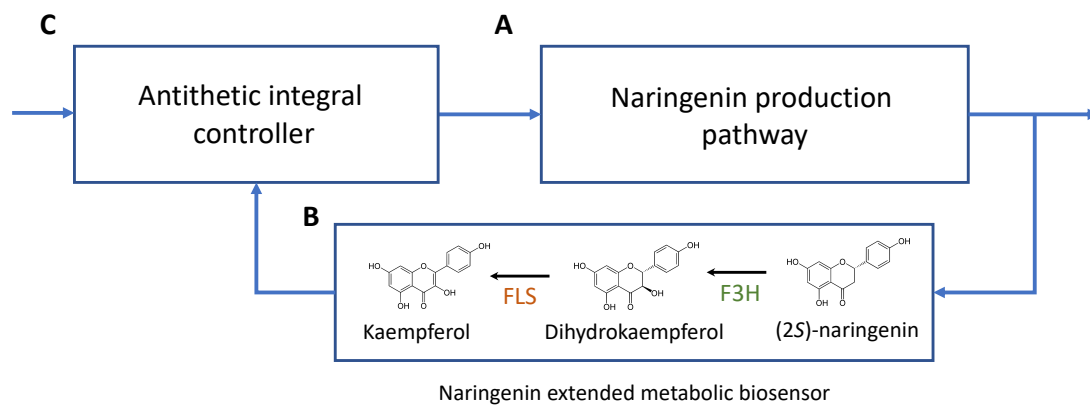


Figure 1: Schema of the dynamically regulated cell factory for naringenin production. A. The production pathway in the forward chain converts *l*-tyrosine into (2*S*)-naringenin, the target product. B. The reverse chain consists of a biosensor that contains two enzymes, FLS and F3H, that convert target naringenin into kaempferol, a molecule that can be detected by means of transcription factor QdoR. C. An antithetic integral controller as in [2] regulates the overexpression of enzymes in the production pathway in order to cope with perturbations such as variations in availability of precursors.

[1, 9]. The antithetic mechanism relies on the mutual annihilation between σ protein and its co-factor anti- σ , leading to an integral-like behavior controller.

This predictive model has been used in this study varying three main parameters which are explained in the following sections.

3.2 EFFECT OF GENE VARIANT SELECTION ON THE DYNAMIC RESPONSE

As mentioned in section 2, there is a taxonomic relation between the efficiency of the enzyme expression in the host organism and the organism source of the gen. The efficiency of gene expression is determined by many factors, including the solubility of the protein in the organism, the underlying non-toxicity of its production in the host, the maximum amount of protein produced, among other factors [6].

Two organisms having similar genomes (whose taxonomic distance is short) present similar phenotypes. However, the greater the taxonomic distance between two organisms, the more differences there will be in the expression of their genes, resulting in different cellular environments. Therefore, to achieve efficient enzyme expression it is reasonable to consider that the more genetically similar two organisms are, the better the protein is expressed and the higher concentrations can be reached. For this reason, and in order to analyze the trade-off between efficiency of expression and performance of the regulation circuit, we assumed that the closest taxonomic sources should express more enzyme concentration in the host [6]. In this study, we have considered differences among the concentration parameters as function of the enzyme variant for each enzyme involved in the target path-

way, i.e. the kaempferol metabolic pathway.

Table 1 list the kinetic parameters for both enzymes F3H and FLS, and their respective taxonomic distances between the host and the target organism from they were obtained. Kinetic parameters were collected from BRENDA and Uniprot databases. As shown, for K_{cat} values, the available information is scarce. The simulations were therefore performed using a constrained working range for both k_{cat} and K_m , as explained in the next section. It should be noted that these kinetic values do not depend on the affinity between the host organism and the gene source.

Taxonomic distances have been calculated using the enzyme selection tool Selenzyme [4], which allows the entry of both host and enzyme reaction of study. This is considered to be useful for the implementation of more structurally robust genetic circuits. Here, we have introduced the *Escherichia coli* organism entry as a host, but it should be noted that similarly other hosts and organism families could be used for future studies. Table 1 illustrates how the kinetic constants vary depending on the organism to be considered. In the study, a concentration range was also performed. The concentrations of both enzymes were varied in a range from to determine the effects produced on kaempferol production. This will be discussed in more detail in section 4.

Table 1: FLS parameters of interest

| FLS variants | | | | |
|---|--|------------------------------|-----------|-------------|
| Enzyme variant | Organism | Phylogenetic Distance | Km | Kcat |
| Trifunctional UDP-glucose 4,6-dehydratase/UDP-4 -keto-6-deoxy-D-glucose | Arabidopsis thaliana | 29 | 0,116 | 2860 |
| Flavonol synthase/ flavanone 3-hydroxylase | Arabidopsis thaliana | 29 | 0,059 | - |
| Jasmonate-induced oxygenase 2 | Arabidopsis thaliana | 29 | 0,0093 | - |
| Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase | Pyrus communis | 30 | 0,001 | - |
| Flavonol synthase/ flavanone 3-hydroxylase | Citrus unshiu | 31 | 0,045 | - |
| Leucoanthocyanidin dioxygenase | Zea mays | 29 | 0,0584 | 6,6 |
| Bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase | Malus domestica | 31 | 0,01 | - |
| F3H variants | | | | |
| Enzyme variant | Organism | Phylogenetic Distance | Km | Kcat |
| Flavanone 3-dioxygenase 3 | Oryza sativa subsp. japonica (Rice) | 30 | 0,0063 | - |
| Flavone synthase | Petroselinum crispum | 31 | 0,005 | - |
| Naringenin;2-oxoglutarate 3-dioxygenase (Fragment) | Petunia hybrida | 29 | 0,0056 | - |
| Naringenin;2-oxoglutarate 3-dioxygenase | Arabidopsis thaliana | 29 | 0,024 | - |
| flavanone 3-dioxygenase | Artemisa annua | 31 | 0,045 | - |

4 RESULTS

All computational simulations were performed using the metabolic pathway model in equations (4-7), and the model in [2] for the antithetic controller dynamics.

To implement the enzymatic analysis, the study has been approached in two ways: considering the differences between kinetic parameters and, on the other hand, considering the difference in F3H and FLS concentrations underlying the taxonomic distance between host organism and gene-provider.

In terms of phylogeny, the sequences of the enzymes F3H and FLS (coded by the biosensor genes) from organisms phylogenetically closer to the host organism (*E. coli*) are assumed to be the most efficient in terms of expression. As a matter of fact, the similarity between genes from closer organisms is greater than between taxonomically distant organisms, so they will be better candidates in terms of structural viability [3]. In other words, if the host and target organisms are phylogenetically close, the enzyme expression in the host is assumed to be greater and thus more protein concentrations will be reached. For this reason, considering more than one possible enzyme variant for the F3H and FLS enzymes is highly relevant in both naringenin degradation and kaempferol production. To determine the effect of the different variants, we worked in a range of different concentrations for each of the enzymes and, as expected, we obtained variations in both naringenin degradation and kaempferol production, being greater the higher the concentration of protein synthesized.

Regarding enzyme kinetics, multiple kinetic scenarios will be simulated taking into account different possibilities of values depending on the enzyme variant, and several data (kaempferol production) can be obtained generating a range of values for the kinetic parameters of both enzymes. Kaempferol production is used as a substrate for the biosensor explained in [2], high values of this enzyme should be avoided in order to diminish biosensor leakage of the target. Focusing the study on the variation of kaempferol production was therefore considered relevant.

For the study, kinetic variations of F3H and FLS have been taken into account in order to evaluate the amount of kaempferol produced in relation to the k_{cat}/K_m ratios of both enzymes. The kinetic parameters of both enzymes have been assumed to be equivalent, i.e. the efficiency calculated as k_{cat}/K_m is the same for both enzymes in each simulation performed.

The study data were obtained by using different values of the enzymatic parameters of F3H and FLS enzymes into the dynamic model proposed in [2]. The values of $4K_m$ have been varied from 0.001 from 1 mM. The values of k_{cat} have been varied from 0.01 to $10 s^{-1}$.

Figure 2 shows how kaempferol production depends on the biosensor enzymatic efficiency. Enzymatic efficiency which is the ratio k_{cat}/K_m . Results show a kinetic efficiency range between 0.7 and 1, where kaempferol production increases. k_{cat} (s^{-1}) must be at least in the order of 10^3 higher than K_m (mM) to obtain a relevant increase in kaempferol production, and thus biosensor. Therefore, this range is considered decisive for selecting a variant enzyme pair to show good sampling results.

In Figure 3, the variation of naringenin production (g/L) in relation to the variation of kaempferol production (g/L) is shown. It can be seen that the initial and final naringenin amounts are saturated (1.0265 g/L maximum; 0.016 g/L minimum), so there is only one range of kaempferol that actually affects naringenin production.

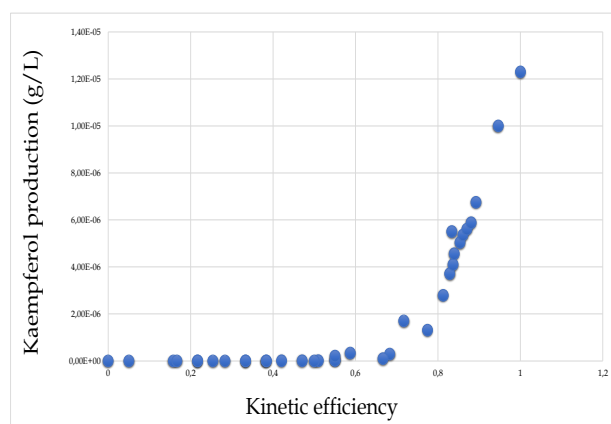


Figure 2: Kaempferol production (g/L) depending on enzymatic efficiency based on k_{cat}/k_m ratio

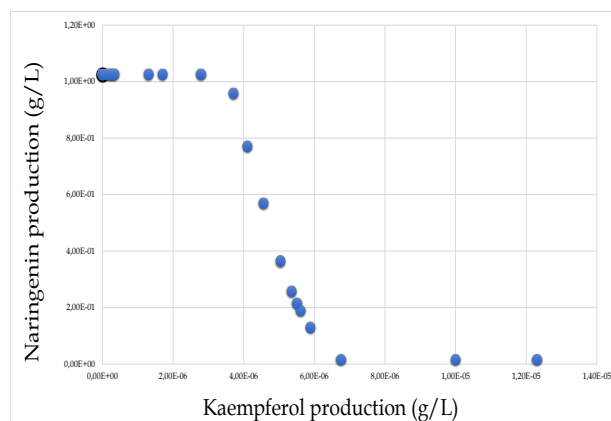


Figure 3: Naringenin production (g/L) depending on Kaempferol production (g/L)

In other words, there is a range where naringenin production is affected as a result of the controller action. The conversion of naringenin as a substrate into kaempferol must be carefully tuned and, whenever possible, to interfere as little as possible with narin-

genin production in order to generate a representative level of sensing.

As the enzymatic efficiency increases, the amount of naringenin consumed to generate kaempferol is considerably small until the amount of kaempferol reaches $4 \times 10^{-6} \text{g/L}$. However, when productions above $5 \times 10^{-6} \text{g/L}$ are reached, the amount of consumed naringenin starts to be significantly high and this can compromise naringenin production. Kaempferol production has to be appropriately tuned, so the selection of enzymes has to be based on the trade-off between enzymatic efficiency and naringenin degradation.

5 CONCLUSIONS

We improved the performance of the kaempferol-biosensor for different enzyme variants in the naringenin metabolic pathway proposed in [2], as an example of efficient dynamic pathway regulation. We have analyzed the impact of selecting different variants, with their associated parameters, in the dynamic response of the regulated circuit. Interestingly, we observed that increasing the kinetic efficiency of the biosensor led to a higher production of kaempferol, which, in turn, decreased the production of the target molecule naringenin. These results show that an appropriate selection of the biosensor enzymes is necessary in order to avoid excessive cross-talk and leakage.

However, efficiency of enzymes can often be considered as a competing objective with expression efficiency, as enzyme expression efficiency depends directly on organisms genetic similarity and consequently on their taxonomic distance. Such distant sequences are generally less efficient in terms of expression in the host. Therefore, a trade-off needs to be achieved by means of a design tool. To that end, Selenzyme, the enzyme selection tool [4], has been used to obtain the phylogenetic distances of the target organisms to *E. Coli*, in order to qualify the gene expression efficiency. This can ensure the highest possible similarity between the genetic structure of the host and the target is of crucial relevance to ensure the expression efficiency of the introduced gene.

As a future work and using new available data, this gene-variant analysis could be improved by including other gene indicators like the genetic sustainability, enzyme efficiency ratio, among other parameters. The potential for automation of enzyme selection offers great prospects for the future in the chemical and biochemical industry. Without automated processes, enzyme selection is costly and inefficient due to the testing requirements needed to select the correct enzymes. Through automation, only the enzymes of greatest interest are selected as the object of study.

Acknowledgement

MC was supported by an AI2-UPV studentship. PC acknowledges funding from UPV Talento Programme and European Union's Horizon 2020 research and innovation programme under grant agreement number 814408 (ShikiFactory100). This research was partially supported by grants MINECO/AEI, EU DPI2017-82896-C2-1-R and MICINN/AEI, EU PID2020-117271RB-C21.

References

- [1] Stephanie K. Aoki, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570(7762):533–537, jun 2019.
- [2] Yadira Boada, Alejandro Vignoni, Jesús Picó, and Pablo Carbonell. Extended metabolic biosensor design for dynamic pathway regulation of cell factories. *iScience*, 23(7):101305, 2020.
- [3] Pablo Carbonell, Guillaume Lecointre, and Jean-Loup Faulon. Origins of specificity and promiscuity in metabolic networks. *Journal of Biological Chemistry*, 286(51):43994–44004, 2011.
- [4] Pablo Carbonell, Jerry Wong, Neil Swainston, Eriko Takano, Nicholas J Turner, Nigel S Scrutton, Douglas B Kell, Rainer Breitling, and Jean-Loup Faulon. Selenzyme: Enzyme selection tool for pathway design. *Bioinformatics*, 34(12):2153–2154, 2018.
- [5] James M Clomburg, Anna M Crumbley, and Ramon Gonzalez. Industrial biomanufacturing: the future of chemical production. *Science*, 355(6320), 2017.
- [6] Gerhard Hannig and Savvas C Makrides. Strategies for optimizing heterologous protein expression in escherichia coli. *Trends in biotechnology*, 16(2):54–60, 1998.
- [7] Cynthia Ni, Christina V. Dinh, and Kristala L.J. Prather. Dynamic Control of Metabolism. *Annual Review of Chemical and Biomolecular Engineering*, 12(1):annurev-chembioeng-091720-125738, jun 2021.
- [8] Jens Nielsen. Metabolic engineering. *Applied microbiology and biotechnology*, 55(3):263–283, 2001.
- [9] Noah Olsman, Ania-Ariadna Baetica, Fangzhou Xiao, Yoke Peng Leong, Richard M. Murray, and John C. Doyle. Hard Limits and Performance

- Tradeoffs in a Class of Antithetic Integral Feedback Networks. *Cell Systems*, 9(1):49–63.e16, jul 2019.
- [10] Bahare Salehi, Patrick Valere Tsouh Fokou, Mehdi Sharifi-Rad, Paolo Zucca, Raffaele Pezzani, Natália Martins, and Javad Sharifi-Rad. The therapeutic potential of naringenin: a review of clinical trials. *Pharmaceuticals*, 12(1):11, 2019.
- [11] I. Schomburg, L. Jeske, M. Ulbrich, S. Placzek, A. Chang, and D. Schomburg. The BRENDA enzyme information system—From a database to an expert system. *Journal of Biotechnology*, 261:194–206, nov 2017.
- [12] Solvej Siedler, Steen G Stahlhut, Sailesh Malla, Jérôme Maury, and Ana Rute Neves. Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into escherichia coli. *Metabolic engineering*, 21:2–8, 2014.
- [13] Roland Wohlgemuth, Tomasz Twardowski, and Alfredo Aguilar. Bioeconomy moving forward step by step—a global journey. *New Biotechnology*, 61:22–28, 2021.



© 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).