12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Towards Constraint-Based Burden-Aware Models for Metabolic Engineering

M. Maton * **F.N. Santos-Navarro** * **J. Picó** *** **Ph. Bogaerts** **
A. Vande Wouwer *

* *Systems, Estimation, Control and Optimization (SECO) Laboratory, University of Mons, Mons 7000, Belgium.*
** *3BIO-BioControl, Université Libre de Bruxelles. Av. F.D. Roosevelt 50 C.P. 165/61, Brussels 1050, Belgium.*
*** *Synthetic Biology and Biosystems Control Lab, I.U. de Automática e Informática Industrial (ai2), Universitat Politècnica de Valencia. 46022, Camino de Vera S/N, Valencia, Spain.*

Abstract: Over the years, hundreds of applications have proved the effectiveness of constraint-based methods to validate the definition of metabolic networks, determine the robustness of metabolic models, and analyze the flow of metabolites through a network. However, stoichiometric models do not include information on flux capacity via enzymatic activity. Methods combining biological data from genome to metabolome have been developed to obtain improved flux predictions and constrain the range of possible flux distributions. Yet, these models still lack relevant information to design de novo metabolic pathways. Expressing the exogenous enzymes induces a cell burden due to competition for cell resources between the exogenous genes and the endogenous host ones, compromising the performance of the designed pathway. Thus, optimal selection of the expression strength of the pathway enzymes is still a challenge. Host-aware models have been developed to tackle cell burden in the context of designing increasingly complex synthetic genetic circuits in synthetic biology. This paper suggests a method to integrate host-aware gene expression models with constraint-based modeling to maximize the flux through an exogenous pathway by optimizing promoter and ribosome binding site strengths, crucial parameters that define the required transcription and translation strengths of the pathway enzymes' genes. This study considers the formation of p-coumaric acid, shows promising results, and paves the way for further investigations.

Keywords: Metabolic flux analysis, burden-aware modeling, synthetic-gene circuit, multi-scale modeling, enzymatic activity, synthetic biology.

1. INTRODUCTION

Microbial cell factory development through metabolic engineering seeks to obtain high levels of products of interest through genetic modification of microorganisms. Although microbial cell factories are a promising solution to achieve this objective, traditional approaches in metabolic engineering based on extensive combinatorial trial-and-error experiments are expensive, time-consuming, and undirected. The design of complex biological systems with novel functions, done rationally and systematically, aims at eliminating this bottleneck by introducing principles and tools elucidated by engineering, including mathematical modeling and computational optimization.

With the advances in sequencing techniques and genome annotation methods, several mathematical modeling approaches have been developed to analyze the integrated behavior of microbial cells on a genome scale (Varner, 2000; Nielsen and Olsson, 2002). In such integrative analysis, biological data from genome to metabolome are combined and the prediction of cellular responses to different kinds of genetic and environmental perturbations are evaluated. At the metabolic level, stoichiometric modeling, also known as constraint-based modeling, is now routinely applied. This class of methods relies on mass balances over intracellular metabolites and the assumptions of pseudo-steady-state conditions to estimate intracellular metabolic fluxes. Among the realm of possible methods, flux balance analysis (FBA) and flux variability analysis (FVA) are very popular approaches. While the first method reduces the solution space considering that the cell behaves to satisfy a biological objective (Orth et al., 2010b), flux variability analysis determines the range of possible reaction fluxes that still satisfies the original FBA problem within some optimality factor (Gudmundsson and Thiele, 2010). Constraint-based analyses are implemented in well-established software such as the COBRA Toolbox (Schellenberger et al., 2011).

However, stoichiometric models do not include information on metabolic regulation such as gene expression via translation or enzyme activity. Henceforth, the integration of transcriptome data into stoichiometric metabolic models has been considered to obtain improved flux predictions and constrain the range of possible flux distributions. In this context, Covert et al. (2001) extend the stoichiometric modeling framework with a transcriptional regulatory

network using a Boolean logic formalism and Covert and Palsson (2002) use extreme pathways analysis and reduce the solution space by removing a pathway when it becomes inconsistent with the imposed regulatory constraints. Later, Akesson et al. (2004) extend the approach by constraining a metabolic flux to zero in the absence of any essential component (sub-unit, assembly factor or translational activator). Yet, many regulatory phenomena cannot be accurately described by Boolean logic, and the previous methods are limited by the available knowledge of regulatory processes.

To tackle this issue, dynamic methods have been developed to account for feedback of FBA on the regulatory network by considering metabolite concentrations. Covert et al. (2008) integrate a set of ODEs to their previous approach (at the expense of kinetic parameters to identify) to predict the phenotype of diauxic growth of E. coli. Chandrasekaran and Price (2010) propose a data-driven method to limit a reaction to a percentage of its maximal flux value, corresponding to the probability of activation of the gene related to the reaction. Marmiesse et al. (2015) develop the first metabolic flux analysis tool that integrates regulatory networks directly to constrain the FBA. Nevertheless, regulatory networks are supposed to be known and must be provided by the user. To guarantee the optimality of the predicted phenotype, Machado et al. (2016) suggest a model transformation resulting to a stoichiometric representation of gene-protein-reaction associations that can be directly integrated into the stoichiometric matrix. Then, a two-step variant of FBA determines the flux distribution that satisfies an optimal objective while also ensuring the minimization of enzyme usage.

In another vein, methods enabling the prediction of optimal enzyme levels and reaction fluxes under changing environmental conditions have emerged. Waldherr et al. (2015) introduce the dynamic enzyme-cost flux balance analysis (deFBA) to predict all reaction rates and enzyme levels dynamically over time for well-defined deterministic settings in which dynamics of the environment are precisely known. Reimers et al. (2017) solve genome-scale deFBA models on large time horizons (computationally costly) and thereafter, Lindhorst et al. (2018a) utilize the concept of receding prediction horizon and develop the short-term deFBA (sdeFBA). Finally, to handle uncertainty in nutrient availability, Lindhorst et al. (2018b) introduce the robust deFBA (rdeFBA) combining deFBA with multi-stage MPC to predict robust optimal gene expression levels for rapidly changing environments.

The previous methods deal with enzyme capacities as parameters to optimize with no consequences for the host cell but limiting the metabolic fluxes. However, the enzymes catalyzing engineered metabolic pathways consist of exogenous proteins [1] that must be expressed. Indeed, their expression, on the one hand, consumes substrates. This is accounted for in constraint-based models. On the other hand, the expression of exogenous genes competes for cell expression resources (e.g. ribosomes) with the endogenous ones, inducing a cell burden. This aspect is not accounted for in constraint-based models. Yet, this competition for cell expression resources affects the behavior of both the host cell and the exogenous genes, compromising the performance of the designed pathway and thus the production rate of the product of interest. As a result, optimal selection of the expression strength of the pathway enzymes is still challenging.

In the past, a large amount of studies has been conducted for the design of gene circuits with the desired behavior (Boada et al., 2016) and to model and optimize the regulation of engineered metabolic pathways (Wehrs et al., 2019; Boada et al., 2022). Also recently, host-aware models, also called burden-aware models, have been developed to tackle cell burden in the context of designing increasingly complex synthetic genetic circuits in synthetic biology (Liao et al., 2017).

Nevertheless, to the best of the authors' knowledge, none of the above-mentioned methods consider the problem of conciliation of constraint-based metabolic and host-aware gene expression models when exogenous enzymes are expressed so as to optimize the flux through a de novo metabolic pathway. The contribution of this study is therefore to address this problem and to suggest a method providing the following benefits: (i) constraining the solution space of the flux distribution taking into account the enzyme activity on the one hand and on the other hand, the cell burden caused by the expression of exogenous enzymes, (ii) having a better prediction of the production rate of the metabolite of interest and (iii) assessing/tuning essential enzymes' gene expression parameters: transcription and translation strengths. This approach combines several preexisting tools, making the methodology easy to understand and apply by non-expert users.

This paper is organized as follows. The next section reviews the main concepts relative to metabolic network analysis and constraint-based methods and introduces the burden-aware model considered in the context of this study. Section 3 presents the methodology to connect stoichiometric models and burden-aware models of gene expression, particularly when exogenous proteins are expressed. An application of the method is described in section 4 and finally, conclusions are drawn in section 5.

2. MODELING FRAMEWORKS

This section is devoted to the main concepts relative to metabolic network analysis and introduces the burden-aware model, developed in (Santos-Navarro et al., 2021).

2.1 Constraint-Based Models

Cellular metabolism is defined as a set of chemical reactions, possibly catalyzed by enzymes, taking place within the cell and forming metabolic pathways. These intracellular reactions may be translated into a matrix representation defining a metabolic network as a $m \times n$ stoichiometric matrix N where m represents the number of internal metabolites and n stands for the number of reactions. Assuming the pseudo-steady state, the following system of linear equations is obtained:

$$N\underline{v} = 0 \tag{1}$$

where $\underline{v} \in \mathcal{R}^n$ gathers the metabolic fluxes of the network. Moreover, network fluxes are often constrained to positivity assuming that the direct reactions prevail over their reverse counterparts:

$$\underline{v} \geq 0 \tag{2}$$

[1] Exogenous proteins do not contribute to the specific growth rate of the cell but to its mass only - unlike endogenous proteins.

However, in any realistic large-scale metabolic model, there are more reactions than compounds, leading to under-determined systems for which no unique solution exists. Constraint-based methods may alleviate this issue. For instance, FBA is a method for identifying an optimal flux distribution $\underline{v}$ which maximizes or minimizes an objective function $Z = \underline{c}^T \underline{v}$ such that:

$$\underline{v}^{opt} = \max_{\underline{v}} Z \ s.t. \ \{N\underline{v} = 0 \ ; \ \underline{v} \geq 0 \ ; \ \underline{v}_l \leq \underline{v} \leq \underline{v}_u\} \quad (3)$$

where $\underline{c}$ is a vector of weights translating how much each reaction contributes to the objective function and $\underline{v}_l$ and $\underline{v}_u$ are vectors of lower and upper bounds. Another approach is provided by FVA, which computes the maximal and minimal values of the reaction fluxes. This means solving two optimization problems for each flux:

$$\begin{cases} \underline{v}_i^{min} = \min_{\underline{v}} \underline{v}_i \ s.t. \ \{N\underline{v} = 0 \ ; \ \underline{v} \geq 0 \ ; \ \underline{v}_l \leq \underline{v} \leq \underline{v}_u\}, \\ \underline{v}_i^{max} = \max_{\underline{v}} \underline{v}_i \ s.t. \ \{N\underline{v} = 0 \ ; \ \underline{v} \geq 0 \ ; \ \underline{v}_l \leq \underline{v} \leq \underline{v}_u\} \end{cases} \quad (4)$$

2.2 Burden-Aware Models

Models of gene expression accounting for host-circuit interactions are particularly interesting for understanding how the cell growth and the expression of cell endogenous genes evolve when exogenous ones are introduced. The model proposed in (Santos-Navarro et al., 2021) accounts for the dynamics of the expression of the cell host endogenous protein-coding genes (contributing to cell mass and cell growth) on the one hand and the expression of protein-coding exogenous genes (contributing to cell mass only) on the other hand. The gene expression dynamics for a generic k-th protein-coding gene in prokaryote cells, under the assumption that transcription is faster than translation and that ribosomes are the limiting shared resource required for protein expression, is given by:

$$\dot{p}_k = \frac{\nu_t(s)}{l_{pk}} J_k(\mu, r) r - (d_k + \mu) p_k \quad (5)$$

where p_k is the number of copies of the k-th protein, l_{pk} represents the protein length, d_k is the protein degradation rate, μ stands for the specific growth rate, r is the number of free ribosomes and $\nu_t(s)$ symbolizes the substrate-dependent effective peptide elongation rate. As a first approximation, the latter is expressed using a Michaelis-Menten expression which is considered as organism- and substrate-dependent but does not depend on the nucleotide sequence. The key functional coefficient $J_k(\mu, r)$ is defined as the resource recruitment strength and allows explaining the distribution of resources (free ribosomes) between the host cell and the genes of interest. It takes into account lab-accessible parameters such as promoter and ribosome binding site (RBS) strengths characterizing the steps of genetic transcription and translation, respectively. This term is defined as follows:

$$J_k(\mu, r) \triangleq E_{mk}(l_{pk}, l_e) \ \omega_k \ (\frac{d_{mk}}{K_C^k(s)} + \mu r)^{-1} \quad (6)$$

where E_{mk} represents the ribosome density-related term and can be approximated to $0.62 \ l_{pk} \ (l_e)^{-1}$ with $(l_e)^{-1}$ being the specific ribosome density, ω_k represents the mRNA production rate (i.e. the promoter strength), d_{mk} is the mRNA degradation rate and $K_C^k(s)$ symbolizes the RBS strength and depends on the availability of the substrate. That latter parameter is expressed as follows:

$$K_C^k(s) \triangleq \frac{K_b^k}{K_u^k + K_e(s)} \quad (7)$$

where K_b^k and K_u^k represent respectively the association and dissociation rate between a free ribosome and the RBS and $K_e(s)$ is the translation initiation rate.
Furthermore, cell growth might be defined as the variation of the protein fraction of the total cell mass, knowing not all protein mass contributes to cell growth (e.g. exogenous ones). Under the assumption that the protein mass for each cell rapidly reaches the steady state, the cell-specific growth rate is obtained:

$$\mu(s) = \frac{m_{aa}}{m_h} \ \nu_t(s) \ \Phi_t^h \ \Phi_m \ r_T \quad (8)$$

where m_{aa} is the average amino-acid mass, m_h is the mass of the native host cell, $\Phi_m r_T$ represents the number of available mature ribosomes and Φ_t^h can be defined as a modulation function which is an image of the cell burden and represents the fraction of available mature ribosomes used to express endogenous genes (those contributing to cell growth). Henceforth, this modulation function can be expressed as follows:

$$\Phi_t^h = \frac{N_r \ J_r(\mu, r) + N_{nr} \ J_{nr}(\mu, r)}{1 + \Psi(\mu, r)} \quad (9)$$

with the function $\Psi(\mu, r)$ defined as:

$$\Psi(\mu, r) = N_r J_r \left(1 + \frac{1}{E_{mr}}\right) + N_{nr} J_{nr} \left(1 + \frac{1}{E_{mnr}}\right) + \sum_{k=1}^{N_{exo}} J_k \left(1 + \frac{1}{E_{mk}}\right) \quad (10)$$

where N represents the number of protein-coding genes. The indexes $_r$ and $_{nr}$ stand for ribosomal and non-ribosomal protein-coding endogenous genes and N_{exo} allows the existence of exogenous genes. More information and details can be found in (Santos-Navarro et al., 2021).

3. CONCILIATING CONSTRAINT-BASED AND BURDEN-AWARE MODELS

The objective of this study consists of finding a simple way to connect the two modeling frameworks described in the previous section, i.e., constraint-based models (metabolic scale) and models of gene expression (genetic level) when exogenous enzymes associated with engineered metabolic pathways are expressed. The burden-aware model is used to predict the enzyme concentration and the cell growth rate for a given amount of substrate and promoter/RBS strengths. In turn, the constraint-based model and FVA are exploited to predict the flux through the pathway of interest for a given growth rate, availability of limiting substrate and its capacity. The main objective of this paper is thus, given some amount of limiting substrate, to estimate the promoter and RBS strengths required to express the enzymes of the metabolic pathway under design that maximizes the flux of the output metabolite. In the following, we consider that a metabolic flux is limited by the availability of the enzyme catalyzing the corresponding reaction. Henceforth, the following relationship holds:

$$v_E^{ub} = k_{cat,E} \, [E] \quad (11)$$

where v_E^{ub} is the upper bound of the flux related to the reaction catalyzed by the enzyme E, $k_{cat,E}$ is a catalytic constant, and $[E]$ is the enzyme concentration at steady-state (s.s.).

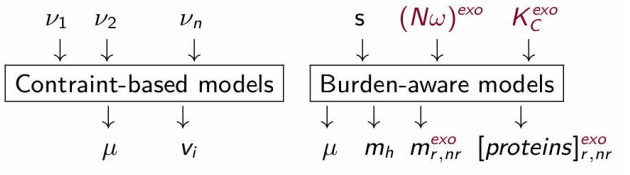

Fig. 1. Stoichiometric and burden-aware models in terms of input-output characteristics.

As illustrated in Fig. 1, it is worth noting that the stoichiometric and genetic models differ from an input-output perspective. Indeed, constraint-based models require experimental uptake and excretion rates as inputs and provide as outputs the growth rate of the cell and the value of the metabolic fluxes. The burden-aware model considers as input an abstract quantity s and gives as outputs the growth rate of the host cell as well as the mass and the concentration of proteins/enzymes for ribosomal and non-ribosomal endogenous genes and exogenous genes. The quantity s can be considered as the most limiting substrate (i.e. glucose in this case). Therefore, as a first step, the mapping between the description of substrates in both kinds of models is obtained so that it minimizes the difference between the predicted cell growths.
The proposed conciliation procedure is given below, for the case of one pathway enzyme E and a limiting substrate:

Algorithm 1: Conciliation procedure

Input: s, genetic parameters/constants
Output: v_i, promoter strength ^{opt}

1 Select s $\leftrightarrow$ ν_{Glc} ;
2 Initial guess for promoter strength (RBS strength fixed) ;
3 Compute m_E and m_h in s.s. from genetic model ;
4 **while** $m_E < 0.4 m_h$ **do**
5 Increase promoter strength ;
6 Get m_E, m_h and μ in s.s. from genetic model ;
7 **end**
8 Set $\mu_{min} = \mu$;
9 Get [E] in s.s. using m_E, m_h and M_E ;
10 Compute v_E^{ub} using Eq. 11 ;
11 Do FVA s.t. $\{N\underline{v} = 0$; $\underline{v} \geq 0$; $0 \leq v_E \leq v_E^{ub}$; μ_{min} fixed$\}$
12 **if** *feasible solution* **then**
13 Get v_i from metabolic analysis ;
14 Get promoter strength ^{opt} ;
15 **else**
16 Decrease promoter or RBS strength ;
17 Back to **3**
18 **end**

As mentioned in Sec. 2.2, several genetic functions are substrate-dependent. The first step of the procedure consists of selecting an appropriate/realistic value of the substrate s in the burden-aware model that corresponds to a specific uptake rate ν_{Glc}. Next, initial values for the promoter strength (and/or RBS strength) characterizing the stages of transcription and translation of the gene expressing the enzyme E are given. However, the present method enables estimating their optimal value to optimize the production of the metabolite of interest. Thereupon, only an initial guess of the parameters is required and an optimal value is found such that the mass of exogenous proteins is at most 40% of the total mass of the host cell. Beyond this limit, it can be shown that the cell does not have enough resources to grow correctly. It is important to highlight that the remaining parameters defining the burden-aware model and the exogenous protein E (e.g. its catalytic constant) can be found in the literature or approximated. Afterwards, the concentration of the protein of interest can be obtained using the mass of the cell m_h, m_E and the molar mass of the enzyme, M_E. Then, the upper bound of the corresponding flux is computed using Eq. 11, and finally, an FVA is conducted taking advantage of the outcomes from the burden-aware model, i.e. fixing the growth rate of the cell μ, limiting the flux v_E and restricting the value of other fluxes if experimental data are at disposal. If there is a feasible solution of the FVA, the algorithm stops and the values of the intracellular fluxes are obtained taking into account the expression of the exogenous protein. Otherwise, it means that the constraints are too severe and the promoter/RBS strengths of the exogenous enzyme are adjusted for the next iteration.

4. APPLICATION

4.1 Case Study

The simple case of the production of the metabolite p-coumaric acid is considered. The latter is an organic compound that decreases low-density lipoprotein peroxidation, shows antioxidant and antimicrobial activities, and plays an essential role in human health. Furthermore, some studies have shown that coumaric acid could have a role in reducing the risk of stomach cancer by limiting the formation of carcinogenic nitrosamines. It is also a precursor of phenolic acids, flavonoids, lignin precursors, and other secondary metabolites. Regarding its biosynthesis, p-coumaric acid can be produced from cinnamic acid by the action of the enzyme 4-cinnamic acid hydroxylase or from L-tyrosine by the action of tyrosine ammonia lyase, denoted enzyme TAL. It is worth noting that L-tyrosine is secreted in most microorganisms, as in E.coli.

4.2 Engineered Metabolic Pathway

From the metabolic point of view, considering an engineered metabolic pathway to produce p-coumaric acid will change the size of the host metabolic network. In this study, the E. coli core model from (Orth et al., 2010a) is considered, which consists of 1805 internal metabolites and 2583 reactions where L-tyrosine is involved. Henceforth, it is required to extend the network by adding the following reaction and the corresponding exchange reactions:

$$\mathrm{tyrL} \xrightarrow{\mathrm{TAL}} \mathrm{pC} + \mathrm{NH_3} + \mathrm{H^+} \tag{12}$$

where tyrL represents the metabolite L-tyrosine, pC is the p-coumaric acid and NH_4^+ is the ammonium cation. This

network modification is done using the COBRA Toolbox in MATLAB (Schellenberger et al., 2011).
From the genetic point of view, adding a pathway also demands the definition of the corresponding exogenous genes required to synthesize the enzymes catalyzing the added reactions. In this case, the added reaction is catalyzed by the enzyme TAL. The expression of this enzyme and the burden it induces on the cell must be considered to optimize the metabolic flux through the p-coumaric pathway. To this end, the burden-aware model is used and implemented in the open-source modeling tool OneModel (Santos-Navarro et al., 2022). To define the exogenous protein in OneModel, information related to the enzyme is required such as the length of the enzyme l_{pTAL}, the mRNA production rate ω_{TAL} (related to the promoter strength), and the copy number of protein-coding exogenous genes N_{TAL}. In addition, it requires the parameters K_b^{TAL} and K_u^{TAL} corresponding to the RBS strength. The remaining parameters of the model are obtained from (Santos-Navarro et al., 2021) for the endogenous genes of the host-aware model or using the enzyme database BRENDA for the case of the enzyme TAL.

4.3 Results

Following the procedure presented in Sec. 3, the first step consists of selecting a value for the substrate. For this study, the quantity s is chosen equal to 3.6 g.L^{-1}, a value for which the burden-aware model developed in (Santos-Navarro et al., 2021) fits the best to experimental data. Thereafter, it is required to provide parameters defining the stages of genetic transcription and translation. Fig. 2 illustrates the evolution of the mass fractions of the cell, the growth rate, and the concentration of the exogenous protein in steady-state for different values of the promoter strength. This figure shows the impact of increasing promoter strength (i.e. transcription rate of the exogenous protein) on the concentration of the enzyme and the growth rate of the cell. Therefore, the larger the promoter strength, the higher the concentration of the enzyme and the smaller the growth rate of the cell because of the cell burden. Moreover, when the promoter strength increases, a larger number of ribosomes is involved in the genetic synthesis of the exogenous protein so that less mature ribosomes will be available for the synthesis of ribosomal and non-ribosomal endogenous proteins. A correct first estimate is to choose a promoter strength such that the doubling time of the bacteria remains acceptable despite the cell burden. Then, an optimal value for the transcription and translation strengths may be obtained knowing that the minimal growth rate of the cell is such that the mass of exogenous proteins is at most 40% of the total mass of the host cell. In this study, only the promoter strength is optimized (the RBS strength is fixed). When the promoter strength is set, it is possible to compute the concentration of the exogenous protein, TAL enzyme in this case, using its mass, its molar mass and the total mass of the host cell in steady-state given by the burden-aware model. Then, the upper bound of the corresponding flux is calculated using Eq. 11 and the flux variability analysis can be executed. Table 1 summarizes the value of some parameters for the optimal value of the promoter strength (RBS strength fixed) and provides some essential genetic constants, such as l_{pTAL} and M_{TAL}.

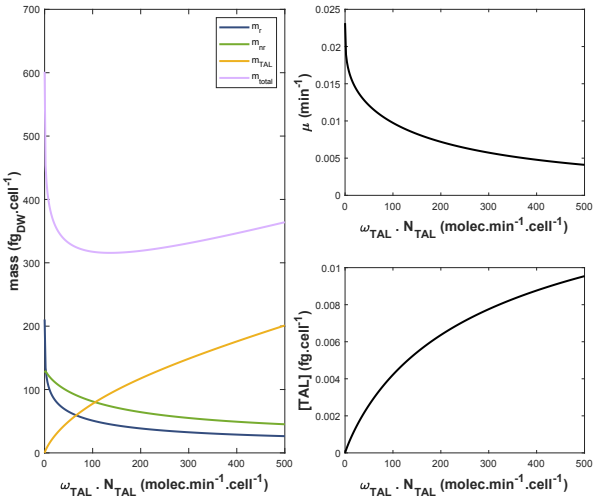

Fig. 2. Evolution of the mass fractions of the cell, the growth rate, and the concentration of the enzyme TAL, in steady-state, for different values of the promoter strength of the exogenous protein TAL.

For FVA, it is required to provide the value of the uptake rate of glucose, which can be deduced from the substrate s by regression. Also, the growth rate of the cell is known and is given by the burden-aware model, and finally, the upper bound of the engineered flux is computed using Eq. 11 and is equal to 0.4965 $\text{mmol.g}_{DW}^{-1}\text{.h}^{-1}$. Thereupon, to assess the merits of the procedure, the same FVA analysis can be conducted but without imposing any upper bound, related to the enzyme availability, for the flux of interest. In this case, it is shown that the actual value of the flux should be within [0 ; 4.8483] $\text{mmol.g}_{DW}^{-1}\text{.h}^{-1}$. That means that taking into account the enzyme activity allows reducing significantly the interval for the actual value of the flux, as depicted in Fig. 3.

Table 1. Parameter values and constants

	Value	Unit
l_{pTAL}	531	aa
M_{TAL}	57'895	Da
$(\omega_{TAL}N_{TAL})^{opt}$	234	$\text{molec.min}^{-1}\text{.cell}^{-1}$
μ_{min}	0.0066	min^{-1}
m_h	322.3596	$\text{fg}_{DW}\text{.cell}^{-1}$
$[TAL]$	0.0069	fg.cell^{-1}
$k_{cat,TAL}$	1.2	min^{-1}
ν_{Glc}	-15.8	$\text{mmol.g}_{DW}^{-1}\text{.h}^{-1}$

For the sake of clarity, intervals of possible values of other metabolic fluxes are not presented in this work but are in concordance with existing studies. Fig. 3 shows that combining both constraint-based and burden-aware models leads to a better prediction of the production rate of the metabolite of interest (smaller intervals of values). To further constrain the solution space of the flux distribution, existing methods as the ones mentioned in Sec. 1 can be exploited additionally or the procedure presented in this paper can be extended for the expression of non-ribosomal endogenous proteins. Although it will demand significant effort to evaluate the corresponding genetic parameters, it is most likely that the solution space will be highly reduced. Further research should entail this issue.

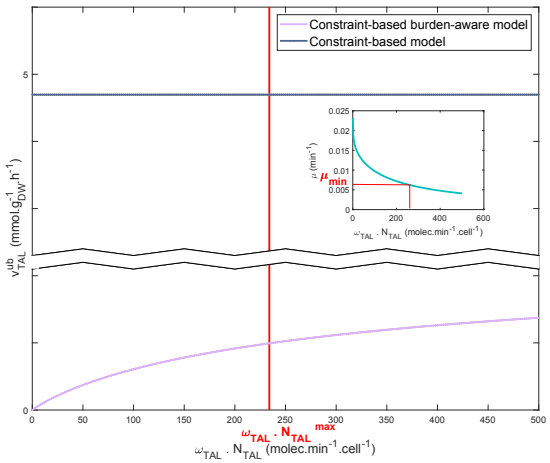

Fig. 3. Evolution of v^{ub}_{TAL} for different values of the promoter strength of the exogenous protein TAL.

5. CONCLUSION

This work proposes a procedure for obtaining a better estimation of some metabolic fluxes in the particular case for which an engineered metabolic pathway is catalyzed by an exogenous protein. The method enables linking two modeling levels, namely metabolic models and models of gene expression accounting for host-circuit interactions (i.e. the cell burden), to further constrain the FVA problem and allows getting the optimal value for the promoter/RBS strengths for a specific amount of substrate. The procedure is effective and provides promising results. Further research entails the consideration of endogenous proteins to further reduce the solution space and an optimal tuning of genetic parameters as part of research in synthetic biology.

REFERENCES

Akesson, M., Forster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.*, 6, 285–293.

Boada, Y., Reynoso-Meza, G., Pico, J., and Vignoni, A. (2016). Multi-objective optimization framework to obtain model-based guidelines for tuning biological synthetic devices : an adaptative network case. *BMC Syst. Biol.*, 10 (27), doi:10.1186/s12918–016–0269–0.

Boada, Y., Santos-Navarro, F., Pico, J., and Vignoni, A. (2022). Modeling and optimization of a molecular biocontroller for the regulation of complex metabolic pathways. *Front. Mol. Biosci.*, 9:801032.

Chandrasekaran, S. and Price, N. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in escherichia coli and mycobacterium tuberculosis. *PNAS*, 107 (41), 17845–17850.

Covert, M. and Palsson, B. (2002). Transcriptional regulation in constraints-based metabolic models of e. coli. *JBC*, 277, 28058–28064.

Covert, M., Schilling, C., and Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, 213, 73–88.

Covert, M., Xiao, N., Chen, T., and Karr, J. (2008). Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli. *Bioinform.*, 18, 2044–2045.

Gudmundsson, S. and Thiele, I. (2010). Computationally efficient flux variability analysis. *BMC Bioinform.*, 11:489.

Liao, C., Blanchard, A., and Lu, T. (2017). An integrative circuit-host modelling framework for predicting synthetic gene network behaviours. *Nat. Microbiol.*, 2, 1658–1666.

Lindhorst, H., Lucia, S., Findeisen, R., and Waldherr, S. (2018a). Modeling enzyme controlled metabolic networks in rapidly changing environments by robust optimization. *IEEE Control Syst. Lett.*, 99:1-1, doi:10.1109/LCSYS.2018.2866234.

Lindhorst, H., Reimers, A., and Waldherr, S. (2018b). Dynamic modeling of enzyme controlled metabolic networks using a receding time horizon. *Preprints on 10th IFAC International Symposium on Advanced Control of Chemical Processes.*

Machado, D., Herrgard, M., and Rocha, I. (2016). Stoichiomotric representation of gene-protein-reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. *PLoS Comput. Biol.*, 12 (10), e1005140. doi:10.1371/journal.pcbi.1005140.

Marmiesse, L., Peyraud, R., and Cottret, L. (2015). Flexflux : combining metabolic flux and regulatory network analyses. *BMC Syst. Biol.*, 9:93.

Nielsen, J. and Olsson, L. (2002). An expanded role for microbial physiology in metabolic engineering and functional genomics : moving forward systems biology. *FEMS Yeast Res.*, 2, 175–181.

Orth, J., Fleming, R., and Palsson, B. (2010a). Reconstruction and use of microbial metabolic networks: the core e.coli metabolic model as an educational guide. *EcoSal Plus*, 4 (1), 10.1128.

Orth, J., Thiele, I., and Palsson, B. (2010b). What is flux balance analysis. *Comput. Biol.*, 3, 245–248.

Reimers, A., Knoop, H., Bockmayr, A., and Steuer, R. (2017). Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *PNAS*, 114 (31), E6457–E6465.

Santos-Navarro, F., Navarro, J., Boada, Y., Vignoni, A., and Pico, J. (2022). Onemodel : an open-source sbml modeling tool focused on accessibility, simplicity and modularity. *IFAC-PapersOnLine*, 55 (7), 125–130.

Santos-Navarro, F., Vignoni, A., Boada, Y., and Pico, J. (2021). Rbs and promoter strengths determine the cell-growth-dependent protein mass fractions and their optimal synthesis rates. *ACS Synth. Biol.*, 10 (12), 3290–3303.

Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J., and Feist, A. (2011). Quantitative prediction of cellular metabolism with constraint-based models : the cobra toolbox v2.0. *Nat. Protoc.*, 6 (9), 1290–1307.

Varner, J. (2000). Large-scale prediction of phenotype : concept. *Biotechnol. Bioeng.*, 69, 664–678.

Waldherr, S., Oyarzun, D., and Bockmayr, A. (2015). Dynamic optimization of metabolic networks coupled with gene expression. *J. Theor. Biol.*, 365, 469–485.

Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T., and Mukhopadhyay, A. (2019). Engineering robust production microbes for large-scale cultivation. *Trends Microbiol.*, 27, 524–537.