Flexible Nets to Improve GEM Cell Factories by Combining Kinetic and Proteomics Data

Jorge Lázaro[1][0009−0009−5732−9150], Jorge Júlvez[2][0000−0002−7093−228X], and Jürgen Zanghellini[3][0000−0002−1964−2455]

[1] Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain, jorgelazaro@unizar.es
[2] Aragón Institute for Engineering Research, University of Zaragoza, Zaragoza, Spain julvez@unizar.es
[3] Department of Analytical Chemistry, University of Vienna, Vienna, Austria juergen.zanghellini@univie.ac.at

Abstract. Alzheimer's disease is expected to reach a prevalence of 152 million people worldwide caused by the aggregation of amyloid β-proteins leading to apoptosis of neurons and loss of cognitive function. Although there is no effective treatment for this disease, molecules such as *scyllo*-inositol have been shown to be promising. *Bacillus subtilis* has been proposed as a suitable organism for the production of *scyllo*-inositol. Metabolic computational models have proven useful in the prediction of the production of a metabolite. However, most genome-scale metabolic models lack detailed parameters and tend to overestimate the production of a metabolite with respect to the consumption of medium resources. In order to reduce the solution space and, hence, obtain a more realistic model, additional constraints from experimental data can be added to the model. This work exploits the modeling capabilities of Flexible Nets to model the production of *scyllo*-inositol in a genome-scale metabolic model of *Bacillus subtilis* that has been previously enriched with proteomic and enzymatic data. We assess how these constraints limit the *scyllo*-inositol production to more realistic levels. Moreover, the integration of different types of constraints allowed us to uncover which one of them limits the production of *scyllo*-inositol for a given growth rate.

Keywords: Metabolic modeling · Flexible Nets · genome-scale models · *scyllo*-inositol · proteomic constraints · enzymatic constraints.

1 Introduction

Genome-scale metabolic models (GEMs) are mathematical representations of the metabolism of an organism that account for the stoichiometry and flux bounds of the metabolic reactions (Gu et al. [2019]). GEMs have various practical applications such as strain development for producing bio-based chemicals and materials (Senger [2010], designing bio-processes Gotsmy et al. [2023], Lázaro et al. [2022]), drug targeting pathogens (Kim et al. [2010]), predicting enzyme functions (Orth et al. [2011]), modeling interactions among multiple cells or organisms (Zorrilla

et al. [2021]), and understanding human diseases (O'brien et al. [2013]). The most widely-used method for analyzing GEMs is Flux Balance Analysis (FBA) (Orth et al. [2010]), which consists of solving a linear programming problem.

Although GEMs have proven useful in many cases, they are still far from reliably predicting growth rates or product efficiency because they are exclusively based on stoichiometric and mass balance constraints. Consequently, the FBA optimal solution still tends to be relatively large and usually fails to correctly approximate reality (Durot et al. [2008], Shaw and Cheung [2021], Mao et al. [2015]). The implementation of additional constraints to GEMs is gaining more and more attention, since they can reduce the solution space and, hence, result in a more realistic prediction of the metabolic fluxes. It must be noticed that, in general, the bottleneck that limits the rate at which a reaction occurs is the enzyme abundance and its kinetic parameters (Massaiu et al. [2019]).

This work focuses on the implementation of enzymatic and proteomic constraints. The metabolic flux of each reaction strongly depends on the availability of the enzyme that catalyzes that reaction. These properties can be integrated into the GEM in the form of parameters such as protein abundance and catalytic constants. Here, we show how to exploit the modeling formalism of Flexible Nets, an extension of Petri nets, to integrate enzymatic constraints into a GEM of *Bacillus subtilis*.

B. subtilis, which is a model organism for Gram-positive bacteria, stands out for its extensively mapped genome and remarkable adaptability to diverse environments and substrates. *B. subtilis* efficiently secretes products without genetic engineering. In particular, its ability to grow in minimal medium and produce *scyllo*-inositol, a promising therapeutic for Alzheimer's disease which inhibits β-amyloid aggregation, has been well-characterized. Recently, a method was devised by Michon et al. [2020] to synthesize *scyllo*-inositol from glucose, making *B. subtilis* a competitive producer compared to *E. coli*.

The remainder of this paper is organized as follows: Section 2 introduces Flexible Nets and shows their capabilities to model complex dynamics in biological systems. Sections 3, 4 and 5 introduce two methods to implement kinetic and proteomic constraints as well as their alternative using FNs. Sections 6 and 7 introduce the case study and discuss how we enriched the model and assessed its performance. The main conclusions are drawn in Section 8.

2 Flexible Nets to model biological systems

2.1 Flexible Nets

Flexible Nets (FNs) (Júlvez et al. [2018]) is a modeling formalism for dynamic systems inspired by Petri Nets (Murata [1989], Silva [Chapman & Hall, London, 1993]). An FN models the interaction between the state and the processes of a dynamic system through two interconnected nets: the event net and the intensity net. While the event net describes how processes alter the state variables, the intensity net captures how these variables determine the process rates.

In contrast to Petri Nets, both the event and intensity nets are tripartite graphs whose vertices are places, transitions, and handlers. In the event net, event handlers trigger changes in metabolite concentration resulting from reactions, while in the intensity net, intensity handlers model how metabolite concentrations affect reaction rates. Places (depicted as circles) are linked with metabolites; transitions (depicted as rectangles) are linked with reactions; event handlers (depicted as dots) monitor metabolite concentration changes; and intensity handlers (also dots) model the impact of metabolite concentrations on reaction rates.

If only the stoichiometry of a metabolic network is taken into account, i.e., no reaction rate is specified, then, an event net is enough to model the behavior of the network. Figure 1 shows a simple event net modeling a system consisting of 4 metabolites (places $\{A, B, C, D\}$), 2 reactions or transitions ($\{R_1, R_2\}$), and 2 event handlers ($\{v_1, v_2\}$) that connect places and transitions. In this simple system, A and B react (reaction R_1) to produce 2 molecules of C; and B is also substrate of a reaction (R_2) that transforms this metabolite into D. The reaction network modeled by the net is:

$$R_1\colon A + B \to 2C \tag{1}$$

$$R_2\colon B \to D \tag{2}$$

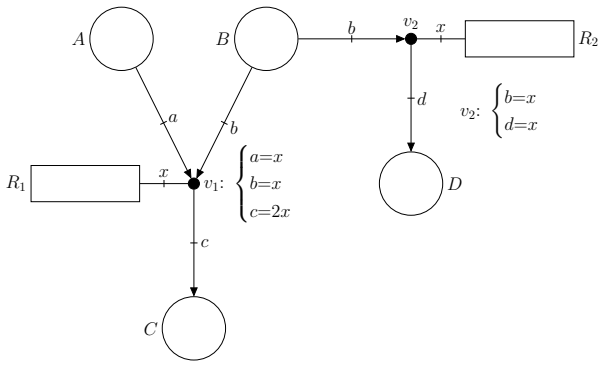

Fig. 1. Event net that only takes into account the stoichiometry of the reactions R_1 : $A + B \to 2C$ and $R_2 : B \to D$.

The equalities in v_1, $a{=}x$; $b{=}x$; and $c{=}2x$ imply that the occurrence of reaction R_1 triggers the consumption of 1 unit of metabolite A and 1 unit of metabolite B to produce 2 units of metabolite C. Similarly, the equation of v_2 models that when reaction R_2 happens, 1 unit of metabolite B is converted into 1 unit of metabolite D.

An appealing feature of FNs is that its intensity net can model complex dynamics in an intuitive way. For instance, assume that reaction R_2 is enzymat-

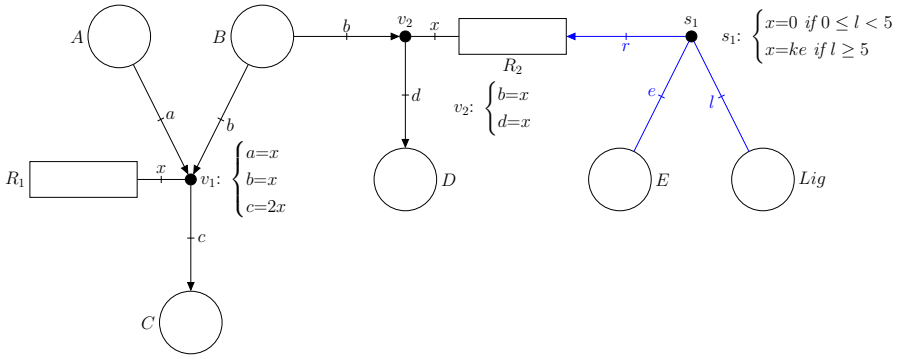

Fig. 2. Flexible Net modeling the reactions $R_1 : A + B \rightarrow 2C$ and $R_2 : B + E + L \rightarrow D + E + Lig$. The event net is represented by black arcs and edges. In contrast, blue arcs and edges depict the intensity net.

ically catalyzed by an enzyme E that needs to incorporate a ligand, Lig, to be active (Figure 2). This phenomenon is quite common in biology as an enzymatic regulation process. The reaction formulation would look like:

$$R_1\text{: } A + B \rightarrow 2C \tag{3}$$
$$R_2\text{: } B + E + L \rightarrow D + E + Lig \tag{4}$$

The regulation of R_2 is captured by the intensity net which, for clarity, is colored in blue. The enzyme, depicted by E catalyzes the reaction R_2 which will be activated depending on the ligand concentration, modeled by l. This relationship is determined by the inequalities that define the intensity handler s_1. The first condition, $r = 0$ if $0 \leq l < 5$, indicates that the reaction rate, r, of R_2 is null when the concentration of Lig, l, is less than a threshold set to 5 units. If the concentration of Lig is greater than or equal to 5 units, then the rate of R_2 is equal to a constant, k, multiplied by the enzyme abundance, e. This dependency is captured by the inequality $r = ke$ if $l \geq 5$.

In FNs, the rate of a reaction, R_i, is captured by the variable $\lambda[R_i]$. This variable is equal to $\lambda_0[R_i]$, which is the default intensity in a transition, plus the intensities associated with its input arcs minus the intensities of its output arcs.

If no $\lambda_0[R_i]$ is explicitly associated with a reaction R_i, then it is assumed that $\lambda_0[R_i] = 0$. Thus, in the presented example, the intensity of R_2 in Figure 2 can be expressed as:

$$\lambda[R_2] = \lambda_0[R_2] + \Delta\lambda[(s_1, R_2)] = \Delta\lambda[(s_1, R_2)] \tag{5}$$

where, according to the equations associated with s_1, $\Delta\lambda[(s_1, R_2)]$ is equal to k times the enzyme abundance, e, if the concentration of Lig is greater than or equal to 5, and equal to 0 otherwise.

2.2 Modeling GEMs with FNs

GEMs are mathematically defined as a tuple $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ L,\ U\}$, where $\mathcal{R}$ is the set of reactions, $\mathcal{M}$ is the set of metabolites, $\mathcal{S} \in \mathbb{R}^{|\mathcal{M}|\times|\mathcal{R}|}$ refers to the stoichiometric matrix whose rows(columns) are associated with the metabolites(reactions) of the model, and $L, U \in \mathbb{R}^{|\mathcal{R}|}$ are the lower and upper flux bounds that constrain the fluxes of the reactions. Since GEMs mainly account for mass balance and stoichiometric constraints, these flux bounds are usually extremely loose. A common assumption when analyzing GEMs is that the model is in steady state. Under this condition, all the metabolites have a constant concentration and, hence, its production rate is equal to its consumption rate. This is expressed by:

$$\mathcal{S} \cdot v = 0 \tag{6}$$

where v is a one-column vector with the fluxes of the reactions. The addition of the flux bounds, L and U, and an objective function to (6) results in the FBA linear programming problem:

$$\begin{aligned} \max\ & c^T \cdot v \\ & \mathcal{S} \cdot v = 0 \\ & L \leq v \leq U \end{aligned} \tag{7}$$

where c is a vector of weights indicating the contribution of each reaction to the objective function. Notice that the FBA of GEMs can be mimicked by FNs by imposing constant concentrations of all metabolites, adding flux bounds and introducing an objective function (see Júlvez and Oliver [2020] for details on steady state analysis in general FNs).

Some methods, such as GECKO (Domenzain et al. [2022]) and sMOMENT (Bekiaris and Klamt [2020]), have been developed recently to restrict the metabolic fluxes of an organism with enzymatic and proteomic data. Nevertheless, they rely on the addition of various artificial reactions and metabolites that expand the original GEM stoichiometric matrix. In addition to modeling all the reactions in a GEM, FNs can capture complex enzymatic relationships that are difficult to accommodate with current methods.

3 GECKO method

The GECKO method consists of constraining the metabolic flux of each catalyzed reaction for which there is information about the abundance of a specific enzyme. In particular, the flux of a given reaction R_i, v_i ($mmol\ g_{DW}^{-1}\ h^{-1}$), cannot exceed the product between the catalytic constant, $k_{cat,i}$ ($1/h$) and the enzyme abundance, e_i ($mmol\ g_{DW}^{-1}$) (Equation (8)). This is applied to each metabolic enzyme, e_i, in the model:

$$v_i \leq k_{cat,i} \cdot e_i \tag{8}$$

GECKO constraints are integrated into the GEM as follows:

1. Add a metabolite, E_i, as a reactant with stoichiometric coefficient $\frac{1}{k_{cat,i}}$ to each metabolic enzyme-catalyzed reaction R_i. For instance, if R_i:$A \rightarrow B$, then the addition of E_i results in:

$$R_i'\colon A + \frac{1}{k_{cat,i}} E_i \rightarrow B$$

2. Add a reaction, R_i'' for each implemented enzyme abundance, e_i. For instance:

$$R_i''\colon \emptyset \rightarrow E_i$$

The upper bound of R_i'' is set to the enzyme abundance, i.e. $U[R_i''] = e_i$.

3.1 Modeling the GECKO method with FNs

The GECKO method can be easily modeled by FNs. Consider a reaction R_1 catalyzed by an enzyme E that transforms metabolite A into B:

$$R_1\colon A + E \rightarrow B + E$$

The FN in Figure 3 shows how the reaction rate of R_1 is constrained by enzymatic limitations according to Equation (8).

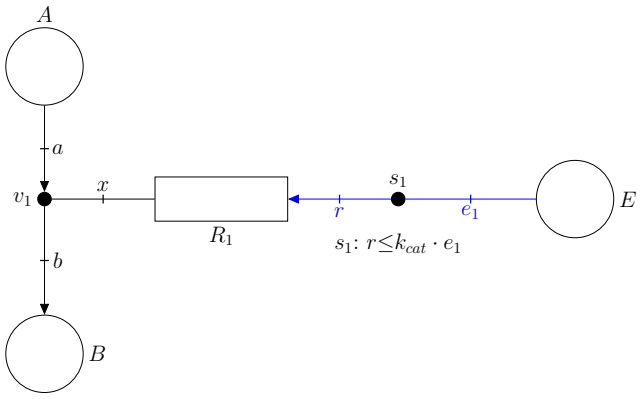

Fig. 3. Flexible Net that models the dynamics of a system where the reaction rate of R_1 is limited by a GECKO-like constraint. Since the intensity transferred from s_1 to R_1 satisfies that $r \leq k_{cat} \cdot e_1$, the intensity of R_1 will satisfy $\lambda[R_1] \leq k_{cat} \cdot e_1$.

In Figure 3, the inequatility presented in the intensity handler s_1 is equivalent to Equation (8). Thus, the flux of a reaction must be less or equal than the catalytic constant multiplied by the enzyme abundance. This strategy is applied to all the reactions for which information about their kinetic parameters and specific enzyme abundances is available.

4 sMOMENT method

sMOMENT is a method to constrain metabolic fluxes of the reactions in a GEM when the k_{cat} and molecular weights of the enzymes that catalyze the reactions are known. As GECKO, sMOMENT also assumes steady state conditions, see Equation (6). Additionally, the specific constraint added in this method corresponds to the total concentration, P, of all the enzymes experimentally measured. A threshold is set so that the total abundance of enzymes cannot exceed this value, P (g/g_{DW}):

$$\sum_i e_i \cdot MW_i \leq P \tag{9}$$

where MW_i is the molecular weight of an enzyme. Additionally, Equation (8) can be alternatively expressed as:

$$\frac{v_i}{k_{cat,i}} \leq e_i \tag{10}$$

This leads to Equation (9):

$$\sum_i v_i \cdot \frac{MW_i}{k_{cat,i}} \leq P \tag{11}$$

which can be rewritten as:

$$v_{pool} - \sum_i v_i \frac{MW_i}{k_{cat,i}} = 0; v_{pool} \leq P \tag{12}$$

where v_{pool} is an auxiliary variable that corresponds to the total mass of metabolic enzymes expressed in g/g_{DW}.

This constraint is integrated into a GEM through the following steps:

1. Add an artificial metabolite called M_{pool} in all the reactions catalyzed by metabolic enzymes. For instance, if R_i:$A \rightarrow B$, then the addition of M_{pool} results in:

$$R_i'''\text{: } A + \frac{MW_i}{k_{cat,i}} M_{pool} \rightarrow B$$

2. Add a new artificial reaction, R_{pool}:

$$R_{pool}\text{: } \emptyset \rightarrow M_{pool}$$

whose upper flux bound is equal to the sum of the abundances of all the metabolic enzymes, P.

This way the resulting model is constrained by the total abundance of metabolic enzymes.

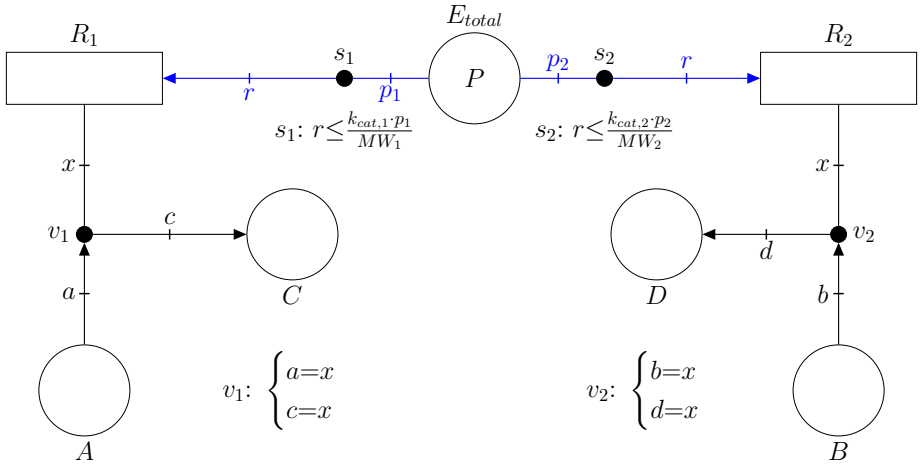

Fig. 4. FN that models the dynamics relationships of a system where the reaction rates of R_1 and R_2 are limited by an sMOMENT-like global constraint. The net determines that the intensity transferred from s_1 to R_1 satisfies that $r \leq \frac{k_{cat,1} \cdot p_1}{MW_1}$, thus, the intensity of R_1 will satisfy that $\lambda[R_1] \leq \frac{k_{cat,1} \cdot p_1}{MW_1}$. The same is applied to R_2.

4.1 Modeling the sMOMENT method with FNs

In contrast to the sMOMENT method which requires the addition of artificial net elements, i.e. a metabolite and a transition, to account for the effect of the available enzymes, FNs can model the same behavior in a more natural and simple way by means of the intensity net. For instance, assume that we know the kinetic parameters of two reactions, R_1 and R_2, and the sum of the abundances of the enzymes that catalyze them, Y and Z, respectively. R_1 is catalyzed by an enzyme that transforms metabolite A into C and R_2 is catalyzed by a different enzyme converting B into D:

$$R_1\colon A + Y \rightarrow C + Y$$
$$R_2\colon B + Z \rightarrow D + Z$$

Both enzymes are part of a pool of enzymes E_{total} whose total amount is constrained by experimental data.

The global constraint referring to the total measured abundance of enzymes is implemented as in Figure 4. The place E_{total} accounts for all the metabolic enzymes, two in this case, and it is assigned a marking, P, that represents the total enzyme concentration ($g\ g_{DW}^{-1}$) which is expressed in (12). The edges connected to E_{total} allow enzyme abundances to be distributed between reactions R_1 and R_2. The dynamics of the system are captured by the intensity handlers s_1 and s_2. These net elements, E_{total}, s_1 and s_2, model the following inequalities,

which are derived from (12):

$$v_i \leq \frac{k_{cat,i} \cdot p_i}{MW_i} \tag{13}$$

$$\sum_{i=1}^{20} p_i \leq P \tag{14}$$

where v_i is the metabolic flux of reaction R_i, $k_{cat,i}$ is the catalytic constant of the enzyme that catalyzes reaction R_i, MW_i is its molecular weight, and p_i is the abundance of an individual metabolic enzyme from the pool directed to the reaction R_i.

5 Combining the sMOMENT and GECKO methods

This section shows how FNs can accommodate simultaneously the global and individual enzymatic constraints (sMOMENT and GECKO methods) of a metabolic network by means of its intensity network.

Assume that we know the kinetic and proteomic information of two enzymes Y and Z that catalyze two reactions R_1 and R_2:

$$R_1\colon A + Y \rightarrow C + Y$$
$$R_2\colon B + Z \rightarrow D + Z$$

In the FN in Figure 5, the intensity handlers s_1 and s_2 collect the enzymatic constraints that limit the fluxes of R_1 and R_2 respectively. This is extended to all the reactions for which enzymatic information is available.

The flux of the reaction R_1, r, is doubly constrained. On the one hand, it has to be less than or equal to the product of the k_{cat} ($k_{cat,1}$ in the figure) of the enzyme Y by its abundance from the pool of enzymes directed to the reaction that Y catalyzes, p_1, divided by its molecular weight, MW_1. This would be the implementation of the global constraint. On the other hand, the flux of R_1 must satisfy a second inequality, being r less or equal than the k_{cat} of Y enzyme multiplied by its specific abundance, e_1 (equivalent to GECKO). The same procedure is applied to the reaction R_2.

6 *Bacillus subtilis* models

In this work, we enriched a *B. subtilis* GEM (iYO844 available in the BIGG Models database (Norsigian et al. [2019])) with enzymatic constraints using FNs to computationally simulate the work presented in Michon et al. [2020] where a mutant strain was genetically modified to produce *scyllo*-inositol from glucose.

Since a wild-type strain of *B. subtilis* cannot endogenously produce *myo*-inositol from glucose, Norsigian et al. [2019] introduced the gene *ino1* that encodes the inositol-1-phosphate (MI1P) synthase (MI1PS) from *Mycobacterium tuberculosis* to convert glucose-6-phosphate (G6P) into MI1P. Subsequently, the

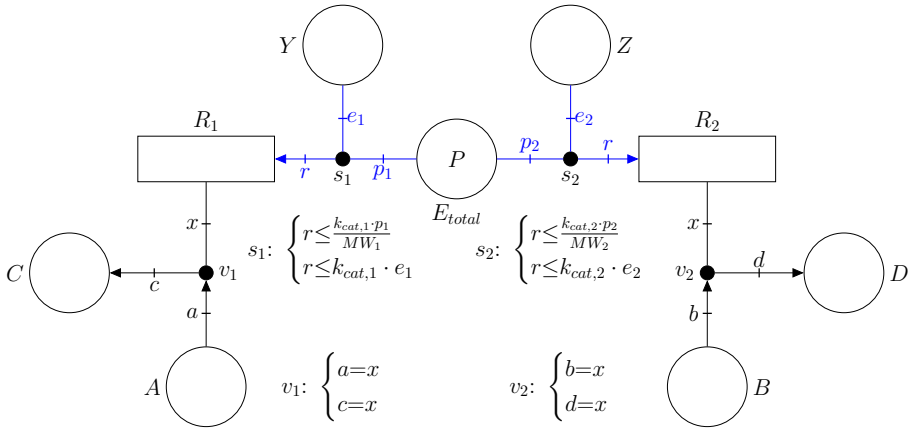

Fig. 5. Flexible Net that sets the dynamics relationships of a system where the reaction rates of R_1 and R_2 are limited by GECKO-like and an sMOMENT-like global constraints. Consider that the the intensity transferred from s_1 to R_1 satisfies the two inequalities associated to s_1. As a result, the intensity of R_1 will satisfy both conditions: $\lambda[R_1] \leq \frac{k_{cat,1} \cdot p_1}{MW_1}$ and $\lambda[R_1] \leq k_{cat,1} \cdot e_1$. The same is applied for s_2.

inositol monophosphatase (IMP) cleaves-off a phosphate from the MI1P originating a molecule of *myo*-inositol. This *myo*-inositol is the substrate of the inositol 2-dehydrogenase IolG encoded by the gene *iolG*. The last step of the *scyllo*-inositol production pathway is catalyzed by the IolW enzyme, the scyllo-inositol 2-dehydrogenase encoded by the gene *iolW* (Figure 6). However, the metabolite produced by IolG, 2-inosose, can be degraded by a series of catabolic reactions. To avoid that the 2-inosose follows its catabolic pathway, researchers inhibited it, so all the 2-inosose goes through the *scyllo*-inositol pathway.

The above genetic modifications can be computationally integrated in the GEM of *B. subtilis* by means of the following steps:

1. Add the reaction catalyzed by the MI1PS. This way, the *B. subtilis* model can transform the G6P into MI1P (R_1 in Figure 6).
2. Set a null flux of the reaction that degrades 2-inosose, *INSCR*:

$$\lambda[R_2] = 0 \tag{15}$$

3. Include the reaction that catalyzes the conversion of 2-inosose into *scyllo*-inositol (R_3 in Figure 6).

For clarity, in Figure 6, the transitions of the model that are not relevant for this case study have been omitted. The equations corresponding to the event handlers were also omitted because the stoichiometry of the represented metabolites is one to one in all the reactions of the *scyllo*-inositol production pathway, and hence, all the labels in the event handlers are equal (e.g. in v_1, $a = x$ and $b = x$).

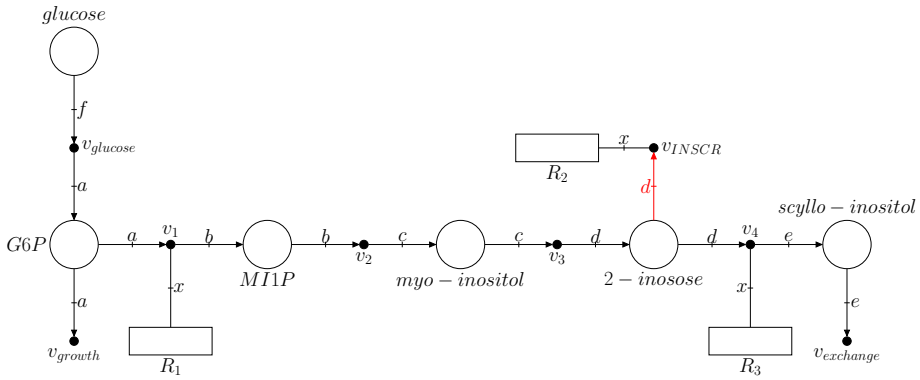

Fig. 6. Schematic representation of the analyzed pathways in the *B. subtilis* modified model. Glucose (*glc_c* in the scheme) can be either converted into MI1P, following the *scyllo*-inositol synthesis pathway or employed for the growth of the cell. The red arrow pointing to v_{INSCR} indicates the inhibited reaction in the model, *INSCR*, that simulates the experimental knock-out of the multiple reactions degrading the intermediate 2-inosose in Michon et al. [2020].

Moreover, the event handler v_{growth} models that the G6P can be consumed by a series of reactions to generate growth and $v_{exchange}$ models an exchange reaction that allows *scyllo*-inositol to leave the system in order to satisfy the steady state condition.

As a result, we obtained a *B. subtilis* model with the capacity to synthesize *scyllo*-inositol from glucose as carbon source (sc_iYO844). These modifications were included using the library COBRApy (Ebrahim et al. [2013]).

7 Enriching iYO844 with kinetic and proteomic constraints using FNs

7.1 Collection of enzymatic parameters

To obtain the enzymatic data for the GEM enrichment, we followed the process introduced in Massaiu et al. [2019]. This process consists in a manual search in several databases and literature. The kinetic parameters were obtained from the BRENDA (Chang et al. [2020]) and SABIO-RK (Wittig et al. [2017]) databases and the literature. The molecular weights (kDa) of the enzymes were obtained from the SubtiWiki (Pedreira et al. [2021]), a database that collects exclusively information about *B. subtilis* proteins and genes. The information for each enzyme is in Table 1 (see Annex).

Protein abundances expressed in $mmol\ g_{DW}^{-1}$ for each enzyme were quantified in Goelzer et al. [2015] at exponential phase. As stated in Massaiu et al. [2019], when the protein levels were under the detection limit, we took the minimum value among the protein level measurements in the same condition (6.8 x 10^{-8}

mmol g_{DW}^{-1}). Since we also integrated the sMOMENT global constraint, total protein abundance data were required. On the one hand, according to Milo [2013], the total amount of enzymes in an *E. coli* cell is 0.55 $g\ g_{DW}^{-1}$. On the other hand, the mass fraction of metabolic enzymes (the 20 enzymes taken into account) is 0.0194 (Massaiu et al. [2019]). This value was computed by summing the abundance expressed in parts per million of all the 20 enzymes retrieved in PaxDB (Huang et al. [2023]). Thus, the total enzyme abundance is $P = 0.0194 \cdot 0.55\ g\ g_{DW}^{-1} = 0.01067\ g\ g_{DW}^{-1}$.

7.2 Integrating kinetic and proteomic constraints

Once all the necessary enzymatic data was collected, we applied the specific constraints on the enzyme-catalyzed reactions (17 in the glucose metabolism pathway and 3 that belong to the *scyllo*-inositol production pathway) and over the total concentration of metabolic enzymes.

For the GECKO-like constraints, we followed Equation (8) for the 17 enzyme-catalyzed reactions participating in glucose metabolism. Conversely, for the 3 enzymes participating in the *scyllo*-inositol pathway, we assumed the following: the flux of the *scyllo*-inositol synthesis reactions will satisfy Equation (16) because these enzymes are overexpressed and, consequently, their abundances must be strictly higher when *scyllo*-inositol production is maximized than when they were experimentally measured (when the growth rate is maximized).

$$v_i \geq k_{cat,i} \cdot e_i \tag{16}$$

For the constraint that accounts for the total amount of enzymes modeled by FNs as showed in Figure 4, we applied the corresponding restriction, i.e. inequality (13), to each of the 20 enzyme-catalyzed reactions.

7.3 Assessing the effect of the integrated constraints

The aim of this section is to assess the effect of kinetic, proteomic and glucose uptake rate constraints on the *scyllo*-inositol production and growth rate. Several models, which depend on the integrated constraints, were generated from the original GEM (iYO844):

- **sc_ec_iYO844**. This model implements the kinetic constraints for the 17 reactions of the carbon source metabolism (the GECKO-like constraints).
- **sc_ec_iYO844 + uptake constraint**. This model integrates the kinetic constraints for the 17 reactions and the constraint for the glucose uptake rate.
- **sc_ec_iYO844 +** v_{pool}. This model integrates the proteomic constraints that end up limiting the total amount of enzymes as a pool (sMOMENT-like constraint).
- **sc_ec_iYO844 +** v_{pool} **+ uptake constraint**. This model includes all the above presented constraints: kinetic, proteomic and glucose uptake rate constraints.

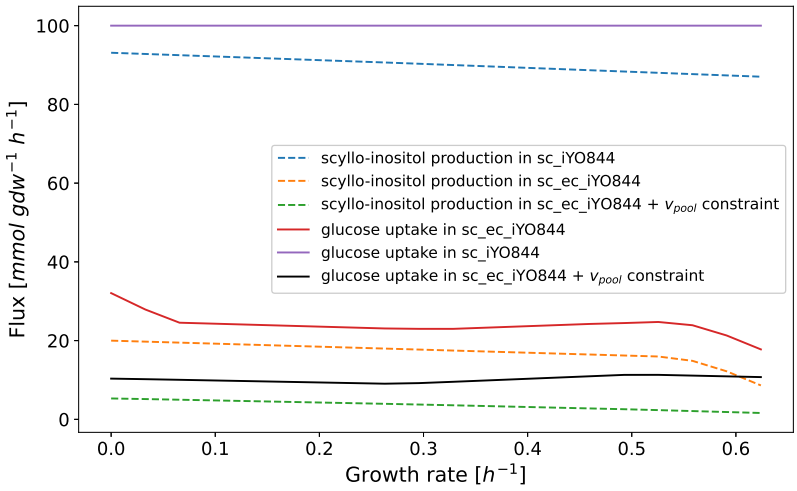

Fig. 7. Fluxes of *scyllo*-inositol synthesis (dashed lines) and glucose uptake (continuous lines) for growth rates in the interval $[0,\ 0.624]\ h^{-1}$.

If the steady state condition is imposed to the sc_iYO844 FN and the synthesis of *scyllo*-inositol is maximized, then the obtained growth rate is null. Conversely, if the growth rate is maximized ($0.624\ h^{-1}$), *scyllo*-inositol production is zero. This is biologically non-realistic and shows that, when there is no kinetic or proteomic constraints, the carbon source is either directed to the growth or to *scyllo*-inositol synthesis. In other words, there are pathways that compete for the carbon source, and the maximization of one implies zero flux in the other.

In further simulations, we ran the steady state condition applied in Equation 7 on each of the presented models converted to FN, setting the *scyllo*-inositol synthesis as the objective function to be maximized. The maximization of *scyllo*-inositol synthesis was performed while fixing the growth rate to 20 different values, equally distributed in the interval between 0 and $0.624\ h^{-1}$. The obtained results are shown in Figure 7.

The purple and blue lines refer to the glucose uptake and the *scyllo*-inositol production, respectively, in the model sc_iYO844. The red and orange lines correspond with the glucose uptake and *scyllo*-inositol production, respectively, in sc_ec_iYO844. Finally, the black and green lines represent the glucose uptake and the *scyllo*-inositol production, respectively, when including all the constraints introduced in Section 7.2 (sc_ec_iYO844 + v_{pool} constraint). These results suggest that the inclusion of the kinetic and proteomic constraints results in a reduction of the solution space, and consequently predicts a more realistic production of *scyllo*-inositol. The initial production of *scyllo*-inositol is high (between 93.118 $mmol\ g_{DW}^{-1}\ h^{-1}$ and 87.034 $mmol\ g_{DW}^{-1}\ h^{-1}$ when the growth rates are minimum

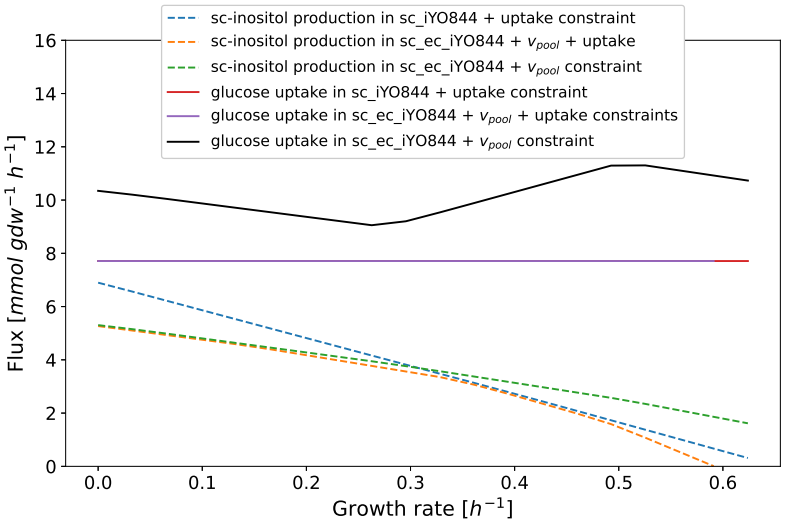

Fig. 8. Fluxes of *scyllo*-inositol synthesis (dashed lines) and glucose uptake (continuous lines) for growth rates in the interval $[0,\ 0.624]\ h^{-1}$ with the experimental glucose uptake in *B. subtilis* as an additional constraint.

and maximum, respectively) due to the lack of tight constraints in sc_iYO844. Notice that the experimental uptake rate of glucose in *B. subtilis* is 7.71 $mmol\ g_{DW}^{-1}\ h^{-1}$ (Massaiu et al. [2019]) and that, according to the reaction network in Figure 6, one molecule of glucose can produce at most one molecule of *scyllo*-inositol. Hence, 7.71 $mmol\ g_{DW}^{-1}\ h^{-1}$ is an upper bound for the *scyllo*-inositol synthesis flux and, therefore, the reported initial productions are unreachable by a real cell.

As shown in Figure 7, the glucose uptake rate varies in an interval between 9.06 and 11.30 $mmol\ g_{DW}^{-1}\ h^{-1}$ when implementing the whole set of enzymatic constraints in the model, so our enriched model predicts a similar but slightly higher uptake rate than the experimental one even if it is not explicitly constrained in the model. Therefore, this adjusted uptake rate solely limited by the enzyme availability leads to a more reduced flux for the *scyllo*-inositol production and a more realistic value in comparison to the one obtained with the original model sc_iYO844. To further support the predictive capacity of the enriched sc_ec_iYO844 + v_{pool}, we calculated the conversion efficiency from glucose to *scyllo*-inositol (the ratio between the *scyllo*-inositol synthesis flux and the glucose uptake rate) and compared it with the available bibliography. In Tanaka et al. [2017], a conversion efficiency of 0.55 with *myo*-inositol as substrate was reported, and in Michon et al. [2020], the reported value using glucose as substrate was 0.10. Notice that these values are closer to the conversion efficiencies predicted using sc_ec_iYO844 + v_{pool}, see Table 2 (Annex).

In the end, we conducted a further analysis to uncover how the glucose uptake rate and the kinetic and proteomic constraints affect the *scyllo*-inositol production.

In Figure 8, the green and black lines refer to the *scyllo*-inositol production and the glucose uptake rate, respectively, in the sc_ec_iYO844 + v_{pool} model; the blue and red lines correspond with the *scyllo*-inositol production and the glucose uptake rate, respectively, in the sc_ec_iYO844 + uptake constraint model; the orange and purple lines represent the *scyllo*-inositol production and glucose uptake rate when utilizing the sc_ec_iYO844 + v_{pool} + uptake constraint model. Here, the results indicate that when higher levels of *scyllo*-inositol are produced, which implies low growth rates, the enzymatic constraints are predominant over the glucose uptake constraint, being the enzyme availability the bottleneck for the *scyllo*-inositol synthesis. However, there is a point, when the growth rate is approximately 0.32 h^{-1}, where the glucose uptake rate constraint imposes over the enzymatic constraints. This happens exactly when the blue and green lines intersect. This can also be noted in the black line: when growth rate reaches 0.32 h^{-1}, the glucose uptake rate needs to increase because it becomes the limiting constraint instead of the kinetic and proteomic ones.

The Python script that performs the reported optimizations is available at: `https://github.com/jlazaroibanezz/scyllo-inositol-bacillus`.

8 Conclusions

Flexible Nets (FNs) can integrate naturally different enzymatic constraints in order to produce a realistic model of metabolism. In this work, two different types of enzymatic constraints were implemented in the same FNs model: the constraints that limit individually the availability of an enzyme (GECKO) and a global constraint that restricts the total amount of enzymes (sMOMENT). Among other properties such as having a simple and intuitive graphical representation, FNs offer the possibility of modeling the activation or inhibition of a certain reaction depending on the concentration of a metabolite, see Figure 2, and can approximate non-linear dynamics (Júlvez et al. [2018]), making this approach more versatile in several aspects than its recently updated analogue GECKO 3.0 (Chen et al. [2024]). Moreover, the integration of the constraints seems more natural in FNs because they directly constrain the rates of real biological processes (biochemical reactions). In contrast, GECKO 3.0 does the same but less directly and consequently, in a harder to interpret way.

An FN has been developed to study the production of *scyllo*-inositol in a genome-scale metabolic model of *Bacillus subtilis*. It has been shown that more realistic levels of *scyllo*-inositol production are computed when the model is enriched with enzymatic constraints. In particular, the model predicts that the *scyllo*-inositol production is limited by the enzyme availability and kinetics when the metabolic flux is mainly directed to maximize *scyllo*-inositol production. Nevertheless, the limiting constraint changes to the glucose uptake rate as the growth rate increases and, consequently, the *scyllo*-inositol production decreases.

9 Annex

Reaction name	**EC**	$k_{cat}[s^{-1}]$	$[E][mmol/g_{DW}]$	**Organism of kcat data**
PGI	5.3.1.9	126	$1.55 \cdot 10^{-5}$	*E. coli*
TPI	4.1.1.31	150	$1.28 \cdot 10^{-5}$	*E. coli*
GAPD_NAD	1.2.1.12	70	$5.77 \cdot 10^{-5}$	*B. subtilis*
PGK	2.7.2.3	329	$3.60 \cdot 10^{-5}$	*E. coli*
PGM	5.4.2.12	765.9	$8.85 \cdot 10^{-6}$	*B. subtilis*
ENO	4.2.1.11	130.4	$3.17 \cdot 10^{-5}$	*B. subtilis*
G6PDH	1.1.1.49	174	$8.05 \cdot 10^{-6}$	*E. coli*
CS	2.3.3.16	49	$2.51 \cdot 10^{-5}$	*B. subtilis*
ICDHy	1.1.1.42	82	$1.10 \cdot 10^{-4}$	*B. subtilis*
FUM	4.2.1.2	283.3	$7.29 \cdot 10^{-6}$	*E. coli*
MDH	1.1.1.37	177.1	$1.06 \cdot 10^{-4}$	*B. subtilis*
PTAr	2.3.1.8	651.6	$8.49 \cdot 10^{-6}$	*B. subtilis*
LDH_L	1.1.1.27	6416.6	$3.60 \cdot 10^{-6}$	*B. subtilis*
PGCDr	1.1.1.95	14.56	$1.90 \cdot 10^{-5}$	*B. subtilis*
OXADC	4.1.1.2	59	$6.21 \cdot 10^{-7}$	*B. subtilis*
MICITL	4.1.3.30	19	$6.80 \cdot 10^{-8}$	*E. coli*
OXGDC	4.1.1.71	0.2	$6.80 \cdot 10^{-8}$	*B. subtilis*
IMP	3.1.3.25	6.5	$6.80 \cdot 10^{-8}$	*E. coli*
IOLG	1.1.1.18	21.84	$6.80 \cdot 10^{-8}$	*B. subtilis*
IOLW	1.1.1.371	84.63	$6.80 \cdot 10^{-8}$	*B. subtilis*

Table 1. This table gathers data from Massaiu et al. [2019], and includes the kinetic and proteomic information of three of the enzymes participating in the *scyllo*-inositol synthesis pathway (IMP, IolG and IolW).

	Conversion efficiency	
Growth rate (h^{-1})	**sc_iYO844**	**sc_iYO844 + vpool**
0.0	0.931	0.513
0.032	0.928	0.505
0.065	0.924	0.496
0.098	0.921	0.487
0.131	0.918	0.477
0.164	0.915	0.468
0.197	0.912	0.457
0.229	0.909	0.447
0.262	0.906	0.436
0.295	0.903	0.409
0.328	0.899	0.374
0.361	0.896	0.341
0.394	0.893	0.309
0.426	0.890	0.280
0.459	0.886	0.253
0.492	0.883	0.227
0.525	0.880	0.207
0.558	0.876	0.189
0.591	0.873	0.170
0.624	0.870	0.150

Table 2. Conversion efficiency values in the sc_iYO844 and the sc_ec_iYO844 + *vpool* models depending on the fixed growth rate.

Bibliography

Pavlos Stephanos Bekiaris and Steffen Klamt. Automatic construction of metabolic models with enzyme constraints. *BMC bioinformatics*, 21:1–13, 2020.

Antje Chang, Lisa Jeske, Sandra Ulbrich, Julia Hofmann, Julia Koblitz, Ida Schomburg, Meina Neumann-Schaal, Dieter Jahn, and Dietmar Schomburg. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research*, 49(D1):D498–D508, 11 2020. ISSN 0305-1048. https://doi.org/10.1093/nar/gkaa1025. URL `https://doi.org/10.1093/nar/gkaa1025`.

Yu Chen, Johan Gustafsson, Albert Tafur Rangel, Mihail Anton, Iván Domenzain, Cheewin Kittikunapong, Feiran Li, Le Yuan, Jens Nielsen, and Eduard J Kerkhoven. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using gecko toolbox 3.0. *Nature Protocols*, pages 1–39, 2024.

Iván Domenzain, Benjamín Sánchez, Mihail Anton, Eduard J Kerkhoven, Aarón Millán-Oropeza, Céline Henry, Verena Siewers, John P Morrissey, Nikolaus Sonnenschein, and Jens Nielsen. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using gecko 2.0. *Nature communications*, 13(1):3766, 2022.

Maxime Durot, Pierre-Yves Bourguignon, and Vincent Schachter. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, 33(1):164–190, 12 2008. ISSN 0168-6445. https://doi.org/10.1111/j.1574-6976.2008.00146.x. URL `https://doi.org/10.1111/j.1574-6976.2008.00146.x`.

Ali Ebrahim, Joshua A. Lerman, Bernhard O. Palsson, and Daniel R. Hyduke. COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Systems Biology*, 7(1):74+, August 2013. ISSN 1752-0509. https://doi.org/10.1186/1752-0509-7-74. URL `http://dx.doi.org/10.1186/1752-0509-7-74`.

Anne Goelzer, Jan Muntel, Victor Chubukov, Matthieu Jules, Eric Prestel, Rolf Nölker, Mahendra Mariadassou, Stéphane Aymerich, Michael Hecker, Philippe Noirot, Dörte Becher, and Vincent Fromion. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243, 2015. ISSN 1096-7176. https://doi.org/https://doi.org/10.1016/j.ymben.2015.10.003. URL `https://www.sciencedirect.com/science/article/pii/S1096717615001317`.

Mathias Gotsmy, Florian Strobl, Florian Weiß, Petra Gruber, Barbara Kraus, Juergen Mairhofer, and Jürgen Zanghellini. Sulfate limitation increases specific plasmid dna yield and productivity in e. coli fed-batch processes. *Microbial Cell Factories*, 22(1):242, 2023.

Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models. *Genome biology*, 20:1–18, 2019.

Qingyao Huang, Damian Szklarczyk, Mingcong Wang, Milan Simonovic, and Christian von Mering. Paxdb 5.0: curated protein quantification data suggests adaptive proteome changes in yeasts. *Molecular & Cellular Proteomics*, 22(10), 2023.

J. Júlvez, D. Dikicioglu, and S. G. Oliver. Handling variability and incompleteness of biological data by flexible nets: a case study for Wilson disease. *npj Systems Biology and Applications*, 4(1):7, 1 2018.

Jorge Júlvez and Stephen G Oliver. Steady state analysis of flexible nets. *IEEE Transactions on Automatic Control*, 65(6):2510–2525, 2020. https://doi.org/10.1109/TAC.2019.2931836.

Hyun Uk Kim, Tae Yong Kim, and Sang Yup Lee. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen acinetobacter baumannii aye. *Molecular BioSystems*, 6(2):339–348, 2010.

Jorge Lázaro, Giorgio Jansen, Yiheng Yang, Mario A Torres-Acosta, Gary Lye, Stephen G Oliver, and Jorge Júlvez. Combination of genome-scale models and bioreactor dynamics to optimize the production of commodity chemicals. *Frontiers in Molecular Biosciences*, 9:855735, 2022.

Longfei Mao, Averina Nicolae, Miguel A.P. Oliveira, Feng He, Siham Hachi, and Ronan M.T. Fleming. A constraint-based modelling approach to metabolic dysfunction in parkinson's disease. *Computational and Structural Biotechnology Journal*, 13:484–491, 2015. ISSN 2001-0370. https://doi.org/https://doi.org/10.1016/j.csbj.2015.08.002. URL `https://www.sciencedirect.com/science/article/pii/S2001037015000379`.

Ilaria Massaiu, Lorenzo Pasotti, Nikolaus Sonnenschein, Erlinda Rama, Matteo Cavaletti, Paolo Magni, Cinzia Calvio, and Markus J Herrgård. Integration of enzymatic data in bacillus subtilis genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains. *Microbial cell factories*, 18:1–20, 2019.

Christophe Michon, Choong-Min Kang, Sophia Karpenko, Kosei Tanaka, Shu Ishikawa, and Ken-ichi Yoshida. A bacterial cell factory converting glucose into scyllo-inositol, a therapeutic agent for alzheimer's disease. *Communications Biology*, 3(1):93, 2020.

Ron Milo. What is the total number of protein molecules per cell volume? a call to rethink some published values. *Bioessays*, 35(12):1050–1055, 2013.

T. Murata. Petri Nets: Properties, Analysis and Applications. *Procs. of the IEEE*, 77(4):541–580, 1989.

Charles J Norsigian, Neha Pusarla, John Luke McConn, James T Yurkovich, Andreas Dräger, Bernhard O Palsson, and Zachary King. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Research*, 48(D1):D402–D406, 11 2019. ISSN 0305-1048. https://doi.org/10.1093/nar/gkz1054. URL `https://doi.org/10.1093/nar/gkz1054`.

Edward J O'brien, Joshua A Lerman, Roger L Chang, Daniel R Hyduke, and Bernhard Ø Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular systems biology*, 9 (1):693, 2013.

Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, 2010.

Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism—2011. *Molecular systems biology*, 7(1):535, 2011.

Tiago Pedreira, Christoph Elfmann, and Jörg Stülke. The current state of SubtiWiki, the database for the model organism Bacillus subtilis. *Nucleic Acids Research*, 50(D1):D875–D882, 10 2021. ISSN 0305-1048. https://doi.org/10.1093/nar/gkab943. URL `https://doi.org/10.1093/nar/gkab943`.

Ryan S Senger. Biofuel production improvement with genome-scale models: The role of cell composition. *Biotechnology journal*, 5(7):671–685, 2010.

Rahul Shaw and C Y Maurice Cheung. Integration of crop growth model and constraint-based metabolic model predicts metabolic changes over rice plant development under water-limited stress. *in silico Plants*, 3(2):diab020, 07 2021. ISSN 2517-5025. https://doi.org/10.1093/insilicoplants/diab020. URL `https://doi.org/10.1093/insilicoplants/diab020`.

M. Silva. Introducing Petri Nets. *Practice of Petri Nets in Manufacturing*, pages 1–62, Chapman & Hall, London, 1993.

Kosei Tanaka, Ayane Natsume, Shu Ishikawa, Shinji Takenaka, and Ken-ichi Yoshida. A new-generation of bacillus subtilis cell factory for further elevated scyllo-inositol production. *Microbial cell factories*, 16:1–8, 2017.

Ulrike Wittig, Maja Rey, Andreas Weidemann, Renate Kania, and Wolfgang Müller. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Research*, 46(D1):D656–D660, 10 2017. ISSN 0305-1048. https://doi.org/10.1093/nar/gkx1065. URL `https://doi.org/10.1093/nar/gkx1065`.

Francisco Zorrilla, Filip Buric, Kiran R Patil, and Aleksej Zelezniak. metagem: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Research*, 49(21):e126–e126, 2021.