On the computation of the minimum set of reactions for optimal growth in constraint-based models

Alexandru Oarga and Jorge Júlvez

***Abstract*— Technical advances in sequencing have allowed the reconstruction of metabolic models of multiple microorganisms, which have proven useful in advancing metabolic engineering and drug discovery. Optimization methods have provided a way to accurately predict flux phenotypes of various unicellular organisms and their response to gene knockouts. Despite the broad application of these methods, the role that different biochemical reactions have in providing robustness and flexibility has not been studied extensively. In this work, a method is presented to identify those sets of reactions that are essential for growth and those that are redundant and therefore account for the robustness of metabolism. The problem of computing a minimum set of reactions that can produce optimum growth is formally stated. It is proven that such a problem is NP-complete and a technique to reduce the search space of the problem is proposed. The presented approach is experimentally applied in various genome-scale models. The contribution of this work is to provide insight into the roles that different reactions play in the production of growth and to propose methods that can be directly applied in model curation and analysis.**

I. INTRODUCTION

Metabolism is the set of basic life processes that take place in the cell and can be represented as a biomolecular network of chemical reactions. As of 2019, genome-scale models (GEM) of metabolism have been reconstructed for more than 6000 organisms including bacteria, archaea, and eukaryotes. GEMs have proven useful in a wide range of applications, such as expanding knowledge on microorganisms microbial engineering , and drug discovery [1].

One of the main procedures regarding GEMs is the estimation of the optimum growth, which is especially significant as it helps in essentiality identification [2]. The computation of the minimum metabolism is also a recurrent problem in reconstruction and genome design [3], where the produced results are expected to sustain growth. Most of the proposed methods usually employ mixed-integer problems [4]. In areas such as gap-filling cost-effective alternatives also exist such as approximations [5] and topology based methods [6].

Despite the widespread usage of optimisation methods, these cannot help in the systematic identification of reactions, which hampers our ability to analyse and control such systems. Approaches such as elementary modes [7] are impractical due to their combinatorial explosion, and hence, new methods are required [8].

This work was supported by the Spanish Ministry of Science and Innovation [ref. DAMOCLES-PID2020-113969RB-I00] and by a grant from The Engineering Research Institute of Aragon (I3A).

A. Oarga and J. Júlvez are with the Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain, aoarga@unizar.es, julvez@unizar.es

In this work, we characterise formally essential reactions and provide a method that addresses the minimum metabolism problem. The document is organised as follows: Section II defines constraint-based models and methods for growth estimation; in Section III a computational procedure for the identification of minimal metabolism for optimum growth is proposed and its complexity is characterised; Section IV presents the results obtained on various genome-scale models; the main conclusions are drawn in Section V.

II. PRELIMINARY CONCEPTS

This section introduces the preliminary definitions and concepts used in the paper.

Definition II.1. A *constraint-based model* [9] is a tuple $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ L,\ U\}$ where $\mathcal{R}$ is a set of reactions, $\mathcal{M}$ is a set of metabolites, $\mathcal{S} \in \mathbb{R}^{|\mathcal{M}|\times|\mathcal{R}|}$ is the stoichiometric matrix, and $L, U \in \mathbb{R}^{|\mathcal{R}|}$ are lower and upper flux bounds of the reactions.

Without loss of generality, it is assumed that $L[r] \leq U[r] \ \ \forall r \in \mathcal{R}$.

Reactions are associated with a set of reactant metabolites and a set of product metabolites (one of these two sets can be empty). For example, the reaction $r{:}A \rightarrow 2B$ has a reactant metabolite A, and a product metabolite B with stoichiometric weight 2, i.e. two molecules of type B are produced per each molecule of type A that is consumed by r. The stoichiometric matrix $\mathcal{S}$ accounts for all the stoichiometric weights of the reactions, i.e. $S[m, r]$ is the stoichiometric weight of metabolite $m \in \mathcal{M}$ for reaction $r \in \mathcal{R}$.

Constraint-based models can be graphically represented as Petri nets, with metabolites depicted as places and reactions depicted as rectangles [10]. The presence of an arc from a place(transition) to a transition(place) means that the place is a reactant(product) of the reaction modelled by the transition. The weights of the arcs of the Petri net account for the stoichiometry of the constraint- based model. Notice that the stoichiometric matrix of the model and the incidence matrix of the Petri net coincide.

Example II.1. The Petri net in Figure 1 represents a simple constraint-based model with 10 reactions and 6 metabolites, where transition r_5 models the reaction: $r5$: $2m_3 \rightarrow m_4$.

The flux bounds L, U of the model can be used to classify reactions as *dead*, *reversible*, or *non-reversible*:

Definition II.2. A reaction $r \in \mathcal{R}$ is dead if $L[r]{=}U[r]{=}0$.

Definition II.3. A reaction $r \in \mathcal{R}$ is reversible if $L[r]<0<U[r]$.

Definition II.4. A reaction $r \in \mathcal{R}$ is non-reversible if r is not dead and r is not reversible.

The sets of dead, reversible and non-reversible reactions are denoted DR, RR and NR respectively.

Flux Balance Analysis (FBA) [11] is a mathematical procedure for the estimation of steady state fluxes in constraint-based models. FBA can be used, for instance, to predict the maximum growth rate of an organism. Let $v \in \mathbb{R}^{|\mathcal{R}|}$ be the vector of fluxes of reactions and $v[r]$ denote the flux of reaction r. At steady state, it holds that $S \cdot v = 0$, where S is the stoichiometric matrix. Thus, the linear programming problem (LPP) for FBA is:

$$\begin{aligned} &\max\ z \cdot v \\ &st.\ \ S \cdot v = 0 \\ &\qquad L \leq v \leq U \end{aligned} \tag{1}$$

where $z \in \mathbb{R}^{|\mathcal{R}|}$ expresses the objective function.

Let r_g be the reaction that models growth (or biomass production). Without loss of generality, it will be assumed that $L[r_g] \geq 0$. A theoretical optimum for the growth rate can be obtained by the following FBA:

$$\begin{aligned} &\max\ v[r_g] \\ &st.\ \ S \cdot v = 0 \\ &\qquad L \leq v \leq U \end{aligned} \tag{2}$$

The maximum $v[r_g]$ obtained by the above LPP (2) will be denoted μ_{max}.

Two computational methods that are related to FBA are Parsimonious Flux Balance Analysis (pFBA) and Flux Variability Analysis (FVA). pFBA computes a flux distribution that produces the optimum growth rate while minimising the overall sum of fluxes [12]. On the other hand, FVA [13] computes the minimum and maximum fluxes of reactions that are compatible with a given state. For instance, FVA can be used to compute the fluxes that are compatible with a growth $\gamma \cdot \mu_{max}$ where $\gamma \in [0, 1]$. Such an FVA can be computed by solving two independent LPPs per reaction $r \in \mathcal{R}$. One programming problem maximises $v[r]$, and the other minimises $v[r]$. The constraints of both problems are the same: the steady state condition $S \cdot v = 0$, the flux bounds $L \leq v \leq U$, and the constraint $\gamma \cdot \mu_{max} \leq v[r_g]$. The two LPPs for a given $r \in \mathcal{R}$ can be expressed as:

$$\begin{aligned} &\max / \min\ v[r] \\ &st.\ \ S \cdot v = 0 \\ &\qquad L \leq v \leq U \\ &\qquad \gamma \cdot \mu_{max} \leq v[r_g] \end{aligned} \tag{3}$$

Let $lb_\gamma, ub_\gamma \in \mathbb{R}^{|\mathcal{R}|}$ be the result of computing FVA (3) on a constraint-based model $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ L,\ U\}$ for a given γ, i.e. $lb_\gamma[r]$ and $ub_\gamma[r]$ are the minimum and maximum flux given by FVA for reaction r. If the flux bounds L, U of the constrained-based model are replaced by lb_γ, ub_γ, a new constraint-based model, $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ lb_\gamma,\ ub_\gamma\}$, with more restrictive (and realistic) flux bounds is obtained. Given that the flux bounds of the model have changed, new sets of *growth dependent* dead, reversible, and non-reversible reactions must be considered. These sets are denoted DR_γ, RR_γ and NR_γ respectively.

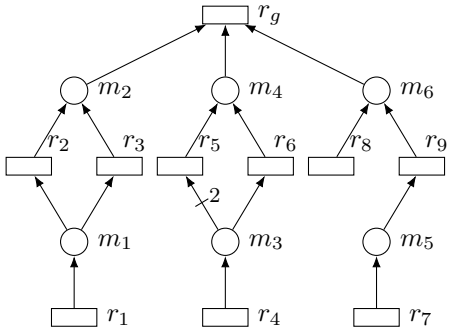

Fig. 1. Example Petri net modelling a constraint-based model.

III. REACTIONS REQUIRED FOR GROWTH

A. Essential reactions

A reaction is said to be essential if it is required by the organism to grow. In other words, the deletion of an essential reaction implies null growth. Consequently, these reactions have the potential to cause the death of the modelled organism.

Definition III.1. [2] A reaction $r \in \mathcal{R}$ is an *essential reaction* if the solution of the following LPP:

$$\begin{aligned} &\max\ v[r_g] \\ &st.\ \ S \cdot v = 0 \\ &\qquad L \leq v \leq U \\ &\qquad v[r] = 0 \end{aligned} \tag{4}$$

is equal to 0 or the LPP is infeasible.

The set of essential reactions, which is denoted ER, can be computed straightforwardly by solving (4) for each $r \in \mathcal{R}$.

Example III.1. In the Petri net in Figure 1, where r_g models growth and $L[r]=0, U[r]=100\ \forall r \in \mathcal{R}$, reactions r_1, r_4 are essential reactions (i.e. $ER = \{r_1, r_4\}$). This is because, if the flux of any of these reactions is set to 0, it is not possible to produce metabolites m_2 and m_4 respectively, which are required for the growth reaction r_g.

Similarly to essential reactions, growth dependent essential reactions are those reactions that are necessary to produce a certain growth on the model.

Definition III.2. Let μ_{max} be the solution of the LPP (2) and $\gamma \in [0, 1]$. Reaction $r \in \mathcal{R}$ is *essential for growth* $\gamma \cdot \mu_{max}$ if the solution of LPP (4) is lower than $\gamma \cdot \mu_{max}$ or the LPP is infeasible.

The set of reactions essential for growth $\gamma \cdot \mu_{max}$ will be denoted ER_γ. This set can be computed straightforwardly by solving LPP (4) for each reaction.

Special attention will be paid to the set of reactions ER_1, as it will consist of those reactions that are necessary to produce the optimum growth of the model. This set will be named *essential reactions for optimum growth* (EROG).

Example III.2. In the Petri net of Figure 1, reactions r_1, r_4, r_6 are essential reactions for optimum growth (i.e. $EROG = \{r_1, r_4, r_6\}$). Reactions r_1, r_4 are in ER and therefore in $EROG$. Reaction r_6 is in $EROG$ because, if its flux is set to 0, metabolite m_4, which is required for growth reaction r_g, is produced through a less optimal reaction (i.e. r_5) and the model is not able to produce the optimum growth.

Notice that in addition to the reactions in EROG, other reactions might be required to produce optimum growth. Let $||v||$ denote the support of $v \in \mathbb{R}^{|\mathcal{R}|}$, i.e., $||v|| = \{r \in \mathcal{R} \ |v[r] \neq 0\}$. The *set of reactions for optimum growth* is defined as follows:

Definition III.3. A set of reactions F is *a set of reactions for optimum growth* (ROG) if $\exists v \in \mathbb{R}^{|\mathcal{R}|}$ such that $S \cdot v = 0$, $L < v < U$, $v[r_g] = \mu_{max}$ and $||v|| = F$.

Notice that $EROG \subseteq ROG$. Moreover, since there can be several flux distributions that produce optimum growth, ROG might not be unique. Given that the reactions in a ROG are sufficient to produce optimum growth, the model can produce the optimum growth even if all the reactions in $\mathcal{R} - ROG$ are inhibited.

B. Minimum set of reactions for optimum growth

This section focuses on the problem of computing a minimum ROG. Let $\mathcal{O}$ be the set of all ROG sets of a model.

Definition III.4. A set of reactions $O_i \in \mathcal{O}$ is a *minimum set of reactions for optimum growth* (MROG) if $|O_i| \leq |O_j| \ \forall O_j \in \mathcal{O}$.

Similarly to ROG, the set $MROG$ might not be unique.

Example III.3. The model in Figure 1 has 2 feasible $MROG$ sets: $\{r_1, r_2, r_4, r_6, r_8\}$ and $\{r_1, r_3, r_4, r_6, r_8\}$. Metabolite m_2 is necessary for growth and can be equally produced by reactions r_1, r_2 or r_1, r_3. Metabolite m_6 can be produced by various reactions, however, the minimum number of reactions required to produce it optimally is 1 (i.e. r_8), thus r_8 is in $MROG$. Finally, reactions r_1, r_4, r_6 are in $EROG$ and therefore are present in any $MROG$ set.

The problem to compute an MROG can be stated as:

Problem III.1. Given a constraint-based model $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ L,\ U\}$, and an objective reaction $r_g \in \mathcal{R}$, the minimum set of reactions for optimum growth problem ($MROGP$) is the problem of finding a minimum set of reactions for optimum growth $MROG$.

It will be shown that $MROGP$ can be solved by a Mixed-Integer Linear Programming problem (MILP) where the objective is to minimise the number of reactions required for optimum growth. We will make use of a vector of *initial* fluxes, $w \in \mathbb{R}^{|\mathcal{R}|}$, and a vector of binary variables, $\delta \in \{0,1\}^{|\mathcal{R}|}$, that indicates which fluxes are canceled out, i.e. $\delta[r] = 0$ implies that there is no flux through r regardless of $w[r]$. Thus, the actual flux of a given reaction, r, is $v[r] = \delta[r] \cdot w[r]$. Let us consider the following programming problem:

$$\begin{aligned} \min \ & \sum_{r \in \mathcal{R}} \delta[r] \\ st. \ & S \cdot v = 0 \\ & v[r] = \delta[r] \cdot w[r] \ \forall \ r \ \mathcal{R} \\ & L \leq w \leq U \\ & v[r_g] = \mu_{max} \end{aligned} \tag{5}$$

Given that the number of reactions with non-null flux is minimized, the support of a vector v that is a solution to the programming problem (5) is an $MROG$.

Equation $v[r]{=}\delta[r]{\cdot}w[r]$ makes the problem (5) non-linear. Such an equation is equivalent to the following inequalities:

$$\begin{aligned} & v[r] \leq U[r]{\cdot}\delta[r] \ \forall r \in \mathcal{R} \\ & v[r] \geq L[r]{\cdot}\delta[r] \ \forall r \in \mathcal{R} \\ & v[r] \leq w[r] - L[r]{\cdot}(1 - \delta[r]) \ \forall r \in \mathcal{R} \\ & v[r] \geq w[r] - U[r]{\cdot}(1 - \delta[r]) \ \forall r \in \mathcal{R} \end{aligned} \tag{6}$$

Thus, the replacement of $v[r]{=}\delta[r] \cdot w[r]$ in (5) by the above inequalities results in a MILP which solves MROGP.

C. Computational complexity

This section proves that a solution for $MROGP$ can not be found in polynomial time. Let us first restate the problem as a decision problem:

Problem III.2. Given a constraint-based model $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ L,\ U\}$, an objective reaction $r_g \in \mathcal{R}$, and integer k, the set of reactions for optimum growth problem ($ROGP$) is the problem of determining whether there exists a ROG set O_i with $|O_i| \leq k$.

We will prove that $ROGP$ is NP-complete. First, it is proved that this problem is in NP.

Lemma III.1. *ROGP is in NP.*

Proof. Given a set of reactions $O_i \subseteq \mathcal{R}$, we can verify that the set is a ROG set for a constraint-based model $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ L,\ U\}$ with objective reaction $r_g \in \mathcal{R}$, by removing all reactions not in O_i from the model and solving the LPP in (2). If the growth obtained is equal to μ_{max} and $|O_i| \leq k$, then the set O_i is a ROG set with size at most k. Since LPPs can be solved in polynomial time, $ROGP$ is in NP. □

Let us now prove that $ROGP$ is NP-hard by reducing the vertex cover problem [14] to $ROGP$. The vertex cover problem is defined as:

Problem III.3. Given an undirected graph $G = (V, E)$, a vertex cover V' is a subset of V such that $uv \in E \rightarrow u \in$

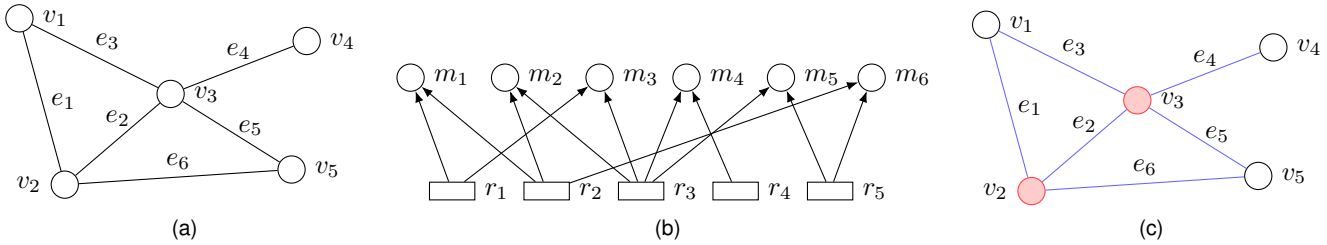

Fig. 2. a) Undirected graph with 5 vertices and 6 edges. b) Net of source reactions and metabolites resulted from transforming the undirected graph in Figure 2a. c) Undirected graph from Figure 2a with a vertex cover of size $k = 2$ highlighted in red.

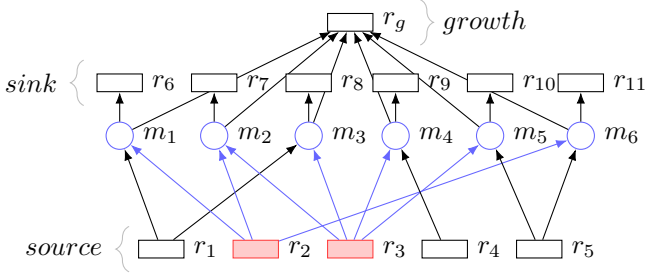

Fig. 3. Constraint-based model resulting from transforming the undirected graph in Figure 2a. Input reactions mapped from the vertex cover of Figure 2c are highlighted in red. Arcs and metabolites highlighted in blue show that reactions r_2 and r_3 are able to produce all the metabolites of the model.

$V' \vee v \in V'$. The vertex cover problem is the problem of determining whether there exists a vertex cover of size at most k.

Lemma III.2. *ROGP is NP-hard.*

Proof. Let us reduce an instance of the vertex cover problem, consisting of an undirected graph $G = (V, E)$, to a $ROGP$. First, the undirected graph is transformed into a bipartite graph (represented graphically as a Petri net) of reactions and metabolites as follows:

- For each vertex $v_i \in V$ create a *source* reaction r_i with $L[r_i]$=0, $U[r_i]$=2.
- For each edge $e_i \in E$ create a metabolite m_i.
- For each adjacent edge e_i of each vertex v_j make the corresponding metabolite m_i a product of the corresponding reaction r_j.

Figure 2b shows an example of a network $\mathcal{R}$=$\{r_1, r_2, r_3, r_4, r_5\}$ and $\mathcal{M}$=$\{m_1, m_2, m_3, m_4, m_5, m_6\}$, resulting from the transformation of the undirected graph in Figure 2a with V=$\{v_1, v_2, v_3, v_4, v_5\}$ and E=$\{e_1, e_2, e_3, e_4, e_5, e_6\}$.

In addition to the previous transformations, the following ones are also performed:

- For each metabolite(edge) $m_i \in \mathcal{M}$ create a *sink* reaction r_j with $|V|$<j≤$|V|+|E|$ with $L[r_j]$=1, $U[r_j]$=$|V|$.
- Add an objective reaction r_g with $L[r_g]$=0, $U[r_g]$=1 that consumes all metabolites $m_i \in \mathcal{M}$.

Figure 3 shows the final constraint-based model resulting from applying the described transformation to the graph in Figure 2a.

In the obtained constraint-based model, each source reaction r_i will act as an input to the network and sink reactions will balance the potential excess of produced metabolite. Notice that in order to achieve the optimum growth, all the metabolites must be produced, and as long as $|E| > 0$, this model will always be able to produce the maximum growth (i.e. μ_{max}=1) with a certain $v \in \mathbb{R}^{|\mathcal{R}|}$ obtained by the LPP in (2).

It can be seen that any ROG set will have the following number of reactions: all sink reactions (the number of sink reactions is $|E|$) since all sink reactions are constrained to have non-null flux; the growth reaction; and a number k of reactions, with $1 \leq k \leq |V|$, that correspond to a set of reactions necessary to produce all the metabolites in the model. To summarise, any ROG set will have a size equal to: k (source reactions) + $|E|$ (sink reactions) + 1 (growth reaction). The set of k source reactions will be used to derive a solution for the vertex cover problem. Let us prove the following claim: a vertex cover of size k exists if and only if a ROG set of size $k + |E| + 1$ exists. We proceed by proving both directions of the claim:

1) If a ROG set of size $k + |E| + 1$ exists, then a vertex cover of size k exists: Let $R_{in} \subseteq \mathcal{R}$ be the set of k source reactions of a given ROG set. A vertex cover $V' \subseteq V$ of the graph G can be built as follows: $v_i \in V'$ if $r_i \in R_{in}$. Here, a source reaction producing metabolites is considered equivalent to a vertex covering its adjacent edges. If we consider any source reaction $r_i \in R_{in}$, it produces a set of metabolites $m_j, ..., m_k \in \mathcal{M}$ that is equivalent to the set of edges $e_j, ..., e_k \in E$ that would be covered by the corresponding vertex $v_i \in V$. Since the k source reactions produce all metabolites in the model, it is guaranteed that the resulting vertex set V' covers all edges of the graph, thus making V' a vertex cover.
2) If a vertex cover of size k exists, then a ROG set of size $k + |E| + 1$ exists: let $V' \subseteq V$ be a vertex cover with $|V'| = k$. Since all edges of the graph are covered by k vertices and given the equivalence between source reactions and vertices, it is guaranteed that the corresponding k source reactions of the model are sufficient to produce all metabolites of the model and achieve optimum growth, hence a ROG of size at most $k + |E| + 1$ exists.

Figure 2c shows highlighted in red a vertex cover V' of

size $k = 2$ (i.e. $V' = \{v_2, v_3\}$) of the undirected graph in Figure 2a. Figure 3 shows that the corresponding source reactions r_2, r_3 (highlighted in red) are able to produce all metabolites of the model (highlighted in blue). Therefore, there exists a ROG set of size at least $k + |E| + 1$ with $k = 2$ and $|E| = 6$.

Clearly, the same reasoning can be applied reversely to obtain a vertex cover from the ROG set. Given the k source reactions of the ROG set (e.g. r_2, r_3 in Figure 3), the corresponding vertices are guaranteed to be a vertex cover (e.g. v_2, v_3 in Figure 2c). □

The following theorem is derived straightforwardly from Lemmas III.1 and III.2.

Theorem III.3. *Reactions for Optimum Growth Problem (ROGP) is in NP-complete.*

D. Problem size reduction

Given that $ROGP$ is NP-complete, we can not expect to find a solution in polynomial time, and consequently, we can not expect to solve $MROGP$ in polynomial time either. It is possible, however, to achieve a reduction in the size of $MROGP$ and along with that a potential reduction in the execution time required to solve it.

As mentioned previously, for any set ROG it holds that $EROG \subseteq ROG$. Consequently, the space to search for reactions in $MROG$ can be reduced from $\mathcal{R}$ to $\mathcal{R} - EROG$. Similarly, the set DR_1 contains all reactions whose only compatible flux with optimum growth is the null flux, hence, this set can also be ignored in the search for $MROG$. The search space of the problem is then reduced from $|\mathcal{R}|$ to $|\mathcal{R} - EROG - DR_1|$.

Note that the MILP in (5) had $|\mathcal{R}|$ binary variables, and the above reasoning reduces the number of binary variables to $|\mathcal{R} - EROG - DR_1|$. The resulting *reduced* MILP is:

$$\begin{aligned} \min \ & \sum_{r \in F} \delta[r] \\ st. \ & S \cdot v = 0 \\ & v[r] = \delta[r] \cdot w[r] \ \forall r \in F \\ & v[r] = w[r] \ \forall r \in EROG \\ & L \leq w \leq U \\ & v[r_g] = \mu_{max} \end{aligned} \tag{7}$$

where $F = \mathcal{R} - EROG - DR_1$.

IV. RESULTS

This section presents the results of computing the sets defined in this work on a total of 30 constraint-based models obtained from the BioModels repository [15]. The size of each set can be found in Table I.

In the table, it can be seen that the coverage of the set $EROG$ varies considerably between models. Size ranges from 8.18% of MODEL1507180006 reactions (32 of 391 reactions), to 60.96% of all MODEL1507180052 reactions (214 of 351). It can also be seen that larger models do not directly imply larger $EROG$ sets.

Let us now consider the size of the dead reaction sets in column DR_1. Contrary to what happened with the $EROG$ set, larger models tend to have a larger number of dead reactions. The coverage of this set ranges from 28.47% of all reactions in smaller models such as MODEL1710040000 to 82% in larger models such as MODEL1507180054.

Section III-D presented how the problem of finding a minimum set ROG could be reduced to minimisation in the set of reactions $\mathcal{R} - EROG - DR_1$. Recall that this size corresponds with the numbers of binary variables of MILP (7). The size of this set has also been included in Table I. In the corresponding column, it can be seen that the set size ranges from 21 reactions in MODEL1507180052 to 804 in MODEL1212060001. If we consider, for instance, the largest of the selected models (i.e. model MODEL1507180017), it can be seen that out of 2546 reactions, set $\mathcal{R} - EROG - DR_1$ has only 84 reactions. This means that the number of binary variables is reduced from 2546 with the MILP (5) to only 84 in MILP (7).

The MILP in (7) was solved for each model. The solver used approximates the solution by linearising all variables in the problem, hence reducing significantly the execution time (maximum wall clock for MILP resolution is 223.66s with model MODEL1507180058). The downside is that the solution is not guaranteed to be optimal. Column $|MROG^*|$ reports the sizes of the computed minimum sets for optimum growth. To compare the results, the table includes sizes of sets ROG computed with FBA and pFBA. $|MROG^*|$ is lower than the sizes of $|ROG|$ computed by FBA and pFBA in 24 models (out of 30). The largest difference is found in model MODEL1507180015 where $MROG^*$ (682 reactions) has 37 reactions less than the ROG computed with pFBA (719 reactions). Note that 2 models (MODEL1507180070, MODEL1507180017) have suboptimal solutions, ($MROG^*$ is larger than the other ROG sets).

The manipulation of the constraint-based models presented and FBA, FVA computation was performed by the Python toolbox COBRApy [16]. The presented MILP for MROG computation was implemented using Pyomo language [17] and solved by the commercial solver Gurobi Optimizer 9.1.2. The maximum wall clock time was 252.48s to compute the results of model MODEL1507180058 in an Intel Core i5-9300H CPU @ 2.40GHz × 8.

V. CONCLUSIONS

This work provided methods for the identification of the different types of reactions involved in the production of growth. Taking into account only the growth-related information available in the model, reactions that are indispensable for the survival of the organism have been identified. Such essential reactions are appealing pharmacological targets in the fight against pathogens. We also computed those reactions that provide redundancy and thus contribute to a robust metabolism. It has also been shown that optimum growth itself involves a combination of the above two types of reactions. All this helps understand how metabolism works in growth production and breaks the black-box conception

TABLE I

SIZES OF REACTION SETS COMPUTED ON MULTIPLE CONSTRAINT-BASED MODELS.

ID	$\|\mathcal{R}\|$	$\|\mathcal{M}\|$	$\|EROG\|$	$\|\mathcal{R} - EROG - DR_1\|$	$\|DR_1\|$	$\|RR_1\|$	$\|NR_1\|$	$\|ROG\|$ (FBA)	$\|ROG\|$ (pFBA)	$\|MROG^*\|$
MODEL1507180052	351	346	**214 (60,96%)**	**21**	116 (33,04%)	4	231	221	222	221 (62,96%)
MODEL1507180006	391	371	**32 (8,18%)**	55	304 (77,74%)	9	78	41	40	40 (10,23%)
MODEL1106080000	469	342	61 (13,00%)	52	356 (75,90%)	13	100	78	74	67 (14,28%)
MODEL1507180007	554	485	305 (55,05%)	41	208 (37,54%)	6	340	314	313	313 (56,49%)
MODEL1507180030	560	479	251 (44,82%)	39	270 (48,21%)	12	278	270	267	263 (46,96%)
MODEL1507180048	645	565	251 (38,91%)	41	353 (54,72%)	7	285	274	273	269 (41,70%)
MODEL1507180070	743	655	264 (35,53%)	169	310 (41,72%)	13	420	300	299	300 (40,37%)
MODEL1710040000	748	737	285 (38,10%)	250	**213 (28,47%)**	69	466	387	389	359 (47,99%)
MODEL1507180024	832	790	369 (44,35%)	90	373 (44,83%)	21	438	406	406	398 (47,83%)
MODEL1507180036	870	713	259 (29,77%)	252	359 (41,26%)	51	460	347	339	324 (37,24%)
MODEL1507180021	900	688	416 (46,22%)	57	427 (47,44%)	20	453	435	435	434 (48,22%)
MODEL1507180044	948	892	314 (33,12%)	47	587 (61,91%)	5	356	343	338	332 (35,02%)
MODEL1507180049	971	496	135 (13,90%)	393	443 (45,62%)	140	388	248	237	229 (23,58%)
MODEL1507180068	1056	911	284 (26,89%)	259	513 (48,57%)	71	472	357	358	344 (32,57%)
MODEL1507180060	1075	761	279 (25,95%)	85	711 (66,13%)	8	356	302	301	299 (27,81%)
MODEL1507180059	1112	1101	222 (19,96%)	278	612 (55,03%)	142	358	366	353	315 (28,32%)
MODEL1507180013	1245	987	325 (26,10%)	114	806 (64,73%)	49	390	372	368	361 (28,99%)
MODEL1507180058	1285	943	150 (11,67%)	494	641 (49,88%)	97	547	279	264	242 (18,83%)
MODEL1507180022	1333	1243	239 (17,92%)	251	843 (63,24%)	63	427	341	322	318 (23,85%)
MODEL1507180012	1379	796	214 (15,51%)	515	650 (47,13%)	227	502	328	322	321 (23,27%)
MODEL1507180027	1462	1253	285 (19,49%)	70	1107 (75,71%)	11	344	326	321	317 (21,68%)
MODEL1507180011	1576	1913	294 (18,65%)	54	1228 (77,91%)	17	331	326	326	317 (20,11%)
MODEL1507180033	1577	1228	255 (16,16%)	141	1181 (74,88%)	47	349	303	302	300 (19,02%)
MODEL1507180015	1681	1381	480 (28,55%)	521	680 (40,45%)	334	667	**740**	**719**	**682 (40,57%)**
MODEL1507180064	1785	2087	366 (20,50%)	18	1401 (78,48%)	0	384	373	372	372 (20,84%)
MODEL1212060001	1845	1008	272 (14,74%)	**804**	769 (41,68%)	0	1076	324	308	306 (16,58%)
MODEL1507180054	2262	1658	279 (12,33%)	128	**1855 (82,00%)**	10	397	326	321	317 (14,01%)
MODEL1105030000	2378	1669	374 (15,72%)	114	1890 (79,47%)	10	478	410	405	399 (16,77%)
MODEL1507180010	2477	1747	362 (14,61%)	124	1991 (80,37%)	16	470	414	409	397 (16,02%)
MODEL1507180017	2546	1802	479 (18,81%)	**84**	1998 (78,47%)	13	535	486	483	498 (19,56%)

of growth in constraint-based models. Moreover, it has been shown that finding a minimum set of reactions that supports optimal growth is computationally complex and a method to optimise this procedure has been proposed.

REFERENCES

[1] C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee, "Current status and applications of genome-scale metabolic models," *Genome Biology*, vol. 20, no. 1, pp. 1–18, jun 2019.

[2] J. Oyelade, I. Isewon, E. Uwoghiren, O. Aromolaran, and O. Oladipupo, "In Silico Knockout Screening of Plasmodium falciparum Reactions and Prediction of Novel Essential Reactions by Analysing the Metabolic Network," *BioMed Research International*, vol. 2018, 2018.

[3] L. Wang and C. D. Maranas, "Mingenome: an in silico top-down approach for the synthesis of minimized genomes," *ACS synthetic biology*, vol. 7, no. 2, pp. 462–473, 2018.

[4] J. L. Reed, T. R. Patel, K. H. Chen, A. R. Joyce, M. K. Applebee, C. D. Herring, O. T. Bui, E. M. Knight, S. S. Fong, and B. O. Palsson, "Systems approach to refining genome annotation," *Proceedings of the National Academy of Sciences*, vol. 103, no. 46, pp. 17 480–17 484, 2006.

[5] N. Vlassis, M. P. Pacheco, and T. Sauter, "Fast reconstruction of compact context-specific metabolic network models," *PLoS Comput Biol*, vol. 10, no. 1, p. e1003424, 2014.

[6] S. Prigent, C. Frioux, S. M. Dittami, S. Thiele, A. Larhlimi, G. Collet, F. Gutknecht, J. Got, D. Eveillard, J. Bourdon *et al.*, "Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks," *PLoS computational biology*, vol. 13, no. 1, p. e1005276, 2017.

[7] J.-M. Schwartz, C. Gaugain, J. C. Nacher, A. de Daruvar, and M. Kanehisa, "Observing metabolic functions at the genome scale," *Genome biology*, vol. 8, no. 6, pp. 1–17, 2007.

[8] A. von Kamp and S. Klamt, "Enumeration of smallest intervention strategies in genome-scale metabolic networks," *PLoS computational biology*, vol. 10, no. 1, p. e1003378, 2014.

[9] A. Varma and B. Ø. Palsson, "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use," *Nature Biotechnology*, vol. 12, no. 10, pp. 994–998, Oct. 1994.

[10] M. Heiner, D. Gilbert, and R. Donaldson, "Petri nets for systems and synthetic biology," in *Formal Methods for Computational Systems Biology*, M. Bernardo, P. Degano, and G. Zavattaro, Eds., Springer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 215–264.

[11] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?" *Nature Biotechnology 2010 28:3*, vol. 28, no. 3, pp. 245–248, mar 2010.

[12] N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer *et al.*, "Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models," *Molecular systems biology*, vol. 6, no. 1, p. 390, 2010.

[13] R. Mahadevan and C. H. Schilling, "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models," *Metabolic Engineering*, vol. 5, no. 4, pp. 264–276, oct 2003.

[14] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of computer computations*. Springer, 1972, pp. 85–103.

[15] R. S. Malik-Sheriff, M. Glont, T. V. Nguyen, K. Tiwari, M. G. Roberts, A. Xavier, M. T. Vu, J. Men, M. Maire, S. Kananathan *et al.*, "Biomodels—15 years of sharing computational models in life science," *Nucleic acids research*, vol. 48, no. D1, pp. D407–D415, 2020.

[16] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, "Cobrapy: constraints-based reconstruction and analysis for python," *BMC systems biology*, vol. 7, no. 1, pp. 1–6, 2013.

[17] M. L. Bynum, G. A. Hackebeil, W. E. Hart, C. D. Laird, B. L. Nicholson, J. D. Siirola, J.-P. Watson, and D. L. Woodruff, *Pyomo–optimization modeling in python*, 3rd ed. Springer Science & Business Media, 2021, vol. 67.