Growth Dependent Computation of Chokepoints in Metabolic Networks*

Alexandru Oarga[1][0000−0002−7271−733X], Bridget Bannerman[2][0000−0002−5746−8283], and Jorge Júlvez[1][0000−0002−7093−228X]

[1] Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain {718123,julvez}@unizar.es

[2] Department of Medicine, University of Cambridge, Cambridge, UK bpc28@cam.ac.uk

Abstract. Bacterial infections are among the major causes of mortality in the world. Despite the social and economical burden produced by bacteria, the number of new drugs to combat them increases very slowly due to the cost and time to develop them. Thus, innovative approaches to identify efficiently drug targets are required. In the absence of genetic information, chokepoint reactions represent appealing drug targets since their inhibition might involve an important metabolic damage. In contrast to the standard definition of chokepoints, which is purely structural, this paper makes use of the dynamical information of the model to compute chokepoints. This novel approach can provide a more realistic set of chokepoints. The dependence of the number of chokepoints on the growth rate is assessed on a number of metabolic networks. A software tool has been implemented to facilitate the computation of growth dependent chokepoints by the practitioners.

Keywords: Chokepoint Reactions · Metabolic Networks · Petri Nets · Flux Balance Analysis

1 Introduction

Diseases caused by bacteria are one of the main causes of mortality in both developed and in-development countries. According to the World Health Organisation (WHO) in 2016, tuberculosis was the tenth cause of death worldwide which makes its pathogen, *Mycobacterium tuberculosis*, the infectious agent with the higher caused mortality. Moreover, upper respiratory system's diseases caused by microorganisms, like virus and bacteria, were the fourth cause of mortality [20]. In 2010 pneumonia was the leading cause of child mortality causing nearly 1.4 million deaths among children younger than 5 years of age [7].

Despite the high mortality caused by bacteria the development of new antibiotics is slow and challenging. Furthermore, bacteria have evolved complex

* This work was supported by the Spanish Ministry of Science, Innovation and Universities [ref. Medrese-RTI2018-098543-B-I00], and by the Medical Research Council, UK, MR/N501864/1.

mechanisms which make them difficult to fight. Thus, there is an urgent need to design novel methods for the development of new drugs. A promising possibility is to consider basic cellular processes as targets for antibiotic development [12].

Metabolism is the set of basic life processes that take place in the cell, and it is the means by which cells can maintain life and grow from their environment. The metabolism of a cell can be represented by a *metabolic network* that accounts for all the metabolic reactions that take place in the cell. A possible strategy for drug discovery is to find and damage critical vulnerabilities of the metabolic network that could stop the growth and replication of the bacteria.

Metabolism as a target has been proven to be an interesting approach in other areas like oncology [9] or viral diseases [10]. A number of methods have been proposed in order to find vulnerabilities in the metabolism that may lead to therapeutic results. Some of these methods consider topological properties of the metabolism with the purpose of finding possible critical spots, as for example: determine the importance of a metabolite based on the k-shortest paths between metabolites [15], or consider the inter-reactions dependence to find out how much influence a reaction has on metabolism [16]. Other methods focus on the genetic information associated with the metabolism, and compute, for instance, the set of genes that are essential for the survival of the cell [21].

Although a number of genome-scale models (GEMs) have been developed recently, most of them just account for the stoichiometry of their reactions and lack genetic information. This is usually the case in GEMs of bacteria. This dearth of data hampers the analysis of models, namely those based on gene essentiality, and calls for the design of computational methods that exploit as much as possible the available biological information. Here, we focus on the computation of chokepoints in metabolic networks [18], where a chokepoint is a reaction that is either the only producer or the only consumer of a given metabolite. Hence, the inhibition of a chokepoint would lead to the depletion or unlimited accumulation of metabolites, thus, potentially leading to an important disruption in the cellular metabolism. Chokepoints are, therefore, appealing drug targets of the bacterial metabolism.

The current approaches to compute chokepoints are based exclusively on the topology of the metabolic network and disregard the dynamic information that might be available. This dynamic information usually refers to the flux bounds of some metabolic reactions. As it will be shown, ignoring such an information can lead to the misidentification of chokepoints. The approach presented in this paper exploits the available flux bounds and computes, for a given growth rate of the cell, the set of chokepoints of the metabolic network. Such chokepoints are potential drug targets whose inhibition could involve a metabolic burden at the given growth rate.

The rest of the paper is organized as follows: Section 2 introduces the basic concepts and definition that will be used in the paper. Section 3 describes the computational method to obtain growth dependent chokepoints. Section 4 analyses the relationship between growth rate and number of chokepoints in the GEM of *Mycobacterium leprae*. The main conclusions of the paper are drawn in

section 5. Finally, Appendix A introduces the software tool developed to compute growth dependent chokepoints, and Appendix B reports the number of chokepoints found in the GEMs of different microorganisms.

2 Preliminary concepts and definitions

2.1 Constraint-based models

A *constraint-based model* [19, 13] is a tuple $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ lb,\ ub\}$ 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 $lb, ub : \mathcal{R} \to \mathbb{R}$ are lower and upper flux bounds of the reactions.

Each reaction is associated with a set of reactant metabolites and a set of product metabolites (one of these sets can be empty). For instance, the reaction $r_1 : A \to 2B$ has one reactant, A, and one product, B. The number 2 expresses the stoichiometric weight, i.e. two units of B are produced per each unit of A that is consumed. 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 $r \in \mathcal{R}$. Thus, if $S[m, r] < 0$ then m is consumed when r occurs; if $S[m, r] > 0$ then m is produced when r occurs; and if $S[m, r] = 0$ then m is neither consumed nor produced when r occurs.

Constraint-based models can be represented graphically as Petri nets [11, 5] where places, which are drawn as circles, are associated with metabolites, and transitions, which are drawn as rectangles, are associated with reactions. An arc from a place(transition) to a transition(place) means that the place is a reactant(product). The weights of the arcs of the Petri net correspond to the stoichiometric weights, in other words, the stoichimetric matrix of a constraint-base model and the incidence matrix of its corresponding Petri net coincide.

Example 1. The Petri net in Figure 1 represents a simple contraint-based model that consists of 13 reactions and 9 metabolites. As an example, transition r_6 models the reaction $r_6 : m_a \to 2m_d$.

2.2 Topological definitions.

Borrowing the usual Petri net notation (given a node x of a Petri net, ${}^\bullet x$ and $x^\bullet$ denote the sets of the input and output nodes of x respectively), we define the following sets for constraint-based models:

- Set of products of r: $r^\bullet = \{m \in \mathcal{M} | S(m, r) > 0\}$
- Set of reactants of r: ${}^\bullet r = \{m \in \mathcal{M} | S(m, r) < 0\}$
- Set of consumers of m: $m^\bullet = \{r \in \mathcal{R} | S(m, r) < 0\}$
- Set of producers of m: ${}^\bullet m = \{r \in \mathcal{R} | S(m, r) > 0\}$

A chokepoint is a reaction that is the only producer or the only consumer of a metabolite. More formally:

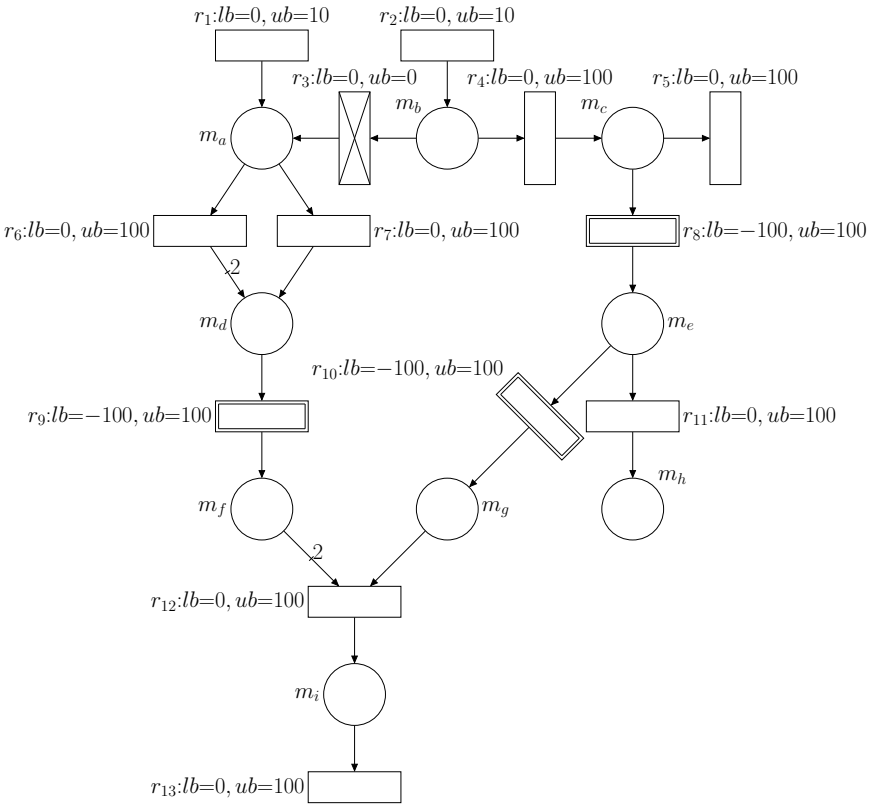

Fig. 1: Petri net modelling a constraint-based model. The values lb and ub are the lower and upper flux bounds of reactions. Non-reversible reactions are represented by simple rectangles, reversible reactions by double rectangles and dead-reactions by rectangles with a cross.

Definition 1. *A reaction $r \in \mathcal{R}$ is a chokepoint if there exists $m \in \mathcal{M}$ such that $m^\bullet = \{r\}$ or ${}^\bullet m = \{r\}$.*

The set of chokepoint reactions will be denoted as CP. Notice that the inhibition of the enzymes associated with a chokepoint will lead either to the depletion of metabolites (which might be essential for the cell) if the chokepoint is the only producer, or to the indefinite accumulation of metabolites (which will not be used as expected or might be toxic) if the chokepoint is the only consumer. Thus, a chokepoint is an attractive drug target because in both cases, essential functions of the cell can be affected by its inhibition [18].

A dead-end metabolite (DEM) is a metabolite that lacks either producing or consuming reactions:

Definition 2. *A metabolite $m \in \mathcal{M}$ is a dead-end metabolite (DEM) if $m^\bullet = \{\}$ or ${}^\bullet m = \{\}$.*

The presence of a DEM in the network reflects an incompleteness in the model, which might require further curation [8].

Example 2. In the Petri net in Figure 1, r_1 is a producer of m_a, i.e. $r_1 \in {}^\bullet m_a$; r_4 is a chokepoint because it is the only producer of m_c, i.e. $\{r_4\} = {}^\bullet m_c$ and $r_4 \in CP$; and m_h is a dead-end metabolite, i.e. $m_h \in DEM$.

2.3 Flux dependent definitions.

The functions lb and ub establish lower and upper steady state flux bounds on the reactions, where flux is the rate of turnover of molecules through the reaction. These functions must satisfy that $lb(r) \leq ub(r)$ for every $r \in \mathcal{R}$. Lower and upper bounds provide useful information about the system and might alter the sets of consumer and producer reactions previously defined. Such bounds will be used in the following to improve the analysis of constraint-based models.

In contrast to Petri nets, these bounds can be negative, and hence, the flux of a reaction can also be negative. A negative flux implies that the metabolites on the left-hand side of the reaction (which in principle are "reactants") are produced, and the metabolites on the right-hand side of the reaction (which in principle are "products") are consumed. A reaction whose flux can be both negative and positive is called reversible. Functions lb and ub will be used to define the sets of flux dependent reversible reactions (RR_d), dead reactions (DR_d), and non-reversible reactions (NR_d), where the subindex d indicates that the sets are flux *dependent*:

Definition 3. *A reaction $r \in \mathcal{R}$ is reversible if $lb(r) < 0 < ub(r)$.*

The set of reversible reactions is denoted RR_d, i.e. $RR_d = \{r \in \mathcal{R} |\ r$ is reversible$\}$.

Definition 4. *A reaction $r \in \mathcal{R}$ is dead if $lb(r) = ub(r) = 0$.*

The set of dead reactions is denoted DR_d, i.e. $DR_d = \{r \in \mathcal{R} |\ r$ is dead$\}$.

Definition 5. *A reaction $r \in \mathcal{R}$ is non-reversible if $(0 \leq lb(r) \wedge 0 < ub(r)) \ \vee (lb(r) < 0 \wedge ub(r) \leq 0)$.*

The set of non-reversible reactions is denoted NR_d, i.e. $NR_d = \{r \in \mathcal{R} |\ r$ is non-reversible $\}$.

Clearly, the sets RR_d, DR_d and NR_d partition the set of reactions $\mathcal{R}$, i.e. $RR_d \cup DR_d \cup NR_d = \mathcal{R}$, $RR_d \cap DR_d = \emptyset$, $DR_d \cap NR_d = \emptyset$, and $RR_d \cap NR_d = \emptyset$.

Non-reversible, reversible and dead reactions will be represented graphically as rectangles, double rectangles, and rectangles with a cross inside respectively.

Example 3. In Figure 1, the above defined sets are: $RR_d = \{r_8, r_9\, r_{10}\}$, $DR_d = \{r_3\}$, and $NR_d = \{r_1, r_2, r_4, r_5, r_6, r_7, r_{11}, r_{12}, r_{13}\}$.

Given that, in constraint-based models, reactions can be reversible or can proceed only backwards, i.e. $lb(r) \leq ub(r) < 0$, the concepts related to the consumption and production of metabolites must be revisited. Thus, new sets of reactants, products, consumers, and producers which take into account the flux bounds are defined as follows:

- Set of products of r:
$$r_d^\bullet = \{m \in \mathcal{M} | (S(m,r) > 0 \wedge ub(r) > 0) \vee (S(m,r) < 0 \wedge lb(r) < 0)\}$$
- Set of reactants of r:
$${}^\bullet r_d = \{m \in \mathcal{M} | (S(m,r) < 0 \wedge ub(r) > 0) \vee (S(m,r) > 0 \wedge lb(r) < 0)\}$$
- Set of consumers of m:
$$m_d^\bullet = \{r \in \mathcal{R} | (S(m,r) < 0 \wedge ub(r) > 0) \vee (S(m,r) > 0 \wedge lb(r) < 0)\}$$
- Set of producers of m:
$${}^\bullet m_d = \{r \in \mathcal{R} | (S(m,r) > 0 \wedge ub(r) > 0) \vee (S(m,r) < 0 \wedge lb(r) < 0)\}$$

Flux dependent definitions of chokepoints and dead-end metabolites can be written as:

Definition 6. *A reaction $r \in \mathcal{R}$ is a flux dependent chokepoint if there exists $m \in \mathcal{M}$ such that $m_d^\bullet = \{r\}$ or ${}^\bullet m_d = \{r\}$.*

Definition 7. *A metabolite $m \in \mathcal{M}$ is a flux dependent dead-end metabolite if $m_d^\bullet = \{\}$ or ${}^\bullet m_d = \{\}$.*

The sets of flux dependent chokepoint reactions and dead-end metabolites will be denoted as CP_d and DEM_d respectively.

Example 4. In Figure 1, r_{12} is a flux dependent chokepoint, i.e. $r_{12} \in CP_d$; and m_h is a flux dependent dead-end metabolite, i.e. $m_h \in DEM_d$.

3 Growth dependent chokepoints

In GEMs, unknown flux bounds are given default values, e.g. $lb(r) = -1000\ mmol\ g^{-1}h^{-1}$ and $ub(r) = 1000\ mmol\ g^{-1}h^{-1}$ (recall that flux bounds establish the direction in which the reaction can proceed). Thus, all the reactions that are given default values are considered as reversible. However, not all the fluxes in the ranges given by the flux bounds of GEMs models are compatible with a positive growth rate. By using Flux Balance Analysis (FBA)[14] and Flux Variability Analysis (FVA)[4] it is possible to obtain tighter flux bounds for a given growth rate. Such tighter bounds could imply that, reactions which were initially considered as reversible, are in fact non-reversible for the given growth rate. This might alter the original set of chokepoints, i.e. the set of chokepoints depend on the growth rate. This section describes how growth dependent chokepoints can be computed.

Flux Balance Analysis (FBA) is a mathematical procedure for the estimation of steady state fluxes in constraint-based models. FBA can be used, for instance, to predict the growth rate of an organism or the rate of production of a given

metabolite. Mathematically, FBA is expressed as a linear programming problem that maximises an objective function subject to steady state constraints. In the case of estimating the growth rate, the objective function is biomass production, a reaction that defines the ratios at which metabolites are converted into basic constituents of the cell as nucleic acids or proteins [14].

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. The steady state fluxes of reactions are also lower and upper bounded by lb and ub. Thus, the FBA linear programming problem is:

$$\begin{aligned} &\max\ z \cdot v \\ &st.\ \ S \cdot v = 0 \\ &\qquad lb(r) \leq v[r] \leq ub(r) \ \ \forall r \in \mathcal{R} \end{aligned} \tag{1}$$

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

It is a common assumption that the metabolism of prokaryotes has evolved to maximize the growth of the cells. Hence, the growth rate given by the biomass production is an empirically reasonable choice for the objective function of FBA applied to bacteria [17].

A given growth, i.e. a given flux through the reaction modelling biomass production, can be achieved by different fluxes of the reactions. This means that each reaction can have a range of fluxes that is compatible with a given growth. Flux Variability Analysis (FVA) can be used to compute such range of fluxes for each reaction. [2].

More precisely, FVA[4] is a mathematical procedure to compute the minimum and maximum fluxes of reactions that are compatible with some state, e.g. supporting 90% of the maximum growth yielded by FBA. Among other applications, FVA can be used to study the network flexibility, and studying the network response under suboptimal conditions.

Let μ_{max} be the maximum growth calculated by FBA. FVA is computed by solving two independent linear programming problems per reaction $r \in \mathcal{R}$. One programming problem maximizes the flux of r, $v[r]$, and the other minimizes $v[r]$. The constraints of both problems are the same: the steady state condition $S \cdot v = 0$, the flux bounds $lb(r) \leq v[r] \leq ub(r)$, and the maintenance of the optimum value given by FBA to a certain degree. This last constraint is expressed as $\gamma \cdot \mu_{max} \leq z \cdot v$ where z is the same vector as in (1) and $\gamma \in [0, 1]$ represents the fraction of optimal value that must be satisfied. Thus, the two programming problems for a given reaction $r \in \mathcal{R}$ can be expressed as:

$$\begin{aligned} &\max / \min\ v[r] \\ &\qquad st.\ \ S \cdot v = 0 \\ &\qquad\qquad lb(r) \leq v[r] \leq ub(r) \ \ \forall r \in \mathcal{R} \\ &\qquad\qquad \gamma \cdot \mu_{max} \leq z \cdot v \end{aligned} \tag{2}$$

Let $lb_\gamma, ub_\gamma : \mathcal{R} \to \mathbb{R}$ be the result of running FVA on a constraint-based model $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ lb,\ ub\}$ for a given γ. If the flux bounds lb, ub 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 constrained flux bounds is obtained.

Given γ, the sets of flux dependent products, reactants, consumers, and producers of the model $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ lb_\gamma,\ ub_\gamma\}$ are denoted as $r^\bullet_\gamma, {}^\bullet r_\gamma, m^\bullet_\gamma, {}^\bullet m_\gamma$ respectively. Similarly, the sets of flux dependent reversible reactions, dead reactions, and non-reversible reactions are denoted as RR_γ, DR_γ, NR_γ respectively. The sets of flux dependent chokepoint reactions and dead-end metabolites are denoted as CP_γ and DEM_γ.

In Algorithm 1, an iterative procedure is proposed that, given an input constrained-based model and γ, it produces a list of pairs $(reactant, reaction)$ $((reaction, product))$ where $reaction$ is a growth dependent chokepoint and $reactant(product)$ is the metabolite whose only consumer(producer) is $reaction$.

Algorithm 1 Growth dependent chokepoint reactions computation

INPUT: $\{\mathcal{R},\ \mathcal{M},\ \mathcal{S},\ lb,\ ub\}$, γ.
OUTPUT: List of tuples $(reactant, reaction)$ and $(reaction, product)$ such that $reaction \in CP_\gamma$ and $reaction$ is the only consumer of $reactant$ or the only producer of $product$.

```
 1: procedure FindChokepointReactions
 2:     lb_γ, ub_γ ← FVA({R, M, S, lb, ub}, γ)
 3:     {R, M, S, lb, ub} ← {R, M, S, lb_γ, ub_γ}
 4:
 5:     result ← empty list
 6:     for reaction in R do
 7:         for reactant in •reaction_γ do
 8:             if reactant•_γ = {reaction} then
 9:                 result ← result + (reactant, reaction)
10:             end if
11:         end for
12:         for product in reaction•_γ do
13:             if •product_γ = {reaction} then
14:                 result ← result + (reaction, product)
15:             end if
16:         end for
17:     end for
18:     return result
19: end procedure
```

Prior to the computation of chokepoints (lines 5-18), the algorithm refines the input model as explained by replacing the initial flux bounds by the flux values computed with FVA (lines 2-3). To compute chokepoints, the algorithm iterates over the reactions of the model. For each reaction, the reactants and products involved are iterated. For each pair $(reactant, reaction)$ and $(reaction, product)$,

if Definition 6 is satisfied, the reaction is considered a chokepoint reaction with the given metabolite.

Example 5. Let us assume that r_{13} in Figure 1 represents growth, i.e. the component of z in (1) that corresponds to r_{13} is equal to 1 and the rest of components of z are 0. Let us also assume that it is desired to assess the directionality of the reactions and compute the set of chokepoints when the growth is maximum. This can be achieved by applying Algorithm 1 with $\gamma = 1$. The new flux bounds yielded by the algorithm are shown in Figure 2.

As a result of the new flux bounds given to the model, reactions r_5, r_7 and r_{11} become dead reactions, i.e. $r_5, r_7, r_{11} \in DR_1$, and reactions r_8, r_9, r_{10}, which where reversible reactions in Figure 1, become non-reversible reactions, i.e. $r_8, r_9, r_{10} \in NR_1$. This change in the directionality of the reactions involves changes in the set of flux-dependent chokepoints, in particular r_6 becomes a chokepoint, i.e. $r_6 \in CP_1$, and r_{11}, which was a chokepoint in Figure 1, becomes a dead-reaction, i.e. $r_{11} \in DR_1$.

4 Chokepoint Analysis

The computation of flux bounds by means of FVA can be carried out with an optimal state, i.e. $\gamma = 1$ in (2), or with suboptimal states, i.e. $0 \leq \gamma < 1$. While in an optimal state all the fluxes must be optimally directed towards growth, in suboptimal states fluxes are allowed to be diverted towards other functionalities. This section analyses the impact of γ in the sets of reversible, non-reversible, dead and chokepoint reactions of a constraint-based model. To achieve this goal the flux bounds of the model will be refined according to different values of γ and the mentioned sets of reactions will be computed.

All the results presented in this section have been obtained by a software tool which implements Algorithm 1 and computes the sets RR_γ, DR_γ, NR_γ and CP_γ. A description of the tool can be found in Appendix A. The tool has been executed on a number of constraint-based models yielding the results presented in Appendix B.

4.1 Case study: Mycobacterium leprae

This subsection presents the results obtained for the constraint-based model of the *in vivo* GEM of *M. leprae* [1, 6]. This model is composed of 998 metabolites and 1228 reactions. The sizes of the flux-dependent sets of reactions are $|RR_d|$=288, $|NR_d|$=938, $|DR_d|$=2, and the number of flux-dependent CP is $|CP_d| = 667$. In order to assess the dependence of these sets on γ, the flux bounds of the model have been refined for different values of γ in the interval $[0, 1]$.

Figure 3 shows the sizes of the sets DR_γ, NR_γ and CP_γ, in plot (a), (b) and (c) respectively, for different values of γ. Notice that if $\gamma = 0$ then the constraint $\gamma \cdot \mu_{max} \leq z \cdot v$ in (2) does not impose a minimum growth on the model, and

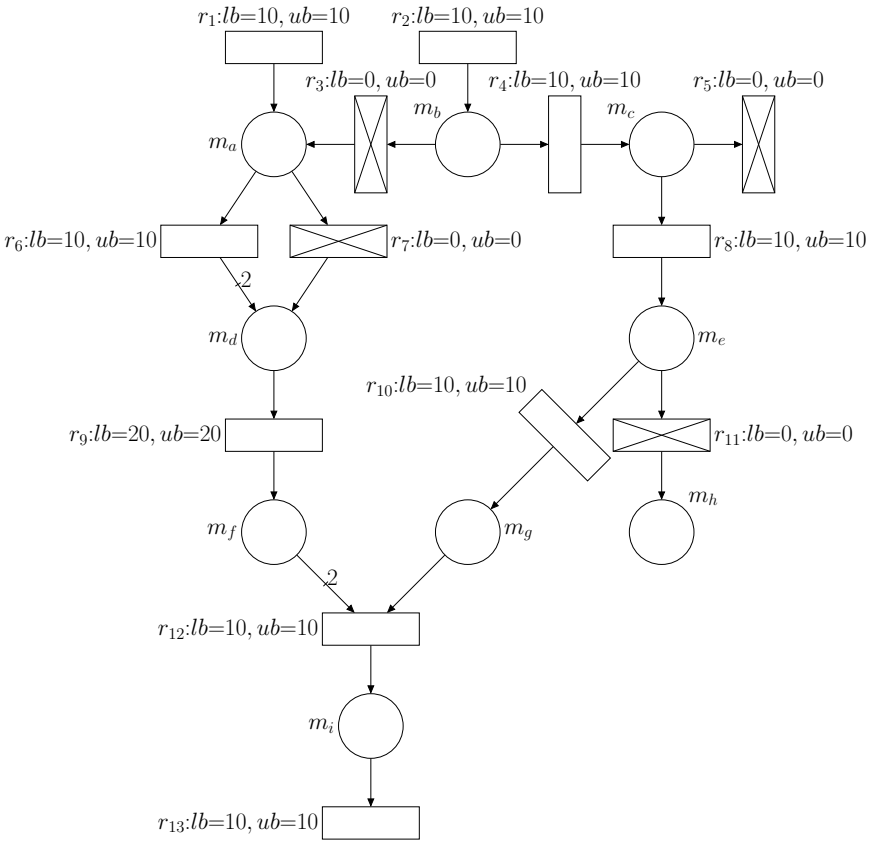

Fig. 2: Petri net resulting from the application of Algorithm 1 to the Petri net in Figure 1 with r_{13} as the objective function and $\gamma = 1$.

only the steady state condition $S \cdot v = 0$ must be satisfied. In addition to the γ dependent sets, the leftmost value of each plot (depicted in green) in Figure 3 represents the sizes of the flux-dependent sets prior to FVA, i.e. DR_d, NR_d and CP_d.

It can be seen that the set of dead reactions exhibits two major increases, the first one from $|DR_d| = 2$ to $|DR_0| = 219$, and the second one at $\gamma = 1$ ($|DR_1| = 667$), see Figure 3(c). Given that only the steady state condition is active for $\gamma = 0$, the first increase implies that the fluxes of all the reactions in DR_0 must necessarily be 0 in the long run regardless of the growth rate. Thus, this increase can be due to a shortcoming or incompleteness in the model.

The second increase in the set of dead reactions takes place at $\gamma = 1$ and can be caused by the existence of alternative pathways that consume nutrients, one of them being more efficient than the others in terms of biomass production.

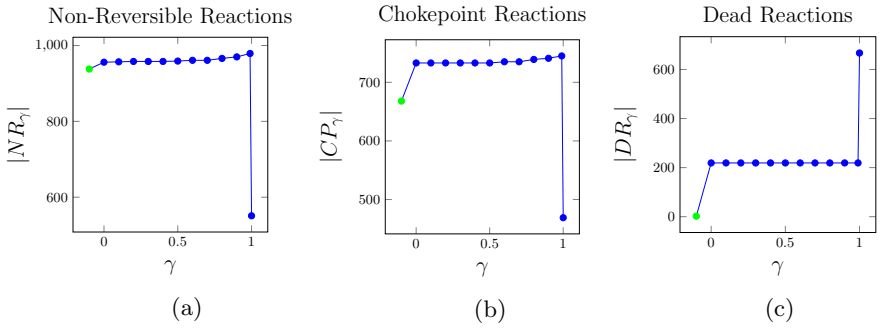

Fig. 3: Sizes of the sets of reactions DR_γ, NR_γ and CP_γ of *M. leprae* for $\gamma \in [0, 1]$. The leftmost value of each plot corresponds to DR_d, NR_d and CP_d respectively.

The next subsection proposes simplified models that illustrate how these sudden increases take place.

With respect to chokepoints, see Figure 3(b), the number of flux-dependent chokepoints is $|CP_d| = 668$. This number increases to $|CP_0| = 733$ when the steady state constraint is forced, and increases slowly with γ, at $\gamma = 0.9$ it holds $|CP_{0.9}| = 741$. That is, Algorithm 1 identifies more chokepoints than the ones that are present originally in the model. The sudden drop of chokepoints at $\gamma = 1$, $|CP_1| = 469$, is due to the fact that at the optimal state many chokepoints become dead reactions as discussed previously.

In a similar way to the set of chokepoints, the number of non-reversible reactions increases slowly with γ and it falls abruptly at $\gamma = 1$. This means that the sets NR_γ and CP_γ decrease at the optimal state as many of these reactions become dead reactions.

Figure 4 presents a Sankey diagram showing how reactions are distributed among the sets NR, RR and DR and the flow of transformations that takes place among these sets from the initial model to a model refined with a suboptimal state of $\gamma = 0.9$, and from this suboptimal state model to an optimal state model refined with $\gamma = 1.0$. Notice that at $\gamma = 1.0$ the set of dead reactions becomes the largest set of reactions, and that the set of reversible reactions is vastly reduced at both suboptimal and optimal growth.

As it is reported in Appendix B, similar trends to the ones discussed here for the sets NR, DR and CP are exhibited by other models.

4.2 Dead reactions and growth rate

It has been shown that refining a model with a suboptimal growth can cause the set of dead reactions, DR, to increase, and also that this set increases further with an optimal growth refinement. These changes in DR are caused by particular network structures that can appear in a metabolic network. This subsection

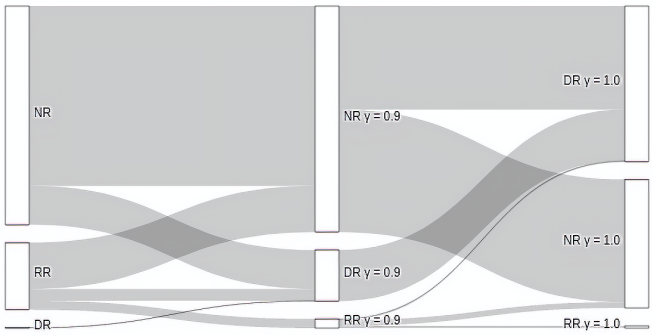

Fig. 4: Sankey diagram showing the dependence of the sets NR, RR and DR on the growth rate.

illustrates through an abstract model the types of structures that can produce such changes in DR.

Let us consider the constraint-based model depicted as a Petri net in Figure 5. The lower and upper flux bounds of the model are reported in Table 1 of the figure. According to such flux bounds all the reactions are non-reversible, i.e. $DR = \{\}$, see column *type*. The exchange reactions are r_1 and r_{10}, which could model nutrient uptake and secretion of a metabolite respectively. The reaction modelling biomass production is r_7, i.e. the flux of r_7 represents the growth rate of system being modelled. Hence, the objective funcion of FBA, see (1), is the maximization of $v[r_7]$. For the given net structure and flux bounds, the maximum growth rate yielded by FBA (1) is $\mu_{max} = 200$.

Table 2 reports the refined flux bounds and the type of each reaction for $\gamma = 0$, $\gamma = 0.9$ and $\gamma = 1$. If $\gamma = 0$ then only the steady state constraint, $S \cdot v = 0$, of FVA (2) is taken into account to compute the refined flux bounds. Thus, any possible steady state must satisfy the bounds, lb_0 and ub_0, of the rows associated with $\gamma = 0$. For such γ, all the lower bounds, lb_0, are kept to 0 and the upper bounds, ub_0, are constrained by the upper flux bound of r_1 which is $ub(r_1) = 100$. It should be noted that $ub_0(r_8) = ub_0(r_9) = ub_0(r_{10}) = 0$, that is reactions r_8, r_9 and r_{10} are dead at any steady state, i.e. $DR_0 = \{r_8, r_9, r_{10}\}$.

Recall that this increase in the number of dead reactions, DR_0, also took place in the previous subsection for the *M. leprae* model. The reason why r_8 becomes dead is because it is producing a DEM m_g, if the flux of r_8 was positive then the concentration of m_g would increase indefinitely which contradicts the existence of a steady state.

Let us now focus on r_9 and r_{10}. The steady state condition, $S \cdot v = 0$, for metabolite m_h imposes $v[r_9] = v[r_{10}]$ (i.e. $input\ flux = output\ flux$), while for m_i the steady state condition is $2 \cdot v[r_9] = v[r_{10}]$ (see weight 2 in the arc

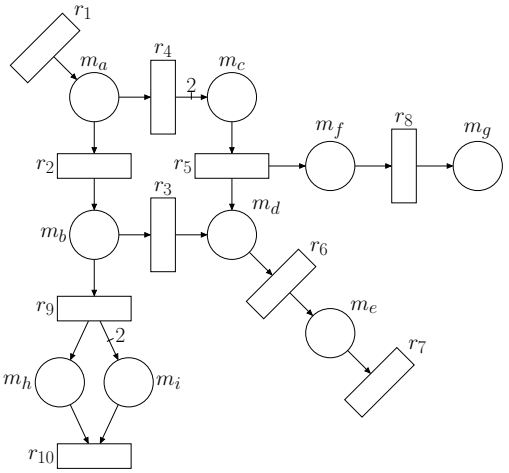

Table 1: Initial flux bounds

r	$lb(r)$	$ub(r)$	type
r_1	0.0	100.0	NR
r_2	0.0	1000.0	NR
r_3	0.0	1000.0	NR
r_4	0.0	1000.0	NR
r_5	0.0	1000.0	NR
r_6	0.0	1000.0	NR
r_7	0.0	1000.0	NR
r_8	0.0	1000.0	NR
r_9	0.0	1000.0	NR
r_{10}	0.0	1000.0	NR

Table 2: Refined model flux bounds

γ	r	$lb_\gamma(r)$	$ub_\gamma(r)$	type
0.0	r_1	0.0	100.0	NR
	r_2	0.0	100.0	NR
	r_3	0.0	100.0	NR
	r_4	0.0	100.0	NR
	r_5	0.0	200.0	NR
	r_6	0.0	200.0	NR
	r_7	0.0	200.0	NR
	r_8	0.0	0.0	**DR**
	r_9	0.0	0.0	**DR**
	r_{10}	0.0	0.0	**DR**
0.9	r_1	90.0	100.0	NR
	r_2	0.0	20.0	NR
	r_3	0.0	20.0	NR
	r_4	80.0	100.0	NR
	r_5	160.0	200.0	NR
	r_6	180.0	200.0	NR
	r_7	180.0	200.0	NR
	r_8	0.0	0.0	**DR**
	r_9	0.0	0.0	**DR**
	r_{10}	0.0	0.0	**DR**
1.0	r_1	100.0	100.0	NR
	r_2	0.0	0.0	**DR**
	r_3	0.0	0.0	**DR**
	r_4	100.0	100.0	NR
	r_5	200.0	200.0	NR
	r_6	200.0	200.0	NR
	r_7	200.0	200.0	NR
	r_8	0.0	0.0	**DR**
	r_9	0.0	0.0	**DR**
	r_{10}	0.0	0.0	**DR**

Fig. 5: Petri net illustrating the evolution of dead reactions at $\gamma = 0$ and $\gamma = 1$.

(r_9, m_i). The only fluxes that satisfy these two conditions simultaneously are $v[r_9] = v[r_{10}] = 0$, i.e. the reactions are dead at any steady state.

At $\gamma = 0.9$, the flux of r_7 (biomass production) must be at least $0.9 \cdot \mu_{max} = 180$. Notice that, although $DR_0 = DR_{0.9}$, the lower bounds of some reactions are higher at $\gamma = 0.9$ than at $\gamma = 1$, and the upper bounds of some other reactions are lower at $\gamma = 0.9$ than at $\gamma = 1$. In general it holds:

$$ub_{0.9}[r] - lb_{0.9}[r] \geq ub_1[r] - lb_1[r] \quad \forall r \in \mathcal{R}$$

In other words, the range of steady state fluxes allowed for each reaction decreases with γ. This is due to the existence of alternative paths in the metabolic network. In the present example, there are two alternative pathways to biomass production, namely $p_1 = (r_1, r_2, r_3, r_6, r_7)$ and $p_2 = (r_1, r_4, r_5, r_6, r_7)$. Notice that p_2 is more advantageous than p_1 for biomass production as 2 metabolites m_c are produced per each metabolite m_a. Thus, near optimal solutions will tend to exploit p_2 instead of p_1. This implies strictly positive lower bounds, $lb_{0.9}$, for all the reactions in p_2, and decreased upper bounds, $ub_{0.9}$, for the reactions that belong exclusively to p_1, i.e. r_2 and r_3.

The interval $[lb_\gamma[r], ub_\gamma[r]]$ shrinks for every reaction $r \in \mathcal{R}$ as γ increases, and at $\gamma = 1$ the intervals become points. Moreover, at $\gamma = 1$ only the most favourable path p_2 can be used, and hence reactions r_2 and r_3 become dead. This type of phenomenon causes the sudden increase of DR_1 in the *M. leprae* model of the previous subsection.

5 Conclusions

This work has introduced a computational method to incorporate dynamic information, namely flux bounds, in the computation of chokepoints in metabolic networks expressed as constraint-based models. The goal behind this approach is to obtain more realistic chokepoints, which are known to be potential drug targets, than those based only on the net topology. Given that flux bounds depend on the growth rate of the organism, the concept of growth dependent chokepoints has been defined and an algorithm to compute such chokepoints has been designed.

It was found that the number of chokepoints was seriously affected by the number of dead reactions, i.e. by reactions with null lower and upper flux bounds. Although the number of dead reactions is not relevant in most of the original models, this number increases significantly when dynamic information is accounted for. A major increase takes place when the steady state constraint is enforced on the metabolic network, i.e. at $\gamma = 0$. As discussed in subsection 4.2, we hypothesize that such an increase is a sign of some incompleteness in the model such as the existence of dead-end metabolites, or to missing reactions that lead to the incompatibility of positive fluxes with the network stoichiometry. Another major increase of dead reactions takes place when the growth rate is maximum, i.e. at $\gamma = 1$. At such a rate, all the flux is diverted towards the optimal paths for biomass production, and hence, the fluxes of non-optimal alternative paths leading to biomass production and other non-essential paths become 0, i.e. such paths contain dead reactions at $\gamma = 1$. Thus, in this case dead reactions might indicate the existence of alternative, or redundant, paths leading to biomass production. Notice that such redundant paths have the potential to make the metabolism more robust to attacks.

The protocol for chokepoint computation presented on this paper can reduce the time spent identifying drug targets in the process of drug discovery. Drug discovery is a time-consuming process, which involves the identification and validation of drug targets, optimisation, lead discovery and testing before the production of drug candidates. Our protocol can contribute reducing the drug target identification time by prioritising the selection of potential top targets of the pathogenic organism for subsequent validation and optimisation protocols of the drug discovery process.

References

1. Bannerman, B.P., Vedithi, S.C., Júlvez, J., Torres, P., Waman, V.P., Munir, A., Mendes, V., Malhotra, S., Skwark, M.J., Oliver, S.G., Blundell, T.L.: Analysis

of metabolic pathways in mycobacteria to aid drug-target identification. bioRxiv (2019). https://doi.org/10.1101/535856
2. Burgard, A.P., Vaidyaraman, S., Maranas, C.D.: Minimal reaction sets for escherichia coli metabolism under different growth requirements and uptake environments. Biotechnology progress **17**(5), 791–797 (2001). https://doi.org/10.1021/bp0100880
3. Glont, M., Nguyen, T.V.N., Graesslin, M., Hälke, R., Ali, R., Schramm, J., Wimalaratne, S.M., Kothamachu, V.B., Rodriguez, N., Swat, M.J., et al.: Biomodels: expanding horizons to include more modelling approaches and formats. Nucleic acids research **46**(D1), D1248–D1253 (2018). https://doi.org/10.1093/nar/gkx1023
4. Gudmundsson, S., Thiele, I.: Computationally efficient flux variability analysis. BMC Bioinformatics **11**(1), 489 (sep 2010). https://doi.org/10.1186/1471-2105-11-489
5. Heiner, M., Gilbert, D., Donaldson, R.: Petri nets for systems and synthetic biology. vol. 5016, pp. 215–264 (06 2008). https://doi.org/10.1007/978-3-540-68894-5_7
6. Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S.M., Subhraveti, P.: The BioCyc collection of microbial genomes and metabolic pathways. Briefings in Bioinformatics **20**(4), 1085–1093 (08 2017). https://doi.org/10.1093/bib/bbx085
7. Lamberti, L.M., Zakarija-Grković, I., Fischer Walker, C.L., Theodoratou, E., Nair, H., Campbell, H., Black, R.E.: Breastfeeding for reducing the risk of pneumonia morbidity and mortality in children under two: A systematic literature review and meta-analysis (2013). https://doi.org/10.1186/1471-2458-13-S3-S18
8. Mackie, A., Keseler, I.M., Nolan, L., Karp, P.D., Paulsen, I.T.: Dead End Metabolites - Defining the Known Unknowns of the E. coli Metabolic Network. PLoS ONE **8**(9), e75210 (sep 2013). https://doi.org/10.1371/journal.pone.0075210
9. Mazurek, S.: Pyruvate kinase type M2: A key regulator of the metabolic budget system in tumor cells. International Journal of Biochemistry and Cell Biology **43**(7), 969–980 (jul 2011). https://doi.org/10.1016/j.biocel.2010.02.005
10. Munger, J., Bennett, B.D., Parikh, A., Feng, X.J., McArdle, J., Rabitz, H.A., Shenk, T., Rabinowitz, J.D.: Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nature Biotechnology **26**(10), 1179–1186 (oct 2008). https://doi.org/10.1038/nbt.1500
11. Murata, T.: Petri Nets: Properties, Analysis and Applications. Procs. of the IEEE **77**(4), 541–580 (1989). https://doi.org/10.1109/5.24143
12. Murima, P., McKinney, J.D., Pethe, K.: Targeting bacterial central metabolism for drug development (nov 2014). https://doi.org/10.1016/j.chembiol.2014.08.020
13. Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.Ø.: A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. Molecular Systems Biology **7**(1) (2011). https://doi.org/10.1038/msb.2011.65
14. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? (mar 2010). https://doi.org/10.1038/nbt.1614
15. Rahman, S.A., Schomburg, D.: Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. Bioinformatics (Oxford, England) **22**(14), 1767–74 (jul 2006). https://doi.org/10.1093/bioinformatics/btl181
16. Raman, K., Vashisht, R., Chandra, N.: Strategies for efficient disruption of metabolism in Mycobacterium tuberculosis from network analysis. Molecular bioSystems **5**(12), 1740–51 (dec 2009). https://doi.org/10.1039/B905817F

17. Segre, D., Vitkup, D., Church, G.M.: Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences **99**(23), 15112–15117 (2002). https://doi.org/10.1073/pnas.232349399
18. Singh, S., Malik, B.K., Sharma, D.K.: Choke point analysis of metabolic pathways in E.histolytica: A computational approach for drug target identification. Bioinformation **2**(2), 68–72 (aug 2007). https://doi.org/10.6026/97320630002068
19. Varma, A., Palsson, B.Ø.: Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Nature Biotechnology **12**(10), 994–998 (Oct 1994). https://doi.org/10.1038/nbt1094-994
20. WHO: WHO — Causes of death. WHO (2018)
21. Zhang, R., Lin, Y.: DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Research **37**(suppl_1), D455–D458 (10 2008). https://doi.org/10.1093/nar/gkn858

A Appendix

The software tool *findCPcli* developed in this work consists of a command line application that, given an input model provided by the user, computes the sizes of the sets of non-reversible reactions, reversible reactions, dead reactions and chokepoint reactions for different values of γ. The results are saved in a spreadsheet file with a format similar to the one presented in Table 3.

The tool *findCPcli* is distributed as a *Python* package and requires *Python 3.5* or a higher version. The source can be found at `github.com/findCP/findCPcli`. *findCPcli* can be installed with the *pip* package management tool:

```
pip install findCPcli
```

Once installed, the results for a given SBML model can be computed running:

```
findCPcli -i <input_file> -cp <output_file>
```

, where:

- `<input_file>` is the path of the input SBML model file to be used. The supported file formats are *.xml*, *.json* and *.yml*.
- `<output_file>` is the path of the spreadsheet file that will be saved with the results computed on the model. The available file formats for the spreadsheet file are *.xls*, *.xlsx* and *.ods*.

When the above command is executed, the command line application will inform about the task that will be computed. If the task finishes successfully and the spreadsheet file has been saved, the application will inform about it and will end the execution.

Further information about the operations provided by the application can be found by executing: `findCPcli -h` .

B Appendix

Table 3 reports the sizes of the sets of reversible, non-reversible, dead and chokepoint reactions for several constraint-based models of the *Biomodels* repository [3]. All the results were computed by the tool *findCPcli*. The maximum CPU time was 82.776s to compute the results of model MODEL1507180017 in an Intel Core i5-9300H CPU @ 2.40GHz $\times$ 8.

Table 3: Sizes of the sets of reversible, non-reversible, dead and chokepoint reactions for several constraint-based models.

				γ										
model	set	initial	set	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
mlep_inVivo_media	RR_d	288	RR_γ	53	52	51	51	51	50	48	48	43	39	10
reactions: 1228	NR_d	938	NR_γ	956	957	958	958	958	959	961	961	966	970	551
metabolites: 998	DR_d	2	DR_γ	219	219	219	219	219	219	219	219	219	219	667
	CP_d	668	CP_γ	733	733	733	733	733	733	735	735	739	741	469
MODEL1507180021	RR_d	208	RR_γ	72	72	72	72	72	72	65	64	63	60	20
reactions: 900	NR_d	692	NR_γ	710	710	710	710	710	710	717	718	719	722	417
metabolites: 688	DR_d	0	DR_γ	118	118	118	118	118	118	118	118	118	118	463
(M. tuberculosis)	CP_d	506	CP_γ	504	504	504	504	504	504	510	510	510	512	350
MODEL1507180007	RR_d	183	RR_γ	27	27	27	26	26	24	24	24	22	20	6
reactions: 554	NR_d	371	NR_γ	377	377	377	378	378	380	380	380	382	384	304
metabolites: 485	DR_d	0	DR_γ	150	150	150	150	150	150	150	150	150	150	244
	CP_d	277	CP_γ	315	315	315	315	315	317	317	317	319	319	271
MODEL1507180063	RR_d	184	RR_γ	32	32	31	31	31	31	28	26	26	13	6
reactions: 556	NR_d	370	NR_γ	399	399	400	400	400	400	403	405	405	418	316
metabolites: 549	DR_d	2	DR_γ	125	125	125	125	125	125	125	125	125	125	234
	CP_d	274	CP_γ	316	316	316	316	316	316	318	321	321	330	291
MODEL1507180030	RR_d	193	RR_γ	28	28	28	28	28	27	27	27	27	27	12
reactions: 560	NR_d	305	NR_γ	317	317	317	317	317	318	318	318	318	318	278
metabolites: 479	DR_d	62	DR_γ	215	215	215	215	215	215	215	215	215	215	270
	CP_d	310	CP_γ	257	257	257	257	257	258	258	258	258	258	258
MODEL1507180070	RR_d	216	RR_γ	25	21	20	18	18	18	17	17	17	17	13
reactions: 743	NR_d	513	NR_γ	425	429	430	432	432	432	433	433	433	433	420
metabolites: 655	DR_d	14	DR_γ	293	293	293	293	293	293	293	293	293	293	310
	CP_d	324	CP_γ	330	333	333	335	335	335	335	335	335	335	336
MODEL1507180045	RR_d	354	RR_γ	171	171	171	171	171	171	171	171	171	169	149
reactions: 778	NR_d	424	NR_γ	487	487	487	487	487	487	487	487	487	489	295
metabolites: 662	DR_d	0	DR_γ	120	120	120	120	120	120	120	120	120	120	334
	CP_d	336	CP_γ	387	387	387	387	387	387	387	387	387	389	229
MODEL1507180024	RR_d	318	RR_γ	75	75	75	75	75	75	75	75	74	62	21
reactions: 832	NR_d	514	NR_γ	510	510	510	510	510	510	510	510	511	523	438
metabolites: 790	DR_d	0	DR_γ	247	247	247	247	247	247	247	247	247	247	373
	CP_d	397	CP_γ	406	406	406	406	406	406	406	406	406	414	380
MODEL1507180036	RR_d	277	RR_γ	103	103	103	103	103	103	99	99	93	89	51
reactions: 870	NR_d	593	NR_γ	593	593	593	593	593	593	597	597	603	607	460
metabolites: 713	DR_d	0	DR_γ	174	174	174	174	174	174	174	174	174	174	359
	CP_d	433	CP_γ	402	402	402	402	402	402	406	406	409	411	346
MODEL1507180049	RR_d	271	RR_γ	189	189	189	189	189	189	189	188	188	184	140
reactions: 971	NR_d	700	NR_γ	717	717	717	717	717	717	717	718	718	722	388
metabolites: 496	DR_d	0	DR_γ	65	65	65	65	65	65	65	65	65	65	443
	CP_d	265	CP_γ	273	273	273	273	273	273	273	273	273	273	156

MODEL1507180068	RR_d	361	RR_γ	84	84	84	84	84	80	80	77	77	77	71
reactions: 1056	NR_d	695	NR_γ	568	568	568	568	568	572	572	575	575	575	463
metabolites: 911	DR_d	0	DR_γ	404	404	404	404	404	404	404	404	404	404	522
	CP_d	549	CP_γ	469	469	469	469	469	471	471	474	474	474	395
MODEL1507180060	RR_d	254	RR_γ	57	57	57	57	54	54	52	50	50	48	8
reactions: 1075	NR_d	821	NR_γ	610	610	610	610	613	613	615	617	617	619	356
metabolites: 761	DR_d	0	DR_γ	408	408	408	408	408	408	408	408	408	408	711
	CP_d	441	CP_γ	363	363	363	363	364	364	366	368	368	370	304
MODEL1507180020	RR_d	256	RR_γ	50	48	48	48	47	47	45	45	45	45	12
reactions: 1110	NR_d	751	NR_γ	556	558	558	558	559	559	561	561	561	561	367
metabolites: 879	DR_d	103	DR_γ	504	504	504	504	504	504	504	504	504	504	731
	CP_d	455	CP_γ	396	398	398	398	398	398	400	400	400	400	319
MODEL1507180059	RR_d	630	RR_γ	203	203	203	203	203	203	202	202	200	196	142
reactions: 1112	NR_d	482	NR_γ	511	511	511	511	511	511	512	512	514	518	358
metabolites: 1101	DR_d	0	DR_γ	398	398	398	398	398	398	398	398	398	398	612
	CP_d	470	CP_γ	436	436	436	436	436	436	436	436	436	438	306
MODEL1507180013	RR_d	551	RR_γ	249	249	249	249	249	249	249	249	247	242	49
reactions: 1245	NR_d	694	NR_γ	708	708	708	708	708	708	708	708	710	715	390
metabolites: 987	DR_d	0	DR_γ	288	288	288	288	288	288	288	288	288	288	806
	CP_d	484	CP_γ	533	533	533	533	533	533	533	533	534	536	346
MODEL1507180058	RR_d	452	RR_γ	155	155	155	155	155	155	155	155	150	147	97
reactions: 1285	NR_d	833	NR_γ	904	904	904	904	904	904	904	904	909	912	547
metabolites: 943	DR_d	0	DR_γ	226	226	226	226	226	226	226	226	226	226	641
	CP_d	584	CP_γ	611	611	611	611	611	611	611	611	614	616	409
MODEL1507180015	RR_d	1093	RR_γ	510	510	510	510	510	510	510	510	510	510	334
reactions: 1681	NR_d	588	NR_γ	822	822	822	822	822	822	822	822	822	822	667
metabolites: 1381	DR_d	0	DR_γ	349	349	349	349	349	349	349	349	349	349	680
	CP_d	473	CP_γ	612	612	612	612	612	612	612	612	612	612	551
MODEL1507180054	RR_d	546	RR_γ	85	85	85	82	80	80	77	77	75	71	10
reactions: 2262	NR_d	1716	NR_γ	1138	1138	1138	1141	1143	1143	1146	1146	1148	1152	397
metabolites: 1658	DR_d	0	DR_γ	1039	1039	1039	1039	1039	1039	1039	1039	1039	1039	1855
	CP_d	1039	CP_γ	748	748	748	750	750	750	752	752	754	759	319
MODEL1507180017	RR_d	606	RR_γ	85	85	85	81	80	78	78	77	77	75	13
reactions: 2546	NR_d	1923	NR_γ	1504	1504	1504	1508	1509	1511	1511	1512	1512	1514	525
metabolites: 1802	DR_d	17	DR_γ	957	957	957	957	957	957	957	957	957	957	2008
	CP_d	1112	CP_γ	984	984	984	986	986	988	988	988	988	990	478