Bioinformatics, 39(2), 2023, btad053
https://doi.org/10.1093/bioinformatics/btad053
Advance Access Publication Date: 24 January 2023
Original Paper

Systems biology

CONTRABASS: exploiting flux constraints in genome-scale models for the detection of vulnerabilities

Alexandru Oarga[1,*], Bridget P. Bannerman[2,3] and Jorge Júlvez [1,*]

[1]Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza 50018, Spain, [2]Lucy Cavendish College, Biological Sciences, University of Cambridge, Cambridge CB3 0BU, UK and [3]Science Resources Foundation, Health Unit, London EC1V 2NX, UK

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on February 16, 2022; revised on November 16, 2022; editorial decision on January 17, 2023; accepted on January 23, 2023

Abstract

Motivation: Despite the fact that antimicrobial resistance is an increasing health concern, the pace of production of new drugs is slow due to the high cost and uncertain success of the process. The development of high-throughput technologies has allowed the integration of biological data into detailed genome-scale models of multiple organisms. Such models can be exploited by means of computational methods to identify system vulnerabilities such as chokepoint reactions and essential reactions. These vulnerabilities are appealing drug targets that can lead to novel drug developments. However, the current approach to compute these vulnerabilities is only based on topological data and ignores the dynamic information of the model. This can lead to misidentified drug targets.

Results: This work computes flux constraints that are consistent with a certain growth rate of the modelled organism, and integrates the computed flux constraints into the model to improve the detection of vulnerabilities. By exploiting these flux constraints, we are able to obtain a directionality of the reactions of metabolism consistent with a given growth rate of the model, and consequently, a more realistic detection of vulnerabilities can be performed. Several sets of reactions that are system vulnerabilities are defined and the relationships among them are studied. The approach for the detection of these vulnerabilities has been implemented in the Python tool CONTRABASS. Such tool, for which an online web server has also been implemented, computes flux constraints and generates a report with the detected vulnerabilities.

Availability and implementation: CONTRABASS is available as an open source Python package at https://github.com/openCONTRABASS/CONTRABASS under GPL-3.0 License. An online web server is available at http://contrabass.unizar.es.

Contact: contrabass@unizar.es or julvez@unizar.es

Supplementary information: A glossary of terms are available at *Bioinformatics* online.

1 Introduction

Antimicrobial resistance (AMR) occurs when bacteria, viruses, fungi and parasites evolve to become unresponsive to conventional antibiotics, thus threatening our ability to treat common infections and increasing the risk of spread of severe illnesses. The emergence of multidrug-resistant bacteria (MDR) is particularly alarming, as they can cause infections untreatable with existing antibiotics. According to the World Health Organization, AMR is one of the primary global public health threats of humanity due to its high increasing rate (Leung *et al.*, 2011).

Despite the increasing emergency of AMR, the pipeline of drug development involves a huge cost of time and budget and, in addition, most drug candidates fail the clinical stages. Therefore, novel approaches for the development of drugs that tackle the rising problem of AMR should be established urgently (Jubeh *et al.*, 2020).

The emergence of high-throughput technologies has led to the construction of large datasets with structural and transcriptomic data of different pathogens. The integration of multidimensional data has the potential to offer a more rapid and cost-effective strategy compared to traditional screening methods in the process of drug discovery. In this context, the integration of metabolic functions obtained from genome annotations enabled the reconstruction of metabolic networks of multiple microorganisms. These reconstructions allow the modelling of metabolism, which can provide non-intuitive insights into biological systems that *in vivo* assays alone cannot provide (Ramos *et al.*, 2018; Richelle *et al.*, 2020).

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

© The Author(s) 2023. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023

from their environment. Metabolism can be represented as a metabolic network, which includes all the metabolic reactions that can occur in a cell. A possible strategy for drug development is to find vulnerabilities in the metabolism that can stop the growth and replication of the bacterium, and thus its damaging effects.

As of 2019, genome-scale models (GEMs) of metabolism have been reconstructed for more than 6000 organisms including bacteria, archaea and eukaryotes by either automatic or manual reconstruction (Gu *et al.*, 2019). GEMs have proven useful in a wide range of applications, such as expanding knowledge on microorganisms (Henry *et al.*, 2006; Navid and Almaas, 2012) and microbial communities (Kumar *et al.*, 2019; Zomorrodi and Maranas, 2012), microbial engineering (McAnulty *et al.*, 2012; Roberts *et al.*, 2010) and drug discovery (Ramos *et al.*, 2018). Furthermore, these models have proven to be interesting in areas such as oncology, by studying drug targets in cancer metabolism (Folger *et al.*, 2011), and viral diseases (Bannerman *et al.*, 2021). Drug targeting in pathogens is usually performed by considering essential genes, reactions or metabolites, whose inhibition can effectively kill a pathogen (Gu *et al.*, 2019) or through metabolic network topology analysis (Ramos *et al.*, 2018).

Despite the number of GEMs developed recently, most of these models lack genetic information, which could hamper the model analysis 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 critical reactions in metabolic networks such as essential reactions, dead reactions, and chokepoints (Chukualim *et al.*, 2008; Yeh *et al.*, 2004).

A reaction is essential if it is required by the organism to grow. The deletion of one of these reactions has the potential to cause a malfunction in the metabolism and stop the growth of the organism. As this eventually leads to the death of the organism, essential reactions are widely accepted as drug target candidates (Oyelade *et al.*, 2018).

Dead reactions are those reactions that cannot carry any flux. These reactions will receive special attention since they may reflect an incompleteness of the model or non-preferred pathways of the organism. As it will be shown, these reactions directly affect the number of critical reactions computed for a model.

A chokepoint is a reaction that is the only consumer or the only producer of a given metabolite. Thus, the inhibition of a chokepoint may lead to the unlimited accumulation of potentially toxic metabolites or the lack of production of an essential compound. Chokepoints are, therefore, appealing drug targets in microbial metabolism.

The current methods for the computation of chokepoints only make use of the topological properties of the metabolic network. Although the directionality of the network accounts for the sets of reactants and products of each reaction, it disregards the fact that such sets depend on the fluxes of the reactions.

Reaction fluxes define the direction in which they take place. To establish a particular direction, all reactions are given flux bounds that delimit the range of flux values that a reaction can take. Reactions that are given positive upper flux bounds and negative lower flux bounds are said to be reversible. These flux bounds are used to model reactions that can take place in both directions or simply to model reactions whose direction is unknown (Thiele and Palsson, 2010). For instance, let us say that a chemical reaction $r: A \leftrightarrow B$ is reversible. This means that r can happen in a forward direction ($r: A \rightarrow B$) when a positive flux is assigned, or a backward direction when a negative flux is assigned ($r: A \leftarrow B$). Notice that this means that metabolites in reversible reactions can act both as reactants and as products. In the reaction $r: A \rightarrow B$, metabolite A is the reactant and metabolite B is the product, however, if the flux of the reaction is negative, then A becomes the product and B the reactant.

In most GEMs, unknown flux bounds are given default flux bounds, e.g. $b(r) = -1000\,\text{mmol}\,\text{gDW}^{-1}\text{h}^{-1}$ and $\text{ub}(r) = 1000\,\text{mmol}\,\text{gDW}^{-1}\text{h}^{-1}$, that is, reactions are defined as reversible by default. However, not all these default flux bounds are compatible with a positive growth rate. This work exploits the flux constraints of the model to compute sets of reactants and products that are consistent with a given growth rate, and in turn, to compute growth-dependent critical reactions. Such approach improves the detection of vulnerabilities and is applied to compute chokepoints, essential reactions and dead reactions. The dependence of these sets with respect to growth is also inferred theoretically. To facilitate the computation of vulnerabilities in GEMs by integrating dynamic constraints, a software tool (*CONTRABASS*) has been developed.

The remainder of the article is organized as follows: Sections 2.1 and 2.2 recall preliminary concepts and definitions such as constraint-based models and essential reactions. Structural definitions of constraint-based models and flux-dependent definitions are recalled in Sections 2.3 and 2.4, respectively. In Section 2.5, growth-dependent definitions are introduced. Section 2.6 is devoted to dead reactions, blocked reactions and their relationship. Section 2.7 shows how the integration of dynamic constraints affects the computation of vulnerabilities in *Plasmodium falciparum* model. Finally, Section 3 reviews the implemented tool *CONTRABASS* and Section 4 concludes the article.

2 Systems and methods

2.1 Constraint-based models

This subsection recalls definitions (Oarga *et al.*, 2020) and methods that will be used in the next sections.

Definition 2.1. A *constraint-based model* (Orth *et al.*, 2011; Varma and Palsson, 1994) 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 $L[r] \leq U[r]$ $\forall r \in \mathcal{R}$.

All 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 $r \in \mathcal{R}$. Thus,

- if $S[m, r] < 0$ then m is a reactant and is consumed when r occurs.
- if $S[m, r] > 0$ then m is a product and is produced when r occurs.
- if $S[m, r] = 0$ then m is neither consumed nor produced when r occurs.

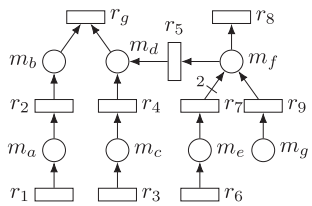
Constraint-based models can be expressed as bipartite graphs with two types of nodes representing reactions and metabolites. Hence, constraint-based models can be represented graphically as Petri nets (Heiner *et al.*, 2008; Murata, 1989), where places, drawn as circles, model metabolites, and transitions, drawn as squares, model reactions. 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. In other words, the stoichiometric matrix of a constraint-based model and the incidence matrix of its corresponding Petri net coincide.

Example 2.1.The Petri net in Figure 1 represents a simple constraint-based model that consists of 10 reactions and 7 metabolites. As an example, transition r_7 models the reaction $r_7: m_e \rightarrow 2m_f$.

Flux balance analysis (FBA) (Orth *et al.*, 2010) is a mathematical procedure for the estimation of steady-state fluxes in constraint-based models. Among other possibilities, FBA can be used to predict the growth rate of an organism or the rate of production of a given metabolite.

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023

(a)



(b)

r	$L(r)$	$U(r)$	type
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	100.0	NR
r_6	0.0	100.0	NR
r_7	0.0	100.0	NR
r_8	0.0	100.0	NR
r_9	0.0	100.0	NR
r_g	0.0	100.0	NR

Fig. 1. (**a**) Example Petri net modelling a constraint-based model. (**b**) Flux bounds of the model

Let $v \in \mathbb{R}^{|\mathcal{R}|}$ be the vector of fluxes of reactions and $v[r]$ denote the flux of reaction r. The system is assumed to be mass balanced at steady state, this is, the rate of production and rate of consumption of metabolites is constant. This constraint is given by the expression $S \cdot v = 0$, where S is the stoichiometric matrix. The steady-state mass balance fluxes of reactions, v, must also satisfy the lower and upper bounds L and U. Thus, the linear programming problem (LPP) for FBA is:

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

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

Let r_g be the reaction that models growth (or biomass production) in a constraint-based model. Without loss of generality, it will be assumed that $L[r_g] \geq 0$. An optimistic estimate for the growth rate of the modelled microorganism can be obtained by:

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

The solution of the above LPP (Equation (2)) will be denoted by $\mu_{\max}$.

2.2 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 2.2. (Oyelade *et al.*, 2018) A reaction $r \in \mathcal{R}$ is an *essential reaction* if the solution of the following LPP:

$$\begin{array}{ll} \max & v[r_g] \\ \text{st.} & S \cdot v = 0 \\ & L \leq v \leq U \\ & v[r] = 0 \end{array} \tag{3}$$

is equal to 0 or the LPP is infeasible.

The set of essential reactions, which is denoted ER, can be computed straightforwardly by solving Equation (3) for each $r \in \mathcal{R}$.

Example 2.2. Let us say that in the Petri net of Figure 1 reaction r_g models growth. In this model, reactions r_1 and r_2 are essential reactions. This is because, if the flux of any of these reactions is set to 0, then it is not possible to produce metabolite m_b, which is necessary for the growth reaction r_g.

2.3 Structural definitions

This section recalls how structural definitions of chokepoints and dead-end metabolites can be defined by making use of Petri net notation. Let X denotes a vertex of the net, i.e. a reaction or a metabolite. Then, ${}^\bullet X(X^\bullet)$ denotes the set of input (output) vertices of X. For instance, for a given reaction $r \in \mathcal{R}$, ${}^\bullet r$ denotes its set of reactants and $r^\bullet$ its set of products; for a given metabolite $m \in \mathcal{M}$, ${}^\bullet m$ denotes its set of producing reactions and $m^\bullet$ its set of consuming reactions.

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

Definition 2.3. (Chukualim *et al.*, 2008; Yeh *et al.*, 2004) 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 a chokepoint reaction can lead either to the infinite accumulation of a substrate (which will not be used as expected or might be toxic) or to the depletion of a metabolite (which might be essential for the cell) (Chukualim *et al.*, 2008; Yeh *et al.*, 2004). The identification of chokepoints is, therefore, relevant to the identification of potential drug targets.

A dead-end metabolite is a metabolite without input or output reactions:

Definition 2.4. (Mackie *et al.*, 2013) A metabolite $m \in \mathcal{M}$ is a dead-end metabolite if $m^\bullet = \{\}$ or ${}^\bullet m = \{\}$.

Dead-end metabolites can indicate a potential shortcoming or incompleteness of the model since their concentration can only increase or decrease.

2.4 Flux-dependent definitions

The previous definitions only take into account the structure of the network and disregard the flux bounds of the reactions. In order to capture the fact that reactions can proceed forwards or backwards, e.g. reversible reactions, new sets of reactants, products, consumers and producers that take into account flux bounds were defined (Oarga *et al.*, 2020):

- Flux-dependent set of reactants of r:

$${}^\star r = \{m \in \mathcal{M} | (S(m,r) < 0 \wedge U[r] > 0) \vee (S(m,r) > 0 \wedge L[r] < 0)\}$$

- Flux-dependent set of products of r:

$$r^\star = \{m \in \mathcal{M} | (S(m,r) > 0 \wedge U[r] > 0) \vee (S(m,r) < 0 \wedge L[r] < 0)\}$$

- Flux-dependent set of producers of m:

$${}^\star m = \{r \in \mathcal{R} | m \in r^\star\}$$

- Flux-dependent set of consumers of m:

$$m^\star = \{r \in \mathcal{R} | m \in {}^\star r\}$$

Flux-dependent chokepoints and dead-end metabolites can be defined accordingly (Oarga *et al.*, 2020):

Definition 2.5. A reaction $r \in \mathcal{R}$ is a flux-dependent chokepoint if there exists $m \in \mathcal{M}$ such that $m^\star = \{r\}$ or ${}^\star m = \{r\}$. The set of flux-dependent chokepoints is denoted $\mathrm{CP}_\star$.

Definition 2.6. A metabolite $m \in \mathcal{M}$ is a flux-dependent dead-end metabolite if $m^\star = \{\}$ or ${}^\star m = \{\}$. The set of flux-dependent dead-end metabolites is denoted $\mathrm{DEM}_\star$.

Example 2.3. In the Petri net in Figure 1, r_1 is a producer of m_a, i.e. $r_1 \in {}^\bullet m_a$; r_2 is a flux-dependent chokepoint because it is the only consumer of m_a, i.e. $m_a{}^\bullet = \{r_2\}$ and $r_2 \in \mathrm{CP}_\star$; and m_g is a flux-dependent dead-end metabolite, i.e. $m_g \in \mathrm{DEM}_\star$.

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023

In addition to chokepoints, flux bounds can be used to classify reactions as *dead*, *reversible* or *non-reversible*:

Definition 2.7. A reaction $r \in \mathcal{R}$ is dead if $L[r] = U[r] = 0$.

Definition 2.8. A reaction $r \in \mathcal{R}$ is reversible if $L[r] < 0 < U[r]$.

Definition 2.9. A reaction $r \in \mathcal{R}$ is non-reversible if r is not dead and r is not reversible.

From the above definitions, it can be deduced that r is non-reversible if $(0 \leq L[r] \wedge 0 < U[r]) \vee (L[r] < 0 \wedge U[r] \leq 0)$.

Notice that dead reactions will never have a non-zero flux. Since this might indicate a deficiency of the model or a blocked pathway of the organism, special attention will be paid to dead reactions.

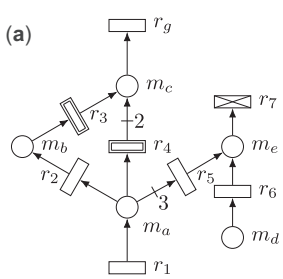
The sets of dead, reversible and non-reversible reactions are denoted DR, RR and NR, respectively. By definition, DR, RR and NR represent a partition of the set of reactions $\mathcal{R}$, i.e. $\mathrm{DR} \cup \mathrm{RR} \cup \mathrm{NR} = \mathcal{R}$, $\mathrm{DR} \cap \mathrm{RR} = \varnothing$, $\mathrm{RR} \cap \mathrm{NR} = \varnothing$ and $\mathrm{DR} \cap \mathrm{NR} = \varnothing$.

For the Petri net representation of a constraint-based model, dead reactions, reversible reactions and non-reversible reactions will be represented as rectangles with a cross inside, as double rectangles, and as rectangles, respectively.

Example 2.4. The Petri net in Figure 2a represents a constraint-based model with five metabolites and eight reactions. In this model, reactions r_3, r_4 are reversible reactions (i.e. $\mathrm{RR} = \{r_3, r_4\}$), reaction r_7 is a dead reaction (i.e. $\mathrm{DR} = \{r_7\}$) and reactions r_1, r_2, r_5, r_6, r_g are non-reversible reactions (i.e. $\mathrm{NR} = \{r_1, r_2, r_5, r_6, r_g\}$).

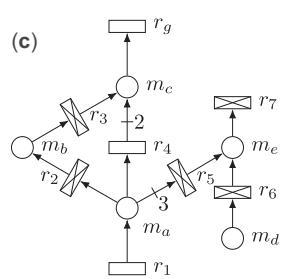
2.5 Growth-dependent definitions

In contrast to the previous flux-dependent definitions of reactions, this section introduces growth-dependent definitions, i.e. the sets of reactions that depend on the growth constraints that are imposed on the model.



(b)

r	$L(r)$	$U(r)$	type
r_1	0.0	100.0	NR
r_2	0.0	1000.0	NR
r_3	−1000.0	1000.0	NR
r_4	−1000.0	1000.0	RR
r_5	0.0	1000.0	NR
r_6	0.0	1000.0	RR
r_7	0.0	0.0	DR
r_g	0.0	1000.0	NR



(d)

r	$lb(r)$	$ub(r)$	type
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	0.0	0.0	DR
r_6	0.0	0.0	DR
r_7	0.0	0.0	DR
r_g	200.0	200.0	NR

Fig. 2. (**a**) Petri net modelling a constraint-based model. (**b**) Initial flux bounds. (**c**) Petri net modelling a constraint-based model obtained with $\gamma = 1$ and r_g as objective. (**d**) Constrained flux bounds obtained by FVA with $\gamma = 1$

2.5.1 Essential reactions

Similar to essential reactions, which are those reactions that are necessary to produce non-null growth on the model, growth-dependent essential reactions are those reactions that are necessary to produce a certain growth on the model. This certain growth will be expressed as $\gamma \cdot \mu_{\max}$ where $\gamma \in [0, 1]$ and $\mu_{\max}$ is the solution of Equation (2).

A reaction is said to be a growth-dependent essential reaction for a given growth $\gamma \cdot \mu_{\max}$ if its deletion implies that the maximum possible growth is below $\gamma \cdot \mu_{\max}$. More formally,

Definition 2.10. Let $\mu_{\max}$ be the solution of the LPP in Equation (2). Given $\gamma \in [0, 1]$, a reaction $r \in \mathcal{R}$ is a *growth-dependent essential reaction* if the solution of LPP (Equation (3)) is lower than $\gamma \cdot \mu_{\max}$ or the LPP is infeasible.

The set of growth-dependent essential reactions for a given growth specified by $\gamma \in [0, 1]$ will be denoted ER_γ. This set can be computed straightforwardly by solving LPP (Equation (3)) for each reaction.

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

Example 2.5. In the Petri net of Figure 1 where r_g models growth, reactions r_1, r_2, r_3, r_4 are EROG (i.e. $r_1, r_2, r_3, r_4 \in \mathrm{EROG}$). Reactions r_1, r_2 are essential reactions, and thus, they are also EROG. If one, or both reactions, r_3, r_4 are forced to have flux equal to 0, metabolite m, which is essential for growth, can only be produced through a non-optimal path. Then, the model will not be able to achieve optimal growth. Therefore, these reactions are EROG.

2.5.2 Dead, reversible and chokepoint reactions

Flux variability analysis (FVA) (Mahadevan and Schilling, 2003) is a mathematical procedure to compute the minimum and maximum fluxes of reactions that are compatible with some state. Let $\mu_{\max}$ be the maximum growth calculated by FBA (Equation (2)). FVA is computed by solving two independent LPPs 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 $L[r] \leq v[r] \leq U[r]$, and the maintenance of $\mu_{\max}$ to a certain degree $\gamma \in [0, 1]$. The two programming problems for a given reaction $r \in \mathcal{R}$ can be expressed as:

$$\begin{array}{ll} \max/\min & v[r] \\ \text{st.} & S \cdot v = 0 \\ & L \leq v \leq U \\ & \gamma \cdot \mu_{\max} \leq v[r_g] \end{array}, \tag{4}$$

where r_g is the reaction that models growth.

The computation of the flux bounds by means of FVA (Equation (4)), can be carried out in an optimal state, i.e. $\gamma = 1$, or in a suboptimal state, i.e. $0 \leq \gamma < 1$. In the optimal state, all fluxes must be optimally directed towards growth, whereas in suboptimal states, fluxes are allowed to deviate towards other functionalities.

Let $lb_\gamma, ub_\gamma \in \mathbb{R}^{|\mathcal{R}|}$ be the result of computing FVA (Equation (4)) 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 fluxes 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 $\gamma \in [0, 1]$, the sets of growth-dependent products, reactants, consumers, and producers of the model $\{\mathcal{R}, \mathcal{M}, \mathcal{S}, L, U\}$, which are denoted $r^\gamma, {}^\gamma r, m^\gamma, {}^\gamma m$, respectively, are defined as the flux-dependent products, reactants, consumers and products of $\{\mathcal{R}, \mathcal{M}, \mathcal{S}, lb_\gamma, ub_\gamma\}$ as discussed in Section 2.4.

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023

Similarly, given $\{\mathcal{R}, \mathcal{M}, \mathcal{S}, L, U\}$ and $\gamma \in [0,1]$, the sets of growth-dependent dead, reversible and non-reversible reactions, which are denoted DR_γ, RR_γ and NR_γ, the sets of growth-dependent chokepoint reactions and dead-end metabolites, which are denoted as CP_γ and DEM_γ, are also defined as the corresponding flux-dependent elements of $\{\mathcal{R}, \mathcal{M}, \mathcal{S}, lb_\gamma, ub_\gamma\}$.

Example 2.6. Let us assume that r_g in Figure 2a represents growth, i.e. the component of z in Equation (1) that corresponds to r_g is equal to 1 and the rest of components of z are 0. In order to obtain growth-dependent sets, the flux bounds computed by FVA with $\gamma = 1$ are assigned to the reactions and the net in Figure 2c is obtained (notice that other values of γ can be considered). In this new net, r_2, r_3, r_5, r_6 and r_7 are dead reactions, i.e. $r_2, r_3, r_5, r_6, r_7 \in \mathrm{DR}_1$, and r_4, which was a reversible reaction in Figure 2a, becomes a non-reversible reaction, i.e. $r_4 \in \mathrm{NR}_1$.

In the model of Figure 2a, reaction r_3 is the only flux-dependent consumer of metabolite m_b and, thus, is a flux-dependent chokepoint reaction, i.e. $r_3 \in \mathrm{CP}_\star$. However, in the growth-dependent model of Figure 2c, r_3 becomes a dead reaction, i.e. $r_3 \in \mathrm{DR}_1$, and hence, it is no longer considered a chokepoint reaction, i.e. $r_3 \notin \mathrm{CP}_1$.

On the other hand, reaction r_4 is not the only producer or consumer of any metabolite in the model of Figure 2a, and thus, it is not a flux-dependent chokepoint reaction, i.e. $r_4 \notin \mathrm{CP}_\star$. However, in the growth-dependent model of Figure 2c, reactions r_2, r_3 and r_5 become dead reactions and reaction r_4 becomes the only consumer of m_a and the only producer of m_c, consequently, r_4 becomes a growth-dependent chokepoint reaction $r_4 \in \mathrm{CP}_1$.

Given their importance in model tuning and drug discovery, the following sections focus on dead reactions, essential reactions and chokepoint reactions.

2.6 Dead reactions

This subsection explores the relationship between dead reactions and blocked reactions, and shows that the set of growth-dependent dead reactions is the same for any suboptimal state, i.e. for any growth strictly lower than the maximum growth $\mu_{\max}$. Such a set coincides with the set of blocked reactions. Recall that a reaction $r \in \mathcal{R}$ is said to be blocked if its flux is 0 at any possible steady state. More formally:

Definition 2.11. (Vlassis *et al.*, 2014) A reaction $r \in \mathcal{R}$ is a blocked reaction if for every $v \in \mathbb{R}^{|\mathcal{R}|}$ such that $S \cdot v = 0$, it holds $v[r] = 0$.

Example 2.7. In the Petri net in Figure 2a, reaction r_6 is a blocked reaction. This is because r_6 consumes m_d, which is a DEM. If the flux of r_6 were positive (negative), the amount of m_d would decrease(increase) indefinitely, which contradicts the steady-state constraint, therefore the only possible steady-state flux for r_6 is $v[r_6] = 0$.

In Vlassis *et al.* (2014) blocked reactions are obtained by solving the following LPPs that compute the maximum and minimum feasible fluxes of the reactions subject to the steady state constraint $S \cdot v = 0$:

$$\begin{array}{ll} \max/\min & v[r] \\ \text{st.} & S \cdot v = 0 \\ & L \leq v \leq U \end{array} \tag{5}$$

A reaction with null maximum and minimum feasible flux is a blocked reaction. Note that this procedure is equivalent to computing FVA, see Equation (4), with $\gamma = 0$. Thus, the set of blocked reactions is equal to DR_0.

Interestingly, the set of growth-dependent dead reactions, DR_γ is the same for any γ such that $0 \leq \gamma < 1$. In other words, the set of dead reactions in suboptimal states, regardless of the growth rate imposed on the model, is equivalent to the set of blocked reactions. This fact will be proved through several steps. Let us first prove that the range of feasible fluxes of a reaction r, i.e. the interval $[lb_\gamma[r], ub_\gamma[r]]$, cannot increase as γ increases.

Lemma 2.1. $[lb_{\gamma_2}[r], ub_{\gamma_2}[r]] \subseteq [lb_{\gamma_1}[r], ub_{\gamma_1}[r]] \, \forall r \in \mathcal{R} \wedge \forall \gamma_1, \gamma_2$ *such that* $0 \leq \gamma_1 < \gamma_2 \leq 1$.

Proof.
Given $\gamma \in [0,1]$, the range of feasible fluxes of $r \in \mathcal{R}$, $[lb_\gamma[r], ub_\gamma[r]]$, is given by the solutions of FVA (Equation (4)). Notice that the constraints of such LPP define a convex set of possible solutions which can only shrink as γ increases, i.e. as the constraint $\gamma \cdot \mu_{\max} \leq z \cdot v$ becomes more restrictive. Thus, if $\gamma_1 < \gamma_2$ then $lb_{\gamma_1}[r] \leq lb_{\gamma_2}[r]$ and $ub_{\gamma_1}[r] \geq ub_{\gamma_2}[r]$. □

Then, the set of growth-dependent dead reactions cannot decrease with γ:

Lemma 2.2.$\mathrm{DR}_{\gamma_1} \subseteq \mathrm{DR}_{\gamma_2} \, \forall \gamma_1, \gamma_2$ *such that* $0 \leq \gamma_1 < \gamma_2 \leq 1$.

Proof. Let $r \in \mathrm{DR}_{\gamma_1}$, i.e. $lb_{\gamma_1}[r] = ub_{\gamma_1}[r] = 0$, then by Lemma 2.1, it follows that $lb_{\gamma_2}[r] = ub_{\gamma_2}[r] = 0$ and hence $r \in \mathrm{DR}_{\gamma_2}$. □

Let us now show that the set of growth-dependent dead reactions cannot increase with γ in suboptimal states, i.e. with $\gamma < 1$:

Lemma 2.3. $\mathrm{DR}_{\gamma_1} \supseteq \mathrm{DR}_{\gamma_2} \, \forall \gamma_1, \gamma_2$ *such that* $0 \leq \gamma_1 < \gamma_2 < 1$.

Proof. The convex set of possible solutions defined by the constraints in Equation (4) can only decrease as γ increases, see Lemma 2.1. Assume there exist γ_a, γ_b such that $0 \leq \gamma_a < \gamma_b < 1$, $lb_{\gamma_a}[r] < ub_{\gamma_a}[r]$ and $lb_{\gamma_b}[r] = ub_{\gamma_b}[r] = 0$, i.e. r becomes dead when γ_a is increased to γ_b. Then, given that the constraints in Equation (4) are linear, if γ_b is further increased by $\epsilon \in \mathcal{R}$ such that $\epsilon > 0$ and $\gamma_b + \epsilon < 1$, then $lb_{\gamma_b+\epsilon}[r] > ub_{\gamma_b+\epsilon}[r]$ should hold which is not possible because a lower bound cannot exceed an upper bound. □

From Lemmas 2.2 and 2.3, the following theorem can be derived straightforwardly:

Theorem 2.4.$\mathrm{DR}_{\gamma_1} = \mathrm{DR}_{\gamma_2} \, \forall \gamma_1, \gamma_2 \in [0,1)$.

Thus, in particular, the set of blocked reactions coincides with the set of dead reactions in suboptimal states.

Corollary 2.5.$\mathrm{DR}_\gamma = \mathrm{DR}_0 \, \forall \gamma \in [0,1)$.

Notice that DR_1 can be strictly greater than DR_γ with $\gamma \in [0,1)$. As an example, the next subsection analyses a constrained-based model in which DR_1 is strictly greater than DR_γ with $\gamma \in [0,1)$, see Figure 3d.

2.7 Case study: *Plasmodium falciparum*

This section presents the results obtained for the constraint-based model of *P.falciparum* (iAM-Pf480) (Abdel-Haleem *et al.*, 2018). The model includes a total of 1083 reactions, 909 metabolites, and 480 genes. The sizes of the sets of flux-dependent reactions are $|\mathrm{RR}| = 493, |\mathrm{NR}| = 590, |\mathrm{DR}| = 0$. The number of initial flux-dependent chokepoint reactions is $|\mathrm{CP}_\star| = 453$. The sets of essential reactions and EROG were also computed, and the following sizes were obtained: $|\mathrm{ER}| = 192$ and $|\mathrm{EROG}| = 317$.

Figure 3 shows the sizes of the growth-dependent sets NR_γ, RR_γ, CP_γ, DR_γ, ER_γ, $\mathcal{R} - \mathrm{ER}_\gamma - \mathrm{DR}_\gamma$ in plots 3(a), 3(b), 3(c), 3(d), 3(e) and 3(f), respectively. To assess the impact of γ on these sets, different values of γ in the interval $[0,1]$ have been used. In addition to the sizes of the sets obtained with γ, the leftmost value (depicted in green) of plots 3(a), 3(b), 3(c) and 3(d) refers to the size of the flux-dependent set prior to FVA, i.e. NR, RR,$\mathrm{CP}_\star$ and DR. Similarly, the leftmost value (in green) of plots 3(e) and 3(f) includes

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023

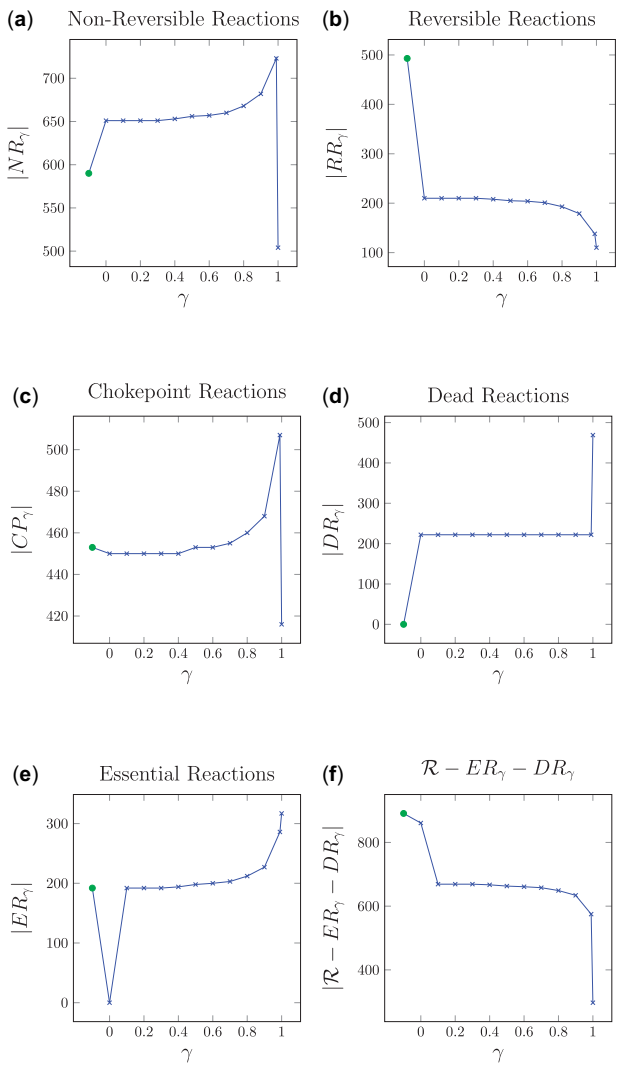


Fig. 3. Sizes of the sets of reactions NR_γ, RR_γ, CP_γ, DR_γ, ER_γ and $\mathcal{R} - ER_\gamma - DR_\gamma$ of *P.falciparum* for $\gamma \in [0, 1]$. The leftmost value of each plot corresponds to NR, RR, $CP_\star$, DR, ER and $\mathcal{R} - ER - DR$, respectively

the sizes of the set of essential reactions ER and the set $\mathcal{R} - ER - DR$, respectively.

Notice that if $\gamma = 0$, the constraint $\gamma \cdot \mu_{max} \leq z \cdot v$ in Equation (4) does not impose a minimum growth on the model, and only the steady state condition $S \cdot v = 0$ must be satisfied. It can be seen that at $\gamma = 0$, the set of dead reactions exhibits an increase from the initial set of flux-dependent dead reactions: $|DR| = 0$ and $|DR_0| = 222$. As noted in Section 2.6, this set DR_0 is equal to the set of blocked reactions, and keeps the same for every $\gamma \in [0, 1)$, see Figure 3d. The second increase in the set of dead reactions takes place at $\gamma = 1$ and it is due to the fact that in the optimal growth state, the flux must be necessarily distributed through optimal paths for biomass production and no flux can be diverted through other paths. Thus, non-optimal reactions for biomass production become dead reactions.

With respect to reversible reactions, the steady-state constraint $S \cdot v = 0$ reduces the size of this set from $|RR| = 493$ to $|RR_0| = 210$, see Figure 3b. Such a reduction is caused by the blocked reactions that belonged to RR and become dead reactions, and by the reversible reactions that become non-reversible with the steady-state constraint.

Similarly, blocked reactions that belonged to the set of flux-dependent non-reversible reactions NR become dead reactions with the steady-state constraint. However, due to the significant amount of reversible reactions that become non-reversible reactions, the size of the set increases from $|NR| = 590$ to $|NR_0| = 651$, see Figure 3a.

With regard to chokepoint reactions, see Figure 3c, the number of flux-dependent chokepoints is $|CP_\star| = 453$. At $\gamma = 0$, there is a decrease to $|CP_0| = 450$, and then the set increases slowly until $|CP_{0.99}| = 507$. As in the sets of non-reversible reactions and reversible reactions, the set of chokepoints decreases at $\gamma = 1$ as many reactions become dead reactions.

Notice that the set of chokepoints at $\gamma = 1$, CP_1, is smaller than the set of flux-dependent chokepoints, $CP_\star$. Moreover, CP_1 is not contained in $CP_\star$. This is due to the changes produced in the sets of non-reversible reactions and reversible reactions as γ increases. Recall that a reversible reaction is considered simultaneously both as a consumer and as a producer of a given metabolite. If this reaction becomes non-reversible, then it might not be a producer and consumer of a given metabolite. This fact can lead to the appearance of new growth-dependent chokepoints in CP_1 that were not chokepoints in $CP_\star$. On the other hand, dead reactions are neither consumers nor producers and, hence, they are not chokepoints. Thus, the increase of dead reactions in DR_1 can imply changes in CP_1 with respect to $CP_\star$.

Figure 3e reports the size of the set of essential reactions ER_γ with respect to γ. Recall that the leftmost value of this graph refers to the ER set, i.e. the set of essential reactions with no growth constraints.

Notice that no reaction is mandatory to produce a null growth, hence at $\gamma = 0$, $|ER_0|$ is always 0. Furthermore, for positive values of γ the size of ER_γ increases as the size of RR_γ decreases. This is due to the fact that, as γ increases, some reversible reactions become non-reversible, and the flux of these reactions is necessary, i.e. essential for growth. On the other hand, as expected, the amount of growth-dependent essential reactions increases with γ.

Finally, Figure 3f shows the size of the set $\mathcal{R} - ER_\gamma - DR_\gamma$ with respect to γ, with the leftmost point corresponding to $\mathcal{R} - ER - DR$. This set is of interest as it is composed of those reactions that can contribute to growth, that is, they are not dead reactions, and at the same time are not growth-dependent essential reactions, that is, if one of these reactions is knocked out the growth can still be kept the same (notice that this does not mean that all the reactions in this set can be knocked simultaneously without affecting growth). The fact that knocking out one of these reactions does not reduce the growth might then imply the presence of redundancy in the metabolism. Therefore, this set of reactions provides the metabolism with resilience and flexibility. As it can be seen, the size of this set decreases with γ. This fact means that producing higher growths is more demanding as it requires a higher amount of reactions to sustain it. In other words, the higher the growth the more reactions are necessary to produce such growth, and therefore, less flexibility is given to the metabolism.

3 Implementation

CONTRABASS is a software tool for the computation of vulnerabilities in GEMs. CONTRABASS is distributed as a Python command line tool but can also be executed through an online web server at http://contrabass.unizar.es. The CONTRABASS web server is designed to offer an intuitive interface to access the operations of the tool.

The tool takes as an input a model in systems biology markup language (SBML) (Hucka *et al.*, 2019) of Levels 1, 2 or 3, and computes the set of chokepoints reactions, essential reactions, dead reactions and dead-end metabolites on the model by taking into account the dynamic constraints for the model as explained in the article. The results are then exported as a spreadsheet file and as an interactive HTML report. The operations that CONTRABASS allows include the computation of sets of chokepoints, dead, reversible, non-reversible and essential reactions with different values of γ; computation and removal of dead-end metabolites from a model; and the update of the flux bounds of the reactions according to FVA.

In addition to the above, through the interactive HTML report users can also access the data available in the model, this is, reactions, genes and metabolites along with their databases identifiers if

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023

available; explore the reaction sets defined in this article; and also explore the intersection of different sets of critical reactions.

The documentation of the tool is available at https://contrabass.readthedocs.io. CONTRABASS makes use of the Python toolbox COBRApy (Ebrahim *et al.*, 2013). The source code of the command line tool and the web server is available at https://github.com/openCONTRABASS/CONTRABASS and https://github.com/openCONTRABASS, respectively. All the code is released under GPL-3.0 License.

4 Discussion

The development of new drugs is an expensive and costly process. This process needs to be streamlined in order to face current health threats as those posed by MDR. In this work, computational methods to identify critical reactions, i.e. vulnerabilities, in the metabolism of microorganisms are proposed. When applied on models of pathogenic bacteria, such critical reactions are potential drug targets.

The developed methods classify the reactions of the models in different sets, e.g. essential reactions, chokepoint reactions, blocked reactions, etc., according to their role in the metabolism. In order to exploit the information available in the model as much as possible, flux- and growth-dependent sets of reactions have been defined. Such sets capture the role of the reactions as a function of the growth rate of the organism. Thus, it is possible to unveil how the cell uses its reactions to achieve its growth. The relationship among these sets has been explored and it has been shown that the set of dead reactions keeps constant for any non-optimal growth rate. Moreover, the computation of other flux- and growth-dependent sets, such as dead-end metabolites and dead reactions, can help in the identification of potential shortcomings of the model. In order to facilitate their use, an online web server and a command line tool implementing the discussed methods have been made publicly available.

Funding

This work was partially 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). The work was also partially supported by the grant number [TED2021-130449B-I00] funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

Conflict of Interest: none declared.

References

Abdel-Haleem,A.M. *et al.* (2018) Functional interrogation of *Plasmodium* genus metabolism identifies species-and stage-specific differences in nutrient essentiality and drug targeting. *PLoS Comput. Biol.*, **14**, e1005895.

Bannerman,B.P. *et al.* (2021) Integrated human/SARS-COV-2 metabolic models present novel treatment strategies against COVID-19. *Life Sci. Alliance*, **4**, e202000954.

Chukualim,B. *et al.* (2008) Trypanocyc – a metabolic pathway database for *Trypanosoma brucei*. *BMC Bioinformatics*, **9** (Suppl 10), P5.

Ebrahim,A. *et al.* (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for python. *BMC Syst. Biol.*, **7**, 1–6.

Folger,O. *et al.* (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.*, **7**, 501.

Gu,C. *et al.* (2019) Current status and applications of genome-scale metabolic models. *Genome Biol.*, **20**, 1–18.

Heiner,M. *et al.* (2008) Petri nets for systems and synthetic biology, Vol. **5016**. pp. 215–264. http://dx.doi.org/10.1007/978-3-540-68894-5_7.

Henry,C.S. *et al.* (2006) Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.*, **90**, 1453–1461.

Hucka,M. *et al.* (2019) The systems biology markup language (SBML): language specification for level 3 version 2 core release 2. *J. Integr. Bioinformatics*, **16**, 20190021.

Jubeh,B. *et al.* (2020) Antibacterial prodrugs to overcome bacterial resistance. *Molecules*, **25**, 1543.

Kumar,M. *et al.* (2019) Modelling approaches for studying the microbiome. *Nat. Microbiol.*, **4**, 1253–1267.

Leung,E. *et al.*; World Health Organization World Health Day Antimicrobial Resistance Technical Working Group. (2011) The WHO policy package to combat antimicrobial resistance. *Bull. World Health Organ.*, **89**, 390–392.

Mackie,A. *et al.* (2013) Dead end metabolites-defining the known unknowns of the *E. coli* metabolic network. *PLoS One*, **8**, e75210.

Mahadevan,R. and Schilling,C.H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**, 264–276.

McAnulty,M.J. *et al.* (2012) Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. *BMC Syst. Biol.*, **6**, 1–15.

Murata,T. (1989) Petri nets: properties, analysis and applications. *Proc. IEEE*, **77**, 541–580.

Navid,A. and Almaas,E. (2012) Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst. Biol.*, **6**, 150.

Oarga,A. *et al.* (2020) Growth dependent computation of chokepoints in metabolic networks. In: *International Conference on Computational Methods in Systems Biology, Konstanz, Germany*, pp. 102–119. Springer, Heidelberg, Germany.

Orth,J.D. *et al.* (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.

Orth,J.D. *et al.* (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism – 2011. *Mol. Syst. Biol.*, **7**, 535.

Oyelade,J. *et al.* (2018) In silico knockout screening of *Plasmodium falciparum* reactions and prediction of novel essential reactions by analysing the metabolic network. *BioMed Res. Int.*, **2018**, 8985718.

Ramos,P.I.P. *et al.* (2018) An integrative, multi-omics approach towards the prioritization of *Klebsiella pneumoniae* drug targets. *Sci. Rep.*, **8**, 1–19.

Richelle,A. *et al.* (2020) Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. *NPJ Syst. Biol. Appl.*, **6**, 6.

Roberts,S.B. *et al.* (2010) Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst. Biol.*, **4**, 1–17.

Thiele,I. and Palsson,B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5**, 93–121.

Varma,A. and Palsson,B.Ø. (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *Nat. Biotechnol.*, **12**, 994–998.

Vlassis,N. *et al.* (2014) Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol.*, **10**, e1003424.

Yeh,I. *et al.* (2004) Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.*, **14**, 917–924.

Zomorrodi,A.R. and Maranas,C.D. (2012) OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput. Biol.*, **8**, e1002363.

Downloaded from https://academic.oup.com/bioinformatics/article/39/2/btad053/7000333 by guest on 11 April 2023