Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling

Lin Liu *[1] and Alexander Bockmayr †[1]

[1]Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

Abstract

Integrated modeling of metabolism and gene regulation continues to be a major challenge in computational biology. While there exist approaches like regulatory flux balance analysis (rFBA), dynamic flux balance analysis (dFBA), resource balance analysis (RBA) or dynamic enzyme-cost flux balance analysis (deFBA) extending classical flux balance analysis (FBA) in various directions, there have been no constraint-based methods so far that allow predicting the dynamics of metabolism taking into account both macromolecule production costs and regulatory events. In this paper, we introduce a new constraint-based modeling framework named *regulatory dynamic enzyme-cost flux balance analysis* (r-deFBA), which unifies dynamic modeling of metabolism, cellular resource allocation and transcriptional regulation in a hybrid discrete-continuous setting. With r-deFBA, we can predict discrete regulatory states together with the continuous dynamics of reaction fluxes, external substrates, enzymes, and regulatory proteins needed to achieve a cellular objective such as maximizing biomass over a time interval. The dynamic optimization problem underlying r-deFBA can be reformulated as a mixed-integer linear optimization problem, for which there exist efficient solvers.

1 Introduction

Constraint-based modeling approaches have become a powerful tool to analyze genome-scale metabolic network reconstructions [1, 2]. Based on the steady-state assumption for internal metabolites, constraint-based methods like *flux balance analysis* (FBA) [3] use the stoichiometry of the metabolic network to define a feasible solution space of possible steady-state flux distributions. By choosing an objective function such as maximizing biomass, an optimal flux distribution can be predicted by solving a linear optimization problem (LP).

While standard FBA requires very few data, it is not able to capture more complex phenomena such as dynamics, resource allocation, or gene regulation. Extending work by Palsson et al. [4], Mahadevan et al. in 2002 introduced *dynamic flux balance analysis* (dFBA) [5] to maximize biomass production over a time interval, taking into account the dynamics of extracellular metabolites and biomass. To incorporate the synthesis costs of macromolecules, Goelzer et al. developed *resource balance analysis* (RBA) [6, 7], which allows predicting an optimal resource allocation for maximizing the steady-state growth rate. Independently, Lerman et al. introduced *ME-models* [8, 9], a related approach for integrating metabolism and gene expression at steady-state. To combine these two ways of extending FBA, dynamics and resource allocation, several frameworks have been developed during the last years, which include *dynamic enzyme-cost FBA* (deFBA) [10], *conditional FBA* (cFBA) [11, 12], *dynamic resource balance analysis* (dRBA) [13] and *dynamicME* [14].

*linliu@zedat.fu-berlin.de

†Alexander.Bockmayr@fu-berlin.de

Approaches	No regulation		Regulation included	
	No enzyme costs	With enzyme costs	No enzyme costs	With enzyme costs
Static	FBA (1992) [21]	RBA (2011) [6] ME models (2012) [8]	SR-FBA (2007) [19] PROM (2010) [20]	
Iterative	dFBA (SOA) (1994) [4, 5]	dynamicME (2019) [14]	rFBA (2001) [15] iFBA (2008) [17]	idFBA (2008) [18]
Dynamic	dFBA (DOA) (2002) [5]	deFBA (2015) [10] cFBA (2015) [11] dRBA (2018) [13]		r-deFBA (this paper)

Table 1: Constraint-based flux balance approaches

Concerning integrated modeling of metabolism and regulation, there exist approaches such as *regulatory flux balance analysis* (rFBA) [15] and Flexflux [16], which combine Boolean or multi-valued logical rules for transcriptional regulation with a steady-state stoichiometric model of metabolism. Like the SOA variant of dynamic flux balance analysis [4, 5], these techniques iterate flux balance analysis by splitting the growth phase into discrete time steps. At each time step, the updated regulatory states are imposed as bounds on the reaction fluxes, while ignoring the costs for enzyme production. *Integrated FBA* (iFBA) [17] allows combining rFBA with a differential equation model for a specific subnetwork, while *integrated dynamic FBA* (idFBA) [18] brings together metabolism, regulation, and also signal transduction.

In addition to these iterative methods, there exist also static approaches for combining metabolism and gene regulation. *Steady-state regulatory flux balance analysis* (SR-FBA) [19] aims at studying the steady-state behaviors of a metabolic-regulatory network by adding Boolean rules to the linear constraints of FBA, resulting in a mixed-integer linear program [19]. *Probabilistic regulation of metabolism* (PROM) [20] makes use of microarray data sets to constrain the reaction upper bounds with a certain percentage of the maximal upper bound.

To summarize, we classify in Tab. 1 existing flux balance approaches according to whether or not they include dynamics, macromolecule production costs, and gene regulation. As can be seen from Tab. 1, there is currently no approach integrating all those features in a unifying framework.

In previous work [22], we introduced *metabolic-regulatory networks* (MRNs) to formalize the interplay of metabolism, macromolecule synthesis and gene regulation. To specify the dynamics of MRNs, we used a hybrid automata framework, combining continuous dynamics of metabolism with discrete control by regulatory events. In this formalization, the amounts of molecular species are represented by continuous variables. The discrete states of the system correspond to gene expression states of regulated proteins, which include regulatory proteins and regulated enzymes. In each discrete state, the continuous variables evolve according to a system of differential equations that is specific for this state. The guard conditions for the discrete state transitions depend on the amounts of the molecular species and associated thresholds.

In the present paper, we look at dynamic optimization or optimal control of the hybrid automata representing MRNs, which leads us to a new constraint-based modeling framework called *regulatory dynamic enzyme-cost flux balance analysis* (r-deFBA). Like in other flux balance approaches, we apply a quasi steady-state assumption for the internal metabolites. The resulting dynamic optimization problem can be transformed into a mixed-integer linear optimization problem (MILP), for which there exist efficient solvers.

The organization of this paper is as follows: We start in Sect. 2 by recalling the definition of MRNs and the hybrid modeling framework from [22]. In Sect. 3, we formally introduce r-deFBA by formulating the metabolic constraints, the regulatory constraints, and the resulting dynamic optimization problem. To illustrate our approach, we consider two biological applications. In Sect. 4 we analyze the self-replicator model already considered in [22]. Finally, in Sect. 5, we apply our approach to a model of core carbon metabolism, inspired from [15] and [10].

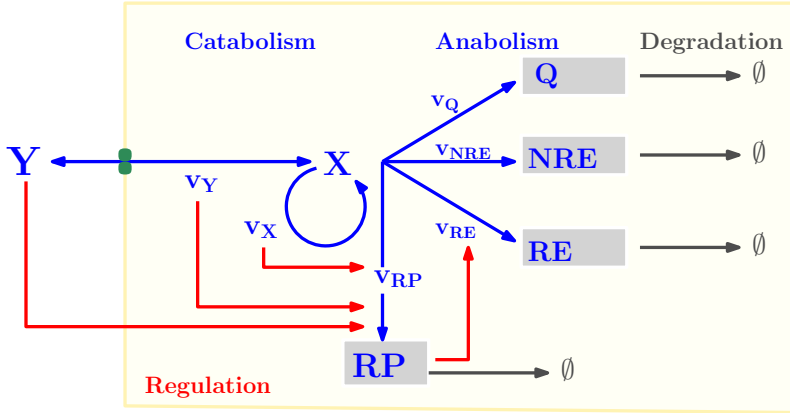

Fig. 1: Schematic model of a metabolic-regulatory network (MRN). We distinguish three types of molecular species: extracellular species $\mathbf{Y}$, intermediate metabolites $\mathbf{X}$, and macromolecules $\mathbf{P}$. The macromolecules $\mathbf{P} = \mathbf{Q} \cup \mathbf{RE} \cup \mathbf{NRE} \cup \mathbf{RP}$ are classified into quota compounds $\mathbf{Q}$, non-regulated enzymes $\mathbf{NRE}$, regulated enzymes $\mathbf{RE}$, and regulatory proteins $\mathbf{RP}$. The reaction fluxes $\mathbf{v}$ includes exchange fluxes $\mathbf{v_Y}$, intracellular fluxes $\mathbf{v_X}$ and macromolecule production fluxes $\mathbf{v_P} = \mathbf{v_Q} \cup \mathbf{v_{NRE}} \cup \mathbf{v_{RE}} \cup \mathbf{v_{RP}}$.

2 Hybrid dynamics of metabolic-regulatory networks

Metabolic-regulatory networks (MRNs) were introduced in [22] to model in an integrated way metabolic reactions, transcriptional regulation, macromolecule production and structural components. In Fig. 1, we illustrate the schematic structure of a MRN model. By $\mathbf{Y}$ we denote the set of extracellular species with corresponding exchange fluxes $\mathbf{v_Y}$. The intermediate metabolites $\mathbf{X}$ are transformed by intracellular fluxes $\mathbf{v_X}$ and utilized to build macromolecules $\mathbf{P} = \mathbf{Q} \cup \mathbf{E} \cup \mathbf{RP}$. For simplicity, we consider here only three types of macromolecules: non-catalytic compounds $\mathbf{Q}$ such as DNA and lipids, catalytic molecules $\mathbf{E}$ including enzymes and ribosomes, and regulatory proteins $\mathbf{RP}$. The corresponding production fluxes are denoted by $\mathbf{v_E}, \mathbf{v_Q}$ and $\mathbf{v_{RP}}$. Within $\mathbf{E} = \mathbf{RE} \cup \mathbf{NRE}$, we distinguish between regulated enzymes $\mathbf{RE}$, and non-regulated enzymes $\mathbf{NRE}$. Overall, the set of molecular species is defined as:

$$\mathcal{M} = \mathbf{Y} \cup \mathbf{X} \cup \mathbf{P} = \mathbf{Y} \cup \mathbf{X} \cup (\mathbf{Q} \cup \mathbf{E} \cup \mathbf{RP}) = \mathbf{Y} \cup \mathbf{X} \cup (\mathbf{Q} \cup \mathbf{RE} \cup \mathbf{NRE} \cup \mathbf{RP}) \quad (1)$$

Continuous dynamics

In a purely continuous modeling approach, the dynamics of the network would be described by a system of ordinary differential equations

$$\dot{\mathcal{M}}(t) = \frac{d\mathcal{M}(t)}{dt} = F(\mathcal{M}, \mathbf{K}, \mathbf{S}, t). \quad (2)$$

Following [10, 23, 22], we assume that $\mathcal{M}(t)$ denotes the molar amounts of the molecular species in $\mathcal{M}$ at time t. Furthermore, $\mathbf{K}$ is the set of kinetic parameters, $\mathbf{S}$ is the stoichiometric matrix, and t denotes time. The function F represents the kinetic laws that govern the dynamics, which could be mass action, Michaelis-Menten, Hill kinetics etc.

Discrete control

Continuous modeling of gene regulatory networks is known to be very difficult due to the lack of the necessary kinetic data. Therefore, we adopt a more qualitative approach to include regulation in our model. In this context, regulatory interactions, as illustrated by the red arrows in Fig. 1, refer to transcriptional regulation, i.e., we do not consider post-translational modifications. We assume that for each regulated protein $p \in \mathbf{RP} \cup \mathbf{RE}$ there are two possible states `on` and `off`, describing whether at a particular time t the gene encoding p is expressed or not.

Formally, we introduce a Boolean variable $\overline{p} = \overline{p}(t) \in \{0, 1\}$ and a logical function $f_p : \mathbb{R}^n \to \{0, 1\}$. Here, the Boolean value 0 corresponds to `off` and the value 1 to `on`. Each function f_p is defined as a Boolean combination (using the Boolean operations $\neg$ (not), $\wedge$ (and), $\vee$ (or)) of atomic formulas of the form $x \geq \theta$, where x is a real variable and θ a threshold value. Overall, the regulation of our MRN is then formalized by a system of Boolean equations of the form

$$\overline{p}(t) = f_p(\mathbf{RP}(t), \mathbf{Y}(t), \mathbf{v_X}(t), \mathbf{v_Y}(t)), \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE}. \tag{3}$$

The logical function f_p indicates how the expression state of the regulated protein p depends on the current amounts of regulatory proteins, extracellular metabolites, and the reaction fluxes.

As an example, consider a regulatory rule stating that an enzyme e is activated by a regulatory protein rp above a certain threshold $\theta > 0$. In this case, the expression state $\overline{e}$ of e is given by the rule $\overline{e}(t) = f_e(rp(t)) = 1$ if and only if $rp(t) \geq \theta$. Thus, whenever $rp(t) \geq \theta$ holds, we have $\overline{e}(t) = 1$ (= `on`), while $\overline{e}(t) = 0$ (= `off`) whenever $rp(t) < \theta$.

Note that in contrast to the metabolic-regulatory networks described in [22], the amounts of intermediate metabolites $\mathbf{X}(t)$ have been replaced with the flux values $\mathbf{v_X}(t)$ and $\mathbf{v_Y}(t)$. This is due to the quasi steady-state assumption for internal metabolites, which is typical for constraint-based modeling approaches, see also [15].

Hybrid system

Combining the continuous dynamics of metabolism in Eq. (2) with the discrete logical control in Eq. (3) leads to a hybrid discrete-continuous system, which we further explore in Sect. 3.

3 Formalization of r-deFBA

The r-deFBA framework that we propose in this paper aims at predicting from some initial conditions the continuous dynamics of metabolism and resource allocation together with discrete state transitions coming from genetic regulation. Compared with our earlier approach deFBA [10], regulatory logical constraints are included in addition to the metabolic constraints. Based on the schematic MRN model in Fig. 1 and the notation in Sect. 2, the metabolic and regulatory constraints of r-deFBA will now be described in detail.

3.1 Metabolic constraints

We start by recalling the constraints on metabolism, which are derived from dynamic enzyme-cost flux balance analysis (deFBA) [10, 23].

3.1.1 Dynamics of external substrates

The dynamics of the extracellular metabolites $\mathbf{Y}$ (nutrients and by-products) is modeled by a system of ordinary differential equations

$$\dot{\mathbf{Y}}(t) = -S_{\mathbf{Y},\mathbf{V_Y}} \cdot \mathbf{v_Y}(t), \tag{4}$$

where $S_{\mathbf{Y},\mathbf{V_Y}}$ is a stoichiometric matrix in which the rows correspond to the extracellular metabolites $\mathbf{Y}$ and the columns to the exchange reactions $\mathbf{v_Y}$. By $\mathbf{v_Y}(t)$ we denote the vector of exchange fluxes at time t.

3.1.2 Dynamic production of macromolecules

The macromolecules $\mathbf{P} = \mathbf{Q} \cup \mathbf{E} \cup \mathbf{RP}$ are assembled from metabolic precursors $\mathbf{X}$ to ensure cellular survival and growth. The synthesis and the degradation of macromolecules is described by a system of differential equations

$$\dot{\mathbf{P}}(t) = S_{\mathbf{P},\mathbf{V_P}} \cdot \mathbf{v_P}(t) - kd_{\mathbf{P}} \circ \mathbf{P}(t), \tag{5}$$

where $S_{\mathbf{P},\mathbf{V_P}}$ is a stoichiometric matrix in which the rows represent macromolecules and the columns macromolecule synthesis reactions. The vector $kd_{\mathbf{P}}$ contains the degradation rates and $\circ$ denotes the component-wise product of vectors.

3.1.3 Steady-state of intermediate metabolites

For the intermediate metabolites $\mathbf{X}$ we assume that they are in quasi steady-state, i.e., the rate of production is equal to the rate of consumption. This leads to a system of algebraic equations

$$S_{\mathbf{X},\mathbf{V_Y}} \cdot \mathbf{v_Y}(t) + S_{\mathbf{X},\mathbf{V_X}} \cdot \mathbf{v_X}(t) - S_{\mathbf{X},\mathbf{V_P}} \cdot \mathbf{v_P}(t) = 0, \tag{6}$$

with stoichiometric matrices $S_{\mathbf{X},\mathbf{V_Y}}, S_{\mathbf{X},\mathbf{V_X}}, S_{\mathbf{X},\mathbf{V_P}}$ and fluxes $\mathbf{v_Y}, \mathbf{v_X}, \mathbf{v_P}$.

3.1.4 Biomass composition constraint

In order to guarantee a sufficient production of non-catalytic macromolecules $\mathbf{Q}$ such as lipids and DNA, which are indispensable for cell growth and proliferation, we require that the mass of these quota compounds has to be at least a given fraction of the total biomass. Mathematically,

$$b_{\mathbf{Q}}^T \cdot \mathbf{Q}(t) \geq \ \Phi_{\mathbf{Q}} \cdot b_{\mathbf{P}}^T \cdot \mathbf{P}(t). \tag{7}$$

Here, $b_{\mathbf{P}}$ is a vector with the molecular weights of all the macromolecules $\mathbf{P}$, $b_{\mathbf{Q}}$ is the subvector of the molecular weights of the quota compounds $\mathbf{Q}$, the operation $\cdot^T$ denotes transposition, and $0 < \Phi_{\mathbf{Q}} < 1$ is a constant.

3.1.5 Enzymatic and translational capacity constraints

Fluxes through enzyme-catalyzed reactions are bound by the amount of the corresponding enzymes. If an enzyme catalyzes more than one reaction, the sum of all the reaction fluxes is limited by the enzyme amount. Formally, we get

$$\sum_{i \in cat(E)} \frac{|\mathbf{v}_i(t)|}{k_{cat}^{i,E}} \leq E(t), \text{ for all } E \in \mathbf{E}, \tag{8}$$

where $cat(E)$ is the set of all reactions i catalyzed by enzyme E and $k_{cat}^{i,E}$ is the corresponding turnover rate. Note that this constraint also holds for protein translation and the ribosome, which is considered to be a special enzyme in our framework (cf. Sect. 2).

3.2 Regulatory logical control constraints

Extending the existing approaches for dynamic metabolic resource allocation such as deFBA, we now add two types of regulatory constraints. The first one describes the control of the discrete state transitions by the continuous variables, which corresponds to the triggering of the discrete jumps in the hybrid system. The second one is the control of the evolution of the continuous variables depending on the discrete state. Taken together, these two types of regulatory constraints specify the interplay between cellular regulation and metabolism.

3.2.1 Control of discrete jumps

The key to the discrete dynamics of a hybrid system is how discrete state transitions are triggered by the continuous variables. According to Sect. 2, the expression state $\overline{p}(t)$ of all regulated proteins p is determined by a logical function $f_p : \mathbb{R}^n \to \{0, 1\}$ depending on the amounts of regulatory proteins, extracellular metabolites and reaction fluxes, i.e.,

$$\overline{p}(t) = f_p(\mathbf{RP}(t), \mathbf{Y}(t), \mathbf{v_X}(t), \mathbf{v_Y}(t)), \tag{9}$$

for all $p \in \mathbf{RP} \cup \mathbf{RE}$.

3.2.2 Control of the continuous dynamics by the discrete states

While the regulatory constraints (3) describe how the discrete states depend on the continuous variables, we also have to specify how the continuous dynamics depends on the discrete state.

For a regulated protein p, the value of $\overline{p}$ determines whether protein p is expressed or not. Therefore, if $\overline{p}(t) = 1$, we impose as constraint that the production flux $\mathbf{v}_p(t)$ should be at least ε_p, while we require $\mathbf{v}_p(t)$ to be zero in the case $\overline{p}(t) = 0$. More formally, we get for all $p \in \mathbf{RP} \cup \mathbf{RE}$ the implications

$$\overline{p}(t) = 1 \quad \Rightarrow \quad \mathbf{v}_p(t) \geq \varepsilon_p, \tag{10}$$

$$\overline{p}(t) = 0 \quad \Rightarrow \quad \mathbf{v}_p(t) = 0, \tag{11}$$

where $\varepsilon_p > 0$ is a lower bound for the production rate $\mathbf{v}_p(t)$ of p in state `on`. We could also allow $\varepsilon_p = 0$ if we want to relax the model and determine the values of $v_p(t)$ by optimization like in deFBA.

Note that the values of the parameters ε_p significantly influence the dynamics of the system. Since the lower bounds constrain the production rates, they directly affect the abundances of the regulatory proteins and the regulated enzymes. Conversely, p is strictly constrained to be degraded whenever $\overline{p}(t) =$ `off`.

3.3 Formulating r-deFBA as a dynamic optimization problem

Up to now, we have specified the metabolic constraints defining a dynamic solution space for cellular metabolism. In addition, we introduced the regulatory constraints to incorporate the dynamic interplay between gene regulation and metabolism. In order to predict how the cell can achieve optimal growth under these constraints, we formulate r-deFBA as a dynamic optimization problem, see Eq. 12. The objective is to compute time courses $\mathbf{v}(t), \mathbf{P}(t), \mathbf{Y}(t), \overline{p}(t)$ that maximize the total biomass production in a given time interval $[t_0, t_f]$.

$$\begin{aligned}
\max_{\mathbf{v}(t),\mathbf{P}(t),\mathbf{Y}(t),\overline{p}(t)} \quad & \int_{t_0}^{t_f} b_{\mathbf{P}}^T \, \mathbf{P}(t) dt \\
s.t. \quad & \dot{\mathbf{Y}}(t) = -S_{\mathbf{Y},\mathbf{V_Y}} \cdot \mathbf{v_Y}(t), \\
& \dot{\mathbf{P}}(t) = S_{\mathbf{P},\mathbf{V_P}} \cdot \mathbf{v_P}(t) - kd_{\mathbf{P}} \circ \mathbf{P}(t), \\
& S_{\mathbf{X},\mathbf{V_Y}} \cdot \mathbf{v_Y}(t) + S_{\mathbf{X},\mathbf{V_X}} \cdot \mathbf{v_X}(t) - S_{\mathbf{X},\mathbf{V_P}} \cdot \mathbf{v_P}(t) = 0, \\
& b_{\mathbf{Q}}^T \cdot \mathbf{Q}(t) \geq \ \Phi_{\mathbf{Q}} \cdot b_{\mathbf{P}}^T \cdot \mathbf{P}(t), \\
& \sum_{i \in cat(E)} \frac{|\mathbf{v}_i(t)|}{k_{cat}^{i,E}} \leq E(t), \text{ for all } E \in \mathbf{E} \\
& \overline{p}(t) = f_p(\mathbf{RP}(t), \mathbf{Y}(t), \mathbf{v_X}(t), \mathbf{v_Y}(t)), \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE} \\
& \overline{p}(t) = 1 \Rightarrow \mathbf{v}_p(t) \geq \varepsilon_p, \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE} \\
& \overline{p}(t) = 0 \Rightarrow \mathbf{v}_p(t) = 0, \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE} \\
& \mathbf{Y}(t_0) = \mathbf{Y}_0, \quad \overline{p}(t_0) = \overline{p}_0, \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE} \\
& \mathbf{P}(t), \mathbf{Y}(t), \mathbf{v_P}(t) \geq 0, \quad \mathbf{v}_{\min} \leq \mathbf{v}(t) \leq \mathbf{v}_{\max}, \\
& \overline{p}(t) \in \{0, 1\}, \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE}.
\end{aligned} \tag{12}$$

By $\mathbf{Y}_0$ and $\overline{p}_0$ we denote the initial values of $\mathbf{Y}(t)$ and $\overline{p}(t)$ at time $t = t_0$. In Eq. (12), the initial amounts $\mathbf{P}(t_0)$ are variables whose values are determined by the dynamic optimization. Alternatively, initial values for $\mathbf{P}(t_0)$ could also be precomputed using RBA.

Involving discrete and continuous variables, the r-deFBA problem in Eq. (12) can be reformulated as a mixed-integer linear optimization problem (MILP), for which there exist efficient solvers. To solve r-deFBA numerically, the dynamic real and Boolean variables are discretized in time like in [12]. The Boolean equations (9) and the logical implications (10)-(11) can be transformed into a system of linear 0-1 inequalities using a standard recursive substitution procedure [19, 24].

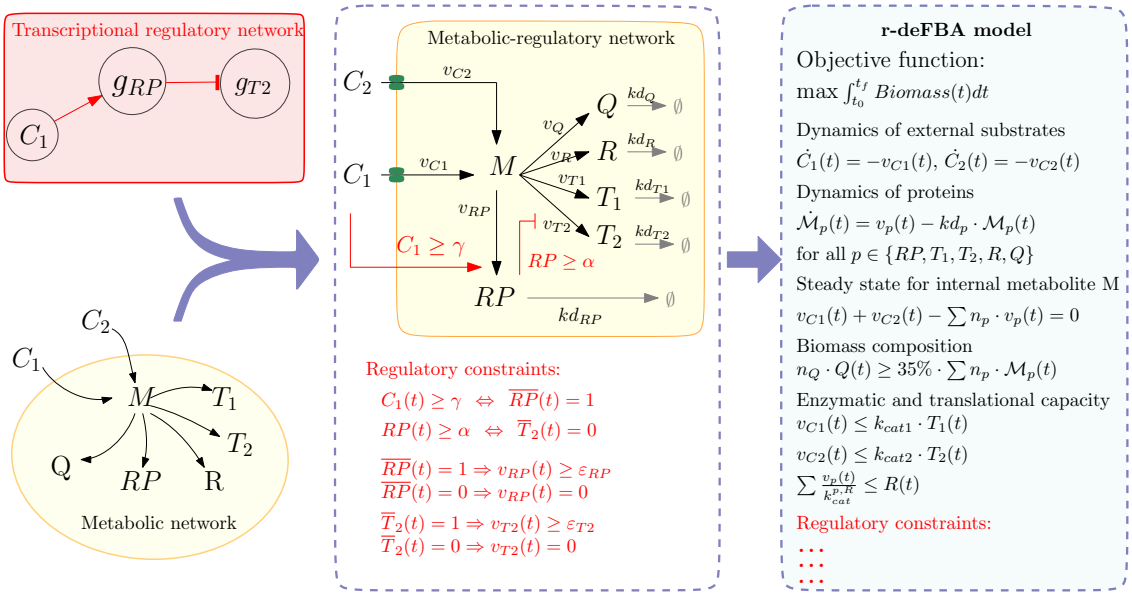

Fig. 2: Workflow to build an r-deFBA model for the regulatory self-replicator with two regulatory rules. The Boolean variables $\overline{RP}, \overline{T}_2$ describe the expression state of the genes g_{RP}, g_{T_2}, which determines the activity of the production reactions v_{RP}, v_{T2}. The thresholds $\varepsilon_{RP}, \varepsilon_{T_2}$ define the minimal expression levels for the regulated proteins RP, T_2 to be in state `on`.

4 Case study 1: Regulatory self-replicator model

Carbon catabolite repression is a common phenomenon in bacteria, especially in *Escherichia coli* [25]. While these bacteria are able to grow on different carbon sources, they do not consume these in parallel, but one after the other. This is called diauxic growth and was described by Monod already in 1942. Mathematical modeling of diauxie has played an important role in understanding these phenomena [26].

As a possible model for diauxie, we built in [22] a small regulatory self-replicator network, extending earlier work in [15, 27]. To illustrate r-deFBA, we construct in Sect. 4.1 an r-deFBA model for this network. In Sect. 4.2 we compare the resulting dynamics for r-deFBA to standard deFBA and to the hybrid automata framework considered in [22].

4.1 Regulatory self-replicator network

The general workflow for building an r-deFBA model is illustrated in Fig. 2. Starting from a metabolic and a transcriptional regulatory network, we first construct a metabolic-regulatory network (MRN), as presented in Sect. 2.

In the metabolic network of Fig. 2, we have two carbon sources $\mathbf{Y} = \{C_1, C_2\}$, which are converted into precursor molecules $\mathbf{X} = \{M\}$. For simplicity, we assume only two uptake reactions $C_1 \to M, C_2 \to M$, catalyzed by enzymes T_1 resp. T_2. The precursor molecules M are used to synthesize five types of macromolecules $\mathbf{P} = \{Q, R, T_1, T_2, RP\}$, which are the enzymes T_1, T_2, regulatory proteins RP, housekeeping proteins Q, and ribosomes R. The stoichiometry of the synthesis reactions and corresponding parameter values are given in Tab. 2. The total biomass $Biomass(t)$ is defined as the sum of the molecular masses

$$Biomass(t) = w \cdot M(t) + \sum_{p \in \mathbf{P}} w \cdot n_p \cdot \mathcal{M}_p(t), \tag{13}$$

where w corresponds to the molar weight of one precursor molecule M and n_p is the number of

precursor molecules needed to build one macromolecule p. By $M(t)$ and $\mathcal{M}_p(t)$ we denote again the molar amounts [mmol] of M resp. $p \in \mathbf{P}$ at time t.

In the regulatory network of Fig. 2, g_{RP} and g_{T_2} denote two genes encoding the regulated proteins RP and T_2. We assume that g_{RP} is activated by the presence of C_1 and that g_{T_2} is inhibited by g_{RP}. This leads to two regulatory rules

$$\overline{RP}(t) = 1 \;\Leftrightarrow\; C_1(t) \geq \gamma \quad \text{and} \quad \overline{T}_2(t) = 0 \;\Leftrightarrow\; RP(t) \geq \alpha, \tag{14}$$

with thresholds $\alpha, \gamma > 0$. The expression states $\overline{RP}, \overline{T}_2$ are linked to the flux variables v_{RP}, v_{T2} by the implications

$$\begin{array}{llll} \overline{RP}(t) = 1 \;\Rightarrow\; v_{RP}(t) \geq \varepsilon_{RP}, & & \overline{T}_2(t) = 1 \;\Rightarrow\; v_{T2}(t) \geq \varepsilon_{T2}, \\ \overline{RP}(t) = 0 \;\Rightarrow\; v_{RP}(t) = 0, & & \overline{T}_2(t) = 0 \;\Rightarrow\; v_{T2}(t) = 0, \end{array} \tag{15}$$

using thresholds $\varepsilon_{RP}, \varepsilon_{T2} > 0$.

r-deFBA and deFBA model

The full r-deFBA model of the regulatory self-replicator in Fig. 2 reads

$$\begin{array}{ll} \max\limits_{v(t), C_1(t), C_2(t), \mathcal{M}(t)} & \displaystyle\int_{t_0}^{t_f} \Big(\sum_{p \in \mathbf{P}} w \cdot n_p \cdot \mathcal{M}_p(t) + w \cdot M(t_0)\Big)\, dt \\ s.t. & \dot{C}_1(t) = -v_{C1}(t), \quad \dot{C}_2(t) = -v_{C2}(t), \\ & \mathcal{M}_p(t) = v_p(t) - kd_p \cdot \mathcal{M}_p(t), \quad \text{for all } p \in \mathbf{P} \\ & v_{C1}(t) + v_{C2}(t) - \displaystyle\sum_{p \in \mathbf{P}} n_p \cdot v_p(t) = 0, \\ & n_Q \cdot Q(t) \geq \Phi_Q \cdot \displaystyle\sum_{p \in \mathbf{P}} n_p \cdot \mathcal{M}_p(t), \\ & v_{C1}(t) \leq k_{cat1} \cdot T_1(t), \quad v_{C2}(t) \leq k_{cat2} \cdot T_2(t), \\ & \displaystyle\sum_{p \in \mathbf{P}} \frac{v_p(t)}{k_{cat}^{p,R}} \leq R(t), \\ & \overline{RP}(t) = 1 \;\Leftrightarrow\; C_1(t) \geq \gamma, \quad \overline{T}_2(t) = 0 \;\Leftrightarrow\; RP(t) \geq \alpha, \\ & \overline{RP}(t) = 1 \;\Rightarrow\; v_{RP}(t) \geq \varepsilon_{RP}, \quad \overline{T}_2(t) = 1 \;\Rightarrow\; v_{T2}(t) \geq \varepsilon_{T2}, \\ & \overline{RP}(t) = 0 \;\Rightarrow\; v_{RP}(t) = 0, \quad \overline{T}_2(t) = 0 \;\Rightarrow\; v_{T2}(t) = 0, \\ & (C_1, C_2, RP, T_1, T_2, R, Q)(t) \geq 0, \quad \overline{RP}(t), \overline{T}_2(t) \in \{0, 1\}, \\ & \qquad \text{for all } t \in [t_0, t_f], \\ & (C_1, C_2, M, RP, T_1, T_2, R, Q)(t_0) = \\ & \qquad (1000, 500, 20, 0, 0.001, 0.001, 0.01, 0.15), \\ & (\overline{RP}, \overline{T}_2)(t_0) = (1, 1). \end{array} \tag{16}$$

For the computations, we used the parameter values given in Tab. 2 and Tab. 3.
The corresponding deFBA model is obtained by omitting the regulatory constraints.

Hybrid automaton

In [22], we also constructed a hybrid automaton for simulating the metabolic-regulatory network. The discrete states or locations correspond to the different 0-1 states of the Boolean variables $\overline{RP}$ and $\overline{T}_2$, which means there are $2 \times 2 = 4$ discrete states. For the uptake of C_1, C_2 we assume a Michaelis-Menten kinetics

$$v_{C_1} = \frac{k_{cat1} \cdot C_1 \cdot T_1}{K_T + C_1}, \quad v_{C_2} = \frac{k_{cat2} \cdot C_2 \cdot T_2}{K_T + C_2}. \tag{17}$$

Metabolic reaction	Flux	Enzyme	Turnover rate [1/min]	
$C_1 \to M$	v_{C1}	T_1	$k_{cat1} = 3000$	
$C_2 \to M$	v_{C2}	T_2	$k_{cat2} = 2000$	
Biomass reaction	Flux	Enzyme	Turnover rate	Degradation rate
$n_p M \to p$	v_p		$k_{cat}^{p,R} = k_r/n_p$ [1/min]	kd_p [1/min]
$300M \to Q$	v_Q	R	4.2	0.01
$7459M \to R$	v_R	R	0.1689	0.01
$400M \to T_1$	v_{T1}	R	3.15	0.01
$1500M \to T_2$	v_{T2}	R	0.84	0.01
$300M \to RP$	v_{RP}	R	4.2	0.2

Table 2: Metabolic and biomass reactions with corresponding parameters

t_0	t_f	w	Φ_Q	α	γ	ε_{RP}	ε_{T2}	K_T	k_r	K_r
0	55	100	0.35	0.03	20	0.01	0.01	1000	1260	7
min	min	mg/mmol		mmol	mmol	mmol/min	mmol/min	mmol	1/min	

Table 3: Additional parameters. Here, k_r denotes the elongation rate and K_r, K_T are Michaelis constants.

The parameter values can be found again in Tab. 2 and Tab. 3. Regarding the synthesis rate v_p of macromolecules p consisting of n_p precursors we assume a Michaelis-Menten type kinetics of the form

$$v_p = \frac{\beta_p}{n_p} \cdot v_M, \text{ with } v_M = \frac{k_r \cdot M \cdot R}{K_r + M} \text{ and } p \in \mathbf{P}. \tag{18}$$

The weights $\beta_p \geq 0$ denote the fraction of cellular resources allocated to protein p, with $\sum_p \beta_p = 1$. For each location of the hybrid automaton, we assume for simplicity that the cellular resources are shared equally between the proteins that are expressed at this location. In other words, $\beta_p = 1/3$ in location $(\overline{RP}, \overline{T}_2) = (0, 0)$, $\beta_p = 1/4$ in location $(1, 0)$ resp. $(0, 1)$, and $\beta_p = 1/5$ in location $(1, 1)$. For further details, we refer to [22].

4.2 Comparing r-deFBA, deFBA, and the hybrid automaton

Next we compare the dynamics of the regulatory self-replicator obtained by r-deFBA, deFBA and the hybrid automaton, see Fig. 3. In all three simulations, we use the same parameter values, given in Tables 2 and 3, and the initial conditions from Eq. (16).

The diauxic shift is predicted successfully by all three approaches. However, the underlying principles are different. By maximizing the biomass production while taking into account only the metabolic constraints, deFBA shows that the diauxic shift is an optimal metabolic behavior. In contrast, r-deFBA computes an optimal trajectory for biomass production, taking into account both the metabolic and the regulatory constraints. Due to the additional regulatory constraints, r-deFBA produces less biomass than deFBA and needs more time to consume the available carbon resources, see Fig. 3D. The continuous metabolic variables of the hybrid automaton evolve according to the Michaelis-Menten kinetics of Eq. (17) and (18). These kinetics depend on the current discrete state, which in turn is determined by the regulatory control, i.e., the jump conditions. As an optimal control strategy for the hybrid system representing the MRN, r-deFBA clearly gains more biomass than the hybrid automaton, but less than deFBA, which does not include regulation.

Both r-deFBA and the hybrid automaton successfully predict the discrete state transitions during diauxie. In Fig. 3A and 3B, the time profiles are divided into three growth phases, corresponding to the discrete state transitions. The transitions of r-deFBA are consistent with those obtained by the hybrid automaton. In the first growth phase (a), expression of RP is activated and $\overline{RP} = \texttt{on}$ because initially $C_1 \geq \gamma$. We also have $\overline{T}_2 = \texttt{on}$ because RP is initialized by 0. Thus, the initial state of the network is $(\overline{RP}, \overline{T}_2) = (\texttt{on}, \texttt{on})$. With time going on, C_1 is consumed

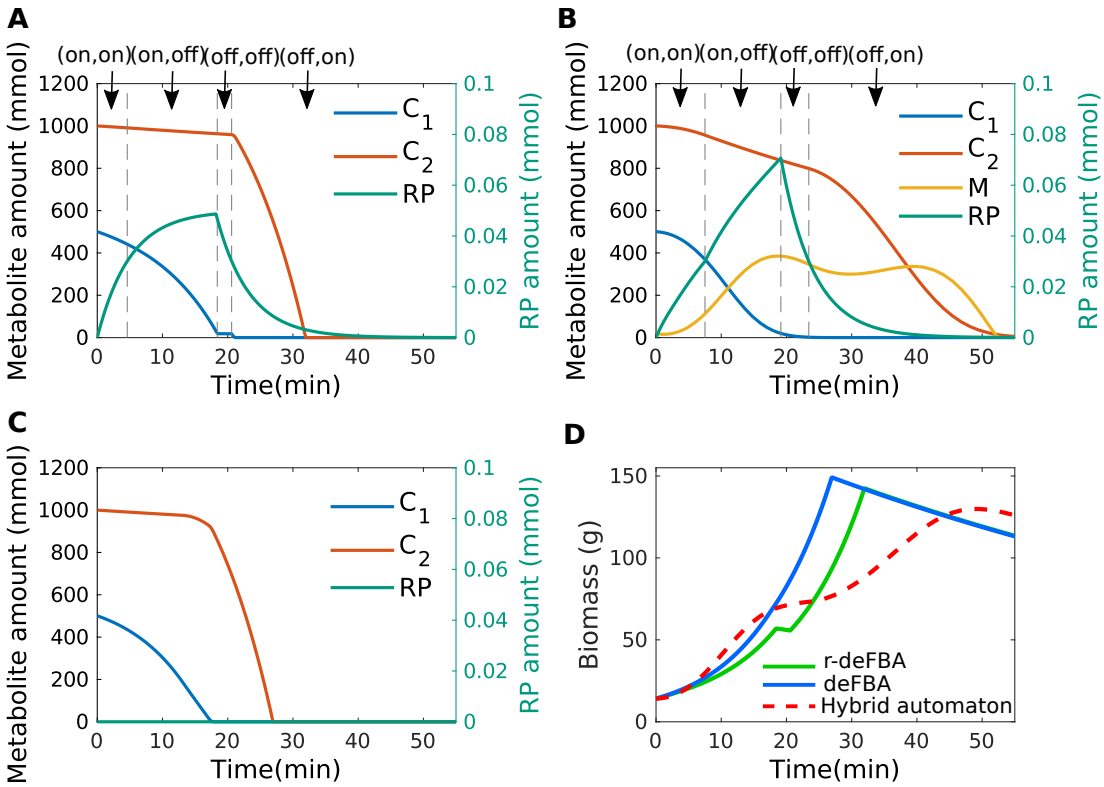

Fig. 3: Time courses of C_1, C_2 (left axis) and RP (right axis) predicted by r-deFBA (**A**), the hybrid automaton (**B**), deFBA (**C**) and corresponding biomass production(**D**). In all three simulations, the same parameter and initial values were used. For r-deFBA and the hybrid automaton, we also indicate the discrete states $(\overline{RP}, \overline{T}_2)$, with the transitions marked by vertical dashed lines. Due to the quasi steady-state assumption, the molar amount of the precursor M remains constant in deFBA and r-deFBA.

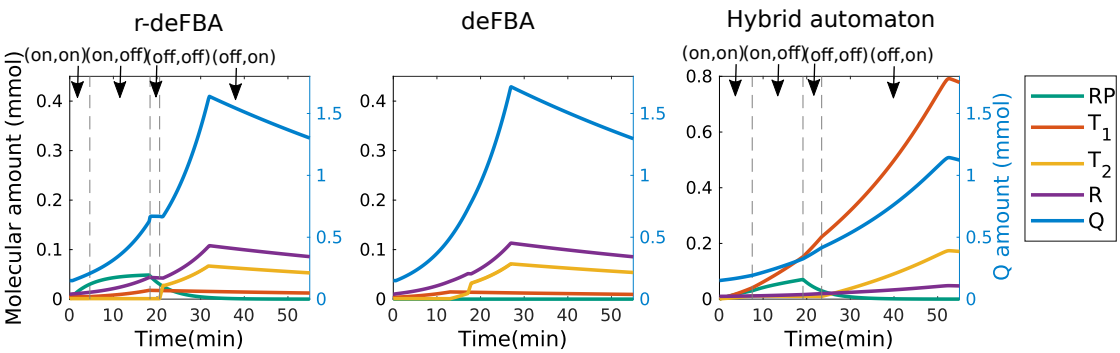

Fig. 4: Time courses of RP, T_1, T_2, R (left axis) and Q (right axis) for r-deFBA, deFBA, and the hybrid automaton.

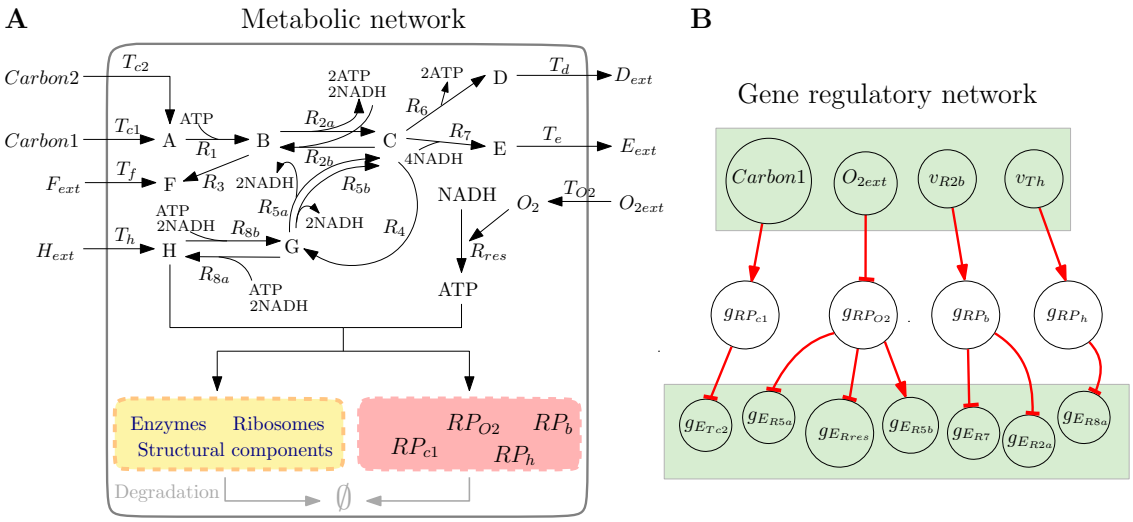

Fig. 5: Metabolic network of core carbon metabolism (including macromolecule production) and corresponding gene regulatory network.

while RP is synthesized and accumulated. When RP reaches the threshold α, the synthesis of T_2 is inhibited and the model jumps to the state $(\texttt{on}, \texttt{off})$. Next, when $C_1 < \gamma$, the discrete state changes to $(\texttt{off}, \texttt{off})$, which represents the lag phase during diauxie. In this phase, enzyme T_2 is still repressed until RP falls below its threshold α. Once this happens, the system switches to the final state $(\texttt{off}, \texttt{on})$, where $RP < \alpha$ and T_2 is produced to metabolize C_2. Overall, the interactions between discrete regulation and continuous metabolism are correctly incorporated in our r-deFBA. In deFBA, the regulatory protein RP remains at the initial value 0, see Fig. 3C. From the optimization perspective, there is no benefit in producing RP because it is a non-catalytic protein and does not sufficiently contribute to biomass.

Another interesting point is the production of macromolecules, see Fig. 4. In deFBA and r-deFBA, the dynamic optimization indicates that the production of T_1 should be stopped once C_1 is exhausted, although there is no regulatory control for T_1. Intuitively, T_1 is not needed anymore for uptake of C_1. In order to increase biomass, it is better to produce T_2 and R. When specifying the dynamics of the hybrid automaton in Eq. (18), we equally share the available resources between all synthesis reactions that are active in the current location. In real cells, this is unlikely to happen and not optimal for biomass production, as can be seen from Fig. 3D. Compared with the hybrid automaton, much more ribosome is produced in r-deFBA and deFBA, leading to a much larger biomass production, see Fig. 4. Clearly, how the cell allocates its resources to different enzymes will affect significantly the cellular growth.

Since r-deFBA is more constrained than deFBA, the maximum biomass predicted by r-deFBA will always be less than or equal to the one for deFBA. However, the two maxima can get very close if the regulatory constraints are consistent with the objective in the dynamic optimization. In our simulation, deFBA successfully predicts the diauxic shift even without regulatory control, showing that this is an optimal strategy for biomass production. However, deFBA fails to provide information about how the cell should be regulated to achieve this result. In contrast, r-deFBA allows predicting both the dynamic evolution of regulatory proteins and the discrete state transitions which together enable the cell to implement an optimal growth strategy.

5 Case study 2: Core carbon metabolism

5.1 Core carbon metabolic network and related gene regulatory network

As before, we first construct a metabolic-regulatory network (MRN), see Fig. 5. Here we combine a metabolic and a regulatory network for core carbon metabolism based on [15, 10]. The

Reaction	Name i	Enzyme j	Turnover rate $k_{cat}^{i,j}$ [1/min]
Exchange reactions			
$Carbon1 \to A$	T_{c1}	E_{Tc1}	3000
$Carbon2 \to A$	T_{c2}	E_{Tc2}	2000
$F_{ext} \to F$	T_f	E_{Tf}	3000
$H_{ext} \to H$	T_h	E_{Th}	3000
$O_{2ext} \to O_2$	T_{O2}	Q	1000
$D \to D_{ext}$	T_d	Q	1000
$E \to E_{ext}$	T_e	Q	1000
Metabolic reactions			
$A + ATP \to B$	R_1	E_{R1}	1800
$B \to C + 2ATP + 2NADH$	R_{2a}	E_{R2a}	1800
$C + 2ATP + 2NADH \to B$	R_{2b}	E_{R2b}	1800
$C \to 2ATP + 3D$	R_6	E_{R6}	1800
$C + 4NADH \to 3E$	R_7	E_{R7}	1800
$B \to F$	R_3	E_{R3}	1800
$C \to G$	R_4	E_{R4}	1800
$G + ATP + 2NADH \to H$	R_{8a}	E_{R8a}	1800
$H \to G + ATP + 2NADH$	R_{8b}	E_{R8b}	1800
$G \to 0.8C + 2NADH$	R_{5a}	E_{R5a}	1800
$G \to 0.8C + 2NADH$	R_{5b}	E_{R5b}	1800
$O_2 + NADH \to ATP$	$Rres$	E_{Rres}	1800

Table 4: Metabolic reactions with catalyzing enzymes and turnover rates

metabolic network in Fig. 5A covers the major carbon pathways including glycolysis, TCA cycle, carbon storage, amino acid synthesis, pentose phosphate pathway, fermentation, and also the macromolecule synthesis.

Using the notation from Sect. 2, we have the following molecular species:

$$\begin{aligned}
\mathbf{X} &= \{A, B, C, D, E, F, G, H, NADH, ATP, O_2\}, \\
\mathbf{Y} &= \{Carbon1, Carbon2, O_{2ext}, D_{ext}, E_{ext}, F_{ext}, H_{ext}\}, \\
\mathbf{Q} &= \{Q\}, \\
\mathbf{RE} &= \{E_{Tc2}, E_{R2a}, E_{R5a}, E_{R5b}, E_{R7}, E_{R8a}, E_{Rres}\}, \\
\mathbf{NRE} &= \{E_{Tc1}, E_{Tf}, E_{Th}, E_{R1}, E_{R2b}, E_{R3}, E_{R4}, E_{R6}, E_{R8b}, E_Q, Q, R\}, \\
\mathbf{RP} &= \{RP_{c1}, RP_{O2}, RP_b, RP_h\}, \\
\mathbf{E} &= \mathbf{RE} \cup \mathbf{NRE}, \quad \mathbf{P} = \mathbf{Q} \cup \mathbf{E} \cup \mathbf{RP}.
\end{aligned} \tag{19}$$

The details on the different metabolic and biomass reactions are given in Tab. 4 and 5. To get reasonable flux bounds on reactions describing diffusive exchange across the plasma membrane, we define the structural component Q as the enzymatic macromolecule for these reactions, together with an appropriate rate constant for diffusion [10].

Regarding the regulatory network of Fig. 5B, we identify again the state of a gene with the activity of the reaction producing the corresponding protein. For example, the gene state g_{RPc1} is identified with the activity state $\overline{RP}_{c1}$ of the reaction producing RP_{c1}. This means that the reaction synthesizing RP_{c1} will be active whenever the amount of external $Carbon1$ exceeds a given threshold. Conversely, the reaction will be blocked if not enough $Carbon1$ is available, see the regulatory rule for RP_{c1} in Tab. 5.

Next we briefly describe the role of the four regulatory proteins $RP_{c1}, RP_{O2}, RP_h, RP_b$, for additional details we refer to [15]. The regulatory protein RP_{c1} is used to operate the diauxic shift between $Carbon1$ and $Carbon2$. Protein RP_{O2} is responsible for switching between the aerobic and anaerobic pathways, which are catalyzed by the isozymes E_{R5a}, E_{R5b} respectively.

Protein RP_h regulates the usage of two primary nutrients of the cell, carbon sources and amino acid, via the reactions R_{8a} and R_{8b}. These two reactions correspond to one reversible reaction

Macromolecule synthesis reaction $n_pH + m_pATP \rightarrow p$	Molar weight b_p [g/mmol]	Turnover rate $k_{cat}^{p,R}$ [1/min]	Regulatory rule
Enzymes			
$400H + 1600ATP \rightarrow E_{Tc1}$	40	2.5	
$1500H + 6000ATP \rightarrow E_{Tc2}$	150	0.67	$\overline{E}_{Tc2} = \texttt{on} \Leftrightarrow RP_{c1} < \vartheta_{RP}$
$400H + 1600ATP \rightarrow E_{Tf}$	40	2.5	
$400H + 1600ATP \rightarrow E_{Th}$	40	2.5	
$500H + 2000ATP \rightarrow E_{R1}$	50	2.0	
$500H + 2000ATP \rightarrow E_{R2a}$	50	2.0	$\overline{E}_{R2a} = \texttt{on} \Leftrightarrow RP_b < \vartheta_{RP}$
$500H + 2000ATP \rightarrow E_{R2b}$	50	2.0	
$1000H + 4000ATP \rightarrow E_{R6}$	100	1.0	
$1000H + 4000ATP \rightarrow E_{R7}$	100	1.0	$\overline{E}_{R7} = \texttt{on} \Leftrightarrow RP_b < \vartheta_{RP}$
$2000H + 8000ATP \rightarrow E_{R3}$	200	0.5	
$500H + 2000ATP \rightarrow E_{R4}$	50	2.0	
$4000H + 16000ATP \rightarrow E_{R8a}$	400	0.25	$\overline{E}_{R8a} = \texttt{on} \Leftrightarrow RP_h < \vartheta_{RP}$
$4000H + 16000ATP \rightarrow E_{R8b}$	400	0.25	
$500H + 2000ATP \rightarrow E_{R5a}$	50	2.0	$\overline{E}_{R5a} = \texttt{on} \Leftrightarrow RP_{O2} < \vartheta_{RP}$
$500H + 2000ATP \rightarrow E_{R5b}$	50	2.0	$\overline{E}_{R5b} = \texttt{on} \Leftrightarrow RP_{O2} \geq \vartheta_{RP}$
$500H + 2000ATP \rightarrow E_{Rres}$	50	2.0	$\overline{E}_{Rres} = \texttt{on} \Leftrightarrow RP_{O2} < \vartheta_{RP}$
$500H + 2000ATP \rightarrow E_Q$	50	2.0	
$4500H + 21000ATP + 1500C \rightarrow R$	600	0.2	
Regulatory proteins			
$300H + 1200ATP \rightarrow RP_{O2}$	30	3.33	$\overline{RP}_{O2} = \texttt{on} \Leftrightarrow O_{2ext} < \vartheta_Y$
$300H + 1200ATP \rightarrow RP_{c1}$	30	3.33	$\overline{RP}_{c1} = \texttt{on} \Leftrightarrow Carbon1 \geq \vartheta_Y$
$300H + 1200ATP \rightarrow RP_h$	30	3.33	$\overline{RP}_h = \texttt{on} \Leftrightarrow v_{Th} \geq \vartheta_v$
$300H + 1200ATP \rightarrow RP_b$	30	3.33	$\overline{RP}_b = \texttt{on} \Leftrightarrow v_{R2b} \geq \vartheta_v$

Structural components		
Synthesis reaction	Molar weight b_Q [g/mmol]	Turnover rate k_{cat}^{Q,E_Q} [1/min]
$250H + 1500ATP + 250C + 250F \rightarrow Q$	75	3.0

Table 5: Macromolecule synthesis reactions with corresponding molar weights, turnover rates and regulatory rules

t_0	t_f^1	t_f^2	kd_E	kd_{RP}	ϑ_{RP}	ϑ_v	ϑ_Y	ε_E	ε_{RP}	Φ_Q
0 min	90 min	50 min	0.01 1/min	0.5 1/min	1.0e-3 mmol	0.1 mmol/min	1 mmol	1.0e-6 mmol/min	1.0e-3 mmol/min	0.35

Table 6: Parameters values for $E \in \mathbf{E}, RP \in \mathbf{RP}$ and end times t_f^1, t_f^2 for Scenario 1 and 2.

that connects the TCA cycle with the uptake of extracellular amino acid H_{ext}. Based on RP_h, the cell will not produce amino acid from carbon sources by R_{8a} if enough amino acid can be taken up from the environment, i.e., if $v_{Th} \geq \vartheta_v$.

The last regulatory protein RP_b is used to balance the key intermediate metabolites denoted by B and C. Since intracellular metabolite concentrations are not available due to the steady-state assumption, the reaction flux v_{R2b} is used as an internal signal. Production of RP_b is activated if $v_{R2b} \geq \vartheta_v$, and RP_b abundance then inhibits the expression of E_{R2a} and E_{R7}.

The complete r-deFBA model reads

$$
\begin{aligned}
\max_{\mathbf{v}(t),\mathbf{P}(t),\mathbf{Y}(t),\overline{p}(t)} \quad & \int_{t_0}^{t_f} b_{\mathbf{P}}^T \mathbf{P}(t) dt \\
s.t. \quad & \dot{\mathbf{Y}}(t) = -S_{\mathbf{Y},\mathbf{V_Y}} \cdot \mathbf{v_Y}(t), \\
& \dot{\mathbf{P}}(t) = S_{\mathbf{P},\mathbf{V_P}} \cdot \mathbf{v_P}(t) - kd_{\mathbf{P}} \circ \mathbf{P}(t), \\
& S_{\mathbf{X},\mathbf{V_Y}} \cdot \mathbf{v_Y}(t) + S_{\mathbf{X},\mathbf{V_X}} \cdot \mathbf{v_X}(t) - S_{\mathbf{X},\mathbf{V_P}} \cdot \mathbf{v_P}(t) = 0, \\
& b_Q^T \cdot Q(t) \geq \ \Phi_Q \cdot b_{\mathbf{P}}^T \cdot \mathbf{P}(t), \\
& \mathbf{v}_i(t) \leq k_{cat}^{i,E_j} \cdot E_j(t), \text{ for all } E_j \in \mathbf{E} \setminus \{Q,R\}, \\
& \sum_{i \in \{T_{O2},T_d,T_e\}} \frac{\mathbf{v}_i(t)}{k_{cat}^{i,Q}} \leq Q(t), \quad \sum_{p \in \mathbf{P} \setminus \{Q\}} \frac{\mathbf{v}_p(t)}{k_{cat}^{p,R}} \leq R(t), \\
& \overline{E}_{Tc2} = \texttt{on} \Leftrightarrow RP_{c1} < \vartheta_{RP}, \quad \overline{E}_{R2a} = \texttt{on} \Leftrightarrow RP_b < \vartheta_{RP}, \\
& \overline{E}_{R7} = \texttt{on} \Leftrightarrow RP_b < \vartheta_{RP}, \quad \overline{E}_{R8a} = \texttt{on} \Leftrightarrow RP_h < \vartheta_{RP}, \\
& \overline{E}_{R5a} = \texttt{on} \Leftrightarrow RP_{O2} < \vartheta_{RP}, \quad \overline{E}_{R5b} = \texttt{on} \Leftrightarrow RP_{O2} \geq \vartheta_{RP}, \\
& \overline{E}_{Rres} = \texttt{on} \Leftrightarrow RP_{O2} < \vartheta_{RP}, \\
& \overline{RP}_{O2} = \texttt{on} \Leftrightarrow O_{2ext} < \vartheta_Y, \quad \overline{RP}_{c1} = \texttt{on} \Leftrightarrow Carbon1 \geq \vartheta_Y, \\
& \overline{RP}_h = \texttt{on} \Leftrightarrow v_{Th} \geq \vartheta_v, \quad \overline{RP}_b = \texttt{on} \Leftrightarrow v_{R2b} \geq \vartheta_v, \\
& \overline{p}(t) = 1 \Rightarrow \mathbf{v}_p(t) \geq \varepsilon_{RP}, \text{ for all } p \in \mathbf{RP} \\
& \overline{p}(t) = 1 \Rightarrow \mathbf{v}_p(t) \geq \varepsilon_E, \text{ for all } p \in \mathbf{RE} \\
& \overline{p}(t) = 0 \Rightarrow \mathbf{v}_p(t) = 0, \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE} \\
& \mathbf{P}(t), \mathbf{Y}(t), \mathbf{v}(t) \geq 0, \quad \overline{p}(t) \in \{0,1\}, \text{ for all } p \in \mathbf{RP} \cup \mathbf{RE}, \\
& \text{for all } t \in [t_0, t_f].
\end{aligned}
\tag{20}
$$

The initial values depend on the specific scenario and will be specified in the next section.

5.2 Comparing r-deFBA and deFBA

In total, there are 11 regulated proteins, which include 4 regulatory proteins and 7 regulated enzymes. The discrete state space thus contains 2^{11} states, which are difficult to explore by the hybrid automaton. In the following, we present two scenarios to show how r-deFBA can be used to predict the integrated dynamics of metabolism and regulation even in a large state space. In each case, we compare r-deFBA with deFBA, which also models metabolism, but does not take into account the regulatory control.

5.2.1 Scenario 1: Diauxie on two carbon sources

Our first scenario focuses again on the diauxie phenomenon. Initially, we set $Carbon1$ and $Carbon2$ to 1000 resp. 500 mmol, oxygen is given in excess, all other extracellular metabolites are set to 0. We do not specify the initial amounts of the macromolecules. Instead, these are computed by the optimization algorithm under the constraint that the initial biomass should be 1g. Thus the initial values for $t = t_0$ are:

$Carbon1$	$Carbon2$	D_{ext}	E_{ext}	F_{ext}	H_{ext}	O_{2ext}	$Biomass$
1000	500	0	0	0	0	$+\infty$	1

(21)

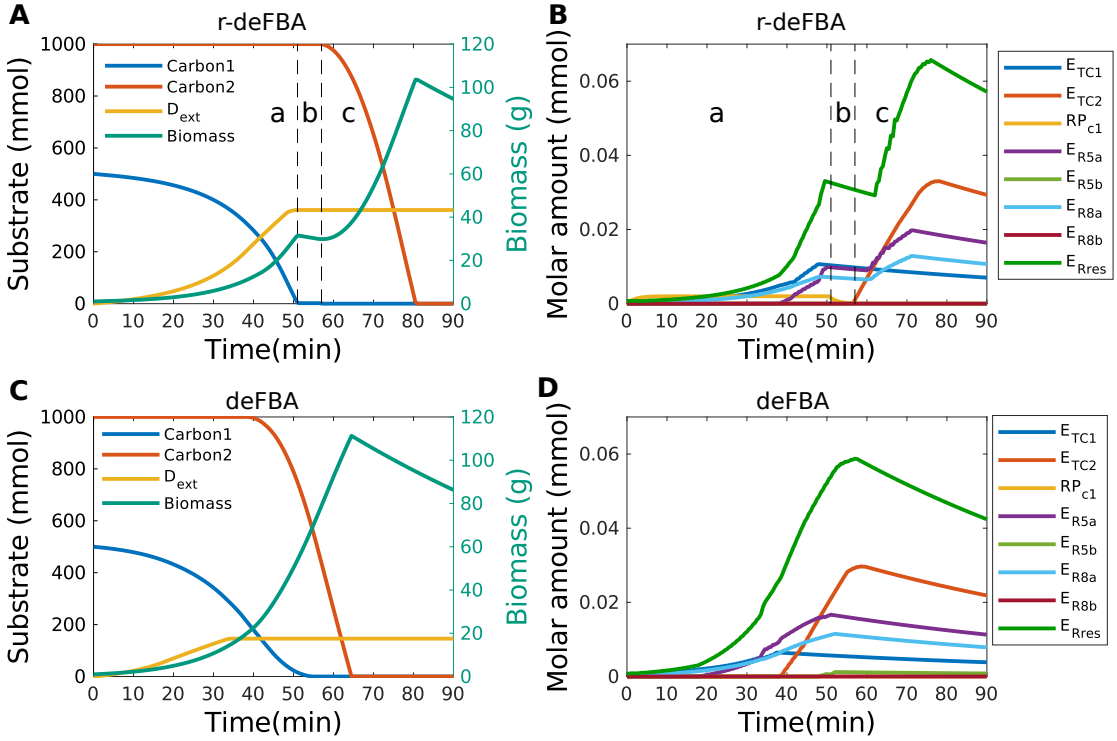

Fig. 6: Dynamics of external substrates (left axis), total biomass (right axis) predicted by r-deFBA (**A**) and deFBA (**C**), and key regulated proteins (**B** and **D**) in Scenario 1.

The Boolean variables are initialized by the 0-1 values for growth phase (a) in Tab. 7.

Comparing the results of r-deFBA and deFBA in Fig. 6A resp. 6C, we note that in both approaches $Carbon1$ is metabolized first. Yet, the biphasic increase of biomass is predicted only by r-deFBA, and not by deFBA. Although most of the available $Carbon1$ is utilized at the beginning, no lag phase is predicted by deFBA. The overall biomass production in the time interval $[t_0, t_f]$ predicted by deFBA amounts to 111.3g, which is 8% more than the 103.0g obtained by r-deFBA.

At the level of individual proteins, $\overline{RP}_{c1}$ is produced in growth phase (a) of r-deFBA, for which $Carbon1 \geq \vartheta_Y$, see Fig. 6B. Here, the expression of $\overline{E}_{Tc2}$ is inhibited since $\overline{RP}_{c1} \geq \vartheta_{RP}$. Thus, only $Carbon1$ supports growth during this period. Once it is exhausted, the growth shifts to phase (b). The indicator variable $\overline{RP}_{c1}$ is triggered to be `off`, implying that RP_{c1} is degraded and not produced anymore. The expression of the transporter E_{Tc2} via $\overline{E}_{Tc2}$ is only activated when $RP_{c1} < \vartheta_{RP}$. So, no carbon can be taken up during phase (b) and the total biomass production shows a lag phase. Finally, in growth phase (c), the transporter E_{Tc2} is produced and biomass production resumes based on $Carbon2$.

Similarly in deFBA, E_{Tc2} is not synthesized as long as $Carbon1$ supports a high growth rate. The protein dynamics for r-deFBA and deFBA in Fig. 6B resp. 6D are also relatively close. However, RP_{c1} is not produced at all in deFBA and there is no lag phase, see Fig. 6D. In deFBA, the uptake of $Carbon2$ starts well before $Carbon1$ is exhausted, while in r-deFBA, $Carbon1$ and $Carbon2$ are metabolized one after the other, due to the regulatory control by RP_{c1}. The synthesis of RP_{c1} generates extra costs in r-deFBA, such that the total biomass in r-deFBA is smaller than in deFBA.

The discrete state transitions for all the regulated proteins as predicted by r-deFBA are given in Tab. 7. Here we group together each of the four regulatory proteins with the corresponding regulated enzymes. The expression of the enzymes regulated by RP_{O2} does not change, since external oxygen is given in excess. Thus, $\overline{RP}_{O2}$ and $\overline{E}_{R5b}$ are always inhibited, while $\overline{E}_{R5a}$ and $\overline{E}_{Rres}$ remain activated. In Scenario 1, with no extracellular H_{ext} in the environment, reaction R_{2a} is constantly activated, consuming $Carbon1, Carbon2$, while T_h is inactive. Thus, the regulatory proteins $\overline{RP}_b, \overline{RP}_h$ are always `off`, and $\overline{E}_{R2a}, \overline{E}_{R7}, \overline{E}_{R8a}$ are `on`. Note that the r-deFBA

Growth phase	$\overline{RP}_{c1}$	$\overline{E}_{Tc2}$	$\overline{RP}_{O2}$	$\overline{E}_{R5a}$	$\overline{E}_{R5b}$	$\overline{E}_{Rres}$	$\overline{RP}_{b}$	$\overline{E}_{R2a}$	$\overline{E}_{R7}$	$\overline{RP}_{h}$	$\overline{E}_{R8a}$
a	1	0	0	1	0	1	0	1	1	0	1
b	0	0	0	1	0	1	0	1	1	0	1
c	0	1	0	1	0	1	0	1	1	0	1

Table 7: Discrete state transitions in Scenario 1

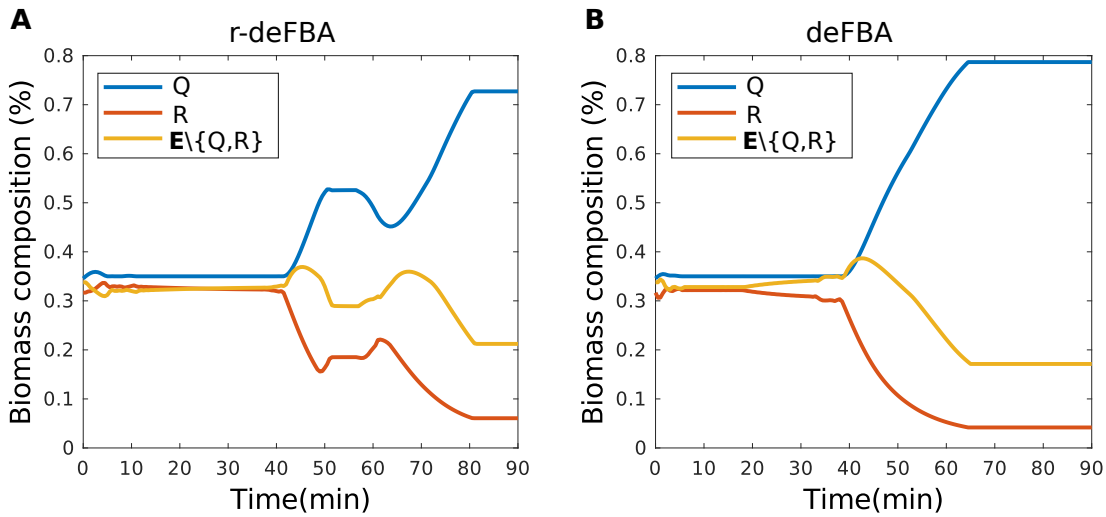

Fig. 7: Dynamics of biomass composition in Scenario 1 with structural components Q, ribosomes R and enzymes $E \in \mathbf{E} \setminus \{Q, R\}$

framework allows computing an optimal regulatory strategy for maximizing growth even though the discrete state space is very large.

The resource allocation during the carbon switch can also be investigated. In Fig. 7, we compare the dynamic biomass composition predicted by r-deFBA and deFBA for three kinds of macromolecules: structural components, enzymes and ribosomes. At the beginning, both approaches exhibit a stable biomass composition. The fraction of structural components initially stays around 35%, which corresponds to the lower bound imposed by the biomass constraint in Eq. (7). As $Carbon1$ is depleted, the structural components reach a rather high level, while the fractions of enzymes and ribosomes are decreasing in both predictions. Interestingly, in r-deFBA, the fractions of ribosomes and enzymes are increasing again while the structural components are going down at the outset of the second growth phase. This means that the cell has to allocate more resources to the ribosomes to start the second growth phase. In the last step, r-deFBA predicts a high fraction of structural components and a low fraction of enzymes and ribosomes, which can also be validated by experiments [28]. Overall, we obtain a biphasic resource allocation in r-deFBA, which is consistent with the two growth phases during diauxie. In deFBA, the quota fraction directly increases to about 80% of the total biomass and then keeps constant.

5.2.2 Scenario 2: growth on carbon and amino acid with amino acid in excess

Scenario 2 explores the dynamic growth on carbon and amino acid, with amino acid in excess. For $t = t_0$ we choose the initial values:

$Carbon1$	$Carbon2$	D_{ext}	E_{ext}	F_{ext}	H_{ext}	O_{2ext}	$Biomass$
100	0	0	0	0	250	$+\infty$	1

(22)

The Boolean variables are initialized by the 0-1 values for growth phase (a) in Tab. 8.

As can be seen from Fig. 8, both r-deFBA and deFBA first predict a co-utilization of $Carbon1$ and extracellular amino acid H_{ext}, followed by the utilization of H_{ext} once $Carbon1$ has been

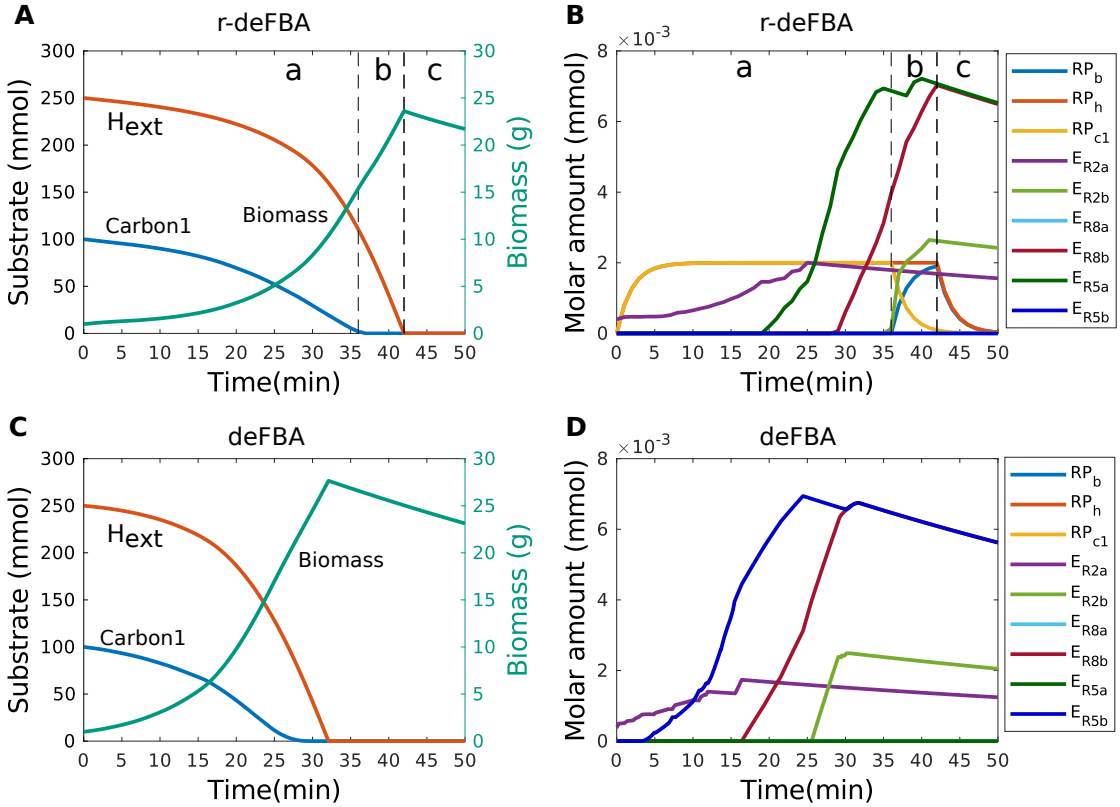

Fig. 8: Dynamics of external metabolites (left axis) and biomass (right axis) predicted by r-deFBA(**A**) and deFBA (**C**), and key regulated proteins (**B** and **D**) in Scenario 2.

exhausted. In the co-utilization phase, R_{2a} instead of R_{2b} is active to metabolize $Carbon1$. Consequently, enzyme E_{R2a} is synthesized in this phase, but not E_{R2b}. When $Carbon1$ gets almost exhausted, enzyme E_{R2b} starts being produced in order to activate reaction R_{2b}, see Fig. 8B and 8D. Now B has to be generated from C, since B is needed for growth. The switch between R_{2a} and R_{2b} is predicted by both approaches because it benefits growth. Although in deFBA the expression of the regulatory protein RP_b is not triggered to inhibit the synthesis of E_{R2a}, the production of E_{R2a} stops in deFBA as well. Like E_{R2a}, enzyme E_{R8b} also has a similar dynamics in both approaches.

An interesting observation in the comparison is that deFBA activates enzyme E_{R5b} responsible for the anaerobic pathway, which is not consistent with the regulation. In contrast, r-deFBA produces enzyme E_{R5a} to catalyze reaction R_{5a}, in agreement with the regulation by RP_{O2}. Intuitively, R_{5a}, R_{5b} are two alternative reactions that play the same role in the network, one in the aerobic, the other in the anaerobic case. Using only optimization without any regulatory information, deFBA cannot guarantee to choose the right pathway. Since optimal solutions are not unique, the solver can choose any of the two reactions or a combination thereof. Indeed, a small amount of E_{R5b} is produced by deFBA in the last phase of Scenario 1, even though this is not significant (see Fig. 6D). Clearly, the consistency between metabolism and regulation cannot be ensured by deFBA without additional regulatory information. In contrast, the dynamics of metabolism in r-deFBA highly depends on the activity of the regulatory proteins, and also the converse is true.

Switching between active reactions by r-deFBA is illustrated in Fig. 9. We can see in Fig. 9A that in the beginning $Carbon1$ and H_{ext} are co-utilized and H is obtained only from H_{ext}. Since the model starts with a small biomass, even the TCA cycle is first inactive. Only when enzyme E_{R5a} has been synthesized, the TCA cycle is activated after 18 min, see Fig. 9B. Next in Fig. 9C, reaction R_{8b} is activated to furnish the TCA cycle with amino acid H, while releasing ATP and NADH. Now $Carbon1$ is not sufficient anymore to provide energy for growth. Finally, in Fig. 9D,

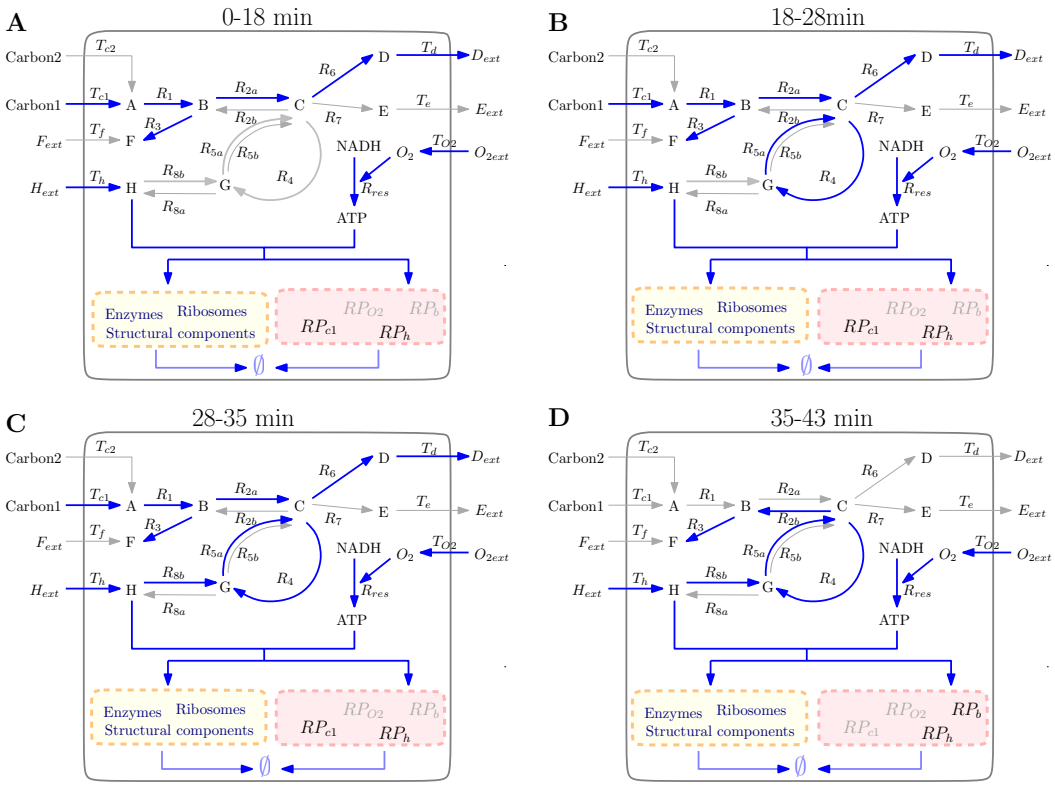

Fig. 9: Patterns of active reactions predicted by r-deFBA in Scenario 2.

Growth section	$\overline{RP}_{c1}$	$\overline{E}_{Tc2}$	$\overline{RP}_{O2}$	$\overline{E}_{R5a}$	$\overline{E}_{R5b}$	$\overline{E}_{Rres}$	$\overline{RP}_b$	$\overline{E}_{R2a}$	$\overline{E}_{R7}$	$\overline{RP}_h$	$\overline{E}_{R8a}$
a	1	0	0	1	0	1	0	1	1	1	0
b	0	0	0	1	0	1	1	0	0	1	0
c	0	1	0	1	0	1	0	0	0	0	0

Table 8: Discrete state transitions in Scenario 2

$Carbon1$ has been exhausted, enzyme E_{R2b} is synthesized and R_{2b} is used to generate B.

Regarding the discrete state transitions, we divide the simulation period of r-deFBA again into three phases (a), (b), and (c), see Fig. 8A and 8B. During the last phase (c), there is no growth since all the nutrients are exhausted. The key regulatory pathways analyzed in Scenario 2 are operated by the regulatory proteins RP_b and RP_h. First, R_{2a} is active rather than R_{2b} for better metabolizing the carbon source. Hence, $\overline{RP}_b$ is `off` until $Carbon1$ is used up. Soon after growth phase (a), R_{2b} has to be activated to use H, so that RP_b is triggered to be produced (see Fig. 8B). The enzymes $\overline{E}_{R2a}, \overline{E}_{R7}$ then switch to `off`. Before H_{ext} has been exhausted, the expression state of RP_h is `on` because T_h has to be active for the uptake. Enzyme $\overline{E}_{R8a}$ is inhibited by RP_h. The state transitions related with RP_{O2} are the same as in Scenario 1 because external oxygen is given in excess during the whole period. $Carbon1$ is given initially and exhausted at the end of growth phase (a). Although $Carbon2$ is set to 0 and the cell cannot use it, the expression state of E_{Tc2} is activated when RP_{c1} is totally degraded, due to the regulatory constraints. In Scenario 2, this happens by chance at the time when H_{ext} is used up. So, $\overline{E}_{Tc2}$ is `on` in phase (c). Meanwhile both $Carbon1$ and H_{ext} have been used up and the two reaction signals $v_{Th} > \vartheta_v$ and $v_{R2b} > \vartheta_v$ are inactive. The indicator variables $\overline{RP}_b, \overline{RP}_h$ are turned `off` and the associated proteins RP_b, RP_h are degraded in the last phase, see Fig. 8B.

6 Conclusion

Overall, r-deFBA computes optimal control strategies for hybrid automata representing metabolic-regulatory networks. Compared to previous approaches, in particular rFBA [15] and deFBA [10], r-deFBA allows for more realistic and accurate predictions by integrating the continuous dynamics of metabolism, including cellular resource allocation, with discrete regulatory control.

In purely discrete modeling frameworks for regulatory networks like the asynchronous logical formalism of R. Thomas [29], it is not possible to quantify the time delay between discrete state transitions. The hybrid automata approach proposed in [22] solves this problem by using continuous variables for regulatory protein amounts together with thresholds that trigger the discrete jumps. However, exploring the dynamics of these hybrid automata is difficult due to the exponentially large discrete state space. By computing an optimal control strategy for the hybrid automaton, r-deFBA is able to predict even in large state spaces the quantitative dynamics of the regulatory proteins together with the sequence of discrete state transitions that are needed to achieve optimal growth.

In summary, r-deFBA allows predicting optimal cellular resource allocation in a dynamic environment by integrating metabolic reactions, enzyme-costs, quota compounds, and transcriptional regulation. Thus, r-deFBA considerably extends the predictive capabilities of current constraint-based modeling approaches as summarized in Tab. 1. Based on a hybrid discrete-continuous dynamics, r-deFBA is able to predict not only the continuous evolution of macromolecules and extracellular metabolites, but also the sequence of regulatory events needed to achieve an optimal growth. Finally, r-deFBA provides a solution for how to share enzymes between different reactions, which includes ribosome allocation in protein synthesis as a special case.

7 Acknowledgments

The authors would like to thank M. Köbis, A. Reimers, and A. Röhl for fruitful discussions about this work. L. Liu gratefully acknowledges support from the China Scholarship Council (CSC) and DFG GRK 1772 Computational Systems Biology.

References

[1] A. Bordbar, J. M. Monk, Z. A. King, and B. Ø. Palsson, "Constraint-based models predict metabolic and associated cellular functions," *Nat. Rev. Genet.*, vol. 15, pp. 107–120, Feb 2014.

[2] N. E. Lewis, H. Nagarajan, and B. Ø. Palsson, "Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods," *Nat. Rev. Microbiol.*, vol. 10, pp. 291–305, Feb 2012.

[3] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?," *Nat. Biotechnol.*, vol. 28, pp. 245–248, Mar 2010.

[4] A. Varma and B. Ø. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110," *Appl. Environ. Microbiol.*, vol. 60, pp. 3724–3731, Oct 1994.

[5] R. Mahadevan, J. S. Edwards, and F. J. Doyle, "Dynamic flux balance analysis of diauxic growth in Escherichia coli," *Biophys. J.*, vol. 83, pp. 1331–1340, Sep 2002.

[6] A. Goelzer, V. Fromion, and G. Scorletti, "Cell design in bacteria as a convex optimization problem," *Automatica*, vol. 47, no. 6, pp. 1210–1218, 2011.

[7] A. Goelzer, J. Muntel, V. Chubukov, M. Jules, E. Prestel, R. Nolker, M. Mariadassou, S. Aymerich, M. Hecker, P. Noirot, D. Becher, and V. Fromion, "Quantitative prediction of genome-wide resource allocation in bacteria," *Metab. Eng.*, vol. 32, pp. 232–243, Nov 2015.

[8] J. A. Lerman, D. R. Hyduke, H. Latif, V. A. Portnoy, N. E. Lewis, J. D. Orth, A. C. Schrimpe-Rutledge, R. D. Smith, J. N. Adkins, K. Zengler, and B. Ø. Palsson, "In silico method for modelling metabolism and gene product expression at genome scale," *Nat Commun*, vol. 3, p. 929, Jul 2012.

[9] E. J. O'Brien, J. A. Lerman, R. L. Chang, D. R. Hyduke, and B. Ø. Palsson, "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction," *Mol. Syst. Biol.*, vol. 9, p. 693, Oct 2013.

[10] S. Waldherr, D. A. Oyarzun, and A. Bockmayr, "Dynamic optimization of metabolic networks coupled with gene expression," *J. Theor. Biol.*, vol. 365, pp. 469–485, Jan 2015.

[11] M. Rügen, A. Bockmayr, and R. Steuer, "Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA," *Sci Rep*, vol. 5, p. 15247, 2015.

[12] A.-M. Reimers, H. Knoop, A. Bockmayr, and R. Steuer, "Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 31, pp. E6457–E6465, 2017.

[13] G. Jeanne, A. Goelzer, S. Tebbani, D. Dumur, and V. Fromion, "Dynamical resource allocation models for bioreactor optimization," *IFAC-PapersOnLine*, vol. 51, no. 19, pp. 20–23, 2018.

[14] L. Yang, A. Ebrahim, C. J. Lloyd, M. A. Saunders, and B. Ø. Palsson, "DynamicME: dynamic simulation and refinement of integrated models of metabolism and protein expression," *BMC Syst Biol*, vol. 13, p. 2, 01 2019.

[15] M. W. Covert, C. H. Schilling, and B. Ø. Palsson, "Regulation of gene expression in flux balance models of metabolism," *J. Theor. Biol.*, vol. 213, pp. 73–88, Nov 2001.

[16] L. Marmiesse, R. Peyraud, and L. Cottret, "FlexFlux: combining metabolic flux and regulatory network analyses," *BMC Syst Biol*, vol. 9, p. 93, Dec 2015.

[17] M. W. Covert, N. Xiao, T. J. Chen, and J. R. Karr, "Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli," *Bioinformatics*, vol. 24, no. 18, pp. 2044–2050, 2008.

[18] J. M. Lee, J. Min Lee, E. P. Gianchandani, J. A. Eddy, and J. A. Papin, "Dynamic analysis of integrated signaling, metabolic, and regulatory networks," *PLoS Comput. Biol.*, vol. 4, p. e1000086, May 2008.

[19] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin, "A genome-scale computational study of the interplay between transcriptional regulation and metabolism," *Mol. Syst. Biol.*, vol. 3, p. 101, 2007.

[20] S. Chandrasekaran and N. D. Price, "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 47, pp. 17845–50, 2010.

[21] J. M. Savinell and B. Ø. Palsson, "Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism," *J. Theor. Biol.*, vol. 154, pp. 421–454, Feb 1992.

[22] L. Liu and A. Bockmayr, "Formalizing metabolic-regulatory networks by hybrid automata," *Acta Biotheor.*, July 2019.

[23] A.-M. Reimers, H. Lindhorst, and S. Waldherr, "A protocol for generating and exchanging (genome-scale) metabolic resource allocation models," *Metabolites*, vol. 7, no. 3, p. 47, 2017.

[24] P. A. Jensen, K. A. Lutz, and J. A. Papin, "TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks," *BMC Syst Biol*, vol. 5, p. 147, Sep 2011.

[25] B. Görke and J. Stülke, "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients," *Nat. Rev. Microbiol.*, vol. 6, pp. 613–624, Aug 2008.

[26] A. Kremling, J. Geiselmann, D. Ropers, and H. De Jong, "An ensemble of mathematical models showing diauxic growth behaviour," *BMC Syst Biol*, vol. 12, no. 1, p. 82, 2018.

[27] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink, "Shifts in growth strategies reflect tradeoffs in cellular economics," *Mol. Syst. Biol.*, vol. 5, p. 323, 2009.

[28] E. Fischer and U. Sauer, "Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism," *Nat. Genet.*, vol. 37, pp. 636–640, Jun 2005.

[29] R. Thomas and M. Kaufman, "Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits," *Chaos*, vol. 11, pp. 180–195, Mar 2001.