

# RBA for eukaryotic cells: foundations and theoretical developments

Anne Goelzer, Vincent Fromion

INRA, UR1404, MaIAGE, Université Paris-Saclay, Jouy-en-Josas, France.  
`firstname.name@inra.fr`

## Abstract

Resource allocation models were recently identified as new ways to investigate cell design principles. In particular, the Resource Balance Analysis (RBA) framework is the first constraint-based modelling method capable of accurate quantitative predictions of the genome-wide resource allocation. Initially developed and validated on bacteria, the objective of this paper is to provide the mathematical fundations of the extension of the RBA framework to eukaryotic cells. We especially investigate the way to handle the cellular compartments in order to formalize eventually the functioning of organelles. It turns out that the final RBA problem for eukaryotic cells is close to the one of prokaryotic cells from a theoretical point of view. The mathematical properties that were already identified on the prokaryotic RBA framework can be easily transposed to eukaryotic cells. In particular, the eukaryotic RBA problem can be solved easily at the cell scale by Linear Programming. This paves the way to future developments of RBA models for eukaryotic cells.

## 1 Introduction

Recently, bacterial whole-cell models based on concepts of parsimonious resource allocation between cellular processes were shown to predict accurately the growth rate and a high number of cellular variables (i.e. metabolic fluxes, abundances of molecular machines including enzymes) across exponential growth conditions [3, 7, 2]. Moreover, they were also shown to reproduce complex genetic regulations without the explicit addition of gene regulations, thus reproducing observed cellular configurations with simple cost-benefit arguments with respect to the use of cellular resources [14]. Such types of models are so-called constraint-based models, and actually formalize the mathematical relationships defining the interactions and allocation of resources between the cellular processes as a set of equality and inequality convex constraints [5, 6]. Satisfying these constraints then led to a linear convex optimization problem. The underlying problem is thus tractable and solvable rapidly even at genome scale in a few seconds [1, 10]. The first genome-scale constraint-based modeling method, named Resource Balance Analysis (RBA) was developed in 2009-2011 [5, 6], where the first versions of the RBA model for the bacterium *Bacillus subtilis* were already developed and simulated. The first (and still only) biological validation of a genome-scale model integrating resource allocation was achieved in 2015 for the RBA model of *B. subtilis* [7]. In parallel, since 2013, other constraint-based modelling methods integrating some of the RBA constraints have been developed [11, 9], confirming that resource allocation is a relevant cell design principle and also an intensive research area within the systems biology community (see [4] and references therein). Currently, the resource allocation framework was used to generate resource allocation models for prokaryotes. There exists published resource allocation based models for the bacterium *B. subtilis* [5, 6, 7, 2], *Escherichia coli* [11, 9, 2] and the cyanobacteria *Synechococcus elongatus PCC 7942* [12]. Extending the

resource allocation framework to eukaryotic cells is now the next step to achieve.

In this document we study the extension of the RBA framework that was originally developed in [5, 6] for bacteria, to eukaryotic cells. Here our purpose is not to develop and implement a RBA model for eukaryotic cells, but to provide the mathematical foundations of the framework, and especially:

- to handle the compartments on RBA constraints in a systematic way;
- to identify clearly the assumptions that are necessary to obtain a convex formulation of the constraints;
- to derive the constraints, and *in fine* the RBA optimization problem for eukaryotic cells.

As we will show in the sequel, under mild assumptions, the existence of compartments does not change the theoretical nature of the RBA problem, implying that the mathematical properties presented in [5, 6] are thus valid in case of eukaryotic cells. In particular, this means that the RBA optimization problem for eukaryotic cells can be efficiently solved at the genome scale. The paper is organized as follows. We first recall in Section 2 the RBA framework for prokaryotes, especially to introduce the notations. Section 2.1 presents the original problem of [7]. In Section 2.2, we extend the constraints to include the turnover rate of proteins and metabolites. Then, we investigate in Section 3.1, 3.2, 3.3 and 3.4 the consequences of compartments on RBA constraints. Finally, in Section 3.5, we present the RBA framework for eukaryotic cells.

**Notation.**  $A^T$  refers to the transpose of the matrix  $A$ .  $\mathbb{R}_{>0}^n \stackrel{\Delta}{=} \{x \in \mathbb{R}^n \mid x_i > 0 \text{ for all } i \in \{1, \dots, n\}\}$ ,  $\mathbb{R}_{>0} \stackrel{\Delta}{=} \mathbb{R}_{>0}^1$ ,  $\mathbb{R}_{\geq 0}^n \stackrel{\Delta}{=} \{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ for all } i \in \{1, \dots, n\}\}$  and  $\mathbb{R}_{\geq 0} \stackrel{\Delta}{=} \mathbb{R}_{\geq 0}^1$ .

## 2 The RBA problem for prokaryotic cells: some recall

In this section, we first recall the RBA optimization problem for prokaryotes, introduce notations and also the constraints. Most of the results are extracted from [5, 6].

### 2.1 The RBA standard formulation: some recalls

A cell is composed of different cellular entities:

- (i)  $N_y$  molecular machines, that can be subdivided in  $N_m$  enzymes and transporters involved in the metabolic network (i.e. enzymes, transporters)  $\mathbb{E} \stackrel{\Delta}{=} (\mathbb{E}_1, \dots, \mathbb{E}_{N_m})$  at the concentrations  $E \stackrel{\Delta}{=} (E_1, \dots, E_{N_m})^T$  and with the fluxes  $\nu \stackrel{\Delta}{=} (\nu_1, \dots, \nu_{N_m})^T$ ; and  $N_p$  macromolecular machines  $\mathbb{M} \stackrel{\Delta}{=} (\mathbb{M}_1, \dots, \mathbb{M}_{N_p})$  involved in non-metabolic cellular processes such as the translation apparatus, at the concentrations  $M \stackrel{\Delta}{=} (M_1, \dots, M_{N_p})^T$ ;
- (ii)  $N_g$  proteins  $\mathbb{P}_G \stackrel{\Delta}{=} \{\mathbb{P}_{G_1}, \dots, \mathbb{P}_{G_{N_g}}\}$  for which the cellular process to which the proteins belong is not specified.  $P_G \stackrel{\Delta}{=} (P_{G_1}, \dots, P_{G_{N_g}})^T$  denotes the set of concentrations of  $\mathbb{P}_G$ ;
- (iii)  $N_s$  metabolites  $\mathbb{S} \stackrel{\Delta}{=} (\mathbb{S}_1, \dots, \mathbb{S}_{N_s})$  at the concentrations  $S \stackrel{\Delta}{=} (S_1, \dots, S_{N_s})^T$ . Among the set  $\mathbb{S}$ , we distinguish a subset  $\mathbb{B} \stackrel{\Delta}{=} (\mathbb{B}_1, \dots, \mathbb{B}_{N_b})$  of metabolites which have fixed concentrations  $\bar{B} \stackrel{\Delta}{=} (\bar{B}_1, \dots, \bar{B}_{N_b})^T$ .

**Remarks.**

1. When a metabolic reaction can be catalyzed by two (respectively,  $n > 2$ ) different enzymes, corresponding to isoenzymes, the metabolic reaction is duplicated (respectively,  $n$  times repeated), i.e., when a metabolic reaction  $r$  is catalyzed by  $\mathbb{E}_1$  or  $\mathbb{E}_2$  then we introduce two reactions  $r_1$  and  $r_2$  that are catalyzed by  $\mathbb{E}_1$  and  $\mathbb{E}_2$  respectively.
2. When an enzymatic complex  $\mathbb{E}_i$  can catalyze two (respectively,  $n > 2$ ) distinct reactions, we duplicate (respectively, repeat  $n$  times) the enzymatic complex: when the enzymatic complex  $\mathbb{E}_i$  catalyzes reactions  $r_1$  and  $r_2$ , then we introduce two enzymatic complexes  $\mathbb{E}_{i_1}$  and  $\mathbb{E}_{i_2}$  that catalyze  $r_1$  and  $r_2$  respectively. In this case, the translation process has to produce both  $\mathbb{E}_{i_1}$  and  $\mathbb{E}_{i_2}$ . ■

The RBA optimization problem for prokaryotic cell ( $P_{rba}^p$ ) can be formalized mathematically as follows.

$P_{rba}^p$ : For a fixed vector of concentrations  $P_G \in \mathbb{R}_{>0}^{N_g}$ , and the growth rate  $\mu \geq 0$ , i.e. the amount of produced biomass per cell per hour,

$$\begin{aligned}
 & \text{find} && Y \in \mathbb{R}_{\geq 0}^{m+p}, \nu \in \mathbb{R}^m, \\
 & \text{subject to} && \\
 (C_1) & -\Omega\nu + \mu(C_Y^S Y + C_B^S \bar{B} + C_G^S P_G) = 0 \\
 (C_{2a}) & \mu(C_Y^M Y + C_G^M P_G) - K_T Y \leq 0 \\
 (C_{2b}) & -K'_E Y \leq \nu \leq K_E Y \\
 (C_3) & C_Y^D Y + C_G^D P_G - \bar{D} \leq 0
 \end{aligned}$$

where all the inequalities are componentwise inequalities,  $Y^T \triangleq (E^T, M^T)$  is the vector of concentrations of molecular machines and:

- $\Omega$  is the stoichiometry matrix of the metabolic network of size  $N_s \times N_m$ , where  $\Omega_{ij}$  corresponds to the stoichiometry of metabolite  $\mathbb{S}_i$  in the  $j$ -th enzymatic reaction;
- $C_Y^S$  (resp.  $C_G^S$ ) is a  $N_s \times N_y$  (resp.  $N_s \times N_g$ ) matrix where each coefficient  $C_{Y_{ij}}^S$  corresponds to the number of metabolite  $\mathbb{S}_i$  consumed (or produced) for the synthesis of one machine  $\mathbb{Y}_j$  (resp.  $\mathbb{P}_{G_j}$ );  $C_{Y_{ij}}^S$  is then positive, negative or null if  $\mathbb{S}_i$  is produced, consumed or not involved in the synthesis of one machine  $\mathbb{Y}_j$  (resp.  $\mathbb{P}_{G_j}$ );
- $C_B^S$  is a  $N_s \times N_b$  matrix where each coefficient  $C_{B_{ij}}^S$  corresponds to the number of metabolite  $\mathbb{S}_i$  consumed (or produced) for the synthesis of one  $\mathbb{B}_j$ ;
- $K_T$  ( $K_E$  and  $K'_E$ , respectively) of size  $N_p \times N_p$  ( $N_m \times N_m$ , respectively) is diagonal matrix where each coefficient  $K_{T_i}$  ( $K_{E_i}$  and  $K'_{E_i}$ , respectively) is positive and corresponds to the efficiency of the molecular machine  $\mathbb{M}_i$ , i.e. the rate of the process per amount of the catalyzing molecular machine, (the efficiency of the enzyme  $\mathbb{E}_i$  in forward and backward sense, respectively);
- $C_Y^M$  (resp.  $C_G^M$ ) is a  $N_p \times N_y$  (resp.  $N_p \times N_g$ ) matrix where each coefficient  $C_{Y_{ij}}^M$  typically corresponds to the length in amino acids of the machine  $\mathbb{Y}_j$  (resp.  $\mathbb{P}_{G_j}$ ). In some cases (for instance for the constraints on protein chaperoning), the length in amino acids can be multiplied by a coefficient, such as the fraction of the whole proteome that necessitates chaperoning;
- $\bar{D}$  is a vector of size  $N^c$ , where  $N^c$  is the number of volume and surface areas for which density constraints are considered.  $\bar{D}^i$  is the density of molecular entity with respect to the volume or surface area. Densities are typically expressed as a number of amino-acid residues by volume or surface area.

- $C_Y^D$  (resp.  $C_G^D$ ) is a  $N^c \times N_y$  (resp.  $N^c \times N_G$ ) matrix where each coefficient  $C_{Y_{ij}}^D$  corresponds to the density of one machine  $\mathbb{Y}_j$  (resp.  $\mathbb{P}_{G_j}$ ) in the compartment  $i$ . By construction, we have one unique localization per machine.

Actually, the RBA method formalizes the mathematical relationships defining the interactions and allocation of resources between the cellular entities. All these relationships take the form of linear growth-rate dependent equalities and inequalities, and form the convex feasibility problem  $P_{rba}^p$  [5, 6]. For cells growing in exponential phase at a growth rate  $\mu$ ,

- (I) the metabolic network has to produce all metabolic precursors necessary for biomass production (equalities  $C_1$  in the optimization problem);
- (II) the capacity of all molecular machines must be sufficient to ensure their function, i.e. to catalyze chemical conversions at a sufficient rate (inequalities  $C_{2b}$  for the enzymes and transporters,  $C_{2a}$  for the molecular machines of macromolecular processes);
- (III) the intracellular density of compartments and the occupancy of membranes are limited (inequalities  $C_3$ );
- (IV) mass conservation is satisfied for all molecule types (equalities  $C_1$ ).

### Remarks.

1. In practice, the vector  $\bar{B}$  contains non-zero values only for the concentrations of macro-components such as DNA, cell wall, and lipid membranes, and for a few set of metabolites (see supplementary data of [7]).
2. In order to model reversible enzymes, we have introduced two diagonal matrices of enzyme efficiencies, i.e.  $K_E$  and  $K'_E$ , describing the constraints attached to each enzyme in both directions. Actually, when an enzyme is assumed irreversible, by convention, the constraints associated to the backward direction is taken equal to 0.
3. In [6, 7], the RBA problem that was built for *Bacillus subtilis* integrates two macromolecular processes in constraint  $C_{2a}$ , the translation and chaperoning of proteins, and two density constraints, the limitation of the cytosolic density and of the membrane occupancy.
4. The RBA model can be refined by integrating for instance other molecular machines or other types of constraints such as the transcription machinery, or the protein secretion (see [7]) to cite very few. ■

To illustrate the facility to integrate new constraints, the following section presents how to integrate the degradation of molecular entities into RBA.

## 2.2 Integration of biological macromolecular turnover into RBA

The objective here is to show how to integrate the degradation of any molecular entity such as proteins, RNAs, metabolites or macromolecular machines into RBA. We restrict our developments to metabolites and proteins, other molecular entities can be handled in the same way.

**Metabolite turnover.** Some bacterial macromolecular components such as the membrane or some metabolites are damaged over time and need to be degraded and replaced. We then consider that a metabolite  $\mathbb{S}_j$  is degraded at rate  $\gamma_{S_j} > 0$ . Then, the dynamics of the concentration  $S_j$  is given by

$$\frac{dS_j(t)}{dt} = \Omega_j \nu(t) - (\mu + \gamma_{S_j}) S_j(t),$$

which leads in steady-state to deduce that for all the metabolites  $\mathbb{B}_j$  in  $\mathbb{B}$ ,  $\Omega_j \nu = (\mu + \gamma_{B_j}) \bar{B}_j$ .

**Protein turnover.** Actually, few of the bacterium's proteins are systematically degraded, with the exception of a small number of proteins generally associated with regulatory or stress mechanisms. The vast majority of proteins are stable and have an average half-life of several hours (days). Eventually, the half-life of this set of stable proteins results from the fact that they are damaged over time (by temperature or specific stresses) and are typically recognized and degraded by dedicated proteases.

We then consider that a protein has a specific turnover rate  $\gamma_{Y_j} > 0$ . The dynamics of the concentration of protein  $Y_j$  is given by

$$\frac{dY_j(t)}{dt} = \lambda_{Y_j}(t) - (\mu + \gamma_{Y_j})Y_j(t),$$

where the flux  $\lambda_{Y_j}(t)$  is the flux of protein (in number of molecules per unit of time and per unit of volume) produced by the ribosome. In steady-state, we thus have:  $\lambda_{Y_j} = (\mu + \gamma_{Y_j})Y_j$ . Moreover, when the protein is degraded, amino-acids are released at a rate of  $\gamma_{Y_j}\alpha_{kj}Y_j$ , where  $\alpha_{kj}$  is the number of the  $k$ -th amino acid in one protein. ATP is also consumed by the protease, at a rate of  $\gamma_{Y_j}\alpha'_jY_j$  where  $\alpha'_j$  is the number of ATP consumed by the protease to degrade the protein.

**Remark.** For fast-growing bacteria such as *E. coli* or *B. subtilis*, the growth rate is usually larger than the turnover rate of most of the proteins. Therefore, protein degradation can be neglected without a significant impact on the predicted resource distribution by RBA. ■

**Constraint on protease processes.** In order to degrade proteins, the cell has to possess enough proteases. That leads to introduce a constraint on the capability of the protease to degrade proteins:

$$\sum_j^{N_y} \gamma_{Y_j} Y_j(t) \leq k_P P(t)$$

where  $P(t)$  corresponds to the concentration of proteases. This concentration is clearly a new decision variable of the RBA problem. It is integrated in the vector  $Y$  since the protease is a new molecular machine.

**RBA problem for prokaryotes with protein and metabolite turnover.** The set of macromolecular machines  $\mathbb{M}$  is now increased with the protease concentration (and  $N_p$  is updated in accordance). The RBA problem is now given by : for a fixed vector of concentrations  $P_G \in \mathbb{R}_{>0}^{N_g}$ , and the growth rate  $\mu \geq 0$ ,

$$\begin{aligned} \text{find} \quad & Y \in \mathbb{R}_{\geq 0}^{m+p}, \nu \in \mathbb{R}^m, \Lambda_Y \in \mathbb{R}_{\geq 0}^y, \Lambda_G \in \mathbb{R}_{\geq 0}^g \\ \text{subject to} \quad & (C_1) \quad -\Omega\nu + C_Y^S \Lambda_Y + C_G^S \Lambda_G + C_Y^{S'} \Gamma_Y Y + C_G^{S'} \Gamma_G P_G + \dots \\ & \quad C_B^S (\mu + \Gamma_B) \bar{B} = 0 \\ & (C_{2a}) \quad C_Y^M \Lambda_Y + C_G^M \Lambda_G + C_Y^{M'} \Gamma_Y Y + C_G^{M'} \Gamma_G P_G - K_T Y \leq 0 \\ & (C_{2b}) \quad -K_E' Y \leq \nu \leq K_E Y \\ & (C_3) \quad C_Y^D Y + C_G^D P_G - \bar{D} \leq 0 \\ & (C_{4a}) \quad \Lambda_Y - (\mu + \Gamma_Y) Y = 0 \\ & (C_{4b}) \quad \Lambda_G - (\mu + \Gamma_G) P_G = 0 \end{aligned}$$

$\Gamma_Y$ ,  $\Gamma_G$  are diagonal matrix of size  $N_y \times N_y$ ,  $N_g \times N_g$  respectively, and where the coefficient  $\Gamma_{Y_{jj}}$ ,  $\Gamma_{G_{jj}}$  correspond to the turnover rate of  $\mathbb{Y}_j$  and  $\mathbb{P}_{G_j}$  respectively.  $\Gamma_B$  is a  $N_s \times N_b$  matrix

where  $\Gamma_{B_{kj}}$  is the turnover rate of  $\mathbb{B}_j$ , and index  $k$  refers to the position of  $\mathbb{B}_j$  in the set  $\mathbb{S}$ . Moreover,  $C_Y^{S'}$  (resp.  $C_G^{S'}$ ) is a  $N_s \times N_y$  (resp.  $N_s \times N_g$ ) matrix where each coefficient  $C_{Y_{ij}}^{S'}$  (resp.  $C_{G_{ij}}^{S'}$ ) corresponds to the number of metabolite  $\mathbb{S}_i$  consumed (or produced) during the degradation of one machine  $\mathbb{Y}_j$  (resp.  $\mathbb{P}_G$ ).

In addition to translation and chaperoning,  $C_{2a}$  now contains the constraint on the capability of proteases to achieve protein degradation (terms  $C_Y^{M'}\Gamma_Y Y$  and  $C_G^{M'}\Gamma_G P_G$ ), and thus the matrix  $C_Y^M$  and  $C_G^M$  have been updated appropriately.

Moreover, constraint  $C_{4a}$  and  $C_{4b}$  correspond to the steady-state regimen of the differential equation of  $Y_j(t)$  for all  $j \in (1 \dots N_y)$  and on the concentration of the  $\mathbb{P}_G$  proteins.

By substituting constraints  $(C_{4a})$  and  $(C_{4b})$  in the other ones, and by noting  $\Gamma_Y^S \triangleq (C_Y^S + C_Y^{S'})\Gamma_Y$ ,  $\Gamma_G^S \triangleq (C_G^S + C_G^{S'})\Gamma_G$ ,  $\Gamma_B^S \triangleq C_B^S\Gamma_B$ ,  $\Gamma_Y^M \triangleq (C_Y^M + C_Y^{M'})\Gamma_Y$ , and  $\Gamma_G^M \triangleq (C_G^M + C_G^{M'})\Gamma_G$ , the previous optimization problem can further be aggregated as

$$\begin{aligned}
 & \text{find} && Y \in \mathbb{R}_{\geq 0}^y, \nu \in \mathbb{R}^m, \\
 & \text{subject to} && \\
 (C_1) & -\Omega\nu + \mu(C_Y^S Y + C_G^S P_G + C_B^S \bar{B}) + \Gamma_Y^S Y + \Gamma_G^S P_G + \Gamma_B^S \bar{B} = 0 \\
 (C_{2a}) & \mu(C_Y^M Y + C_G^M P_G) + \Gamma_Y^M Y + \Gamma_G^M P_G - K_T Y \leq 0 \\
 (C_{2b}) & -K_E' Y \leq \nu \leq K_E Y \\
 (C_3) & C_Y^D Y + C_G^D P_G - \bar{D} \leq 0
 \end{aligned}$$

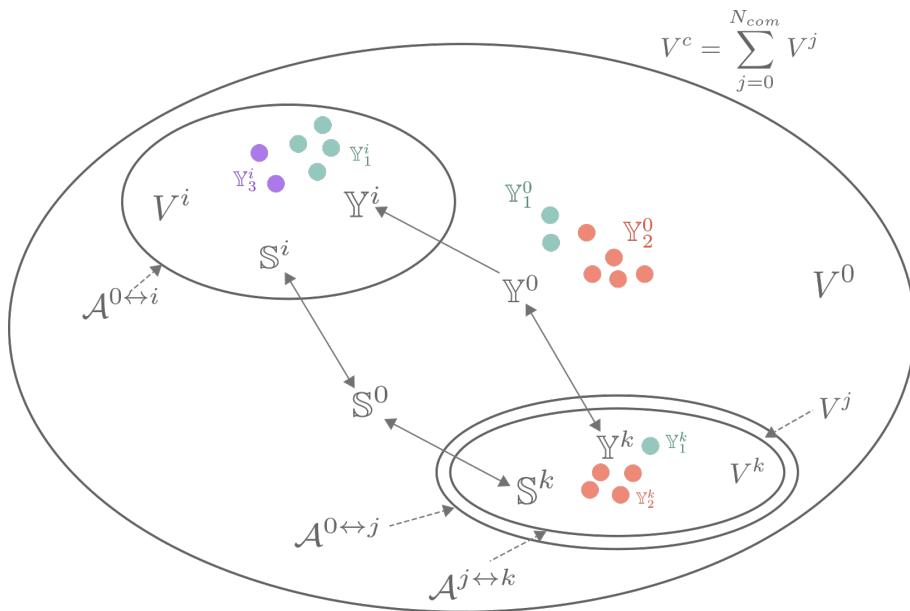
**Remark.** As in the case of protein degradation, machines that degrade cell macro-components or metabolites can also be defined and integrated into the problem in the same way. ■

### 3 RBA for eukaryotes

Eukaryotic cells are obviously much more complex than prokaryotic cells, the precise control of a large diversity of cellular processes are only partially known, even for well-established process like transcription. From a RBA perspective, we have mainly to integrate systematically in the RBA framework the presence of organelles and so to handle the fact that the metabolism, the macromolecular processes, the molecular machines and the associated constraints are now localized in different cell compartments. Actually, we notice that in genome-scale metabolic models, organelles are usually integrated only through the localization of metabolites. In the latest release of the yeast consensus model [8], the localization of some reactions is added in the reaction name as an annotation. However, the biomass reaction typically integrates the composition of the whole-cell, without distinguishing the composition of organelles. The allocation of resources to the organelles is thus fixed and constant for any condition. Hereafter, the formulation will bypass this assumption and the cell investment in the organelle is explicitly included in the resource allocation problem.

#### 3.1 Modeling organelles: first steps

We define a cell as a complex system where its cytoplasm contains organelles which are described through the introduction of  $N^{com}$  compartments, with  $N^{com} \geq 2$  since we always assume that the cell contains at least mitochondria (see Figure 3.1). It is furthermore necessary to add the  $N^{int}$  interfaces between the different compartments. At the time  $t$ , the  $i$ -th compartment has a volume  $V^i(t)$ . It furthermore has interfaces with the other compartments, where each of its interfaces has a given surface area denoted  $\mathcal{A}^{*\leftrightarrow i}$ .  $V^c(t)$  is the volume of the cytoplasm (or the



cell<sup>1</sup>) at time  $t$  and is defined by

$$V^c(t) \triangleq V^0(t) + \sum_{i=1}^{N^{com}} V^i(t)$$

where  $V^0(t)$  corresponds to the volume at time  $t$  of the part of the cell volume not contained in one of  $N^{com}$  compartments, and classically called the cytosol.

### Remarks.

1. In the sequel, the volume of the cytosol is the one which allows to define the concentration of an entity present in this part of the cell.
2. To describe a mitochondrion, two compartments must be defined: the intermembrane space and the matrix. A single interface is associated to the matrix and corresponds to the inner mitochondrial membrane. The intermembrane space has two interfaces, one with the matrix, already mentioned, and a second one with the cytosol corresponding to the outer mitochondrial membrane. ■

In order to model the cell, we then introduce a catalog of entities (defined in detail later) where it is assumed that each entity belongs to a specific compartment or/and to an interface between two compartments. The abundance of each entity is assumed to be known in each compartment and interface:  $n_{X^i}$  corresponds to the number of  $\mathbb{X}$  in the  $i$ -th compartment and  $[X^i]^i$  corresponds to its concentration in the  $i$ -th compartment, i.e.,  $[X^i]^i \triangleq n_{X^i}/V^i$ . Finally,  $[X^i]^c$  defines the concentration of  $\mathbb{X}$  in the  $i$ -th compartment with respect to the cell volume, i.e.,  $[X^i]^c \triangleq n_{X^i}/V^c$ . The total concentration of  $\mathbb{X}$  with respect to the cell volume is then given by

$$[X]^c \triangleq \sum_{i=0}^{N^{com}} \frac{n_{X^i}}{V^c}$$

where  $n_{X^0}$  corresponds to the number of  $\mathbb{X}$  in the cytosol.

For an interface between the  $i$ -th and  $j$ -th compartment, the number of  $\mathbb{X}$  in this interface is denoted by  $n_{X^{i \leftrightarrow j}}$ .

**Modeling the creation of an entity  $\mathbb{X}$  in the  $i$ -th compartment.** We now consider that the creation of  $\mathbb{X}$  in the  $i$ -th compartment is described by this differential equation:

$$\dot{n}_{X^i}(t) = a_{X^i}(t) \tag{1}$$

where  $a_{X^i}$  is then a quantity proportional to a rate, i.e., a number of molecules per unit of time. The evolution of the concentration of this entity in its compartment is obtained after straightforward computations and given by

$$\frac{d}{dt} [X^i(t)]^i = \frac{d}{dt} \left( \frac{n_{X^i}(t)}{V^i(t)} \right) = \frac{a_{X^i}(t)}{V^i(t)} - \mu^i [X^i(t)]^i$$

where  $\mu^i(t) \triangleq \dot{V}^i(t)/V^i(t)$ .

Following the same lines, it is possible to express the evolution of  $\mathbb{X}$  in the  $i$ -th compartment through the evolution of its concentration with respect to the cell volume:

$$\frac{d}{dt} [X^i(t)]^c = \frac{d}{dt} \left( \frac{n_{X^i}(t)}{V^c(t)} \right) = \frac{a_{X^i}(t)}{V^c(t)} - \mu^c(t) [X^i(t)]^c$$

---

<sup>1</sup>With a slight abuse of denomination.

where  $\mu^c \triangleq \dot{V^c}(t)/V^c(t)$ . By introducing  $\lambda_{X^i}(t) \triangleq a_{X^i}(t)/V^c(t)$ , corresponding to the rate of production of  $\mathbb{X}$  in the  $i$ -th compartment per unit of cell volume, the previous differential equation is rewritten as

$$\frac{d}{dt}[X^i(t)]^c = \lambda_{X^i}(t) - \mu^c(t)[X^i(t)]^c. \quad (2)$$

**The density constraints attached to compartments and their interfaces.** The density constraints (i.e. constraints  $(C_3)$  in the RBA problem of the previous section) are related to the fact that biological entities can only occupy a part of the volume of each compartment or of the surface defining the interfaces between the different compartments.

By virtue of a formulation introduced in [5, 6], we then assume that the macromolecules present in a compartment occupy a maximal fraction of the compartment volume (respectively, surface). Thus, we introduce  $\beta^i$ , a global parameter, attached to the  $i$ -th compartment. In the RBA framework, the coefficient  $\beta^i$  is expressed as an equivalent volume occupied by one amino acid in this compartment. This is used to express the volume or the surface occupancy of each macromolecule in terms of an equivalent number of amino acids (see [6, 7] for details). Finally, we can link the volume of the compartment to its content in macro-entities, as e.g. enzymes, chaperones, ribosomes, by this relationship:

$$V^i(t) \geq \beta^i \sum_{k=1}^{N^i} l_k n_{X_k^i}(t) \quad (3)$$

for each  $i \in \{0, 1, \dots, N^{com}\}$  and where  $N^i$  is the number of possible (macro) entities present in the  $i$ -th compartment and  $l_k$  for each  $k \in \{1, \dots, N^i\}$  is the total number of equivalent amino acid residues of the entity  $\mathbb{X}_k$ .

Following the same lines, the density acting on the surface area of the interface between the  $i$ -th and  $j$ -th compartment is given by  $\beta^{ij}$ , expressed as the equivalent surface occupied by one amino acid in this interface, and this relationship:

$$\mathcal{A}_{i \leftrightarrow j}(t) \geq \beta^{ij} \sum_{k=1}^{N^{ij}} l_k n_{X_k^{i \leftrightarrow j}}(t) \quad (4)$$

where  $N^{ij}$  is the number of possible (macro) entities present in the interface between the  $i$ -th and the  $j$ -th compartment and  $l_k$  for each  $k \in \{1, \dots, N^{ij}\}$  is the total number of amino acid residues of the entity  $\mathbb{X}_k$ .

**Remark.** More generally, the volume  $V^i$  and the surface area  $\mathcal{A}_{i \leftrightarrow j}$  of a compartment are usually linked by the so-called volume/surface ratio  $\sigma_i$ . By the definition, the volume/surface ratio  $\sigma_i$  of the  $i$ -th compartment is given by

$$\sigma_i(t) \triangleq \frac{\mathcal{A}^{0 \leftrightarrow i}(t)}{V^i(t)}.$$

We will see in Section 3.5 how the relation between the volume and the surface area of a compartment is taken into account into RBA constraints. ■

**A simplifying assumption.** In order to simplify the presentation in the reminder part of this article, we assume that the volume of each compartment, including cytosol, saturates the density constraint over time, i.e., Inequality (3) is an equality for all compartments. This implies

that the cell volume is related to the (macro) entity content of each compartment by this simple relation:

$$V^c(t) = \sum_{i=0}^{N^{com}} \beta^i \sum_{k=1}^{N^i} l_k n_{X_k^i}(t).$$

The evolution of the cell volume with respect to the time is then given by

$$\dot{V}^c(t) = \sum_{i=0}^{N^{com}} \beta^i \sum_{k=1}^{N^i} \alpha_k a_{X_k^i}(t) = \underbrace{\sum_{i=0}^{N^{com}} \beta^i \sum_{k=1}^{N^i} l_k \lambda_{X_k^i}(t) V^c(t)}_{\triangleq \mu^c(t)}.$$

The expression of the cell volume leads to define a first constraint on the concentrations of the set of all (macro) entities present in the cell:

$$\frac{V^c(t)}{V^c(t)} = 1 = \sum_{i=0}^{N^{com}} \beta^i \sum_{k=1}^{N^i} l_k [X_k^i(t)]^c.$$

If we furthermore introduce the volume of each compartment as a fraction of the cell volume:

$$f^i(t) \triangleq \frac{V^i(t)}{V^c(t)} = \beta^i \sum_{k=1}^{N^i} l_k [X_k^i(t)]^c \quad (5)$$

for all  $i \in \{0, 1, \dots, N^{com}\}$ . We thus conclude that

$$\dot{f}^i(t) = \beta^i \sum_{k=1}^{N^i} l_k \frac{d[X_k^i(t)]^c}{dt} = \beta^i \sum_{k=1}^{N^i} l_k \left( \lambda_{X_k^i}(t) - \mu^c(t) [X_k^i(t)]^c \right).$$

Moreover, the volume fraction  $f^i$  can be related to the surface area of the compartment normalized by the cell volume through Equation (4) as follows:

$$\sigma_i(t) = \frac{\mathcal{A}^{0 \leftrightarrow i}(t)}{V^i(t)} = \underbrace{\frac{\mathcal{A}^{0 \leftrightarrow i}(t)}{V^c(t)}}_{\triangleq f^{0 \leftrightarrow i}(t)} \frac{1}{f^i(t)}$$

This leads to a constraint between the surface area and the volume of the compartment:

$$f^{0 \leftrightarrow i}(t) = \sigma^i(t) f^i(t) \quad (6)$$

**Remark.** Other assumptions can be easily considered, in particular if the active constraints are the interface surface rather than the volume. Finally, if the volume or interface surfaces of compartments are defined differently by other mathematical expressions, it is important to verify that the associated constraints can be formulated as convex constraints. ■

### 3.2 How to handle the flux exchanges between different compartments

We now investigate how the exchange fluxes of metabolites between two compartments are computed. In order to simplify the presentation, we consider the  $i$ -th compartment of volume  $V^i$ , which has an interface with the cytosol of volume  $V^0$ . We will also first assume that the volumes of compartments ( $V^i$  and  $V^0$ ) are constant.

The metabolic network in the  $i$ -th compartment is classically described by a stoichiometric matrix  $\Omega^i$  and the evolution of metabolite concentrations is described by

$$\frac{d[S^i(t)]^i}{dt} = \Omega^i \nu^i(t)$$

where  $[S^i(t)]^i$  is the vector of metabolite concentrations (in the  $i$ -th compartment),  $\nu^i(t)$  is the vector of fluxes (expressed as a number of molecules per unit of time and per unit of volume), concatenating internal fluxes, denoted  $\nu_{int}^i(t)$ , and fluxes of exchange between the compartment and the cytosol, denoted  $\nu_{ex,i \leftrightarrow 0}^i(t)$ . By convention, an exchange flux is positive when the molecule is imported in the compartment and negative when it is exported from it. In such a model, since all the quantities are defined per unit of volume and are attached to a specific compartment, the exchanges with another compartment have to be converted, typically as a number of molecules per unit of time, i.e.,

$$F_{ex,i \leftrightarrow 0}^i(t) \triangleq \nu_{ex,i \leftrightarrow 0}^i(t) V^i.$$

This conversion allows us to write the equation linking the exchanges between the two considered compartments as

$$F_{ex,0 \leftrightarrow i}^0(t) \triangleq -F_{ex,i \leftrightarrow 0}^i(t).$$

However, the previous expression has a general drawback since the fluxes are now expressed as an absolute number of molecules, and this number typically increase exponentially during a growth phase. But that is not really a problem, we have just to normalize all these numbers by a *common quantity*, chosen to be for example the volume of cell. In such a way, we then handle only ‘normalized fluxes’ through all the compartments and these fluxes can now be added or subtracted (at the level of exchange). We then introduce:

$$\tilde{\nu}_k^i(t) \triangleq f^i \nu_k^i(t) \left( \triangleq \frac{V^i}{V^c} \nu_k^i(t) \right)$$

for all  $k \in \{1, \dots, N^i\}$  and all  $i \in \{0, 1, \dots, N^{com}\}$ . In a steady-state regime, the fluxes through metabolic networks satisfy these equality constraints:

$$[\Omega_{in}^0, \Omega_{ex}^0] \begin{bmatrix} \nu_{in}^0 \\ \nu_{ex,0 \leftrightarrow i}^0 \end{bmatrix} = 0 \quad \text{and} \quad [\Omega_{in}^i, \Omega_{ex}^i] \begin{bmatrix} \nu_{in}^i \\ \nu_{ex,i \leftrightarrow 0}^i \end{bmatrix} = 0$$

which can be rewritten with respect to the normalized fluxes as

$$[\Omega_{in}^0, \Omega_{ex}^0] \begin{bmatrix} \frac{1}{f^0} \tilde{\nu}_{in}^0 \\ \frac{1}{f^0} \tilde{\nu}_{ex,0 \leftrightarrow i}^0 \end{bmatrix} = 0 \quad \text{and} \quad [\Omega_{in}^i, \Omega_{ex}^i] \begin{bmatrix} \frac{1}{f^i} \tilde{\nu}_{in}^i \\ \frac{1}{f^i} \tilde{\nu}_{ex,i \leftrightarrow 0}^i \end{bmatrix} = 0$$

where  $f^0$  and  $f^i$  are defined by Relation (5). We conclude that the interconnection of both metabolic networks in steady-state can be written as

$$\begin{bmatrix} \Omega_{in}^0 & 0 & \Omega_{ex}^0 \\ 0 & \Omega_{in}^i & -\Omega_{ex}^i \end{bmatrix} \begin{bmatrix} \tilde{\nu}_{in}^0 \\ \tilde{\nu}_{in}^i \\ \tilde{\nu}_{ex,0 \leftrightarrow i}^0 \end{bmatrix} = 0$$

**Compartments with time-variant volumes.** The previous results are obtained when the volume of compartments is constant. We now consider a more general case. To do this, the effect of dilution due to the increasing compartment volume has to be included. This is classic and leads to the following differential equation:

$$\frac{d[S^i(t)]^i}{dt} = [\Omega_{in}^i, \Omega_{ex}^i] \begin{bmatrix} \nu_{in}^i(t) \\ \nu_{ex,i \leftrightarrow 0}^i(t) \end{bmatrix} - \mu^i(t)[S^i(t)]^i$$

with  $\nu_{in}^i(t)$  and  $\nu_{ex,i \leftrightarrow 0}^i(t)$  are the internal and the exchange fluxes between the  $i$ -th compartment and the cytosol respectively. By definition, we have

$$\tilde{\nu}^i(t) \triangleq \underbrace{\frac{V^i(t)}{V^c(t)}}_{f^i(t)} \nu^i(t)$$

and then

$$\frac{d[S^i(t)]^i}{dt} = [\Omega_{in}^i | \Omega_{ex}^i] \begin{bmatrix} \frac{\tilde{\nu}_{in}^i(t)}{f^i(t)} \\ \frac{\tilde{\nu}_{ex}^i(t)}{f^i(t)} \end{bmatrix} - \mu^i(t)[S^i(t)]^i.$$

Since

$$[S^i(t)]^c \triangleq \frac{n_S^i(t)}{V^c(t)} = \frac{[S^i(t)]^i V^i(t)}{V^c(t)} = [S^i(t)]^i f^i(t),$$

then we have

$$\begin{aligned} \frac{d[S^i(t)]^c}{dt} &= \frac{d[S^i(t)]^i}{dt} f^i(t) + [S^i(t)]^i \frac{df^i(t)}{dt} \\ &= [\Omega_{in}^i | \Omega_{ex}^i] \begin{bmatrix} \tilde{\nu}_{in}^i(t) \\ \tilde{\nu}_{ex}^i(t) \end{bmatrix} - \mu^i(t) \underbrace{[S^i(t)]^i f^i(t)}_{[S^i(t)]^c} \cdots \\ &\quad + [S^i(t)]^i \left[ \frac{\dot{V}_i(t)}{V^c(t)} - V^i(t) \frac{\dot{V}^c(t)}{(V^c(t))^2} \right]. \end{aligned}$$

Finally, since  $\dot{V}^c(t) \triangleq \mu^c(t)V^c(t)$  and  $\dot{V}_i(t) \triangleq \mu^i(t)V^i(t)$ , the terms depending on  $\mu^i(t)$  vanish, so we can rewrite the previous differential equations as

$$\frac{d[S^i(t)]^c}{dt} = [\Omega_{in}^i | \Omega_{ex}^i] \begin{bmatrix} \tilde{\nu}_{in}^i(t) \\ \tilde{\nu}_{ex}^i(t) \end{bmatrix} - \mu^c(t)[S^i(t)]^c.$$

### 3.3 How to link enzyme activities in the different compartments with each other

Typically, the activity of the  $k$ -th enzyme localized in the  $i$ -th compartment is described by

$$\nu_{E_k}^i(t) = [E_k^i(t)]^i F_k([S_*^i(t)]^i)$$

where  $[E_k^i]^i$  is the concentration of the  $k$ -th enzyme in the  $i$ -th compartment,  $\nu_{E_k}^i(t)$  is the flux through the enzyme expressed as a number of molecules per unit of time and per unit of volume.  $F_k$  is the nonlinear characteristic of the enzyme activity which is a function of a set of metabolite concentrations in the  $i$ -th compartment, i.e.,  $[S_*^i(t)]^i \triangleq ([S_1^i(t)]^i, \dots, [S_n^i(t)]^i)^T$ . The

notation  $S_*^i$  will refer in the sequel to a set of metabolite concentrations in the  $i$ -th compartment.

In the RBA framework, the maximal flux through the  $k$ -th enzyme is constrained by its concentration through this simple linear inequality:

$$\nu_{E_k}^i(t) \leq k_{E_k}^i [E_k^i(t)]^i.$$

This last inequality can be rewritten in an equivalent way as an inequality between a normalized flux and the concentration of the enzyme with respect to the cell volume:

$$\tilde{\nu}_{E_k}^i \leq k_{E_k}^i [E_k^i(t)]^c.$$

### 3.4 How to handle protein production in different compartments

**Protein production.** We assume that a protein  $\mathbb{Y}$  is localized in the  $i$ -th compartment, has a given concentration  $[Y^i]^i$ , and that  $\mathbb{Y}$  is produced by the machine<sup>2</sup>  $R^0$  that is located in the cytosol. We also assume that the volumes of compartments and of the cytosol are constant. So  $V^i$ ,  $f^i$ ,  $V^0$ ,  $f^0$  and  $V^c$  are constant.

The total flow of proteins (in number of molecules per unit of time) produced by the  $n_{R^0}$  machines is  $n_{R^0}k_{R^0}$ , where  $k_{R^0}$  is the efficiency of  $R^0$  (in a number of amino acid per unit of time<sup>3</sup>). Actually, like for an enzyme, the efficiency  $k_{R^0}$  can be a nonlinear function that depends on the concentration of a set of metabolites (located in the same compartment than  $R^0$ ),  $k_{R^0}([S_*^0]^0)$ . The machines produce a set of proteins that can be located in the same compartment or translocated in another one.

Let us recall that the creation (in number per unit of time) of  $\mathbb{Y}$  in the  $i$ -th compartment is given by  $a_{Y^i}$  (see Relation (1)). So the capability of machine  $R^0$  has to be sufficient to produce  $a_{Y^i}$ , and, by extension, all the proteins that  $R^0$  must produce. So we have:

$$\sum_{i=0}^{N^{com}} \sum_{k=1}^{N^i} l_k a_{Y_k^i}(t) \leq k_{R^0}([S_*^0(t)]^0) n_{R^0}(t).$$

In the previous equation,  $l_k$  is the length in amino acids of the protein  $\mathbb{Y}_k$ . In order to alleviate the notations, we assume that  $R^0$  produces all proteins of the cell. The formula can be extended directly to a subset of proteins by defining a subset of proteins per compartment. We can finally normalize the previous inequality by  $V^c$  on both sides in order to deduce this inequality:

$$\sum_{i=0}^{N^{com}} \sum_{k=1}^{N^i} l_k \lambda_{Y_k^i}(t) \leq k_{R^0}([S_*^0(t)]^0) [R^0(t)]^c. \quad (7)$$

**Remark.** Normally, the constraints in the compartment 0 is expressed with respect to the concentration in the compartment 0:

$$\sum_{i=0}^{N^{com}} \sum_{k=1}^{N^i} l_k \frac{a_{Y_k^i}(t)}{V^0(t)} \leq k_{R^0}([S_*^0(t)]^0) [R^0(t)]^0.$$

However, since  $[R^0(t)]^0 \triangleq \frac{[R^0(t)]^c}{f^0}$ , and  $\frac{a_{Y_k^i}(t)}{V^0} \triangleq \frac{a_{Y_k^i}(t)}{f^0 V^c}$ , the two inequalities are equivalent. ■

**Impact of protein production on metabolites.** The protein production consumes metabolites such as charged-tRNAs, GTP and releases others such as uncharged-tRNAs, GDP and

<sup>2</sup>Corresponding to the translation apparatus and its associated accessory proteins.

<sup>3</sup>Again, this is used to simplify the formulation of the RBA problem and more general modeling is possible (see [6, 7] for details).

phosphate. In the case of translocation processes, these metabolites can further be located in different compartments.

We first consider the case where all metabolites that are produced/consumed are located in the cytosol. To this end, we consider a metabolite  $S_k$  located in the cytosol, and that  $\gamma_{Y_k}$  copies of  $S_k$  are consumed for the synthesis of the protein  $Y$ . The flux  $\nu_Y^0$  (in number of molecules per unit of time and per unit of volume) of  $S_k$  consumed by  $R^0$  to produce  $Y$  in the cytosol is equal to  $\gamma_{Y_k} \frac{a_{Y^i}(t)}{V^0}$ . So we have:

$$\frac{d[S_k^0(t)]^0}{dt} = \Omega_k^0 \nu^0(t) - \gamma_{Y_k} \frac{a_{Y^i}(t)}{V^0}.$$

### Remarks.

1. If the metabolite is produced, then the term  $\gamma_{Y_k} \frac{a_{Y^i}(t)}{V^0}$  is added and not subtracted in the equation.
2. If the protein  $Y$  is produced in the cytosol, and then remains in the cytosol instead of being translocated in another compartment, the flux  $\nu_Y^0$  is equal to  $\gamma_{Y_k} \frac{a_{Y^0}(t)}{V^0}$ . ■

By using the same notation than the previous section, we have:

$$\frac{d[S_k^0(t)]^0}{dt} = \Omega_k^0 \tilde{\nu}^0(t) - \gamma_{Y_k} \frac{a_{Y^i}(t)}{f^0 V^c}.$$

And for all proteins that are produced by  $R^0$  in the cytosol, we have:

$$\frac{d[S_k^0(t)]^0}{dt} = \Omega_k^0 \tilde{\nu}^0(t) - \sum_{i=0}^{N^{com}} \sum_{j=1}^{N^i} \gamma_{Y_{j,k}} \frac{a_{Y^i}(t)}{f^0 V^c}. \quad (8)$$

**Steady-state regime.** In steady-state, from Equation (2), Inequality (7) becomes

$$\mu^c \sum_{i=0}^{N^{com}} \sum_{j=1}^{N^i} l_j [Y_j^i]^c \leq k_{R^0} ([S_*^0]^0) [R^0]^c,$$

and Equality (8) becomes:

$$\Omega_k^0 \tilde{\nu}^0 - \mu^c \sum_{i=0}^{N^{com}} \sum_{j=1}^{N^i} \gamma_{Y_{j,k}} [Y_j^i]^c = 0.$$

**Compartments with time-varying volumes.** Now let us consider that the volumes are time-varying. The dynamics of  $[S_k^0(t)]^0$  with respect to time is given by

$$\frac{d[S_k^0(t)]^0}{dt} = \Omega_k^0 \frac{\tilde{\nu}^0(t)}{f^0(t)} - \gamma_{Y_k} \frac{a_{Y^i}(t)}{f^0(t) V^c(t)} - \mu^0(t) [S_k^0(t)]^0.$$

Since we have  $[S_k^0(t)]^c \triangleq f^0(t) [S_k^0(t)]^0$  then

$$\frac{d[S_k^0(t)]^c}{dt} = \dot{f}^0(t) [S_k^0(t)]^0 + f^0(t) \frac{d[S_k^0(t)]^0}{dt}$$

and by summing on all proteins:

$$\frac{d[S_k^0(t)]^c}{dt} = \left( \Omega_k^0 \tilde{\nu}^0(t) - \sum_{i=0}^{N^{com}} \sum_{j=1}^{N^i} \gamma_{Y_{j,k}} \frac{a_{Y_j^i}(t)}{f^0(t)V^c(t)} - \mu^0(t)[S_k^0(t)]^0 \right) f^0(t) \dots + [S_k^0(t)]^0 f^0(t)(\mu^0(t) - \mu^c(t))$$

which finally leads to deduce that

$$\frac{d[S_k^0(t)]^c}{dt} = \Omega_k^0 \tilde{\nu}^0(t) - \sum_{i=0}^{N^{com}} \sum_{j=1}^{N^i} \gamma_{Y_{j,k}} \lambda_{Y_j^i}(t) - \mu^c(t)[S_k^0(t)]^c.$$

**Protein translocation.** We now consider the translocation process. A molecular machine, the translocon, translocates proteins from one compartment to another one, and usually consumes and produces metabolites that can be located in both compartments. One metabolite (such as protons) can also be translocated with the protein. In this paragraph, we establish the formulation of constraints related to the translocon capability, and also on the associated metabolites.

We consider (i) 3 metabolites  $\mathbb{S}_1$ ,  $\mathbb{S}_2$  and  $\mathbb{S}_3$ .  $\mathbb{S}_1$  is located in the cytosol,  $\mathbb{S}_2$  in the  $i$ -th compartment, and  $\mathbb{S}_3$  is translocated between compartments; (ii) the translocon  $\mathbb{T}$  that is localized in the membrane; and (iii) a protein  $\mathbb{Y}$  that is translocated. The membrane is considered as the  $0 \leftrightarrow i$  interface. Moreover, we consider that all volumes are constant.

As in the previous sections, we have a number of translocon  $n_{T^{0 \leftrightarrow i}}$  that catalyzes the translocation of  $n_Y$  copies of  $\mathbb{Y}$  from the cytosol to the  $i$ -th compartment. In most of organisms, such translocation steps are irreversible. We thus have one translocon to translocate proteins from the cytosol to the  $i$ -th compartment (e.g. Tim/Tom, Tic/Toc machines). What count in our model will be the active protein  $\mathbb{Y}$  in the  $i$ -th compartment. We will thus note  $a_{Y^i}$  the number of  $\mathbb{Y}$  that will be translocated per unit of time and that will be present finally in the  $i$ -th compartment.

The capability of the translocon (in number per unit of time) must be sufficient to translocate the flow of proteins  $\mathbb{Y}$  (also expressed in unit of time). So we have the constraint:

$$l_Y a_{Y^i}(t) \leq k_T([S_*^0(t)]^0, [S_*^i(t)]^i) n_{T^{0 \leftrightarrow i}}(t)$$

with  $k_T([S_*^0(t)]^0, [S_*^i(t)]^i)$  the efficiency of the translocon (in number of amino acids residues per unit of time), depending on a set of metabolites that can be located in both compartments, and  $l_Y$  the length of amino acid residues of  $\mathbb{Y}$ . We can normalize the previous inequality by  $V^c$  and obtained

$$l_Y \lambda_{Y^i}(t) \leq k_T([S_*^0(t)]^0, [S_*^i(t)]^i) [T^{0 \leftrightarrow i}(t)]^c.$$

And for all proteins that are translocated by  $\mathbb{T}$ , we have:

$$\sum_{j=1}^{N^i} l_j \lambda_{Y_j^i}(t) \leq k_T([S_*^0(t)]^0, [S_*^i(t)]^i) [T^{0 \leftrightarrow i}(t)]^c.$$

**Impact of translocation on metabolites.** For  $\mathbb{S}_1$  and  $\mathbb{S}_2$ , we have the same type of relations than in the previous paragraph. The flux (in number of molecules per unit of volume per unit of time) of  $\mathbb{S}_1$  (located in the cytosol) consumed by  $\mathbb{T}$  to translocate  $\mathbb{Y}$  is equal to  $\delta_{Y_1} \frac{a_{Y^i}(t)}{V^0}$ , where  $\delta_{Y_1}$  is the number of  $\mathbb{S}_1$  consumed per translocated protein. So for all proteins:

$$\frac{d[S_1^0(t)]^0}{dt} = \Omega_1^0 \nu^0(t) - \sum_{j=1}^{N^i} \delta_{Y_{j,1}} \frac{a_{Y_j^i}(t)}{V^0}$$

where  $\delta_{Y_{j,1}}$  is the number of  $\mathbb{S}_1$  consumed to translocate  $Y_j$ . And for  $\mathbb{S}_2$ :

$$\frac{d[S_2^i(t)]^i}{dt} = \Omega_2^i \nu^i(t) - \sum_{j=1}^{N^i} \delta_{Y_{j,2}} \frac{a_{Y_j^i}(t)}{V^i}$$

where  $\delta_{Y_{j,2}}$  is the number of  $\mathbb{S}_2$  consumed to translocate  $Y_j$ .

By normalizing by  $V^c$  instead and by using the same notations than before, we have:

$$\frac{d[S_1^0(t)]^0}{dt} = \Omega_1^0 \tilde{\nu}^0(t) - \sum_{j=1}^{N^i} \delta_{Y_{j,1}} \frac{\lambda_{Y_j^i}(t)}{f^0}$$

and:

$$\frac{d[S_2^i(t)]^i}{dt} = \Omega_2^i \tilde{\nu}^i(t) - \sum_{j=1}^{N^i} \delta_{Y_{j,2}} \frac{\lambda_{Y_j^i}(t)}{f^i}$$

Let us now consider  $\mathbb{S}_3$  and the two concentrations in the cytosol and in the  $i$ -th compartment. The flux of  $\mathbb{S}_3$  (in number of molecules per unit of time) that crosses the interface  $0 \leftrightarrow i$  (from the cytosol to the  $i$ -th compartment) is equal to:

$$F_{0 \leftrightarrow i}^0 = \nu_3^{0 \leftrightarrow i}(t) V^0 = \delta_{Y_3} a_{Y^i}(t).$$

By the mass conservation, the flux of  $\mathbb{S}_3$  that enters is equal to  $-F_{0 \leftrightarrow i}^0$ .

In the cytosol, the dynamics of the concentration of  $\mathbb{S}_3$  when the protein  $\mathbb{Y}$  is translocated, is given by

$$\frac{d[S_3^0(t)]^0}{dt} = \Omega_3^0 \nu^0(t) - \delta_{Y_3} \frac{a_{Y^i}(t)}{V^0},$$

where  $\delta_{Y_3}$  is the number of  $\mathbb{S}_3$  that is translocated together with  $\mathbb{Y}$ .

For the concentration of  $\mathbb{S}_3$  in the  $i$ -th compartment, we have:

$$\frac{d[S_3^i(t)]^i}{dt} = \Omega_3^i \nu^i(t) + \delta_{Y_3} \frac{a_{Y^i}(t)}{V^i}.$$

By using the same transformation than the previous section, we finally obtain:

$$\begin{aligned} \frac{d[S_3^0(t)]^0}{dt} &= \Omega_3^0 \frac{\tilde{\nu}^0(t)}{f^0} - \delta_{Y_3} \frac{a_{Y^i}(t)}{f^0 V^c} \\ \frac{d[S_3^i(t)]^i}{dt} &= \Omega_3^i \frac{\tilde{\nu}^i(t)}{f^i} + \delta_{Y_3} \frac{a_{Y^i}(t)}{f^i V^c} \end{aligned}$$

For all proteins that are translocated, we have:

$$\begin{aligned} \frac{d[S_3^0(t)]^0}{dt} &= \Omega_3^0 \frac{\tilde{\nu}^0(t)}{f^0} - \sum_{j=1}^{N^i} \delta_{Y_{j,3}} \frac{a_{Y_j^i}(t)}{f^0 V^c} \\ \frac{d[S_3^i(t)]^i}{dt} &= \Omega_3^i \frac{\tilde{\nu}^i(t)}{f^i} + \sum_{j=1}^{N^i} \delta_{Y_{j,3}} \frac{a_{Y_j^i}(t)}{f^i V^c} \end{aligned}$$

**Steady state regimen.** In steady state, we thus have

$$\mu^c \sum_{j=1}^{N^i} l_j [Y_j^i]^c \leq k_T([S_*^0(t)]^0, [S_*^i(t)]^i) [T^{0 \leftrightarrow i}]^c,$$

and for the metabolites  $\mathbb{S}_1$ ,  $\mathbb{S}_2$  and  $\mathbb{S}_3$ , these following constraints:

$$\begin{aligned}\Omega_1^0 \tilde{\nu}^0 - \mu^c \sum_{j=1}^{N^i} \delta_{Y_{j,1}} [Y_j^i]^c &= 0 \\ \Omega_2^i \tilde{\nu}^i - \mu^c \sum_{j=1}^{N^i} \delta_{Y_{j,2}} [Y_j^i]^c &= 0 \\ \Omega_3^0 \tilde{\nu}^0 - \mu^c \sum_{j=1}^{N^i} \delta_{Y_{j,3}} [Y_j^i]^c &= 0 \\ \Omega_3^i \tilde{\nu}^i + \mu^c \sum_{j=1}^{N^i} \delta_{Y_{j,3}} [Y_j^i]^c &= 0\end{aligned}$$

**Compartments with time-variant volumes.** For the general case, we follow the same lines than in the previous section, and obtain for  $\mathbb{S}_2$  for instance:

$$\frac{d[S_2^i(t)]^c}{dt} = \Omega_2^i \tilde{\nu}^i(t) - \sum_{j=1}^{N^i} \delta_{Y_{j,2}} \lambda_{Y_j^i}(t) - \mu^c(t) [S_2^i(t)]^c.$$

### 3.5 RBA problem for the $i$ -th compartment

In this section, we write the RBA constraints for the  $i$ -th compartment of an eukaryotic cell. As we said previously, we consider hereafter the concentration of molecular species with respect to the volume of the cell (e.g.  $[Y_j^i]^c$  in Section 3.1). However, for the sake of readability, we will omit the brackets in the notation of concentrations in the sequel. The notation  $Y_j^i$  will thus refer to  $[Y_j^i]^c$ , the concentration of the machine  $\mathbb{Y}_j$  localized in the  $i$ -th compartment with respect to the cell volume. The notation  $P_G^i$  will refer to the concentration of the  $\mathbb{P}_G$  proteins in the  $i$ -th compartment ( $[P_G^i]^c$ ). We also considered the vector of normalized flux  $\tilde{\nu}$ . Moreover, the concentration of metabolites  $\bar{B}_j^i$  will refer to  $[\bar{B}_j^i]^c \triangleq \frac{\bar{n}_{B_j^i}}{V^c}$ , where  $\bar{n}_{B_j^i}$  is constant and given. Again, for the sake of readability, we will not integrate the turnover rate of molecular entities.

Thereafter, we consider the most common case: an organelle that is isolated from the cytosol by a lipidic membrane. We will consider that some metabolic and macromolecular processes occur within the organelle, and that some molecular entities (e.g. metabolites, proteins, to cite a few) are produced in the cytosol and imported into the organelle. Even if the constraints will be formulated for this type of organelle, the formulation is generic enough to describe the exchanges between any compartments.

Let us consider the  $i$ -th compartment and its interface with the cytosol  $0 \leftrightarrow i$ . By convention, the fluxes that enter within the compartments are positive, and the ones that are excreted are negative. Let us consider the  $N_y^i$  machines present in the  $i$ -th compartment of concentrations  $Y^i \geq 0$ , and being synthetized at rate  $\Lambda_Y^i$ . Among the  $N_y^i$  machines, we have  $N_m^i$  enzymes and the vector of their metabolic fluxes  $\tilde{\nu}^i$ . Moreover, we will further specify  $\tilde{\nu}^i$  with the internal metabolic fluxes  $\tilde{\nu}_{in}^i$  and the ones of exchange through the interface  $\tilde{\nu}_{ex}^{0 \leftrightarrow i}$ .

**Metabolic and capacity constraints.** The constraints attached to the metabolic network, and to the production of molecular machines for compartment  $i$  are:

$$\left\{ \begin{array}{ll} (C_1^i) & -\Omega^i \tilde{\nu}^i + C_Y^{S,i} \Lambda_Y^i + C_G^{S,i} \Lambda_G^i + \mu^c C_B^{S,i} \bar{B}^i = 0 \\ (C_{2a}^i) & C_Y^{M,i} \Lambda_Y^i + C_G^{M,i} \Lambda_G^i - K_T^i Y^i \leq 0 \\ (C_{2b}^i) & -K_E^i Y^i \leq \tilde{\nu}_{in}^i \leq K_E^i Y^i \\ (C_{4a}^i) & \Lambda_Y^i - \mu^c Y^i = 0 \\ (C_{4b}^i) & \Lambda_G^i - \mu^c P_G^i = 0 \end{array} \right.$$

where

- $(C_1^i)$  describes the mass conservation for the set of metabolites localized in the  $i$ -th compartment ( $S^i$ ). Moreover, we consider that a subset of metabolites also has a fixed concentration  $\bar{B}^i$  (also given with respect to the cell volume);
- $(C_{2a}^i)$  describes the capability constraints of the molecular machines localized in the  $i$ -th compartment;
- $(C_{2b}^i)$  describes the capability constraints of the enzymes localized in the  $i$ -th compartment. Constraints on transporter are defined in the constraints related to the compartment interface.
- $(C_{4a}^i)$  and  $(C_{4b}^i)$  are the constraints of steady-state regimen for the dynamics of the concentration of molecular machine and  $P_G$  localized in the  $i$ -th compartment.

In all constraints, the matrixes  $C_*^{*,i}$  contains the same type of coefficients than the ones in Section 2.1. Only the structure of the matrices changes with respect to the type of process and to the molecular machines that are present in the  $i$ -th compartment and in the  $j$ -th constraint. For instance, if the  $i$ -th compartment is the mitochondria, then since translation of some mitochondrial proteins occurs only in the mitochondria, then  $(C_{2a}^i)$  contains a constraint of the capability of the mitochondrial translation process. Therefore, we need to identify the proteins that are only translated in the mitochondria to build the constraints. If the vector  $\Lambda_Y^i$  is organized along the machines that are produced in or out the mitochondria  $\Lambda_Y^i = \begin{bmatrix} \Lambda_Y^{i,in} \\ \Lambda_Y^{i,out} \end{bmatrix}$ ,

then  $C_Y^{M,i} = [l_Y^{i,in}, 0]$  where  $l_Y^{i,in}$  contains the length in amino acids of mitochondrial molecular machines produced in the mitochondria.

We also chose to put the constraints  $(C_{4a}^i)$  for all molecular machines of the  $i$ -th compartment whatever is their location of production for readability purpose.

**Remark.** As mentionned previously, the set  $\mathbb{B}$  of metabolites having fixed concentrations usually contains building blocks of macro-components such as DNA, cell wall or membrane, and thus in particular the building blocks for the membrane of organelles (i.e. the interface of compartment). The term  $\mu^c C_B^{S,i} \bar{B}^i$  typically corresponds to the flux necessary to maintain the concentration  $B^i$  constant and equal to  $\bar{B}^i$  at growth rate  $\mu^c$ . If we assume that the set  $\mathbb{B}$  contains at least one metabolite  $\mathbb{B}_k^i$  for the synthesis of any interface (for instance the interface  $0 \leftrightarrow i$ ), then we have:

$$\bar{B}_k^i \triangleq \frac{\bar{n}_{B_k^i}}{V^c} = \frac{\bar{n}_{B_k^i}}{\mathcal{A}_{0 \leftrightarrow i}} \underbrace{\frac{\mathcal{A}_{0 \leftrightarrow i}}{V^c}}_{=f^{0 \leftrightarrow i}} = f^{0 \leftrightarrow i} \frac{\bar{n}_{B_k^i}}{\mathcal{A}_{0 \leftrightarrow i}} = f^{0 \leftrightarrow i} \hat{B}_k^i$$

where  $\hat{B}_k^i \triangleq \frac{\bar{n}_{B_k^i}}{\mathcal{A}_{0 \leftrightarrow i}}$  corresponds to the number of  $\mathbb{B}_k$  per unit of surface area.

In this this case, constraint  $C_1^i$  is rewritten as:

$$-\Omega^i \tilde{\nu}^i + C_Y^{S,i} \Lambda_Y^i + C_G^{S,i} \Lambda_G^i + \mu^c C_B^{S,i} (\bar{B}^i + f^{0 \leftrightarrow i} \hat{B}^i) = 0$$

■

For the interface  $0 \leftrightarrow i$ , we have:

$$\left\{ \begin{array}{l} (C_{2a}^{0 \leftrightarrow i}) \quad C_Y^{M,0 \rightarrow i} \Lambda_Y^i + C_Y^{M,0 \leftarrow i} \Lambda_Y^0 + C_Y^{M,0 \leftrightarrow i} \Lambda_Y^{0 \leftrightarrow i} \dots \\ \quad + C_G^{M,0 \leftrightarrow i} \Lambda_G^{0 \leftrightarrow i} - K_T^{0 \leftrightarrow i} Y^{0 \leftrightarrow i} \leq 0 \\ (C_{2b}^{0 \leftrightarrow i}) \quad -K_E^{0 \leftrightarrow i} Y^{0 \leftrightarrow i} \leq \tilde{\nu}_{ex}^{0 \leftrightarrow i} \leq K_E^{0 \leftrightarrow i} Y^{0 \leftrightarrow i} \\ (C_{4a}^{0 \leftrightarrow i}) \quad \Lambda_Y^{0 \leftrightarrow i} - \mu^c Y^{0 \leftrightarrow i} = 0 \\ (C_{4b}^{0 \leftrightarrow i}) \quad \Lambda_G^{0 \leftrightarrow i} - \mu^c P_G^{0 \leftrightarrow i} = 0 \end{array} \right.$$

Here:

- the term  $C_Y^{M,0 \rightarrow i} \Lambda_Y^i$  corresponds to the flux of machines that are (i) active in the  $i$ -th compartment, (ii) produced in the cytosol and (iii) translocated through the interface.
- The term  $C_Y^{M,0 \leftarrow i} \Lambda_Y^0$  corresponds to the flux of machines that are (i) active in the cytosol, (ii) produced in  $i$ -th compartment and (iii) translocated through the interface<sup>4</sup>.
- The term  $C_Y^{M,0 \leftrightarrow i} \Lambda_Y^{0 \leftrightarrow i}$  corresponds to the flux of the machines localized in the interface  $0 \leftrightarrow i$  and that needs a secretion machine to be inserted in the membrane.

The vector of machines  $Y^{0 \leftrightarrow i}$  typically regroups the systems of import and export (if any) of proteins (e.g. TIM/TOM or TIC/TOC machine), and secretion machines to insert the membranous machines in the interface of the compartment. Only one global secretion machine could be integrated, as well as several machines (e.g. insertion of machines produced in compartment  $i$ , or in compartment 0) depending the level of details that we want to consider.

**Remark.** Often in eukaryotes, a molecular machine localized in the  $i$ -th compartment  $\mathbb{Y}_j^i$  is composed of subunits that are synthesized in different compartments; that thus need to be translocated. This is well represented in the RBA framework through coefficients  $C_Y^{M,0 \leftrightarrow i}$ . In that case, this coefficient will be equal to the length in amino acid residues of the subunits of  $\mathbb{Y}_j^i$  that need to be translocated. ■

**Volume constraints.** It remains to formalize the constraints on the volumes of compartments in steady-state. Some organelles such as the peroxisome, the mitochondria, the chloroplast were ancestrally bacteria. Their internal density is higher than the cytosol, and we could assume that their principle of growth follows the one of prokaryotic cells: the density of the compartment should be kept constant, and thus the increase of volume is driven by an increase in mass. Alternatively, the constraints could arise from the membrane occupancy. In this case, the increase of volume is driven by an increase in the mass of membranous proteins.

Following Section 3.1, a compartment satisfying the density constraint has this following volume:

$$V^i(t) = \beta^i \left( \sum_{j=1}^{N^i} l_j n_{Y_j^i}(t) + \sum_{j=1}^{N_g^i} l_j n_{G_j^i}(t) \right)$$

<sup>4</sup> Again, we introduce this term to be as generic as possible. However, the export of proteins may not occur in all organelles, so this term could vanish.

which can be normalized with respect to the cell volume leading to obtain this steady-state regimen:

$$f^i = \beta^i \left( \sum_{j=1}^{N^i} l_j Y_j^i + \sum_{k=1}^{N_g^i} l_k P_{G_k}^i \right)$$

where  $l_j$  and  $l_k$  are the equivalent length in amino acid residues of the machine  $\mathbb{Y}_j$  and protein  $P_{G_k}$  respectively, and  $\beta^i$  is the volume occupied by one amino acid in the  $i$ -th compartment. The term  $\beta^i l_j Y_j^i$  then correspond to the fraction of the volume of the cell occupied by the machine  $\mathbb{Y}_j^i$ .

**Remark.**  $\beta^i$  would typically correspond to  $\frac{1}{D^i}$  in the RBA for prokaryotic cells. ■

Moreover, we also have the constraint between the volume and the surface area of the compartment (Equality (6)):

$$f^{0 \leftrightarrow i} = \sigma^i f^i$$

We could for instance consider a fixed surface/volume ratio  $\sigma^i$ , so that when the volume of the compartment is higher, so is the surface.

Finally, the user should decide which constraints  $f^i$  or  $f^{0 \leftrightarrow i}$  is the active one:

$$(C_{3a}^i) \triangleq \begin{cases} \beta^i \left( \sum_{j=1}^{N^i} l_j Y_j^i + \sum_{k=1}^{N_g^i} l_k P_{G_k}^i \right) = f^i \\ \beta^{0 \leftrightarrow i} \left( \sum_{j=1}^{N^{0 \leftrightarrow i}} l_j Y_j^{0 \leftrightarrow i} + \sum_{k=1}^{N_g^{0 \leftrightarrow i}} l_k P_{G_k}^{0 \leftrightarrow i} \right) \leq f_{0 \leftrightarrow i} \\ f^{0 \leftrightarrow i} = \sigma^i f^i \end{cases}$$

or

$$(C_{3b}^i) \triangleq \begin{cases} \beta^i \left( \sum_{j=1}^{N^i} l_j Y_j^i + \sum_{k=1}^{N_g^i} l_k P_{G_k}^i \right) \leq f^i \\ \beta^{0 \leftrightarrow i} \left( \sum_{j=1}^{N^{0 \leftrightarrow i}} l_j Y_j^{0 \leftrightarrow i} + \sum_{k=1}^{N_g^{0 \leftrightarrow i}} l_k P_{G_k}^{0 \leftrightarrow i} \right) = f_{0 \leftrightarrow i} \\ f^{0 \leftrightarrow i} = \sigma^i f^i \end{cases}$$

Mitochondria and the thylakoid could typically be modeled by  $(C_{3b}^i)$  while chloroplast and peroxisome by  $(C_{3a}^i)$ .

Moreover, we could also add constraints on the minimal and maximal volume occupied by the compartment such as

$$f^i \leq f^i \leq \bar{f}^i$$

where  $\underline{f}^i$ ,  $\bar{f}^i$  are the minimal and maximal ratio of the  $i$ -th compartment with respect to the cell volume respectively. Having  $\underline{f}^i = 0$  means that the compartment may be absent.

Finally, the volume of the cytosol is defined as the volume of the cell minus the volume of all organelles. In addition, the volume occupied by all machines within the cytosol must be

compatible with the volume of the cytosol:

$$(C_3^0) \triangleq \begin{cases} f^0 = 1 - \sum_{i=1}^{N^{com}} f^i \\ \underline{f}^0 \leq f^0 \\ \beta^0 \left( \sum_{j=1}^{N^0} l_j Y_j^0 + \sum_{k=1}^{N_g^0} l_k P_{G_k}^0 \right) \leq f^0. \end{cases}$$

**Remark.** For complex structures of organelles, such as mitochondria or chloroplast, we expect that other constraints between volumes or surfaces should be added in order to guarantee the structure of the organelle. For instance, for mitochondria, the volume of the membranous interspace between the outer and the inner membrane and of the mitochondrial matrix should be linked such as

$$f^{inter} = \alpha_i f^{matrix}$$

where  $\alpha_i$  is the ratio between both volumes. Thus, if one of the interface or one of the volume is increasing, so are the others.

However, at this stage, the way to tackle the structure of organelles is more a modelling issue, and is thus up to the modeller. The modeler can describe the structure he wants, as soon as the mathematical formulation remains convex. ■

**Formulation of the constraints for the  $i$ -th compartment.** To summarize, we write now the set of constraints  $C_{rba}^i$  attached to the  $i$ -th compartment, composed of the interior and its interface by

$$C_{rba}^i \triangleq \begin{cases} (C_1^i) & -\Omega^i \tilde{\nu}^i + C_Y^{S,i} \Lambda_Y^i + C_G^{S,i} \Lambda_G^i + \mu^c C_B^{S,i} (\bar{B}^i + f^{0 \leftrightarrow i} \hat{B}^i) = 0 \\ (C_2^i) & C_Y^{M,i} \Lambda_Y^i + C_G^{M,i} \Lambda_G^i - K_T^i Y^i \leq 0 \\ & -K_E^{i'} Y^i \leq \tilde{\nu}_{in}^i \leq K_E^i Y^i \\ (C_2^{0 \leftrightarrow i}) & C_Y^{M,0 \leftrightarrow i} \Lambda_Y^i + C_Y^{M,0 \leftrightarrow i} \Lambda_Y^0 + C_Y^{M,0 \leftrightarrow i} \Lambda_Y^{0 \leftrightarrow i} \dots \\ & + C_G^{M,0 \leftrightarrow i} \Lambda_G^{0 \leftrightarrow i} - K_T^{0 \leftrightarrow i} Y^{0 \leftrightarrow i} \leq 0 \\ & -K_E^{0 \leftrightarrow i} Y^{0 \leftrightarrow i} \leq \tilde{\nu}_{ex}^{0 \leftrightarrow i} \leq K_E^{0 \leftrightarrow i} Y^{0 \leftrightarrow i} \\ (C_3^i) & \beta^i \left( \sum_{j=1}^{N^i} l_j Y_j^i + \sum_{k=1}^{N_g^i} l_k P_{G_k}^i \right) \leq f^i \\ & \beta^{0 \leftrightarrow i} \left( \sum_{j=1}^{N^{0 \leftrightarrow i}} l_j Y_j^{0 \leftrightarrow i} + \sum_{k=1}^{N_g^{0 \leftrightarrow i}} l_k P_{G_k}^{0 \leftrightarrow i} \right) = f^{0 \leftrightarrow i} \\ & f^{0 \leftrightarrow i} = \sigma^i f^i \\ & \underline{f}^i \leq f^i \leq \bar{f}^i \\ (C_4^i) & \Lambda_Y^i - \mu^c Y^i = 0 \\ & \Lambda_G^i - \mu^c P_G^i = 0 \\ (C_4^{0 \leftrightarrow i}) & \Lambda_Y^{0 \leftrightarrow i} - \mu^c Y^{0 \leftrightarrow i} = 0 \\ & \Lambda_G^{0 \leftrightarrow i} - \mu^c P_G^{0 \leftrightarrow i} = 0 \end{cases}$$

Here we assume that the constraints on volume follows the case  $(C_{3b}^i)$ . For the other type of organelles and for the cytosol, only the  $(C_3^i)$  constraints need to be substituted by  $(C_{3a}^i)$  and

$(C_3^0)$  respectively.

In  $C_{rba}^i$ , the decision variables are:

- the flux vector  $\tilde{\nu}^i$  of size  $N_m^i + N_m^{0 \leftrightarrow i}$
- the concentration of molecular machines  $Y^i$  and  $Y^{0 \leftrightarrow i}$  of size  $N_m^i + N_p^i$  and  $N_m^{0 \leftrightarrow i} + N_p^{0 \leftrightarrow i}$  respectively
- the synthesis rates of molecular machines ( $\Lambda_Y^{0 \leftrightarrow i}$  and  $\Lambda_Y^i$  of same size as  $Y^i$  and  $Y^{0 \leftrightarrow i}$ ) and of  $\mathbb{P}_G$  proteins ( $\Lambda_G^{0 \leftrightarrow i}$  and  $\Lambda_G^i$  of size  $N_g^i$  and  $N_g^{0 \leftrightarrow i}$  respectively)
- the volume  $f^i$  and  $f^{0 \leftrightarrow i}$ .

The previous constraints can be further aggregated by introducing  $Y^{\bar{i}} \triangleq [Y^i, Y^{0 \leftrightarrow i}]^T$ ,  $\Lambda_Y^{\bar{i}} \triangleq [\Lambda_Y^i, \Lambda_Y^{0 \leftrightarrow i}]^T$ ,  $P_G^{\bar{i}} \triangleq [P_G^i, P_G^{0 \leftrightarrow i}]^T$ ,  $\Lambda_G^{\bar{i}} \triangleq [\Lambda_G^i, \Lambda_G^{0 \leftrightarrow i}]^T$ , and by rewriting constraint  $C_3$  under a matrix form:

$$\left\{ \begin{array}{ll} (C_1^{\bar{i}}) & -\Omega^i \tilde{\nu}^i + C_Y^{S,\bar{i}} \Lambda_Y^{\bar{i}} + C_G^{S,\bar{i}} \Lambda_G^{\bar{i}} + \mu^c C_B^{S,i} (\bar{B}^i + f^{0 \leftrightarrow i} \hat{B}^i) = 0 \\ (C_2^{\bar{i}}) & C_Y^{M,\bar{i}} \Lambda_Y^{\bar{i}} + C_G^{M,\bar{i}} \Lambda_G^{\bar{i}} - K_T^{\bar{i}} Y^{\bar{i}} \leq 0 \\ & -K_E^{\bar{i}} Y^{\bar{i}} \leq \tilde{\nu}^i \leq K_E^{\bar{i}} Y^{\bar{i}} \\ (C_3^{\bar{i}}) & C_Y^{D,i} Y^{\bar{i}} + C_G^{D,i} P_G^{\bar{i}} - f^i \leq 0 \\ & C_Y^{D,0 \leftrightarrow i} Y^{\bar{i}} + C_G^{D,0 \leftrightarrow i} P_G^{\bar{i}} - f^{0 \leftrightarrow i} = 0 \\ & f^{0 \leftrightarrow i} = \sigma^i f^i \\ & \underline{f}^i \leq f^i \leq \bar{f}^i \\ (C_4^{\bar{i}}) & \Lambda_Y^{\bar{i}} - \mu^c Y^{\bar{i}} = 0 \\ & \Lambda_G^{\bar{i}} - \mu^c P_G^{\bar{i}} = 0 \end{array} \right.$$

and where all matrices  $C_*^{*,\bar{i}}$  are deduced by straightforward manipulations.

Finally, if we substitute constraints  $(C_4^{\bar{i}})$  in the other ones, we obtain:

$$C_{rba}^{\bar{i}} \triangleq \left\{ \begin{array}{ll} (C_1^{\bar{i}}) & -\Omega^i \tilde{\nu}^i + \mu^c (C_Y^{S,\bar{i}} Y^{\bar{i}} + C_G^{S,\bar{i}} P_G^{\bar{i}} + C_B^{S,i} (\bar{B}^i + f^{0 \leftrightarrow i} \hat{B}^i)) = 0 \\ (C_2^{\bar{i}}) & \mu^c (C_Y^{M,\bar{i}} Y^{\bar{i}} + C_G^{M,\bar{i}} P_G^{\bar{i}} - K_T^{\bar{i}} Y^{\bar{i}} \leq 0 \\ & -K_E^{\bar{i}} Y^{\bar{i}} \leq \tilde{\nu}^i \leq K_E^{\bar{i}} Y^{\bar{i}} \\ (C_3^{\bar{i}}) & C_Y^{D,i} Y^{\bar{i}} + C_G^{D,i} P_G^{\bar{i}} - f^i \leq 0 \\ & C_Y^{D,0 \leftrightarrow i} Y^{\bar{i}} + C_G^{D,0 \leftrightarrow i} P_G^{\bar{i}} - f^{0 \leftrightarrow i} = 0 \\ & f^{0 \leftrightarrow i} = \sigma^i f^i \\ & \underline{f}^i \leq f^i \leq \bar{f}^i \end{array} \right.$$

The previous formulation is thus very close to the RBA formulation of prokaryotic cells of Section 2.1. Only constraints  $C_3$  differ by having the normalized volume and surface area of the compartment as decision variables, and this, in order to adjust the volume of the organelles to the need of the cell. Moreover, all constraints are linear with respect to the decision variables.

### 3.6 Formulation of the RBA problem for eukaryotic cells

The consequences of the integration of compartments on RBA have been explored in the previous paragraphs. Clearly, by introducing the normalization of all species by the cell volume, the formulation of the RBA problem for eukaryotic cells is very close to the one for prokaryotic cells. We give in this section the RBA optimization problem,  $P_{rba}^e$ , for the eukaryotic cell.

Let us introduce:

- (i) the set of  $N_{com}$  cellular compartments indexed by  $\mathbb{I}_V = \{0, \dots, N_{com}\}$ , and the set of  $N_{int}$  interfaces indexed by  $\mathbb{I}_A = \{e \leftrightarrow 0, 0 \leftrightarrow i, \dots, k \leftrightarrow j\}$ <sup>5</sup>. The set  $\mathbb{I}_A$  contains the indexes of all possible interfaces between compartments. We note  $N_c = N_{com} + 1 + N_{int}$ .
- (ii)  $Y$  the vector of concentrations of all molecular machines of size  $N_y$  (which includes  $N_m$  enzymes), present in all compartments and interfaces.
- (iii) the flux vector  $\nu$  of size  $N_m$  and associated to the  $N_m$  enzymes;
- (iv)  $P_G$  the vector of concentration of the  $\mathbb{P}_G$  proteins of size  $N_g$  present in all compartments and interfaces.
- (v)  $f_V$  is the vector of the normalized volume of compartments with respect to the total cell volume:  $f_V \triangleq ((f^i)_{i \in \mathbb{I}_V})^T$ . Moreover,  $f_A$  is the vector of normalized surface area of compartments with respect to the total cell volume:  $f_A \triangleq ((f^i)_{i \in \mathbb{I}_A})^T$ . Finally,  $f$  is the concatenation of  $f_V$  and  $f_A$ .

The RBA optimization problem can be formalized as follows.

$P_{rba}^e(\mu)$ : For a fixed vector of concentrations  $P_G \in \mathbb{R}_{>0}^{N_g}$ , and the growth rate  $\mu \geq 0$  of the cell,

$$\begin{aligned}
 \text{find} \quad & Y \in \mathbb{R}_{\geq 0}^{N_y}, \nu \in \mathbb{R}^{N_m}, f \in \mathbb{R}_{\geq 0}^{N_c}, \\
 (C_1) \quad & -\Omega \nu + \mu(C_Y^S Y + C_G^S P_G + C_B^S \bar{B} + C_F^S f \hat{B}) = 0 \\
 (C_2) \quad & \mu(C_Y^M Y + C_G^M P_G) - K_T Y \leq 0 \\
 & -K'_E Y \leq \nu \leq K_E Y \\
 (C_{3a}) \quad & C_Y^{D,iq} Y + C_G^{D,iq} P_G - C_F^{D,iq} f \leq 0 \\
 (C_{3b}) \quad & C_Y^{D,eq} Y + C_G^{D,eq} P_G - C_F^{D,eq} f = 0 \\
 (C_{3c}) \quad & C_F^F f - \bar{C} = 0 \\
 (C_{3d}) \quad & \underline{f}_V \leq I_V f \leq \bar{f}_V
 \end{aligned}$$

where:

- $(C_1)$  and  $(C_2)$  have the same meaning than the RBA problem for prokaryotic cells
- $(C_{3a})$  include the volume or surface constraints that may not be saturated. The constraints on the maximal density of the cytosol is typically included here.
- $(C_{3b})$  contain the volume or surface constraints that are assumed to be always saturated
- $(C_{3c})$  contain all the constraints between the surface and the volume of one compartment (e.g.  $f^{0 \leftrightarrow i} = \sigma^i f^i$ ), additional constraints between volumes or surfaces for complex structures of compartments, and the equality constraint determining  $f^0$ :  $f^0 = 1 - \sum_{i=1}^{N_{com}} f^i$ .  $\bar{C}$  is a vector having 1 only for the constraint on  $f^0$  and zero otherwise.

<sup>5</sup>here  $e$  denotes the index of the compartment associated to the extracellular medium

- ( $C_{3d}$ ) regroup the constraints on the minimal and maximal volume of compartments with respect to the total cell volume. The vector  $\underline{f}_V$  and  $\bar{f}_V$  contain the minimal and maximal normalized volumes of compartments respectively, and the matrix  $I_V$  is such that  $f_V = I_V f$ .

The RBA problem for eukaryotic cells is a Linear programming optimization problem, and can thus be solved efficiently at the cell scale. Compared to the RBA problem for prokaryotic cells, the set of  $C_1$  and  $C_2$  constraints are similar, with the same matrices of coefficients and parameters  $C_*$ ,  $K_T$  and  $K'_E, K_E$ . It means that this part of the optimization problem can be built exactly as in the RBA problem for prokaryotes. Only the values of efficiency coefficients ( $K'_E, K_E$ ) of processes could be dependent on the compartment, since they may be modulated by the internal metabolite concentrations of compartments (see Section 3.3 and 3.4).

## 4 Conclusion

In this paper, we investigated how the RBA framework can be extended to eukaryotic cells. Interestingly, it turns out that the RBA framework for eukaryotic cells remains very close to the one for prokaryotic cells. More importantly, the RBA problem for eukaryotic cells for a fixed growth rate is a Linear Programming (LP) problem, and can thus be solved efficiently at the cell scale [1, 10].

Recently, we developed RBApy, a python package for the generation and simulation of RBA models for prokaryotic cells [2]. Because of the similarity between optimization problems, the extension of RBApy to eukaryotic cells should require only a few changes, mainly at the level of density constraints. Only a few additional parameters are required, and they are mainly related to the structure of organelles (such as the surface/volume ratios). Therefore, we expect no major theoretical or operational difficulty at developing RBA models for eukaryotic cells. Running RBA for an eukaryotic cell depends on the number of decision variables and of constraints. In steady-state, the associated optimization problem should be of reasonable size, enabling its efficient resolution.

In practice, developing RBA models for eukaryotic cells strongly relies on the level of knowledge on the functioning or biological processes. The framework is flexible enough to integrate a high level of details, such as all the different chaperoning systems and their targets in mitochondria, if the knowledge is available and if the modeler want to include a high level of details. Or, in contrast, the RBA model can include only an aggregated description of a biological process. Therefore, the final RBA model will result from a compromise between the level of knowledge on the functioning of the organism, the resulting size of the optimization problem, and the choice of the modeler.

RBA model parameters can be estimated from data, and even from existing published data. When we developed the RBA model of *Escherichia coli* in [2], we used existing fluxomics and quantitative proteomic datasets [15, 13] to estimate parameters, and obtained good quantitative predictions. Currently, the real challenge is the biological validation of RBA models of eukaryotic cells as we did for bacteria a few years ago in [7]: to have (or to generate) suitable datasets to learn the parameters in different growth conditions and how they evolve with respect to growth rate, in order to finally perform predictions using varying growth-rate dependent parameters. This will be achieved in a near future, as well as the update of the RBApy package to automatically generates RBA models for eukaryotic cells.

## Acknowledgments

We thank O. Bodeit and A. Bulović for their critical comments on this manuscript.

## References

- [1] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. MPS/SIAM Series on Optimization, 2001.
- [2] A. Bulović, S. Fischer, M. Dinh, F. Golib, W. Liebermeister, C. Poirier, L. Tournier, E. Klipp, V. Fromion, and A. Goelzer. Automated generation of bacterial resource allocation models. *Metabolic engineering*, 2019.
- [3] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochemical Biophysical Acta*, 1810(10):978–988, 2011.
- [4] A. Goelzer and V. Fromion. Resource allocation in living organisms. *Biochemical Society Transactions*, 45(4):945–952, 2017.
- [5] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 4517–4522, December 2009.
- [6] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):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. *Metabolic Engineering*, 32:232–243, 2015.
- [8] H. Lu, F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marcišauskas, P. M. Anton, D. Lappa, C. Lieven, et al. A consensus *S. cerevisiae* metabolic model yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nature Communications*, 10(1):1–13, 2019.
- [9] M. Mori, T. Hwa, O.C. Martin, A. De Martino, and E. Marinari. Constrained allocation flux balance analysis. *PLoS computational biology*, 12(6):e1004913, 2016.
- [10] Y. Nesterov. *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, 2004.
- [11] 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. *Molecular systems biology*, 9(1), 2013.
- [12] A.-M. Reimers, H. Knoop, A. Bockmayr, and R. Steuer. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences*, 114(31):E6457–E6465, 2017.
- [13] A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent *Escherichia coli* proteome. *Nature biotechnology*, 34(1):104, 2016.
- [14] L. Tournier, A. Goelzer, and V. Fromion. Optimal resource allocation enables mathematical exploration of microbial metabolic configurations. *Journal of mathematical biology*, 2017.
- [15] B.R.B.H. Van Rijsewijk, A. Nanchen, S. Nallet, R.J. Kleijn, and U. Sauer. Large-scale <sup>13</sup>C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Molecular systems biology*, 7(1), 2011.