Evaluating proteome allocation of *Saccharomyces cerevisiae* phenotypes with resource balance analysis – Supplementary Text 2

Hoang V. Dinh1,2, Costas D. Maranas1,2,\*

1Department of Chemical Engineering, Pennsylvania State University, University Park, Pennsylvania, USA

2Center for Advanced Bioenergy and Bioproducts Innovation, The Pennsylvania State University, University Park, PA 16802, USA

\*Corresponding author: [costas@psu.edu](mailto:costas@psu.edu) (CDM)

# A. Mathematical formulation

In this section, a formulation overview is first provided. Then, we provide all necessary details for one to reconstruct the yeast RBA model and associated constraints. We will start with “Section A.1. Sets, indexes, and variables” which establish the starting mathematical structure necessary for equations and constraints in Section A.2 – A.4. The equations and constraints in the model are boxed whereas the reaction and mathematical equations used in derivation and explanation are not. Model stoichiometric coefficients and parameters are mentioned alongside with the equations and constraints that they are in (throughout Section A.2 – A.4).

The model *sc*RBA consists of (macro)molecules and reactions for metabolism and cellular machinery production linked through steady-state mass balance constraints as in FBA (1) (see Figure S1.1 for a schematic representation).

A screenshot of a video game

Description automatically generated

**Figure S1.1**. Overview of *sc*RBA mass balance.

An overview of the RBA optimization model that identifies the maximal growth rate is provided in Figure S1.2. By fixing the growth rate the *sc*RBA model is converted into a linear programming LP formulation (i.e., RBA-LP) which can be efficiently solved. A bisection method is employed to iteratively collapse the estimated (infeasible) upper and (feasible) lower growth rates until they are within the tolerance.

Graphical user interface

Description automatically generated

**Figure S1.2**. Overview of mathematical formulation in *sc*RBA.

# A.1. Sets, indexes, and variables

Let be an element in the set of all molecular species, which comprises of following subsets: (i) metabolites , (ii) proteins , (iii) enzymes , and (iv) biomass precursors .

Let be an element in the set of all reactions, which comprises of following subsets: (i) reactions in metabolic network (including metabolic, transport, and exchange reactions), (ii) protein translation reactions , and (iii) ribosome synthesis reactions , (iv) enzyme synthesis reactions , (v) enzyme load reactions , and (vi) biomass synthesis reactions . Variables are create on the index . It is to be noticed that all fluxes have the unit of mmol gDW-1 h-1, whereas the biomass dilution fluxes have the unit of g gDW-1 h‑1. Molar flux is converted to mass flux by applying the molecular weights in g mmol-1 as stoichiometric coefficients for the biomass precursors in their synthesis reactions.

In Section A.3 and A.4, we denote the specific growth rate constraint variable as (in h-1).

# A.2. Mass balance equations formulation

The mass balance constraints can be written as in the matrix-and-vector form, where is the stoichiometric matrix, and is the column vector containing all flux variables. The concepts of reaction stoichiometry, stoichiometric matrix, and mass balance equations have been explained in the a primer on flux balance analysis (FBA) by Orth et al., 2010 (1). In this document, we apply the fundamentals of FBA and reconstruct the components of the matrix in RBA. The rows of are indexed by the set of all molecular species and the columns of or rows of the column vector are indexed by the set of all reactions.

Throughout the text, stoichiometric coefficients (at a particular (row, column) position in the matrix ) can be positive or negative, indicating that a particular metabolite (indexed by a row of ) is being produced or consumed by a particular reaction (indexed by a column of ), respectively. For referencing purposes, we define sub-matrices of and containing coefficients of metabolites in metabolic network and protein/RNA synthesis reactions, respectively. We define sub-matrix containing coefficients of protein subunits in enzyme synthesis reaction. We define sub-matrix containing coefficients of biomass precursors sinking into macromolecule group and subsequently biomass. We also define sub-matrices and mapping proteins and enzymes to their respective synthesis (respectively), mapping enzymes to enzyme load distribution reactions, and mapping metabolites to dilution reactions for biomass precursors, respectively. Finally, we define sub-matrices and containing the molecular weights (in g mmol-1) of metabolites and proteins/RNAs, as biomass precursors. Mentioned terms will be defined and explained as we go along. We provide the Table 1 below to show the locations of the sub-matrices in .

**Table 1**. Location of sub-matrices in the stoichiometric matrix

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Reactions () | | | | | |
|  |  | Metabolic network  () | Biomass synthesis  () | Protein & RNA synthesis () | Ribosome synthesis () | Enzyme synthesis () | Enzyme load distribution ) |
| Molecular species () | Metabolites () |  |  |  |  |  |  |
| Proteins & RNAs () |  |  |  |  |  |  |
| Enzymes () |  |  |  |  |  |  |
| Biomass precursors () |  |  |  |  |  |  |

## A.2.1. Mass balance equations for metabolites

The general form of mass balance equations for metabolites is:

where the four terms, from left to right, correspond to net production or consumption flux of molecular species across the metabolic network, protein/RNA synthesis reactions, and biomass synthesis reactions.

Most metabolites are produced and consumed only by reactions in metabolic network (), thus their mass balance equations are simplified to as in FBA:

Extending from FBA to RBA, some metabolites that are precursors to proteins and RNAs have their stoichiometric coefficients non-zero for their synthesis reactions (i.e., ). The indexes of proteins & RNAs and their synthesis reactions are explained in Section A.2.2. In addition, some metabolites are also consumed for other biomass constituents via biomass synthesis reactions (). Metabolites and corresponding biomass synthesis reactions are mapped using . We will detail how to reconstruct those synthesis reactions sequestering metabolites to establish metabolite mass balance.

### Metabolites to proteins

There are two types of proteins: modeled and dummy. Modeled proteins are subunits of enzymes and ribosomes whereas dummy protein is used to model the requirement of non-enzymatic and non-ribosomal proteins as well as unused enzymatic proteins. We use the term “dummy protein” to be consistent with published literatures (e.g., in ME-model (2) and ETFL-model (3)). The general form of modeled‑protein translation reaction containing only molecular species without stoichiometric coefficients (which are instead shown in Table 2) is (in which denotes amino acid):

**Table 2**. Stoichiometric coefficients in translation reaction of protein of length n.

|  |  |  |
| --- | --- | --- |
| **Species** | **Reaction coefficient** | **Source** |
| Protein | 1 |  |
| Biomass-Protein | Molecular weight of protein (in g mmol-1) | Protein sequence from Saccharomyces Genome Database (SGD) (4) |
| Charged-tRNAs\* | Number of a specific amino acid molecules in the protein sequence |
| Uncharged-tRNAs\* | Negative number of a specific amino acid molecules in the protein sequence |
| Cofactors | Number of cofactor molecules required for the protein | Uniprot (5) |
| GTP | 2n | Translation elongation’s energy requirements (6, 7) |
| GDP | -2n |
| ATP | 1 |
| ADP | -1 |
| H2O | -(2n+1) |
| Pi | 2n+1 |
| H+ | 2n+1 |

\*There are 20 charged-tRNAs and 20 uncharged-tRNAs corresponding to 20 amino acids

Charged-tRNA is the carrier of amino acid for protein synthesis and becomes uncharged-tRNA once used. Per amino acid (out of 20), uncharged-tRNA is converted to charged‑tRNA by the following aminoacyl-tRNA synthetase reaction (that is part of the metabolic network, i.e., ):

Lumped in the protein translation reaction is the energy demand in the formed of GTP and ATP hydrolysis reactions (i.e., ).

Protein translation reaction creates two protein molecular species, () and (), feeding into enzyme and biomass synthesis, respectively. This is because we need to evaluate protein as both enzyme/ribosome precursor (i.e., see Section A.4 for coupling constraints) and capacity load consumer (i.e., see Section A.3 for capacity constraints).

In contrast, dummy protein translation reaction creates only (i.e., different than to track proteome allocation, see Section A.3), because dummy protein does not feed to enzyme/ribosome synthesis. Thus, except for the absence of , the generic reaction form and stoichiometric coefficients of the dummy protein are the same with those of modeled proteins. Dummy protein assumes the median length of all proteins in the model (i.e., 401 amino acids) and has amino acid composition matches experimental measurements (i.e., in (8)).

### Metabolites to RNAs

There are six specific rRNA species (i.e., 18S, 25S, 5.8S, 5S, 15S, and 21S) and three dummy RNA species (i.e., dummy rRNA, tRNA, and mRNA), indexed as . The difference is that the specific rRNAs feed into ribosome synthesis reaction whereas dummy RNAs are created to model mRNA and tRNA requirements and unused rRNA. Their synthesis reactions are indexed as . The general form of their synthesis reactions containing only molecular species without stoichiometric coefficients (which are instead shown in Table 3) is:

**Table 3.** Stoichiometric coefficients in synthesis reaction of an rRNA.

|  |  |  |
| --- | --- | --- |
| **Metabolite** | **Reaction coefficient** | **Source** |
| rRNA | 1 |  |
| Biomass-RNA | Molecular weight of rRNA (in g mmol-1) | RNA sequence from RNAcentral database (9) |
| ATP | Negative number of a specific nucleotide molecules in the rRNA sequence |
| CTP |
| GTP |
| UTP |
| PPi | Length of rRNA sequence |

rRNA synthesis reaction produces two copies of rRNA: () and () which participates in ribosome synthesis (and subsequently protein-ribosome coupling constraint) (see Section A.4) and capacity constraint (see Section A.3), respectively. This is because we need to evaluate rRNA as both ribosome precursor (i.e., see Section A.4 for capacity constraints) and capacity load consumer (i.e., see Section A.3 for capacity constraints).

In contrast, because dummy RNA does not feed to ribosome synthesis, dummy RNA synthesis reaction creates only biomass precursor species , following these reactions (without stoichiometric coefficients):

Since we only need to capture space occupancy in gram per gram dried weight, dummy RNA assumes the length of one and has A/C/G/U composition matches experimental measurements (i.e., as recorded in the GSM *iSace*1144). (Since transcription is not modeled, RNA length is not relevant, unlike translation and protein length)

### Metabolites to biomass precursors

In our RBA model, instead of modeling metabolic growth requirements using a single biomass reaction (see FBA and biomass reaction papers for more details (1, 10)) whose coefficients indicate the ratio of precursor sequestration, we create individual sink reactions and impose on them flux values derived from biomass composition. Our configuration to model biomass precursor requirements is detailed in Section A.3. Here, we mentioned only the necessary information to establish the mass balance equation. For a metabolite that is considered a biomass precursor (see Section A.3), there is a corresponding biomass precursor synthesis reaction . Let us define the sub‑matrix encoding the one-to-one mapping relationship of the pair. if the pair matches the one-to-one mapping and otherwise. The general form of biomass precursor sink reactions for a metabolite with the molecular weight (in g mmol-1) is:

## A.1.2. Mass balance equations for proteins and RNAs

The general form of mass balance equations for protein/RNA is:

where the three terms, from left to right, correspond to protein/RNA synthesis flux (i.e., protein synthesis reaction), consumption flux to enzyme synthesis, and consumption flux to ribosome synthesis.

First, indexes of proteins/RNAs and their synthesis reactions in the model needs to be explained. A protein/RNA is produced by a corresponding protein/RNA synthesis reaction (see Section A.2.1 for reaction reconstruction). Let us define the sub‑matrix encoding the one-to-one mapping relationship of the pair. if the pair matches the one-to-one mapping and otherwise.

On enzyme synthesis, the general form of an enzyme synthesis reaction () containing only proteins and enzymes without stoichiometric coefficients is:

In an enzyme synthesis reaction, stoichiometric coefficients of proteins equal to negative values of numbers of protein subunits in enzyme, sourced from the Uniprot database (5). Due to an incomplete coverage of all enzymes in the model, missing enzyme stoichiometry is established using Oftadeh et al., 2021’s workflow (3). Briefly, enzyme’s subunit stoichiometry is assumed to be the same with its isozyme with subunit information (i.e., obtained from genome-scale metabolic model’s gene-protein-reaction rule), if that applies. Otherwise, enzyme is assumed to be monomeric.

On ribosome synthesis, there are two ribosome synthesis reactions for nucleus (i.e., indexed as ) and mitochondrial (i.e., indexed as ) ribosomes, respectively. This means that . The general form of ribosome synthesis reaction containing only molecular species without stoichiometric coefficients is:

Ribosome is not explicitly modeled as a molecular species as shown in the ribosome synthesis reaction. Under the steady-state mass balance assumption, ribosome synthesis flux strictly equals to the flux of ribosome diluting to daughter cells. Instead of introducing ribosome (as a molecular species) and ribosome sink reaction, mass balance of ribosome is implicitly modeled as shown in the synthesis reaction.

A nucleus ribosome molecule is synthesized from 80 protein and 4 rRNA subunits (rRNAs 18S, 25S, 5.8S, and 5S) (4). Their stoichiometric coefficients are all -1 for the reaction . Two gene paralogs are present for each of 55 protein subunits (denoted using the suffices “A” and “B” in the standard name on SGD) due to the whole genome duplication event. Since the number of combinations of all possible paralog-specific ribosome is too large, only the version “A” presents in the model to calculate resource allocation. A mitochondrial ribosome molecule is synthesized from 74 protein and 2 rRNA subunits (rRNAs 15S and 21S) (4, 11). Their stoichiometric coefficients are all -1 for the reaction .

## A.1.3. Enzyme load distribution reactions of total enzyme pool

The general form of mass balance equations for enzymes () is:

where the two terms, from left to right, correspond to enzyme production flux from protein precursors and enzyme consumption to enzyme load (and subsequently to biomass production).

An enzyme is produced by a corresponding enzyme synthesis reaction (see Section A.2.2 for reaction reconstruction). Let us define the sub-matrix encoding the one-to-one mapping relationship of the pair. if the pair matches the one-to-one mapping and otherwise.

An enzymatic reaction in the model is catalyzed by an enzyme load, which is a partition of whole enzyme pool. This is because an enzyme can catalyze multiple reactions or the forward and reverse direction of a reaction. Enzyme load distribution reactions are used in RBA model to map the oftentimes enzyme-reaction mappings that are not one-to-one. In RBA model, reversible reactions are also split to irreversible forward and reverse versions. In addition, the same enzymatic reactions catalyzed by isozymes are also split into components catalyzed by respective isozymes. To illustrate, we provide examples on reconstruction of enzymatic reactions and enzyme load distribution reactions in RBA model. In those examples, an irreversible reaction is equivalent to and a reversible reaction is split into and , where suffices and indicate forward and reverse reactions, respectively.

**Example 1**: A single enzyme () catalyzing a single irreversible reaction ().

|  |  |  |
| --- | --- | --- |
| **Enzyme** | **Enzyme load** | **Reaction** |
|  |  |  |

**Example 2**: A single enzyme () catalyzing a single reversible reaction ().

|  |  |  |
| --- | --- | --- |
| **Enzyme** | **Enzyme load** | **Reaction** |
|  |  |  |
|  |  |

**Example 3**: Two isozymes ( and ) catalyzing a single reversible reaction (). Here, the two versions of are considered two separate reactions.

|  |  |  |
| --- | --- | --- |
| **Enzyme** | **Enzyme load** | **Reaction** |
|  |  |  |
|  |  |  |

**Example 4**: A single enzyme () catalyzing two irreversible reactions ( and ).

|  |  |  |
| --- | --- | --- |
| **Enzyme** | **Enzyme load** | **Reaction** |
|  |  |  |
|  |  |

**Example 5**: Two isoenzymes ( and ) catalyzing two irreversible reactions ( and ).

|  |  |  |
| --- | --- | --- |
| **Enzyme** | **Enzyme load** | **Reaction** |
|  |  |  |
|  |  |
|  |  |  |
|  |  |

Overall, enzyme-to-load mapping can be one-to-one or one-to-many and load-to-reaction mapping is strictly one-to-one. Now, back to mass balance equations involving total enzyme pools and enzyme loads, let us define enzyme load distribution reaction . The general form of an enzyme load distribution reaction is:

Enzyme load, which is essentially enzyme, is not explicitly modeled as a molecular species. Under the steady-state mass balance assumption, enzyme production strictly equals to the flux of enzyme diluting to daughter cells. Instead of introducing enzyme load (as a molecular species) and enzyme sink reaction, mass balance of enzyme load is implicitly modeled as shown in the synthesis reaction.

We define the sub-matrix encoding the mapping relationship between an enzyme and enzyme load distribution reactions . if the pair matches a recorded enzyme-to-load relationship and otherwise.

## A.1.4. Mass balance of biomass precursors

Through Section A.2.1 and A.2.2, we introduce biomass precursors () coming from metabolic network and protein/RNA synthesis reactions. Those initial precursors are further spooled into their respective macromolecule group and then sink reactions are create for the group. The general form of mass balance equation for biomass precursor is:

where and are the molecular weight matrices. () and () if the index of biomass precursor matches its synthesis reaction . Otherwise, and . Here, is the molecular weight of metabolite for and of protein/RNA for . Stoichiometric coefficients represent the biomass precursor synthesis network shown in Figure S.1.3. We set up biomass synthesis network to incorporate the *S. cerevisiae*’s measurements of biomass composition that changes with growth rate (8). In addition, we also account for proteome pool allocation toward non-enzymatic and non-ribosomal proteins, which is 45% (3), and account for RNA pool allocation toward tRNA and mRNA, which is 15% and 5% (12), respectively.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure S1.3**. Biomass precursor synthesis network. The goal of the network is to facilitate the implementation of capacity constraints and to incorporate biomass composition measurements.

# A.3. Biomass production and capacity constraints formulation

There are six constraints that impose flux value. Their general form is as follows:

The flux imposition indicates that the amount of biomass precursor production (, in g gDW-1 h‑1) equals to the amount diluting to the daughter cells, which is the specific growth rate (, in h-1) times the mass fraction in biomass (g gDW-1). Based on biomass composition measurements (8, 13), protein, RNA, and carbohydrate fractions in biomass varies with growth rate whereas the other fractions remain relatively constant. We thus approximate those fractions using a linear fit to the available experimental composition data points for protein fraction (from , R2 = 0.67) and RNA fraction (from , R2 = 0.72). Carbohydrate fraction is determined by subtracting protein, RNA, and other fractions. The parameters are provided in the following Table 4:

**Table 4**. Mass fraction parameter value (in g gDW-1) or equation

|  |  |
| --- | --- |
| **Mass fraction parameter** | **Value or equation** |
| Protein |  |
| RNA |  |
| Carb |  |
| Lipid |  |
| DNA |  |
| Metal (part of other) |  |
| Cofactor (part of other) |  |
| Sulphate (part of other) |  |
| Phosphate (part of other) |  |

To account for the proteome allocation towards non-enzymatic and non-ribosomal proteins (i.e., modeled as dummy protein), we added the following flux impositions:

where and are the experimentally observed mass fraction of protein allocated to enzymatic & ribosomal protein and other proteins, respectively (3).

To account for the RNA allocation towards mRNA and tRNA, we added the following flux impositions:

where , , are the experimentally observed mass fractions of rRNA, tRNA, and mRNA in total RNA pool, respectively.

Because mass balance constraint enforces the protein & RNA production fluxes to be equal to the protein & RNA sink (see Section A.2.2), and flux constraints impose flux values to the protein & RNA sink fluxes (see Section A.2.4 and this section), the protein and rRNA capacity constraints are implicitly modeled. The equivalent explicit protein and rRNA capacity constraints, respectively, are:

where ) is the set of protein translation flux (which excludes r/t/mRNA synthesis fluxes), and is the set of rRNA synthesis flux (which excludes protein translation fluxes and t/mRNA synthesis fluxes). A protein translation reaction creates two copies: and , thus the flux imposition on sink restrict the protein translation flux and subsequently how much enzyme and ribosome can be made from . The same applies for rRNA synthesis flux creating and .

# A.4. Coupling constraints formulation

## A.4.1. Reaction-Enzyme coupling constraints

The general form of reaction-enzyme coupling constraints is:

where is the flux of enzyme catalyzed reaction , is the apparent kinetic turnover number (unit of h-1) of enzyme on reaction , is the enzyme load synthesis flux , and is the matrix mapping the one-to-one relationship of reaction and enzyme load (see Section A.2.3). if () is the matching pair and otherwise.

To derive the reaction-enzyme coupling constraint, we use the non-indexed variables and parameters for simplicity. Reaction-Enzyme coupling constraints are modeled with the first-order kinetic rate law with respect to enzyme concentration (i.e., ), , where is the apparent kinetic turnover number parameter. Multiplying that inequality by yields , where the enzyme production rate (i.e., ) equals to the rate of enzyme diluting to daughter cells (i.e., ) in steady-state mass balance of enzymes. This is because as the cell maintains the concentration and grows at the rate of , the enzyme dilution rate is . To model enzyme catalyzing multiple reactions, enzyme loads partitioned from total enzyme pool are used instead (by replacing with ), yielding .

## A.4.2. Protein-Ribosome coupling constraints

The protein-ribosome coupling constraints for ribosomal () and mitochondrial () ribosomes are:

where is the number of amino acids in the sequence of the protein being produced by the protein translation reaction (see Section A.1.2 for protein indexing), is the ribosome efficiency (unit of amino acids per ribosome per second for referrals in text, amino acids per ribosome per hour in implementation for unit consistency within the model), is the set of proteins translated by nucleus ribosome, and is the set of proteins translated by mitochondrial ribosome. Ribosome efficiency is 13.2 amino acids per ribosome per second (x3600 for per hour). The number is derived from the literature-reported value 10.5 (14). We found the updated value by increase from 10.5 incrementally by 0.1 until the simulation where growth rate is fixed to the highest experimental value of 0.49 h-1 (15) is feasible. Based on mitochondrial genome annotation (16), the eight proteins in the model being translated by the mitochondrial ribosome are Q0045, Q0080, Q0085, Q0105, Q0130, Q0140, Q0250, and Q0275. The rest of the proteins are translated by the nucleus ribosome.

To derive the protein-ribosome coupling constraint, we use the non-indexed variables and parameters for simplicity. The total amount of proteins is limited by the following protein-ribosome coupling constraint of , where is the length of protein sequence, is the protein translation flux, is the ribosome efficiency, and is the ribosome concentration. Protein-Ribosome coupling constraint indicates that to perform protein translation at a certain rate the cell needs to maintain a corresponding ribosome concentration. Since ribosome dilution rate (i.e., ) equals to production rate (i.e., ), multiplying that equation by yields .

# B. Software implementation

Software is available at <https://github.com/maranasgroup/scRBA>. The scripts are currently set up for reconstruction and simulation of *S. cerevisiae* metabolism, but they can be adapted for other organisms with sufficient molecular biology knowledge and mathematical modeling and programming skill. General instructions for reconstruction and simulation will be described below.

## B.1. Input files requirements

The following data are collected and stored in excel spreadsheets format (.xlsx), except for the genome-scale metabolic model that is stored in COBRApy JSON format (.json) (17).

|  |  |
| --- | --- |
| **Input file** | **Description** |
| **Common path**: ./scRBA/build\_model/input/  (the actual file names are different on the github, but the same (capitalized) prefix tags are used) | |
| GSM.json | Genome-scale model file in COBRApy format |
| BIOMASS.xlsx | Biomass composition |
| PROTEIN.xlsx | Protein precursors stoichiometry and molecular weights |
| PROTEIN\_dummy.xlsx | Parameters on dummy protein |
| ENZYME.xlsx | Enzyme precursors stoichiometry and mappings to reactions |
| RIBOSOME\_nucleus.xlsx | Nucleus ribosome composition |
| RIBOSOME\_mitochondrial.xlsx | Mitochondrial ribosome composition |
| RNA.xlsx | RNA compositions |
| PARAMS.xlsx | Apparent turnover numbers for enzymes |

## B.2. Model reconstruction

Go to the following directory: ./scRBA/build\_model/

**1) Execute the script**: A01\_build\_excel\_stoich\_for\_GAMS.ipynb

This will yield the excel spreadsheet with all reactions compiled:

./model/RBA\_stoichiometry.xlsx

**2) Execute the script**: A02\_build\_GAMS\_Sij\_and\_fluxBounds.ipynb

This will yield the following GAMS model files below. The user needs to manually copy (or write their own scripts to copy) these files to: ./scRBA/GAMS/model/. This was done to prevent overwriting a GAMS file in-working by GAMS file newly created by the python scripts (which could be faulty).

|  |  |
| --- | --- |
| **GAMS file** | **Description** |
| **Common path**:  Files are created in ./scRBA/build\_model/model/  Files are manually copied to and called by GAMS at this path: ./scRBA/GAMS/model/ | |
| RBA\_species.txt | List of all (macro)molecules |
| RBA\_rxns.txt | List of all reactions |
| RBA\_sij.txt | Stoichiometric matrix coefficients |
| RBA\_rxns\_EXREV.txt | A reaction subset. List of nutrient uptake reactions (exchange, reverse). All of them will be turned off by GAMS script to prevent uptake of nutrients that are not in the media. Uptake of nutrients in the media has to be (over)written through “RBA\_rxns\_EXREV\_<media\_id>.txt”. |
| RBA\_rxns\_EXFWD.txt | A reaction subset. List of byproduct secretion reactions (exchange, forward). (Created but not used in the GAMS script. Don’t worry if you cannot figure out the purpose of this file.) |
| RBA\_rxns\_rxnmetabolicnetwork.txt | A reaction subset. List of reactions in the original metabolic network of GSM. (Used in parameterization, to calculate minimal metabolic flux distribution, i.e., minimizing sum of fluxes) |
| RBA\_rxns\_enzsyn.txt | A reaction subset. List of synthesis reaction for total enzyme pool. (Used in parameterization, to calculate enzyme concentration from protein concentration) |
| RBA\_rxns\_enzload.txt | A reaction subset. List of synthesis reaction for individual enzyme loads constituting total enzyme pool. (Used in parameterization, to identify which reactions can be active based on proteomics data). |
| RBA\_rxns\_prowaste.txt | A reaction subset. List of protein waste reactions (which sink the protein to void). These sink reactions are created to debug model protein synthesis and are turned off by default. (Turning these reactions on will create a lot of degrees of freedom to the optimization and significantly inflate the run time) |

The following GAMS model files encoding the nutrient availability in the media has to be written manually by the user:

|  |  |
| --- | --- |
| **GAMS file** | **Description** |
| **Common path**: ./scRBA/GAMS/model/ | |
| RBA\_rxns\_EXREV\_mineralMinimum.txt | All uptake reactions of minerals |
| RBA\_rxns\_EXREV\_YNB.txt | All uptake reactions of minerals and vitamins in YNB media |
| RBA\_rxns\_EXREV\_YP.txt | All uptake reactions of minerals, vitamins in YNB media, and supplemented amino acids |

**3) Execute the script**: A03\_build\_GAMS\_RBA\_constraints.ipynb

Ribosome-Protein coupling constraints are coded directly in runRBA.gms. Thus, this script only writes the following files:

|  |  |
| --- | --- |
| **GAMS file** | **Description** |
| **Common path**:  Files are created in ./scRBA/build\_model/model/  Files are manually copied to and called by GAMS at this path: ./scRBA/GAMS/model/ | |
| RBA\_enzCapacityConstraints\_declares.txt | Declared indexes of reaction-enzyme coupling constraints |
| RBA\_enzCapacityConstraints\_eqns.txt | Inequality reaction-enzyme coupling constraints. This file is created but is not further used because running inequality constraints take significantly longer time that would prevent the model from being practical. |
| RBA\_enzCapacityConstraints\_eqns\_equality\_version.txt | Equality reaction-enzyme coupling constraints (being used). Equality constraint implementation is practical (reduce solving time), valid, and consistent with the RBA-LP objective, which is minimizing protein capacity usage. |
| RBA\_rxns\_prosyn.txt | A reaction subset. List of protein translation reactions. These reactions are called by ribosome-protein coupling constraints. |
| RBA\_proteinLength.txt | Protein sequence lengths. These parameters are called by ribosome-protein coupling constraints. |

**4) kapp parameters**

kapp parameters are stored in ./scRBA/input/ (with the prefix “RBA\_kapp\_”) and are called by reaction-enzyme coupling constraints. The kapp needs to be in per hour for the input files. See parameterization methods section for how they are estimated for *sc*RBA model.

## B.3. Model simulation

Here, only general script structure and guidance are provided since different setup could be involved for different analysis. (Example) Software scripts for *sc*RBA simulation creating the results shown in the manuscript is available at <https://github.com/maranasgroup/scRBA>.

Generic GAMS files are available at ./scRBA/GAMS/. Case studies simulation are recommended to be organized to a different folder. Specific GAMS script needs to be copied to that case study folder and execute within there. Every GAMS script .gms file is accompanied by a GAMS\_settings .txt file that contains a path link to the files called by the .gms script.

Individual RBA-LP optimization is run by runRBA.gms through the command line, with growth rate specified (see below for example run for growth rate of 0.1 h-1).

gams runRBA.gms --mu=0.1 o=/dev/null

Binary search to iterative run RBA-LP is executed through binary\_search.py with run settings specified in binary\_search\_options.py (generic example file is stored in ./scRBA/pycore/binary\_search/).

Raw outputs are processed by python functions (stored in ./scRBA/pycore/) to yield the organized results in JSON format (see examples in <https://github.com/maranasgroup/scRBA>).

The paths to the following GAMS scripts are available for case studies mentioned in the manuscript:

* Flux variability:
  + GAMS core scripts: ./scRBA/GAMS/vary\_flux/
  + Python exec scripts: ./scRBA/vary\_flux/
* Max production analysis:
  + GAMS core scripts: ./scRBA/GAMS/application/
  + Python exec scripts: ./scRBA/application/

## B.4. Parameterization method and software implementation

**Methods**

kapp parameterization workflow is illustrated in Figure S.1.4. To estimate kapp, we first estimated enzyme concentrations (i.e., in mmol gDW-1) from experimental protein concentrations (i.e., in mmol gDW-1). For an enzyme composed of a single protein, its concentration is equal to protein concentration divided by protein subunit stoichiometric coefficient (e.g., 1 for monomer, 2 for homodimer). For an enzyme composed of different proteins, its molar concentration is equal to the least abundant protein subunit. Mathematically, it is the minimum of values, each of which is calculated by dividing protein molar concentration to its enzyme subunit stoichiometric coefficient. In our calculations, a few protein subunits, even for essential enzymes such as ATP synthase, were not measured and thus inference of enzyme concentration relies on the available protein subunit measurements. After that, we calculated intracellular fluxes from experimental extracellular fluxes in a two-step procedure. First, we determine essential (measurement “gap‑filled”) enzyme that was not recorded in experimental data by minimizing the sum of flux of reactions catalyzed by enzymes that are not measured. Then, we determine the flux distribution from an optimization procedure wherein sum of metabolic fluxes catalyzed by measured and “gap-filled” enzymes are minimized. At the end, kapp value is calculated by dividing the metabolic flux by the enzyme concentration. For cases where multiple reactions are catalyzed by the same enzyme or multiple isozymes catalyze the same reaction, since we cannot ascertain individual enzyme-reaction catalysis load, we assume the same kapp value for all reactions and enzymes involved and the sum of metabolic fluxes and the sum of enzyme concentration values are used in kapp calculation.

Graphical user interface, application

Description automatically generated

**Figure S.1.4**. *In vivo* kapp parameterization workflow. Exchange metabolic fluxes and protein concentrations data were used as the inputs to calculate kapp from intracellular metabolic fluxes and enzyme concentrations, accounting for metabolic reaction and protein subunit stoichiometry.

**Software implementation**

Proteomics data needs to be manually processed by users. Scripts are only meant to be used as examples since they are specific to *S. cerevisiae*. After proteomics data processing, then:

Step 1: Estimate enzyme concentrations from protein concentrations, accounting for protein subunit stoichiometry

* GAMS core: ./scRBA/GAMS/parameterization/enz\_from\_proteome/
* Python exec

./scRBA/parameterization/kapp/datasets/<id>/B1\_enz\_from\_proteome.py

Step 2: Rescue reactions from inactive set and put into active set. This is to account for the fact that proteomics data might be incomplete and miss out enzymes performing essential metabolic functions.

* GAMS core: ./scRBA/GAMS/parameterization/min\_flux\_violation/
* Python exec

./scRBA/parameterization/kapp/datasets/<id>/B2\_min\_flux\_violation.py

Step 3: Estimate minimal metabolic flux distributions (only allow reactions in active set to carry fluxes) (analogous to parsimonious FBA (18))

* GAMS core: ./scRBA/GAMS/parameterization/min\_flux\_sum/
* Python exec

./scRBA/parameterization/kapp/datasets/<id>/B3\_min\_flux\_sum.py

Step 4: Estimate kapp from metabolic flux and enzyme concentration.

* Python exec:

./scRBA/parameterization/kapp/datasets/<id>/C1\_calculate\_kapp.ipynb

## B.5. Solver settings

The solver settings need to be recorded in the .opt (text) file and has to be in the same directory as the GAMS script .gms file based on GAMS programming syntax. SoPlex solver was used to solve the RBA-LP linear programming problem, thus the file name is “soplex.opt”. (See examples in <https://github.com/maranasgroup/scRBA>)

# References

1. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).

2. C. J. Lloyd, *et al.*, COBRAme: A computational framework for genome-scale models of metabolism and gene expression. *PLOS Comput. Biol.* **14**, e1006302 (2018).

3. O. Oftadeh, *et al.*, A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics. *Nat. Commun. 2021 121* **12**, 1–10 (2021).

4. J. M. Cherry, *et al.*, Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705 (2012).

5. A. Bateman, *et al.*, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

6. T. E. Dever, T. G. Kinzy, G. D. Pavitt, Mechanism and regulation of protein synthesis in *Saccharomyces cerevisiae*. *Genetics* **203**, 65–107 (2016).

7. R. J. Jackson, C. U. T. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).

8. H. C. Lange, J. J. Heijnen, Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **75**, 334–44 (2001).

9. B. A. Sweeney, *et al.*, RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.* **49**, D212–D220 (2021).

10. J. C. Lachance, *et al.*, BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. *PLoS Comput. Biol.* **15** (2019).

11. N. Desai, A. Brown, A. Amunts, V. Ramakrishnan, The structure of the yeast mitochondrial ribosome. *Science* **355**, 528–531 (2017).

12. J. R. Warner, The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).

13. P. Van Hoek, J. P. Van Dijken, J. T. Pronk, Effect of specific growth rate on fermentative capacity of baker’s yeast. *Appl. Environ. Microbiol.* **64**, 4226–4233 (1998).

14. C. Waldron, R. Jund, F. Lacroute, Evidence for a high proportion of inactive ribosomes in slow-growing yeast cells. *Biochem. J.* **168**, 409–415 (1977).

15. J. Björkeroth, *et al.*, Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21804–21812 (2020).

16. F. Foury, T. Roganti, N. Lecrenier, B. Purnelle, The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* **440**, 325–331 (1998).

17. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).

18. N. E. Lewis, *et al.*, Omic data from evolved *E. coli* are consistent with computed optimal growth from genome‐scale models. *Mol. Syst. Biol.* **6**, 390 (2010).