**Metabolic flux and resource balance in the oleaginous yeast *Rhodotorula toruloides* – Supplementary Text 1**

Eric J. Mooney(1),(2), Patrick F. Suthers(2),(3), Wheaton L. Schroeder(3),(4), Hoang V. Dinh(2),(3), Xi Li(5),(6),(7), Yihui Shen(6),(7),(9), Tianxia Xiao(6),(7),(8), Catherine M. Call(6),(7),(10), Heide Baron(6),(7), Arjuna M. Subramanian(6),(7), Daniel R. Weilandt(5),(6),(7),(12), Felix C. Keber(6),(7), Martin Wühr(7),(11), Joshua D. Rabinowitz(5),(6),(7),(12), Costas D. Maranas(2),(3)\*

1. Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, 16802, USA
2. DOE Center for Advanced Bioenergy and Bioproducts Innovation, The Pennsylvania State University, University Park, PA, 16802, USA
3. Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, 16802, USA
4. Gene and Linda Voiland School of Chemical Engineering and Bioengineering, Washington State University, Pullman, WA, 99164, USA
5. DOE Center for Advanced Bioenergy and Bioproducts Innovation, Princeton University, Princeton, NJ, USA
6. Department of Chemistry, Princeton University, Princeton, NJ, USA
7. Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
8. Current address: State Key Laboratory of Female Fertility Promotion, Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China
9. Current address: Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA
10. Brown University Department of Orthopaedic Surgery, Providence, RI 02912, USA
11. Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
12. Ludwig Institute for Cancer Research, Princeton Branch, Princeton, NJ, USA

\* Correspondence: [costas@psu.edu](mailto:costas@psu.edu)

1. Proteins translated by the mitoribosome in scRBA

|  |  |  |
| --- | --- | --- |
| Protein ID in model | Name | UniProt ID |
| Q0045 | COX1 | P00401 |
| Q0080 | ATP8 | P00856 |
| Q0085 | ATP6 | P00854 |
| Q0105 | COB | P00163 |
| Q0130 | OLI1 | P61829 |
| Q0250 | COX2 | P00410 |
| Q0275 | COX3 | P00420 |
| Q0140 | VAR1 | P02381 |
| Q0065 | Q0065 | Q0065 |
| Q0120 | Q0120 | Q0120 |
| Q0050 | Q0050 | Q0050 |
| Q0055 | Q0055 | Q0055 |
| Q0070 | Q0070 | Q0070 |
| Q0045 | COX1 | P00401 |

1. pFBA formulation

All reactions are split into forward and reverse versions that can only carry nonnegative fluxes.

1. RBA-LP formulation

Identical to scRBA’s original version, aside from the last 2 constraints.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | Mass balance |
|  | Flux bounds |
|  | Non-negative fluxes only |
|  | Enforcing medium composition |
|  | Ribosome-protein coupling |
|  | Protein capacity |
|  | Mitochondrial protein (i.e., set ) capacity |
|  | rRNA capacity |
|  | Reaction-enzyme coupling |
|  | Minimum protein from sink (MPFS) constraints, using the “protein wasting” sink fluxes from the last step of calculations (i.e., ) |
|  | Protein sink fluxes (i.e., ) limited to their values from calculations |

1. Updated GSM model

While preparing the RBA model, updates to the existing RT GSM model were first identified and implemented. The updated base stoichiometric model, *iRhto*1120 contains 2,320 reactions and 1,120 genes, compared to the 2,204 reactions and 1,108 genes in the published model *iRhto*1108 (Dinh et al., 2019). Aside from extensive additions to annotations and other meta-data, key changes to the model are illustrated in Fig. 2 and fall into 3 categories: reaction removals, additions, and modifications. Fatty acid degradation, especially in the peroxisome and mitochondria, was described in greater detail by adding 79 reactions and removing 33 lumped ones, replacing mitochondrial reactions with peroxisomal ones. The model now includes gene *coq4*, which encodes an enzyme that catalyzes two cytoplasmic FMN reductase reactions (one using NADPH, the other using NADH). Both reactions were added to make the cofactor FMNH2. Other modifications include adding spontaneous ADP hydrolysis reactions in the cytosol and mitochondria (to convert ADP to AMP), as well as reactions transporting Cu2+ and Fe3+ via ion channels into the mitochondria for use as cofactors. In total, 150 and 87 reactions were added and removed, respectively, to form *iRhto*1120.

The updated model better accounts for carbon flux apportioning. Five reactions (phosphoketolase, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase, xylulokinase, xylose isomerase, NH3-dependent CTP synthase) mainly involved in xylose metabolism were made inactive under carbon-limited growth, to better reflect *in vivo* observations (Adamczyk et al., 2023; Jagtap et al., 2021; Papini et al., 2012). Two other reactions (l-arabinitol 4-dehydrogenase and NADH-dependent xylulose reductase) were made inactive under carbon-limited conditions since they lacked evidence supporting their activity. Likewise, five acyl-CoA oxidase reactions involved in fatty acid degradation were made inactive to reflect RT’s preference for oxidizing short- and medium-chain fatty acids in the mitochondria.

Furthermore, an updated version of yeastGEM\_hvd was developed, to align it more with the reaction directionalities identified while developing scRBA. The updated GSM model is available at <https://github.com/EarthToMooney/scRBA/blob/main/build_model/input/GSM_iSace1144_rba.json>. The majority of these changes involve changing reaction reversibility.

A diagram of the cycle of a cell

AI-generated content may be incorrect.

Fig. . Visual summary of GSM model updates in an Escher map highlighting the major changes to metabolic pathways between *iRhto*1108 and *iRhto*1120. Dots represent metabolites and lines represent reactions. Reactions added are shown in green and removed in orange. Reactions with GPR, bounds, and/or stoichiometry updates are shown in purple. Specifically, G3PD1r\_c and CITMALta\_m were made reversible, G3PD1i\_m is blocked to prevent a thermodynamically infeasible cycle with G3PD1r\_c, and MTHFD\_m has a new GPR description. ATPS\_m is made irreversible with new protein names in its GPR as well as updated stoichiometry that matches its SC counterpart. Thicker lines denote reactions that were altered between the original *iRhto*1108 and *iRhto*1120 used herein. Pathways include (A) oxidative phosphorylation, (B) FAD and FMNH2 production, (C) 5-formyltetrahydrofolic acid (5FTHF) production, (D) glycerol production, (E) the pentose phosphate pathway, (F) glycerophospholipid metabolism, (G) acetyl-CoA biosynthesis, (H) the citric acid cycle, (I) pyrimidine metabolism.

Because the model content was updated, both GAM and NGAM were recalculated and increased to 187.141 mmol ATP gDW-1 and 1.313 mmol ATP gDW-1h-1, respectively, from their original values (GAM=140.98 mmol ATP gDW-1, NGAM=1.01 mmol ATP gDW-1h-1) (Fig. 2). The predicted maximum growth rate () remains nearly the same as in *iRhto*1108 (i.e., approximately 0.38 h-1). The final GSM model and a detailed network representation of it using the tool Escher (King et al., 2015) are provided in the Supplementary Materials.



Fig. 2. Visual summary of ATP maintenance datasets from [Shen et al. (2013)](#_ENREF_39), with maximum ATP hydrolysis flux calculated via RT’s revised GSM model. Trendline illustrates GAM and NGAM via its slope and intercept, respectively.

1. **Estimation of in vivo kapp (continued)**

When adapting the kapp procedure from scRBA to the proteomics data for RT, growth at the measured rate of 0.38 h-1 was initially infeasible when using the resulting kapp values. One potential reason is that B3 does not force to the minimum value from B2, possibly permitting reactions in to carry more flux than necessary. Therefore, if versions of a reaction from and are both needed to carry sufficient flux through a pathway, the model is not guaranteed to prioritize reducing flux through the set of reactions vs. the ones in set . Thus, because we would prefer to minimize usage of unmeasured reactions in set more so than set , step B2 is instead executed:

|  |  |
| --- | --- |
|  | **(Step B2)** |
|  | 1. Mass balance |
|  | 1. Flux bounds |
|  | 1. Non-negative fluxes only |
|  | 1. Enforcing medium composition |
|  | 1. Enforcing measured growth rate |
|  | 1. Enforcing measured protein translation rates across all cellular compartments (i.e., ) |
|  | 1. Ribosome-protein coupling |
|  | 1. Protein capacity |
|  | 1. Mitochondrial protein (i.e., set ) capacity |
|  | 1. Nonmetabolic protein (i.e., set ) costs |
|  | 1. rRNA capacity |

followed by step B3:

|  |  |
| --- | --- |
|  | **(Step B3)** |
| (1) – (11) |  |
|  | 1. Minimizing use |
|  | 1. Blocking all non-essential reactions outside |

with B3 fixing the value of B2’s objective to its optimal value (i.e., in constraint (12)).

As in scRBA, steady-state mass balance (1) is enforced for all metabolites and macromolecules and reactions . Each reaction has upper and lower flux bounds and , respectively; (2) can only have non-negative fluxes; and (3) reversible reactions are split into forward and reverse versions. Exchange fluxes (, in mmol gDW-1 h-1) and the growth rate (, in h-1) are constrained to their respective experimental levels ( and , in constraints (4) and (5) respectively). In constraints (7) through (10), protein production is limited by ribosome efficiency (, in amino acids/ribosome/s); growth rate (; the measured mass fractions of biomass comprised of proteins (), mitochondrial proteins (), and nonmetabolic proteins (); as well as each protein’s length (, in amino acids) and molecular weight (, in g/mmol). Likewise, (11) uses the product of and the measured mass fraction of biomass comprised of rRNA () to limit the flux of each rRNA type’s production (, in mmol gDW-1h-1) based on its molecular weight (, in g/mmol).

Steps B2 and B3 are identical to their scRBA versions aside from (12) as well as (6), which previously used to only assign proteins to one compartment chosen manually, despite some being able to form enzymes in multiple compartments (i.e., multi-compartment proteins). Also, scRBA did not originally force translation fluxes for ribosome subunits to match their experimental values. Adding this requirement made step B2 infeasible during testing because of insufficient ribosome production, possibly due to other ribosome subunit paralogs being present but not measured and/or not listed in the model as subunits. To address this issue without using unrealistic kribo values while minimizing deviations from measured translation fluxes, the non-negative slack variables and were introduced for each protein in set . These variables allow a protein’s translation flux to go above or below its measured values, respectively. They cannot assume nonzero values by default, and their sum (i.e., is set to the lowest possible value (i.e, ) during step B2 and fixed to this value in future steps, to ensure this feature is utilized only when needed. For example, ribosome subunit proteins were allowed to have nonzero slack values, and only for . Furthermore, the protein abundance data used by scRBA did not include localization information. To better account for multi-compartment proteins, the fluxes producing all compartment-specific versions of each protein (i.e., ) are constrained to , the experimentally determined flux for translation of that protein (in mmol gDW-1h-1).

For each GSM model reaction in set , Steps B2 and B3 can now utilize recommended upper and lower flux bounds ( and , respectively) as inputs, to match experimental data from metabolic flux analysis (MFA) and other methods as closely as possible while allowing minor deviations where needed (e.g., due to using different versions of a GSM model in MFA and RBA). All GSM model reactions with recommended upper or lower bounds are included in sets and , respectively. To enforce these flux bounds over all versions of a reaction in the RBA model, the parameter was added to represent each reaction’s direction:

When recommended flux bounds are provided, step B2 is prefaced by an additional linear programming (LP) problem (i.e., step B1.5) with the same constraints, with Steps B1.5 through 3 including the following constraints:

|  |  |
| --- | --- |
|  |  |
|  | 1. Enforcing recommended flux upper bounds for all reactions that have them (i.e., reactions in set ) |
|  | 1. Enforcing recommended flux lower bounds for all reactions that have them (i.e., reactions in set ) |

Step B1.5 finds the flux distribution that minimizes the sum of thenon-negative slack variables and for all GSM model reactions, while not allowing their sum to exceed these minimal values in steps B2 and B3 respectively.

Another limitation of scRBA’s kapp calculations is that they estimate each enzyme’s synthesis flux (, in mmol gDW-1h-1) without considering that some enzymes catalyze multiple reactions, using equation (16):

|  |  |
| --- | --- |
|  | 1. Estimating enzyme synthesis fluxes |

where is the number of copies of a protein needed per enzyme. All enzymes with positive values are added to set . This strategy allows each protein copy to participate in the simultaneous formation of multiple enzymes, thus overestimating enzyme levels. Here, we address this issue by calculating all enzyme levels simultaneously through (**Steps B4–B5**).

To determine how proteins are apportioned amongst enzymes, we assume that cells aim to minimize production of unused proteins, since doing so would waste less resources crucial to the organism’s fitness. A similar assumption is also invoked when determining flux distributions in scRBA, which minimizes the total protein capacity usage at each growth rate tested (Dinh and Maranas, 2023). rtRBA and scRBA only account for a protein’s usage in creating enzymes, ribosomes, and as a biomass component, though this would likely underestimate protein needed. For instance, some proteins may have other functions, such as how enolase acts as both a metabolic enzyme and a heat-shock protein in SC (Iida and Yahara, 1985). When the quantities of protein needed to support such alternative functions are known, they can be added via constraints forcing flux through a protein’s “protein wasting” sink reaction.

We calculate initialized values (i.e., ) consistent with the “enzyme load” reactions (i.e., enzyme production rate required for meeting the corresponding metabolic flux) determined in Step B1 (i.e., . These values are used in **Steps B4–B5** (see constraint (17) below) to ensure that enzyme loads are apportioned based on the total metabolic flux they support. This ensures that if a reaction has a non-zero flux, its associated enzyme is produced. Not employing constraint (17) results in mismatches between enzyme loads identified in prior steps and the reaction fluxes associated with subsequent steps. As in scRBA, we use the median to approximate the values for reactions in set (Dinh and Maranas, 2023) (i.e., ) lacking proteomics measurements of the associated enzymes. However, this can overestimate values for set . This issue was observed during testing, with some enzyme load values set to the lowest positive value that the solver could distinguish from zero. To address this, we allow departures in the kapp values from via the non-negative slack variables and .

We first minimize production of unused enzyme subunits, represented by the total protein mass consumed by protein wasting reactions (, in g gDW-1h-1), via **Step B4**:

|  |  |
| --- | --- |
|  | **(Step B4)** |
|  | 1. Slacks allowing deviation from the prototype value for each used enzyme. |
|  | 1. Blocking production of enzymes unused by the model. |
|  | 1. Requiring production of all enzymes used by the model. |
| (1) – (11), (14) – (15) |  |

Then, **Step B5** minimizes deviations from the values, forcing the production of all necessary enzymes needed by the model while ensuring each protein copy is allotted to only one enzyme:

|  |  |
| --- | --- |
|  | **(Step B5)** |
| (1) – (11), (14) – (15), (17) – (19) |  |
|  | 1. Enforcing Step B4’s solution. |

Since this step now requires making proteins for non-spontaneous reactions, and thus the metabolites needed to produce them, the model may need to use additional reactions and/or increase its fluxes through ones it already uses. Thus, **Steps B1.5**–**B3** are rerun while enforcing the optimal objective values from B4 and B5. The resulting metabolic and enzyme synthesis fluxes were used to calculate kapp values. The method for doing so is largely identical to the one used in scRBA’s original release (Dinh and Maranas, 2023). However, in that release, the median kapp value was assigned to any reactions with zero flux through either themselves or synthesis of their respective enzymes. This resulted in enzymes with measured subunits (i.e., in set ) having lower kapp values than isozymes without them when the former had kapp values below the median, encouraging the model to rely less on enzymes using measured proteins. To reduce this risk, if an enzyme has isozymes in set, its kapp is assigned to the lowest kapp among them or the median kapp, whichever is lower.

1. References

Adamczyk, P. A., Coradetti, S. T., Gladden, J. M., 2023. Non-canonical d-xylose and l-arabinose metabolism via d-arabitol in the oleaginous yeast Rhodosporidium toruloides. Microb Cell Fact. 22**,** 145.

Dinh, H. V., Maranas, C. D., 2023. Evaluating proteome allocation of Saccharomyces cerevisiae phenotypes with resource balance analysis. Metab Eng. 77**,** 242-255.

Dinh, H. V., Suthers, P. F., Chan, S. H. J., Shen, Y., Xiao, T., Deewan, A., Jagtap, S. S., Zhao, H., Rao, C. V., Rabinowitz, J. D., Maranas, C. D., 2019. A comprehensive genome-scale model for Rhodosporidium toruloides IFO0880 accounting for functional genomics and phenotypic data. Metab Eng Commun. 9**,** e00101.

Iida, H., Yahara, I., 1985. Yeast heat-shock protein of Mr 48,000 is an isoprotein of enolase. Nature. 315**,** 688-690.

Jagtap, S. S., Deewan, A., Liu, J.-J., Walukiewicz, H. E., Yun, E. J., Jin, Y.-S., Rao, C. V., 2021. Integrating transcriptomic and metabolomic analysis of the oleaginous yeast Rhodosporidium toruloides IFO0880 during growth under different carbon sources. Appl Microbiol Biot. 105**,** 7411-7425.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., Palsson, B. O., 2015. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLOS Computational Biology. 11**,** e1004321.

Papini, M., Nookaew, I., Siewers, V., Nielsen, J., 2012. Physiological characterization of recombinant Saccharomyces cerevisiae expressing the Aspergillus nidulans phosphoketolase pathway: validation of activity through 13C-based metabolic flux analysis. Appl Microbiol Biot. 95**,** 1001-1010.