PNAS
www.pnas.org

Supplementary Information for

Yeast has evolved to minimize protein resource cost for synthesizing amino acids

Yu Chen, Jens Nielsen

Jens Nielsen
Email: jni@bii.dk

This PDF file includes:

Supplementary text
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S9

Supplementary Information Text

Construction of yeast metabolic model integrated with enzyme kinetic data. The consensus genome-scale metabolic model (GEM) of *Saccharomyces cerevisiae* Yeast8 (version 8.4.2) (1) was used in this study. To integrate enzyme kinetic data, the GECKO (2) and sMOMENT (3) frameworks with modifications were adopted. All reversible enzymatic reactions in the model were split into forward and reverse reactions. For each enzymatic reaction, a pseudo-metabolite named "protein_cost[c]" was added as a substrate with its coefficient being $-MW/k_{cat}$, in which MW is the molecular weight of the enzyme (or minimum MW if there are multiple isozymes) and k_{cat} is the *in vitro* turnover number of the enzyme (maximum k_{cat} for isozymes) collected from the BRENDA database (4) based on the previous criteria (5). The protein cost of an enzymatic reaction (calculated by MW/k_{cat}) represents protein mass required per unit of flux through the reaction (6). By doing so, the *in vitro* protein cost information can be integrated into the corresponding reaction. Additionally, an exchange reaction of the pseudo-metabolite "protein_cost[c]" was added to estimate total protein mass for a given metabolic state. As a result, the yeast GEM integrated with *in vitro* enzyme kinetic data was constructed.

To improve model performance, the yeast *in vivo* enzyme catalytic rates (k_{app} values) (7) can be integrated into the model by replacing the corresponding *in vitro* k_{cat} values. Note that two types of the *in vivo* data were used, i.e., the maximum k_{app} values (k_{max} values), and condition-specific k_{app} values, which resulted in a k_{max} model and various condition-specific k_{app} models.

Model construction was performed in MATLAB using the COBRA toolbox (8).

Estimation of glucose and protein costs of synthesizing amino acids. Normally, to estimate the cost of synthesizing a metabolite, one should direct the flux from the substrate to the metabolite, which could be achieved by maximizing the metabolite synthesis rate (typically the exchange reaction of the metabolite) while fixing the substrate uptake rate or minimizing the substrate uptake rate while fixing the metabolite synthesis rate. This is however not applicable to the amino acid case as the amino acid exchange reactions in the model are responsible for their secretion rather than the incorporation into the biomass. The flux distribution that an amino acid is incorporated into the biomass can be estimated by simulating two states: one is the reference state in which there is a non-zero synthesis rate of the amino acid and the other is the case state in which the synthesis rate of the amino acid is changed. The case state can be achieved by adding a slight increase or decrease in the stoichiometric coefficient of the amino acid in the biomass equation while maintaining the growth rate the same as that in the reference state, in which there is no modification on the model (9).

The simulations were performed on minimal media with glucose as the carbon source (1). For simulating the reference state, the growth rate was fixed at a feasible value μ so that the synthesis rates of all amino acids have non-zero values, and the glucose uptake rate was minimized. As a result, the flux distribution, i.e., rates of all involved reactions, of the reference state can be obtained. For simulating the case state, a slight increase Δs was added in the stoichiometric coefficient of the amino acid of interest in the biomass composition so that an increased synthesis rate of the amino acid was required for achieving the original growth rate. As a result, the flux distribution of the case state can be also obtained. The change in the flux distributions of the two states (the case state minus the reference state) could be therefore seen as the flux distribution (referred to as "the new flux distribution" for clarification, in which reaction rates are marked with v_i) associated with a biosynthetic flux (the value is $\mu \cdot \Delta s$) of the amino acid. The new flux distribution can be used to estimate the cost of synthesizing the amino acid. The glucose cost of synthesizing the amino acid can be calculated by $v_{glucose}/\mu/\Delta s$, in which $v_{glucose}$ is the value of the glucose uptake reaction in the new flux distribution, i.e., the change in the simulated rates of the glucose uptake reaction of the reference and case states. The protein cost of synthesizing the amino acid can be calculated by the sum of the products of rates v_i in the new flux distribution and the corresponding reaction protein cost pc_i over the biosynthetic flux of the amino acid, i.e., $\Sigma v_i \cdot pc_i/\mu/\Delta s$. The reaction protein cost is calculated by MW/k_{cat}, representing protein mass per unit of flux for the reaction.

Note that there is already an exchange reaction of the pseudo-metabolite "protein_cost[c]" in the model, and the simulated value of the reaction represents the total protein mass for a given metabolic state, which is calculated by the sum of the products of rates and reaction protein costs of all reactions. Therefore, the protein cost of synthesizing the amino acid $\Sigma v_i \cdot pc_i/\mu/\Delta s$ can be converted: $\Sigma v_i \cdot pc_i/\mu/\Delta s = \Sigma(v_{case\_state,i} - v_{ref\_state,i}) \cdot pc_i/\mu/\Delta s = \Sigma(v_{case\_state,i} \cdot pc_i - v_{ref\_state,i} \cdot pc_i)/\mu/\Delta s = (pm_{case\_state} - pm_{ref\_state})/\mu/\Delta s$, in which pm represents the total protein mass.

The simulations were performed using the k_{max} model and condition-specific k_{app} models, which resulted in k_{max}-based and condition-specific k_{app}-based protein costs of amino acids. Note that the k_{max} model and condition-specific k_{app} models resulted in the same glucose costs of synthesizing amino acids as the calculation of glucose costs is independent of protein cost information.

In this study, the growth rate μ was fixed at 0.4 /h, which is a feasible value of *S. cerevisiae* at unlimited conditions, and the slight increase Δs was 0.0001, which ensures a slight perturbation on the simulations. Note that the glucose and protein costs were almost unchanged using other μ values (0.1, 0.2 and 0.3 /h) and Δs values (0.00001 and 0.001).

All simulations were performed in MATLAB and run with the COBRA toolbox using the Gurobi solver (https://www.gurobi.com/).

Amino acid substitution analysis. For each amino acid, the stoichiometric coefficient in the biomass equation was increased by Δs, while the stoichiometric coefficient of each of the other 19 amino acids was decreased by Δs at a time. Subsequently, the glucose uptake rate and the total protein mass were re-estimated, which were normalized by the corresponding values of the reference state in which there was no change in any amino acid's stoichiometric coefficient. The simulations were also performed with μ of 0.4 /h and Δs of 0.0001.

Calculation of relative abundances of amino acids of yeast cells under diverse conditions. The relative abundance of an amino acid in a proteome, i.e., the fraction of the amount of the amino acid in the amount of all 20 amino acids in the proteome, was calculated by:

$$\frac{\sum_{i=1}^{n} N_i \times C_i}{\sum_{i=1}^{n} L_i \times C_i}$$

in which n is the number of all detected proteins in a proteomics dataset, i represents an individual protein among them, N_i is the number of the amino acid of interest in the sequence of the protein, L_i is the length (total amino acid number) of the sequence of the protein, and C_i is the copy number of the protein. The relative abundances of all 20 amino acids were calculated for yeast cells under diverse conditions using various absolute proteomics datasets (10–13).

Calculation of contributions of different enzyme kinetic sources to the estimated protein costs of synthesizing amino acids. This was only performed for the k_{max}-based protein costs of synthesizing amino acids, which were estimated using the k_{max} model. Due to the low coverage of *in vivo* data, the k_{max} model still has a lot of enzymatic reactions that were parameterized by *in vitro* or even non-yeast kinetic data. Based on the sources of enzyme kinetic data, the enzymatic reactions in the k_{max} model were marked as reactions with yeast k_{max}, yeast *in vitro* k_{cat}, and non-yeast k_{cat}. For each amino acid, the increase in the total protein mass from the reference state to the case state, in which the synthesis rate of the amino acid is increased, should be contributed by the changes in the protein demands of individual enzymatic reactions Δpd. The contribution of reactions with yeast k_{max} (can also be yeast *in vitro* k_{cat} or non-yeast k_{cat}) to the estimated protein cost of the amino acid can be calculated by:

$$\frac{\sum_{j=1}^{m} |\Delta pd_j|}{\sum_{i=1}^{n} |\Delta pd_i|}$$

in which n is the number of all enzymatic reactions and m is the number of the reactions with yeast k_{max} (can also be yeast *in vitro* k_{cat} or non-yeast k_{cat}). In other words, the denominator is the

sum of the absolute values of changes in the protein demands of all enzymatic reactions in the model, and the numerator only accounts for the sum of the absolute values of changes in the protein demands of reactions with yeast k_{max} (can also be yeast *in vitro* k_{cat} or non-yeast k_{cat}).

Sensitivity analysis. Given that yeast k_{max} data are of high confidence, the uncertain k_{cat}, i.e., yeast *in vitro* k_{cat} and non-yeast k_{cat}, were investigated. For each uncertain k_{cat}, a value was randomly sampled from a uniform distribution between 0.01 and 100 folds of the original value. All the uncertain k_{cat} were randomly sampled simultaneously while the yeast k_{max} were unchanged, and protein costs of synthesizing 20 amino acids were re-estimated. Subsequently the re-estimated protein costs were correlated with the logarithms of the average relative abundances of amino acids, which resulted in a new Pearson's r value. This workflow was run 1000 times and 1000 Pearson's r values were obtained, which can be compared with the Pearson's r value of correlation between glucose costs of synthesizing 20 amino acids and the logarithms of the relative abundances of amino acids. The 1000 Pearson's r values can be obtained here (https://github.com/SysBioChalmers/Amino_acid).

Dataset S1 (separate file). Glucose and protein costs of synthesizing 20 proteinogenic amino acids in *S. cerevisiae*.

Dataset S2 (separate file). Amino acid substitution analysis.

Dataset S3 (separate file). Growth conditions where absolute proteomics data and *in vivo* enzyme catalytic rates were adopted.

Dataset S4 (separate file). Pearson's r values of correlations between glucose and protein costs of synthesizing amino acids.

Dataset S5 (separate file). Relative abundances of amino acids of yeast cells under diverse conditions.

Dataset S6 (separate file). Pearson's r values of correlations between relative abundances of amino acids of yeast cells under diverse conditions.

Dataset S7 (separate file). Pearson's r values and p values of correlations between amino acid frequencies of individual proteins and cellular average relative abundances of amino acids.

Dataset S8 (separate file). Pearson's r values of correlations between costs of synthesizing amino acids and logarithms of relative abundances of amino acids of yeast cells under various conditions.

Dataset S9 (separate file). Contributions of three types of enzyme kinetic data to the estimated protein cost for each amino acid.

SI References

1. Lu H, et al. (2019) A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* 10(1):1–13.
2. Sánchez BJ, et al. (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol* 13(8):935.
3. Bekiaris PS, Klamt S (2020) Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics* 21(1):19.
4. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D (2019) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* 47(D1):D542–D549.
5. Chen Y, Li F, Mao J, Chen Y, Nielsen J (2021) Yeast optimizes metal utilization based on metabolic network and enzyme kinetics. *Proc Natl Acad Sci U S A* 118(12).

doi:10.1073/pnas.2020154118.
6. Chen Y, Nielsen J (2019) Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc Natl Acad Sci U S A* 116(35):17592–17597.
7. Chen Y, Nielsen J (2021) In vitro turnover numbers do not reflect in vivo activities of yeast enzymes. *Proc Natl Acad Sci* 118(32):2021.
8. Heirendt L, et al. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc*:1.
9. Barton MD, Delneri D, Oliver SG, Rattray M, Bergman CM (2010) Evolutionary systems biology of amino acid biosynthetic cost in yeast. *PLoS One* 5(8):11935.
10. Lahtvee P-J, et al. (2017) Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Syst* 4(5):495-504.e5.
11. Di Bartolomeo F, et al. (2020) Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proc Natl Acad Sci U S A* 117(13):7524–7535.
12. Yu R, et al. (2020) Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. *Nat Commun* 11(1):1881.
13. Yu R, Vorontsov E, Sihlbom C, Nielsen J (2021) Quantifying absolute gene expression profiles reveals distinct regulation of central carbon metabolism genes in yeast. *Elife* 10. doi:10.7554/elife.65722.