PNAS
www.pnas.org

Supplementary Information for

Energy metabolism controls phenotypes by protein efficiency and allocation

Yu Chen[a], Jens Nielsen[a,b,c]

[a]Department of Biology and Biological Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden
[b]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kgs. Lyngby, Denmark
[c]BioInnovation Institute, DK2200 Copenhagen N, Denmark

Jens Nielsen
Email: nielsenj@chalmers.se

This PDF file includes:

Supplementary text
Figs. S1 to S7
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.1906569116

Supplementary Information Text

Modeling energy metabolism for *E. coli*.

In this study, we constructed energy metabolism for *Escherichia coli* (stored in Dataset S1), which is derived from the latest GEM iML1515 (1). In order to determine biomass formation and ATP-producing pathways, we performed FBA with iML1515 assuming minimal medium with glucose as carbon source with oxygen available.

For determining HY pathway, we fixed glucose uptake rate at 1 in the model, then maximized the ATP maintenance (ATPM) reaction with minimizing the calculated total fluxes. As a result, we obtained one unique flux distribution, from which we selected the reactions carrying non-zero fluxes and marked reaction IDs and metabolite IDs with "_HY" and "_hy", respectively.

For determining LY pathway, we changed the objective function to maximization of acetate production rate with minimizing the calculated total fluxes. Likewise, we selected the reactions with non-zero fluxes and marked reaction IDs and metabolite IDs with "_LY" and "_ly", respectively.

By combining the reactions in both the HY and LY pathways and selecting the unique reactions, we obtained energy metabolism. Next, we should determine biomass formation pathways from the unique reactions. In order to achieve so, with iML1515, we fixed growth rate at 1, supplied ATP by fixing the backward rate of the ATPM reaction at 75.55 (this represents the ATP required for producing one biomass, i.e., GAM x mu), and then minimized glucose uptake rate with minimizing the calculated total fluxes. Using exchange rates (e.g., glucose, O2, H2O, and CO2) from the flux distribution, we formulated a net biomass reaction, which shows the amount of glucose needed for producing one biomass. Besides, we should determine which reactions in energy metabolism are needed for producing biomass, i.e., included in biomass formation pathway. To achieve so, we mapped the reactions with non-zero fluxes to the unique reactions in energy metabolism. The overlapped reactions should be included in biomass formation pathway, which were marked with "_Bio" in the model. To consider how much each reaction in the biomass formation pathway contributes to biomass formation, we formulated a drain reaction to produce biomass precursor, in which the flux value of each reaction is included as coefficient.

At last, we added ATP transport reactions between biomass formation and ATP-producing pathways, connecting the independent pathways, as well as exchange reactions.

Modeling energy metabolism for *S. cerevisiae*.

We also constructed energy metabolism model for *Saccharomyces cerevisiae*, stored in Dataset S2. In order to determine biomass and ATP-producing pathways, we used the consensus model Yeast7.6 (2) to perform FBA. Since the reaction IDs in Yeast7.6 do not show any biological information, we changed to the IDs in the GEM iMM904 (3), but still kept the GPR associations from Yeast7.6.

Following the same procedure as shown for *E. coli* modeling, we determined biomass and ATP-producing pathways. A minor change is that when determining LY pathway we maximized ethanol production rate. Additionally, we took into account the fluxes towards glycerol production in the case that experiments show that it cannot be neglected, e.g., growth at high temperature. The glycerol production pathway is not included in the model but we estimated the flux loss. To achieve so, we used the same method as done for determining biomass formation pathway, but we changed to maximize glycerol production rate. As a result, we obtained a net reaction that represents the amount of glucose goes to glycerol, and reactions with the corresponding fluxes contribute to glycerol production, which can then be used to estimate protein cost by producing glycerol.

Likewise, we added exchange and ATP transport reactions.

Protein cost analysis.

In order to calculate protein cost for carrying one unit of flux, we collected molecular weights from the UniProt database (4). Besides, we took into consideration stoichiometry information for each protein, i.e., determined whether it is a monomer or oligomer. This information can be sourced from the BRENDA database (5), the RCSB PDB database (6), and also the UniProt database. The turnover rates used in the two models were collected from published studies and databases. Especially, most of the turnover rates in the *E. coli* model were obtained from the study, which calculated maximal catalytic rates using proteomics data (7). Some other turnover rates were obtained from either the BRENDA database or the SABIO-RK database (8). For the enzyme without available turnover rate, we assumed to be 79 /s, which is the median of CCM (9).

Protein cost is defined in this study as protein mass (g/gCDW) required for carrying one unit of flux (mol/gCDW/h). Based on the kinetic equation of an enzymatic reaction:

$$v = k_{cat} \cdot E \cdot f(X)$$

in which v is reaction rate, k_{cat} is turnover rate, E is enzyme concentration, $f(X)$ is a function of substrate and product concentrations. The calculation of protein cost is just similar to calculate enzyme concentration in the equation:

$$p = E = \frac{v[mol_{substrate}/gCDW/h]}{k_{cat}[mol_{substrate}/mol_{enzyme}/h] \cdot f(X)}$$

When the reaction rate is one mol/gCDW/h, the minimal protein cost can be calculated by assuming $f(X)$ is one. Then we obtained:

$$p = \frac{[mol_{substrate}/gCDW/h]}{k_{cat}[mol_{substrate}/mol_{enzyme}/h]} = \frac{[mol_{enzyme}/gCDW]}{k_{cat}} = \frac{MW}{k_{cat}}[g/gCDW]$$

Accordingly, we can calculate the minimal protein cost by dividing molecular weight (g/mol) of the enzyme by its turnover rate (/h).

We then collected the information, e.g., molecular weight and turnover rate, as described in the main text and calculated protein cost for all the enzymatic reactions in the models. Besides, for the reactions with multiple isozymes, the lowest protein cost was adopted.

The protein cost information can be available in Dataset S1 for *E. coli* and Dataset S2 for *S. cerevisiae*.

Model simulations.
With the two models, we carried out several types of constraint-based simulations in this study following the linear programming (LP) problem:

$$\begin{gathered}\text{minimize } c^T v \\ \text{subject to } S \cdot v = 0, \\ \sum p_i \cdot v_i \leq \emptyset, \\ lb_i \leq v_i \leq ub_i.\end{gathered}$$

We used the function linprog in MATLAB to solve LP problem, which attempt to solve a minimization problem, but can be used to find the maximum by changing the objective function to negative.

Without any experimental data, we performed simulations by changing protein constraint of energy metabolism from 0.05 to 0.15 g/gCDW and glucose uptake rate. We maximized ATP production rate without limiting other exchange reactions. As a result, we can obtain the maximal ATP production rate for each fixed protein constraint and glucose uptake rate, and also flux distribution.

With physiological data, we used the models to estimate the minimal protein mass of energy metabolism. To achieve so, we fixed the corresponding exchange rates to experimental values and then searched for the minimal protein allocation that can still obtain a feasible solution. For the cases that the original experimental data cannot result in a feasible solution even though without limiting protein allocation, we gradually enlarged the bounds of the exchange reactions till we obtained a feasible solution. Then with the changed bounds we searched for the minimal protein allocation. This type of simulation was carried out for investigating strains under batch conditions. Additionally, with fixing the minimal protein allocation, we can then estimate fraction of flux through different pathways as well as proteome allocation among those pathways.

Regarding the simulations of metabolic switch, we firstly determined protein allocation to energy metabolism according to the minimal protein mass estimated for cells in batch conditions. We chose the median of those values estimated with the physiological data from various datasets. Subsequently, we fixed the protein allocation and then minimized glucose uptake rate for a series of growth rates.

In order to predict metabolic engineering targets with the model, we doubled turnover rate of each enzyme in the model one by one, and then maximized the growth rate. The predicted growth rate was then compared to the reference growth rate, which was estimated without changing any turnover rates.

Adjustment of protein efficiency.
The ratio of protein efficiencies between LY and HY pathways can control the simulations of metabolic switch in terms of the critical point and the slope of acetate/ethanol flux versus glucose flux or growth rate. To change the ratio, we just

adjusted the enzyme saturation of HY pathway in this study since adjusting the enzyme saturation of LY pathway can change the maximal glucose flux or growth rate. By adjusting the ratio, we can change simultaneously the critical point and the slope. In this study, the best-fitted ratio with experimental observation was chosen when the difference between the simulated and experimental critical points is small enough (<0.001 /h in growth rate).

Proteomics data analysis.
We used absolute proteomics data from three published datasets (10–12) to estimate protein mass of energy metabolism. Since the datasets use different units, here we show how to convert them.

The unit in the dataset (10) is molecules per fL of cytosol volume, which can be converted to g/gCDW as follow:

$$\frac{[molecules]}{[fL]} = \frac{[mol]}{N_A[fL]} = \frac{MW[g]}{N_A[fL]} = \frac{MW[g]}{N_A \cdot \rho \cdot 10^{-12}[gCWW]}$$
$$= \frac{MW[g]}{N_A \cdot \rho \cdot 10^{-12} \times 0.3[gCDW]}$$

In which is N_A the Avogadro constant, MW is molecular weight (g/mol), ρ is cell buoyant density (1.1 g/mL) (10), 0.3 is from a dry fraction of wet biomass of 30% (10).

The unit in the dataset (12) is molecules per pgCDW, which can be converted to g/gCDW as follow:

$$\frac{[molecules]}{[pgCDW]} = \frac{[mol]}{N_A[pgCDW]} = \frac{MW[g]}{N_A[pgCDW]} = \frac{MW[g]}{N_A \cdot 10^{-12}[gCDW]}$$

Calculating of apparent saturation.
According to (7), the kinetic equation of an enzymatic reaction:

$$v = k_{cat} \cdot E \cdot f(X) = E \cdot k_{app}$$

in which, k_{app} is *in vivo* catalytic rate of the enzyme under a condition. Therefore, the apparent saturation σ_{app} can be calculated by:

$$\sigma_{app} = k_{app}/k_{max}$$

In which, k_{max} is the maximal value of the catalytic rate over many conditions. Therefore, the kinetic equation can be:

$$v = E \cdot k_{app} = \sigma_{app} \cdot k_{max} \cdot E$$

In our models, we used the maximal value k_{max} for each enzyme and did not take the apparent saturation σ_{app} into account. Therefore, we can obtain an estimated total enzyme concentration, which is less than or equal to the real value:

$$E_{est} = \sum_i E_{est,i} = \sum_i \frac{v_i}{k_{max,i}}$$

The real concentration should be:

$$E_{real} = \sum_i E_{real,i} = \sum_i \frac{v_i}{\sigma_{app,i} \cdot k_{max,i}}$$

When we use an average apparent saturation for all the enzymes, we can obtain:

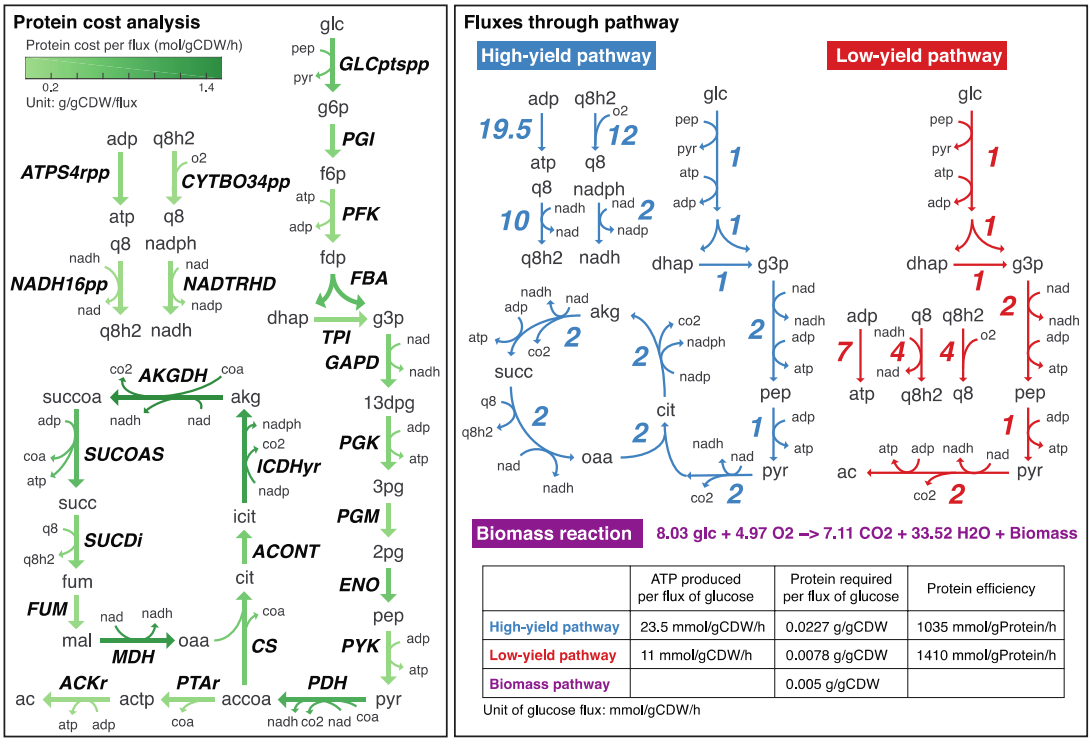
$$E_{real} = \frac{1}{\sigma_{app}} \sum_i \frac{v_i}{k_{max,i}}$$

Therefore, we can calculate the average apparent saturation:

$$\sigma_{app} = E_{est} / E_{real}$$

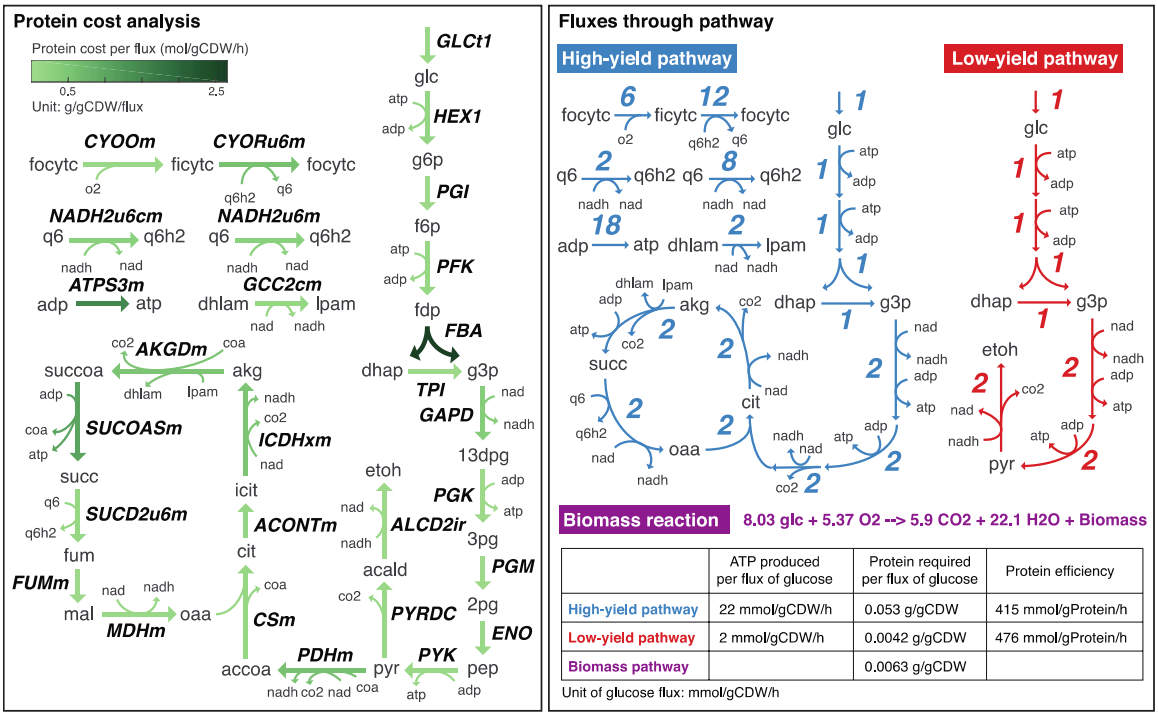
Here we chose as the real value the median of measured proteome levels assigned to energy metabolism, which is around 0.1 g/gCDW in both *E. coli* and *S. cerevisiae*.

A. Model of *E. coli*



	ATP produced per flux of glucose	Protein required per flux of glucose	Protein efficiency
High-yield pathway	23.5 mmol/gCDW/h	0.0227 g/gCDW	1035 mmol/gProtein/h
Low-yield pathway	11 mmol/gCDW/h	0.0078 g/gCDW	1410 mmol/gProtein/h
Biomass pathway		0.005 g/gCDW	

B. Model of *S. cerevisiae*



	ATP produced per flux of glucose	Protein required per flux of glucose	Protein efficiency
High-yield pathway	22 mmol/gCDW/h	0.053 g/gCDW	415 mmol/gProtein/h
Low-yield pathway	2 mmol/gCDW/h	0.0042 g/gCDW	476 mmol/gProtein/h
Biomass pathway		0.0063 g/gCDW	

Fig. S1. Overview of energy metabolism in *E. coli* (A) and *S. cerevisiae* (B). Reaction ID is consistent with that in Dataset S1 and Dataset S2. The estimated protein cost data are available in Dataset S1 and Dataset S2.

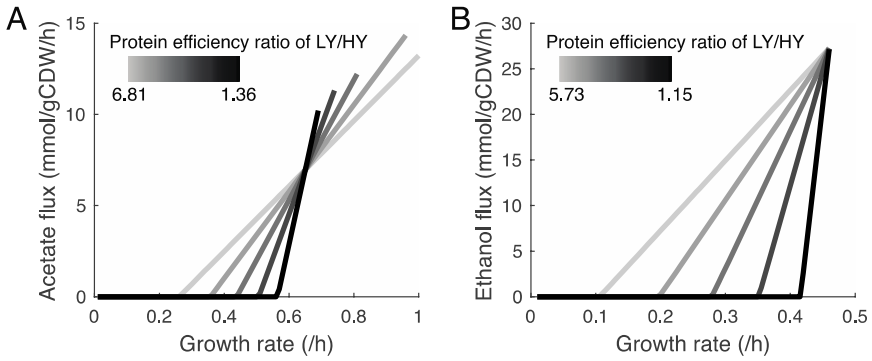


Fig. S2. The effect of the ratio of protein efficiencies between LY and HY on the critical point with increasing growth rates. (A) Simulations for *E. coli*. (B) Simulations for *S. cerevisiae*.

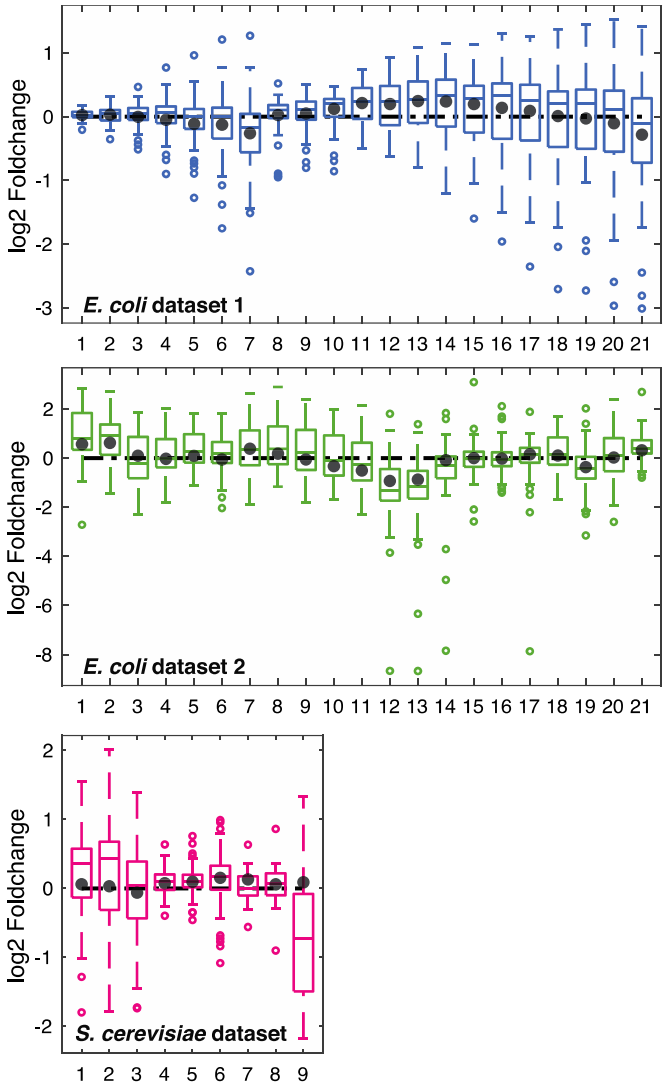


Fig. S3. Change in total protein (black filled dot) and individual proteins (open dot) in energy metabolism of various conditions compared with reference condition. Condition ID is available in Dataset S3. *E. coli* dataset 1 is from (10). *E. coli* dataset 2 is from (11). *S. cerevisiae* dataset is from (12).

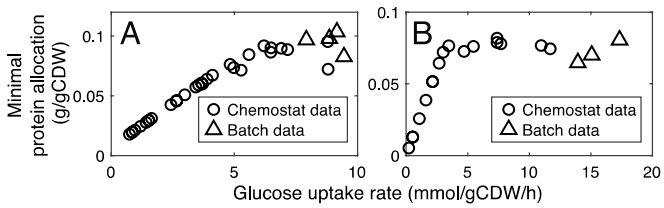


Fig. S4. (A) Predicted minimal protein mass of energy metabolism in *E. coli* using experimental exchange rates as constraints. Data were from the same studies as Fig. 3A. (B) Predicted minimal protein mass of energy metabolism in *S. cerevisiae* using experimental exchange rates as constraints. Data were from the same studies as Fig. 3B.

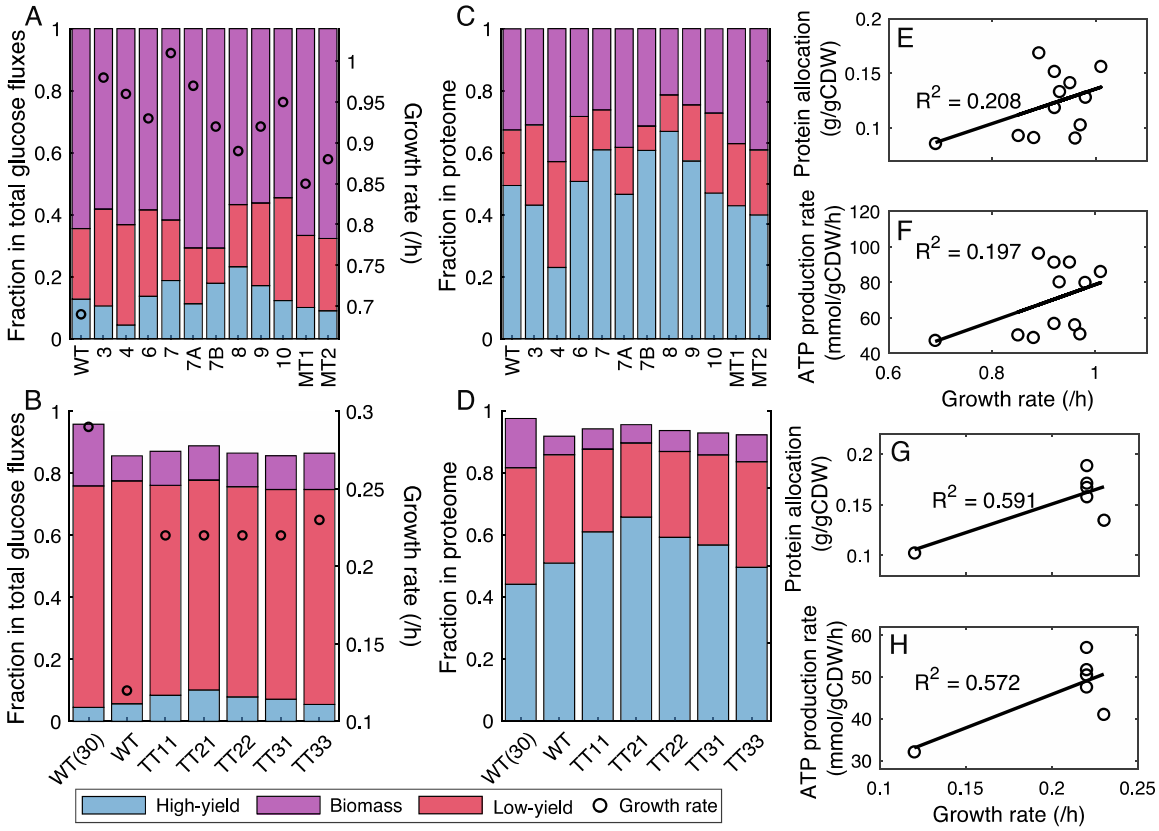
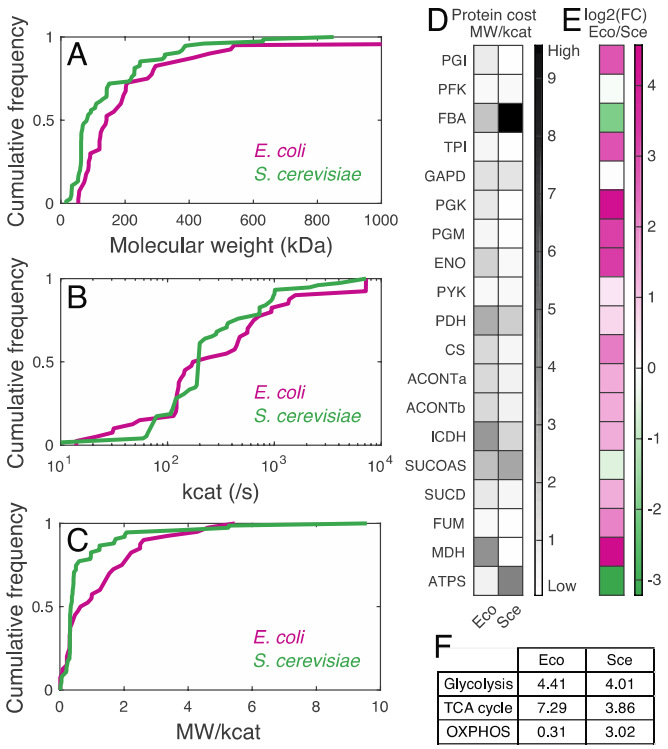


Fig. S5. Simulations of various strains under exponential growth. (ACEF) Simulations for *E. coli*. (BDGH) Simulations for *S. cerevisiae*. For one thing, it shows that cells, even though high ATP production rate is required, still maintain a considerable activity of the HY pathway, which accounts for 40-60% of the protein mass of energy metabolism. For another thing, it shows that the correlation between the protein allocation to energy metabolism and growth rate is quite poor, so is the correlation between the ATP production rate and growth rate.



F

	Eco	Sce
Glycolysis	4.41	4.01
TCA cycle	7.29	3.86
OXPHOS	0.31	3.02

Fig. S6. Comparisons between *E. coli* and *S. cerevisiae*. (A) Cumulative distribution of molecular weights of all the enzymes involved in energy metabolism. It shows that enzymes in *E. coli* are generally a little bigger than those in *S. cerevisiae*. (B) Cumulative distribution of turnover rates of all the enzymes involved in energy metabolism. It shows that *E. coli* and *S. cerevisiae* have considerable turnover rates for the enzymes in energy metabolism. (C) Cumulative distribution of protein costs of all the enzymes involved in energy metabolism, which can be calculated by molecular weight over turnover rate. It shows that *E. coli* has a greater number of costly enzymes than *S. cerevisiae*. (D) Protein costs of shared reactions in *E. coli* (Eco) and *S. cerevisiae* (Sce). It shows that the fructose 1,6-bisphosphate aldolase (FBA) reaction in *S. cerevisiae* is the most expensive among all the reactions. (E) Foldchange of protein costs between *E. coli* and *S. cerevisiae* for the same reaction. It shows that for most reactions *E. coli* should allocate more proteins than *S. cerevisiae*. (F) Comparison of protein costs of pathways between *E. coli* and *S. cerevisiae*. Values in the table are protein costs (g/gCDW) per flux of glucose for glycolysis, pyruvate for the TCA cycle, and NADH for oxidative phosphorylation. The table shows that lower protein efficiency of energy metabolism in *S. cerevisiae* could be contributed by its much higher protein cost of oxidative phosphorylation compared with *E. coli*. More specifically, it is due to higher protein cost of ATP synthase reaction, which is caused by lower turnover rate of the enzyme (Dataset S1 and Dataset S2).

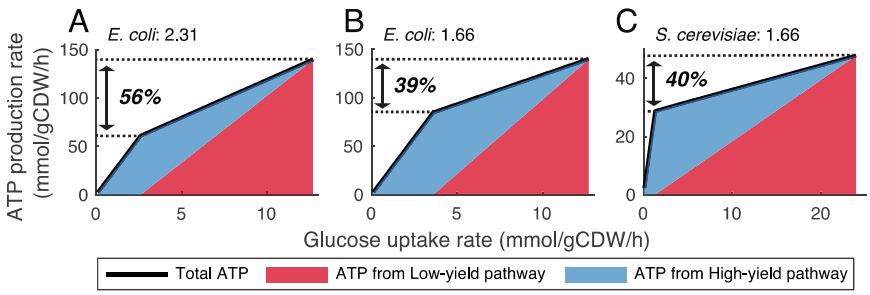


Fig. S7. The effect of the ratio of protein efficiencies between LY and HY on the critical point with increasing glucose uptake rates. The simulations were generated by assuming that the protein mass of energy metabolism is 0.1 g/gCDW. By decreasing the ratio in *E. coli* to the value in *S. cerevisiae*, i.e., from panel A to B, we can see that the critical point increases. This means that increasing the ratio can make cells easier to switch with increasing glucose uptake rates. Besides, we can see that if the ratio in *E. coli* is equal to that in *S. cerevisiae*, i.e., comparing panel B and C, the percentages in the panels are very close, which is calculated by the difference between ATP production rate at the critical point and the maximal ATP production rate over the maximal ATP production rate. Taken together, we can conclude that higher ratio makes *E. coli* easier to switch from HY to LY pathway. The analysis is independent with the protein mass, which has been shown in the main text to be correlated with the maximal ATP production rate and also the critical point. (A) Simulations of *E. coli* with the original ratio, which is 2.31. (B) Simulations of *E. coli* with changing the ratio by increasing the protein efficiency of HY pathway to the value of *S. cerevisiae*, which is 1.66. (C) Simulations of *S. cerevisiae* with the original ratio, which is 1.66.

References

1. Monk JM, et al. (2017) iML1515, a knowledgebase that computes Escherichia coli traits. *Nat Biotechnol* 35(10):904–908.
2. Aung HW, Henry SA, Walker LP (2013) Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind Biotechnol* 9(4):215–228.
3. Mo ML, Palsson BØ, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3(1):37.
4. UniProt Consortium T (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 46(5):2699–2699.
5. 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.
6. Burley SK, et al. (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* 47(D1):D520–D528.
7. Davidi D, et al. (2016) Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proc Natl Acad Sci U S A* 113(12):3401–6.
8. Wittig U, et al. (2012) SABIO-RK--database for biochemical reaction kinetics. *Nucleic Acids Res* 40(D1):D790–D796.
9. Bar-Even A, et al. (2011) The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* 50(21):4402–4410.
10. Peebo K, et al. (2015) Proteome reallocation in Escherichia coli with increasing specific growth rate. *Mol Biosyst* 11(4):1184–1193.
11. Schmidt A, et al. (2016) The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol* 34(1):104–110.
12. 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.