Profiling metabolic flux modes by enzyme cost reveals variable trade-offs between growth and yield in *Escherichia coli*: Supplementary information

Meike T. Wortel[1,†], Elad Noor[2,†], Michael Ferris[3], Frank J. Bruggeman[4], and Wolfram Liebermeister[5,*]

[1]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway
[2]Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule, Zürich, Switzerland
[3]Computer Sciences Department and Wisconsin Institute for Discovery, University of Wisconsin, Madison, USA
[4]Systems Bioinformatics Section, Vrije Universiteit, Amsterdam, The Netherlands
[5]Institute of Biochemistry, Charité – Universitätsmedizin Berlin, Berlin, Germany
[†]M.W. and E.N. contributed equally to this work
[*]wolfram.liebermeister@gmail.com

Contents

S1 Methods details

This section provides some details about the different parts of our modelling approach: parameter balancing for obtaining kinetic parameters, a fast implementation of enzyme cost minimization, the conversion between enzyme-specific biomass production into cell growth rates, and a comparison of flux-analysis ECM to traditional kinetic and constraint-based modelling approaches.

S1.1 Parameter balancing

Parameter balancing [1] is a method for translating incomplete, potentially inconsistent sets of kinetic and thermodynamic constants into a complete, consistent set of model parameters. In brief, parameter balancing works as follows. We collect all relevant quantities that appear in the data or in the model and merge them into a vector $\mathbf{y}$. These quantities must satisfy Wegscheider conditions and Haldane relationships, which defines linear equality constraints between their logarithmic values. Accordingly, to satisfy the constraints in a safe way, we write all these quantities as linear combinations of independent parameters ($\ln k^{\mathrm{V}}$, $\ln k_{\mathrm{M}}$, and μ° values), with the definition $k^{\mathrm{V}} = \sqrt{k^{+}_{\mathrm{cat}}\, k^{-}_{\mathrm{cat}}}$. The independent parameters, which are collected in a vector $\mathbf{x}$, can be varied without violating any constraints. The linear dependence between the complete and the independent parameter sets can be written as $\mathbf{y} = \mathbf{R}\,\mathbf{x}$ with a matrix $\mathbf{R}$ derived from the model structure. Using this equation as a linear regression model, we can convert an experimentally known vector $\mathbf{y}^{\mathrm{data}}$ (which may be incomplete) into a best estimate of the underlying vector $\mathbf{x}$. Using the estimate $\mathbf{x}$, we again apply $\mathbf{R}$ to obtain a completed, consistent version of $\mathbf{y}$. Since this regression problem is usually underdetermined, we employ Bayesian estimation. Priors allow us to obtain plausible estimates even from sparse data. Accordingly, the result is not simply a point estimate of $\mathbf{y}$, but a multivariate Gaussian posterior distribution for possible parameter vectors $\mathbf{y}$. A best estimate is given by the center of the distribution; from the covariance matrix, we obtain uncertainties of individual model parameters as well as the correlations between them.

S1.2 Enzyme cost minimization implemented in the GAMS modeling system

The optimization of enzyme cost for an individual EFM is a convex problem and can therefore be solved with local optimization methods and in polynomial time. This allows us to use a powerful solver that optimizes a single enzyme profile in a few seconds. Enzyme cost minimization has been implemented within the General Algebraic Modeling System (GAMS) modeling system [2], accessible through the NEOS server (`https://neos-server.org/neos/`), which provides a convenient way to write down the optimization problem and uses automatic differentiation techniques to exactly evaluate derivatives, e.g., of the functions defining the constraints and objective function of the model. This is important since it allows the nonlinear optimization solvers that are linked to the modeling system to efficiently use this information for improved solution speed and accuracy. As a default setting, general rate equations as in this paper are used (modular rate laws [3], possibly with simple allosteric regulation). However, the implementation also allows for custom rate equations, such as the biomass equation in this paper (a case study with explanations can be found on `http://www.neos-guide.org/content/enzyme-cost-minimization`). To define an optimization problem, the user provides the metabolic network (in the form of reaction stoichiometries), kinetic constants, fluxes, and possibly enzyme costs. These data can be provided as a single csv file or a collection of csv files, and the user can select a particular optimization solver from within those linked to the GAMS system, and files defining the model and its parameters. A number of additional arguments allow for defining bounds on subsets of the decision variables, changing the solver used for convex optimization, and setting options for

Quantity	Symbol	Physical unit
Metabolite level (metabolite i)	c_i	mM
Metabolite flux (reaction l)	v_l	mM s^{-1}
Enzyme concentration (enzyme l)	E_l	mM
Biomass production	v_{BM}	mg l^{-1} h^{-1}
Total enzyme cost	E_{met}	mg l^{-1}
Enzyme doubling time	τ_{met}	h
Enzyme specific biomass production	r_{BM}	h^{-1}
Cellular protein mass	P_{tot}	mg l^{-1}
Biomass concentration	c_{BM}	mg l^{-1}
Doubling time	T	h
Cell growth rate	μ	h^{-1}

Table S1: Mathematical symbols and physical units used in the formulae.

the solver. Although the method is applied here to EFMs only, it can also be used to compute the enzyme cost of any given flux mode. The calculations for this article were executed on a shared server: a Dell PowerEdge R430 server with the following configuration: CPU - 2x Intel Xeon E5-2698 @ 2.3GHz (32 cores total), HT Enabled, Memory - 192GB RAM, Disk - 4x 300G SAS drives setup in RAID5, Network - 1Gb/s Ethernet.

S1.3 Converting enzyme investments into cell growth rates

Being able to compute enzyme-specific biomass production rates, we next translate these rates into cell growth rates. The growth rate of a cell is given by $\mu = v_{\mathrm{BM}}/c_{\mathrm{BM}}$, where c_{BM} is the biomass concentration, i.e. the amount of biomass per cell volume and v_{BM} is the rate of biomass production (amount of biomass produced per cell volume and per unit time). The cell's growth rate can be approximately computed from the enzyme cost of biomass production, and we will see that higher enzyme-specific biomass production rates entail higher growth rates. Therefore, an assessment of enzyme-specific biomass production rates (in optimization, or in drawing a rate/yield scatter plot) is equivalent to an assessment of growth rates. Here we derive two conversion formulae: a linear formula (by which growth rate and enzyme-specific biomass production only differ by a scaling factor) and a nonlinear one (which takes into account the growth-rate-dependent investment in ribosomes). As shown in Figure S1, even the latter prediction yields an overall picture very similar to the direct assessment of enzyme-specific biomass production.

Linear formula based on fixed enzyme fraction in the proteome To estimate μ from the enzyme-specific biomass production $r_{\mathrm{BM}} = v_{\mathrm{BM}}/E_{\mathrm{met}}$, we need to know the ratio $E_{\mathrm{met}}/c_{\mathrm{BM}}$, i.e., the mass fraction of biomass formed by metabolic enzymes. Empirically, metabolic enzymes (or more specifically: the metabolic enzymes considered in our model) occupy about one eighth of the biomass (in mass units). The mass fraction of protein mass within biomass is: $\alpha_{\mathrm{prot}} = P_{\mathrm{tot}}/c_{\mathrm{BM}} \approx 0.5$ (BioNumber 101955 [4]) relatively constant across cell types. The mass fraction of metabolic enzyme within the proteome, in *E. coli*, varies around $\alpha_{\mathrm{ccm}} = E_{\mathrm{met}}/P_{\mathrm{tot}} \approx 25\%$ (from proteomics data [5]). If these numbers were constant, the growth rate

$$\mu = \frac{v_{\mathrm{BM}}}{c_{\mathrm{BM}}} = \underbrace{\frac{v_{\mathrm{BM}}}{E_{\mathrm{met}}}}_{r_{\mathrm{BM}}} \underbrace{\frac{E_{\mathrm{met}}}{P_{\mathrm{tot}}}}_{\alpha_{\mathrm{ccm}}} \underbrace{\frac{P_{\mathrm{tot}}}{c_{\mathrm{BM}}}}_{\alpha_{\mathrm{prot}}} \tag{S1}$$

would be proportional to the enzyme-specific biomass production r_{BM}, with a prefactor of $\alpha_{\mathrm{ccm}} \cdot \alpha_{\mathrm{prot}} \approx 0.125$. We obtain the formula

$$\mu \approx 0.125\, r_{\mathrm{BM}}. \tag{S2}$$

Nonlinear formula based on growth-dependent enzyme fraction In reality, the proteomic fraction of metabolic enzymes changes with the growth rate. As shown by experiments, and as explained by the allocation of protein resources between metabolic enzymes and ribosomes [6], if the growth rate varies because of varying metabolic efficiency, the fraction of metabolic enzymes decreases with the growth rate. To account for this fact, the formula needs to be modified. In experiments where cell growth is controlled by nutrient quality or by dilution in chemostats, we can assume a linearly decreasing fraction [6]

$$\frac{E_{\mathrm{met}}}{P_{\mathrm{tot}}} = a - b\,\mu. \tag{S3}$$

with positive coefficients a and b. These coefficients can be estimated from proteomics data [5]: the protein fraction devoted to central carbon metabolism decreases from $\approx\ 25\%$ during slow growth (μ = 0.11/h) to $\approx\ 18\%$ during faster growth (μ = 0.48/h), leading to estimates $a \approx 27\%$ and $b \approx 20\%$ h. Inserting (S3) into (S1) and solving for μ, we obtain the formula

$$\mu = \frac{\alpha_{\mathrm{prot}}\ a\ r_{\mathrm{BM}}}{1 + b\ r_{\mathrm{BM}}\ \alpha_{\mathrm{prot}}}. \tag{S4}$$

Inserting the numerical values, we obtain

$$\mu \quad = \quad \frac{0.5 \cdot 0.27 \cdot r_{\mathrm{BM}}}{1 + 0.2[\mathrm{h}] \cdot 0.5\ r_{\mathrm{BM}}} = \frac{0.135\ r_{\mathrm{BM}}}{1 + 0.10[\mathrm{h}] \cdot r_{\mathrm{BM}}}. \tag{S5}$$

In both approximation formulae (S2) and (S5), μ increases with r_{BM}. This is why, in the article, maximizing the growth rate μ is equivalent to maximizing $r_{\mathrm{BM}} = v_{\mathrm{BM}}/E_{\mathrm{met}}$ or minimizing E_{met} at given v_{BM}. As shown in Figure S1, the linear and nonlinear formulae for growth rate calculations yield almost the same results.

It is useful to rewrite formulae (S2)-(S5) also in terms of doubling time. For that, we define the *metabolic enzyme doubling time* as

$$\tau_{\mathrm{met}} \equiv \frac{\ln(2) E_{\mathrm{met}}}{v_{\mathrm{BM}}} = \frac{\ln(2)}{r_{\mathrm{BM}}} \tag{S6}$$

and the cell doubling time (in hours) will thus be

$$T = \frac{\ln(2)}{\mu} = \frac{\tau_{\mathrm{met}}}{\alpha_{\mathrm{prot}}\ a} + \frac{\ln(2) \cdot b}{a} = 7.4\ \tau_{\mathrm{met}} + 0.51[\mathrm{h}]. \tag{S7}$$

An alternative derivation, which directly refers to the sector models by Scott *al.* [6], is given in the following section.

Nonlinear conversion between enzyme-specific biomass production and growth rate, derived from enzyme-ribosome trade-off In the main article, we derived the nonlinear rate/growth relationship from the empirical observation that the metabolic enzyme fraction if the proteome decreases linearly with th growth rate. Here we add an alternative derivation that directly refers to the sector models by Scott *al.* [6]. To derive the nonlinear relationship (S4) between enzyme-specific biomass production and cell growth rate, we assumed that the fraction of metabolic enzymes within the proteome varies with growth rate, and we described this dependence by a simple linear function (with offset), which we extracted from proteome data. This argument agrees with the proteome sector models by Scott *al.* [6], in which the proteomic fraction of metabolic enzymes is given by a constant baseline amount plus a variable amount that is proportional to the growth rate. To clarify the close connection to proteome sector models, we derive our formula again with the terminology of the growth rate models. We assume that the proteome can be split into three mass

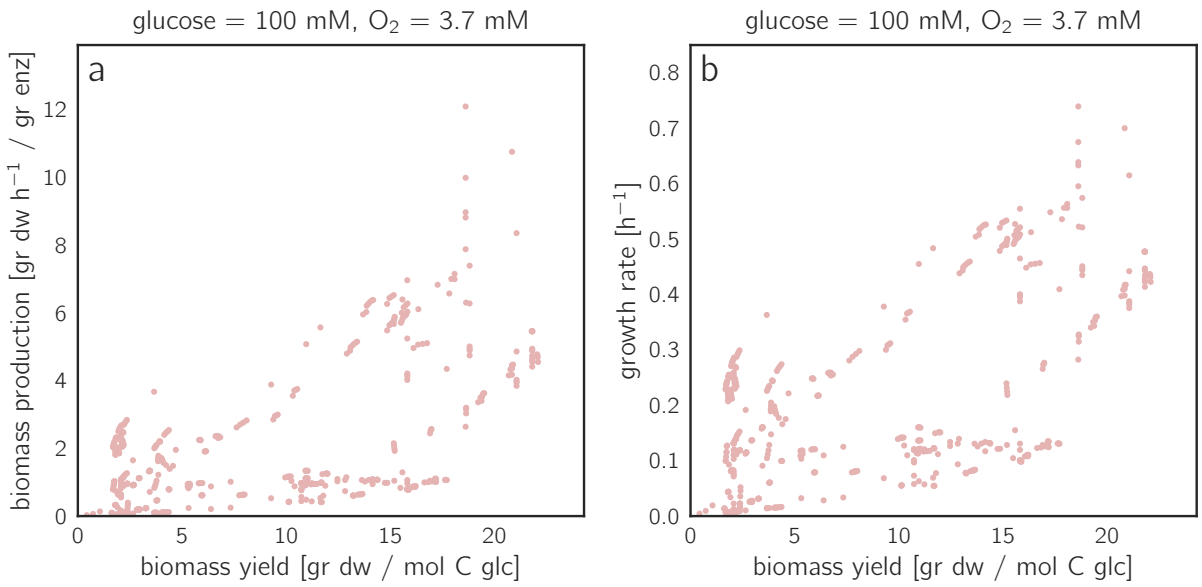

Figure S1: **Biomass production rates and growth rates.** (a) Enzyme-specific biomass production *vs.* biomass yield. (b) Cell growth rate *vs.* biomass yield. The nonlinear scaling (i.e. using equation S5) of the growth rate has only a slight effect on the Pareto plot.

fractions: a constant fraction (about half of the proteome), a variable fraction x (consisting of metabolic enzymes), and a variable fraction y (consisting of ribosomal proteins). We further assume that each of the variable fractions x and y is proportional to the cell growth rate λ, i.e., $\lambda = a_x\, x = a_y\, y$. We further assume that a_x is given by (or proportional to) the enzyme-specific biomass production rate (which we can compute from our model for each EFM) and that a_y is constant (because we do not consider, for example, the effects of translational inhibitors). With these assumptions, the growth rate is given by

$$\lambda = \frac{1}{\frac{1}{a_x} + \frac{1}{a_y}} \underbrace{[x + y]}_{c_{\mathrm{vp}}} \tag{S8}$$

where c_{vp} is the (constant) sum of variable protein fractions. We now set $a_x = r/c_{\mathrm{bm}}$, where r is the biomass production rate (in carbon molar biomass / time) per metabolic enzyme (in carbon molar) and c_{cm} is the biomass concentration in cells (in carbon molar). The value of a_y can be obtained from a proteomics data set by a regression between growth rate and ribosome fraction. Eventually, we obtain

$$\lambda = \frac{1}{\frac{c_{\mathrm{bm}}}{r} + \frac{1}{a_y}} \cdot c_{\mathrm{vp}} = \frac{c_{\mathrm{vp}} \cdot r}{c_{\mathrm{bm}} + \frac{1}{a_y}\, r}. \tag{S9}$$

This is a hyperbolic function just like the one we derived before. By adjusting this formula to proteomics data, we obtain the same parameters as above.

S1.4 Flux-analysis ECM provides advantages over common kinetic or constraint-based modelling methods

Flux-analysis enzyme cost minimization includes full kinetic information and a realistic description of enzyme costs into a flux optimization setting and thereby closes the gap between kinetic and stoichiometry-based modeling. It provides a clear theoretical link between kinetic models and other network-based approaches, which have incorporated some of the kinetic information. In contrast to these existing methods,

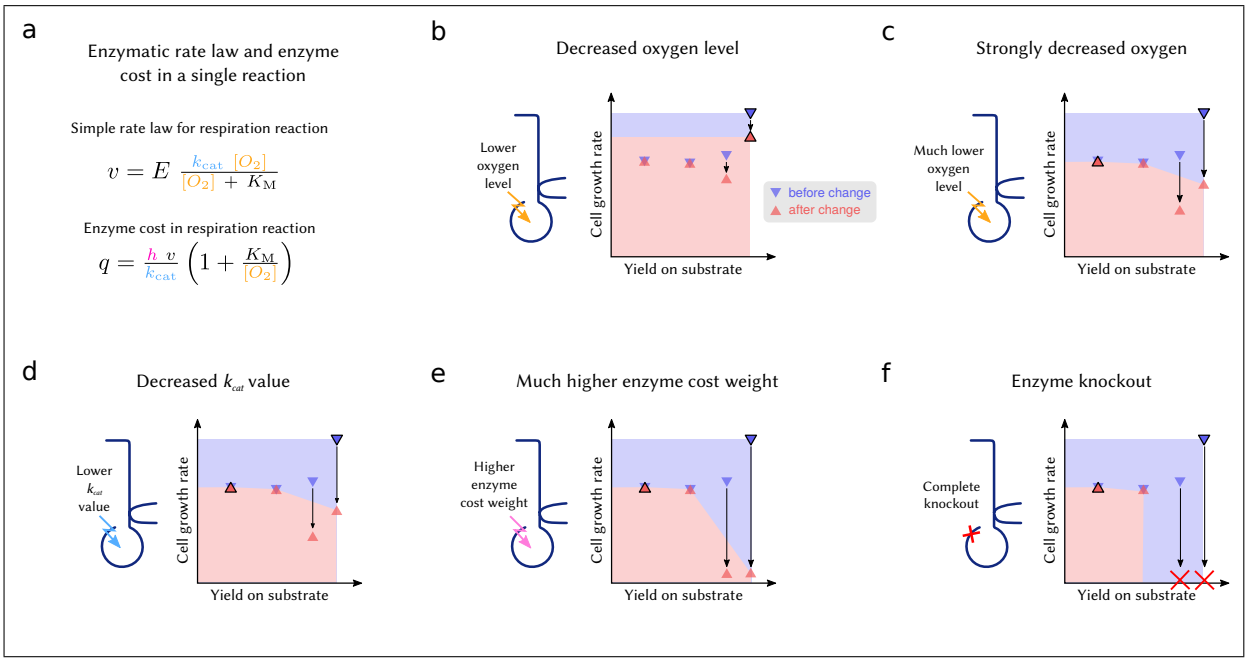

Figure S2: **Flux-analysis enzyme cost minimization can be used to simulate a wide range of enzyme perturbations and their effects on growth and metabolic strategies** (a) From the rate law of a reaction (in this schematic example, a simplified hypothetical Michaelis-Menten rate law with oxygen as a substrate), we obtain a formula for the enzyme cost as a function of flux (which depends only on the EFM), k_{cat} value, enzyme burden h, and oxygen level $[O_2]$. Changing these parameters affects the enzyme cost (and therefore the growth rate) of each EFM differently. For example, EFMs where there is no flux in the perturbed reaction, will not be affected at all and their growth rate will remain the same (in this case, these are the anaerobic EFMs). The remaining panels (b-f) show, schematically, the resulting changes in the growth/yield diagram for different parameter perturbations. In each panel, we mark the EFM with the highest growth rate with a black frame. The blue and red shaded polygons highlight the Pareto front before and after the change, respectively. If we lower the oxygen level (b), the growth-maximizing EFM shows a lower growth rate, but remains optimal. However, at a much lower oxygen level (c) there is another (oxygen-independent) EFM whose growth rate is higher. A similar change can be achieved by decreasing the k_{cat} value (d) or increasing enzyme cost weight (e). (f) When the enzyme is knocked out, all EFMs that use this reaction become infeasible and disappear from the plot. The effect resembles the effect of an extreme decrease in oxygen or k_{cat} value, or an extreme increase in the enzyme cost weight.

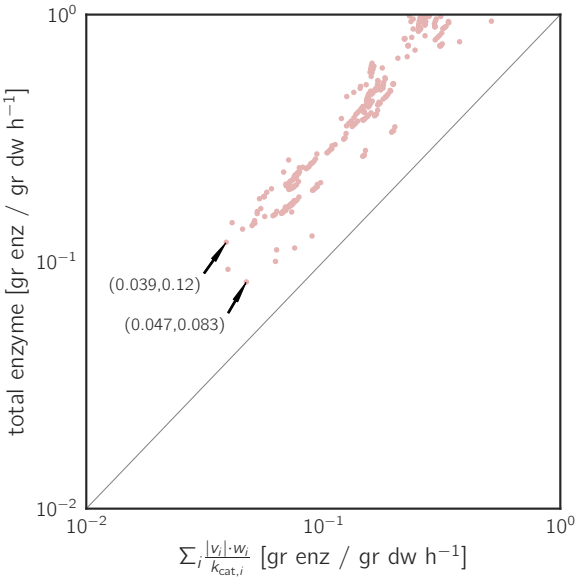

Figure S3: **Total enzyme demand of the naïve estimate versus fECM.** The "naïve" capacity-based approximation $q^{\mathrm{cat}} = \sum_i (|v_i| \cdot w_i)/k_{\mathrm{cat},i}$ of the total enzyme demand assumes that all enzyme molecules work at full speed (i.e., full substrate saturation, no reverse fluxes or saturation by products). By construction, q^{cat} is a lower bound on the estimate of the enzyme demand calculated by fECM, therefore all points in the plot are above the $x = y$ line. The two optimal EFMs (i.e. the EFM with the lowest q^{cat} and the EFM with the lowest enzyme demand) are highlighted and the x and y values are shown in parentheses. This comes to show, that using the approximation (q^{cat}) would yield different results compared to the full enzyme demand model we use throughout this work.

our method allows for systematic studies of parameter sensitivities and uncertainties. Due to the screening of EFMs, its numerical effort is much higher, but on the contrary, gene knock-outs can be easily studied after a single run without any additional numerical effort. The main advantages over existing kinetic or constraint-based methods are as follows.

- **Advantage over a direct optimization of enzyme levels in kinetic models** Our optimization procedure is equivalent to a direct optimization of enzyme levels, which would often be numerically impossible: imagine that we treat the enzyme levels E_l in a kinetic model as free variables to be optimized for minimizing the ratio $v_{\mathrm{BM}}(\mathbf{E})/E_{\mathrm{met}}(\mathbf{E})$. This would be computationally hard: the objective function $v_{\mathrm{BM}}(\mathbf{E})/E_{\mathrm{met}}(\mathbf{E})$ would not only be hard to compute (because this entails the calculation of a steady state in a kinetic model), but is also likely to be non-convex and non-concave, with potentially many local minima. In fECM, in contrast, all calculation steps are computationally tractable for medium-sized models. An advantage of our "flux-first" optimization approach is that fluxes and metabolite constraints (e.g. stationarity, kinetics, thermodynamics) can be easily imposed. Moreover, it is instructive to consider the set of EFMs and to compute the growth rates even for the non-optimal ones. By inspecting the growth/yield scatter plot, one can study which EFMs become growth-optimal under what conditions, and how the optimum switches between them upon parameter changes. This also makes it easy to understand the effects of varying external conditions, changing enzyme parameters or cost weights, and enzyme knockouts (see Figure S2).

- **Advantage over constraint-based methods with linear flux cost functions** On the contrary, our method could be compared to variants of flux balance analysis that employ flux cost functions mimicking enzyme cost. For example, the total enzyme investment $E_{\mathrm{tot}} = \sum_l E_l$ in a pathway or network can be approximated by $E_{\mathrm{tot}} \approx E_{\mathrm{lb}} = \sum_l \frac{v_l}{k_{\mathrm{cat},l}}$, which actually puts a lower bound[1] on the true value of E_{tot}. In flux-analysis enzyme cost minimization, we could use this linear function $E_{\mathrm{lb}}(\mathbf{v})$ instead of the true enzyme

[1] A similar idea underlies FBA with molecular crowding. In this method, fluxes are bounded by assuming a bound $v_l \leq v_l^{\max} = E_l\, k_{\mathrm{cat},l}$ on every reaction flux, and an upper limit on the sum of enzyme levels $E_{\mathrm{tot,max}} \geq E_{\mathrm{tot}} = \sum_l E_l$ representing the limited available space for enzymes. In this case, the linear enzyme cost function is not used as an objective to be minimized, but as a value to be constrained during optimization.

cost derived from ECM[2]. However, as shown in Figure S3, the approximation would not only decrease the predicted enzyme cost in general, but would do so to different extents across EFMs, and would therefore distort the prediction of growth-optimizing metabolic strategies.

S2 Model description

This section provides details about our model of *E. coli* central metabolism: its network structure, the choice of kinetic equations and enzyme parameters, the usage of physical units and the practical calculation of growth rates, the choice of external conditions, and a bit of statistics about the elementary flux modes.

S2.1 Network structure

The network structure of our *E. coli* model (see Figure S4) is based on the model by Carlson (2004) [7]. We changed several details of the model: We split the lumped reactions R7r, R10, R54r and R55r into separate reaction steps, we added the Entner-Doudoroff pathway (reactions R60 and R61r) and merged some reactions (the new reactions R27 and R27b) (see Table S2). For a kinetic model it is important to split linear chains of reactions, since each reaction in the chain might have different kinetics. We kept the ethanol (R90) and CO_2 (R97r) export in the model, but did not consider it in the kinetic calculation of enzyme cost because this is a passive process that is not catalyzed by enzymes. Finally, we set the internal concentrations of CO_2 and ethanol to 1 mM, which we assume is enough to have the desired export flux through diffusion. The stoichiometric constants for the biomass reaction (R70) come from the original model (Table II in [7] for a doubling time of 30 minutes). We added the measured maintenance flux to the ATP stoichiometry in the biomass equation, which changed the stoichiometry constant for ATP to -1641. An SBML version of the model is provided in the Supplementary Files. The EFMs were calculated with EFMtool [8]. The model contains a total of 2772 EFMs, out of which 1566 produce biomass. 760 of the biomass producing EFMs are not feasible because they both need oxygen and use an oxygen sensitive reaction, 97 are facultatively aerobic, 470 are strictly aerobic, and 239 strictly anaerobic (see Figure 2b). Some statistics about reactions used by the biomass producing EFMs can be found in Figure S7.

S2.2 Kinetic equations

We used the same general rate equations for all reactions [3]. Reversible (v_r) and irreversible (v_i) reactions are modeled as follows:

$$v_r = e_r \cdot k_{\text{cat},r} \frac{\prod_j \left(\frac{s_j}{K_{M,s_j,r}}\right)^{n_j} \left(1 - \frac{\prod_k p_k^{n_k} / \prod_j s_j^{n_j}}{K_{\text{eq,r}}}\right)}{\prod_j \left(1 + \frac{s_j}{K_{M,s_j,r}}\right)^{n_j} + \prod_k \left(1 + \frac{p_k}{K_{M,p_k,r}}\right)^{n_k} - 1} \tag{S10}$$

$$v_i = e_i \cdot k_{\text{cat},i} \frac{\prod_j \left(\frac{s_j}{K_{M,s_j,i}}\right)^{n_j}}{\prod_j \left(1 + \frac{s_j}{K_{M,s_j,i}}\right)^{n_j} + \prod_k \left(1 + \frac{p_k}{K_{M,p_k,i}}\right)^{n_k} - 1}. \tag{S11}$$

With e_i the enzyme concentration, s_j the substrate concentrations, p_k the product concentrations, and n the (absolute value of the) stoichiometric constant. The K_M values are the Michaelis constants, the k_{cat}

[2]The resulting method would be equivalent to a flux balance analysis with fixed biomass production rate, no other flux constraints, and a minimization of a weighted sum of fluxes, namely $E_{\text{lb}}(\mathbf{v})$.

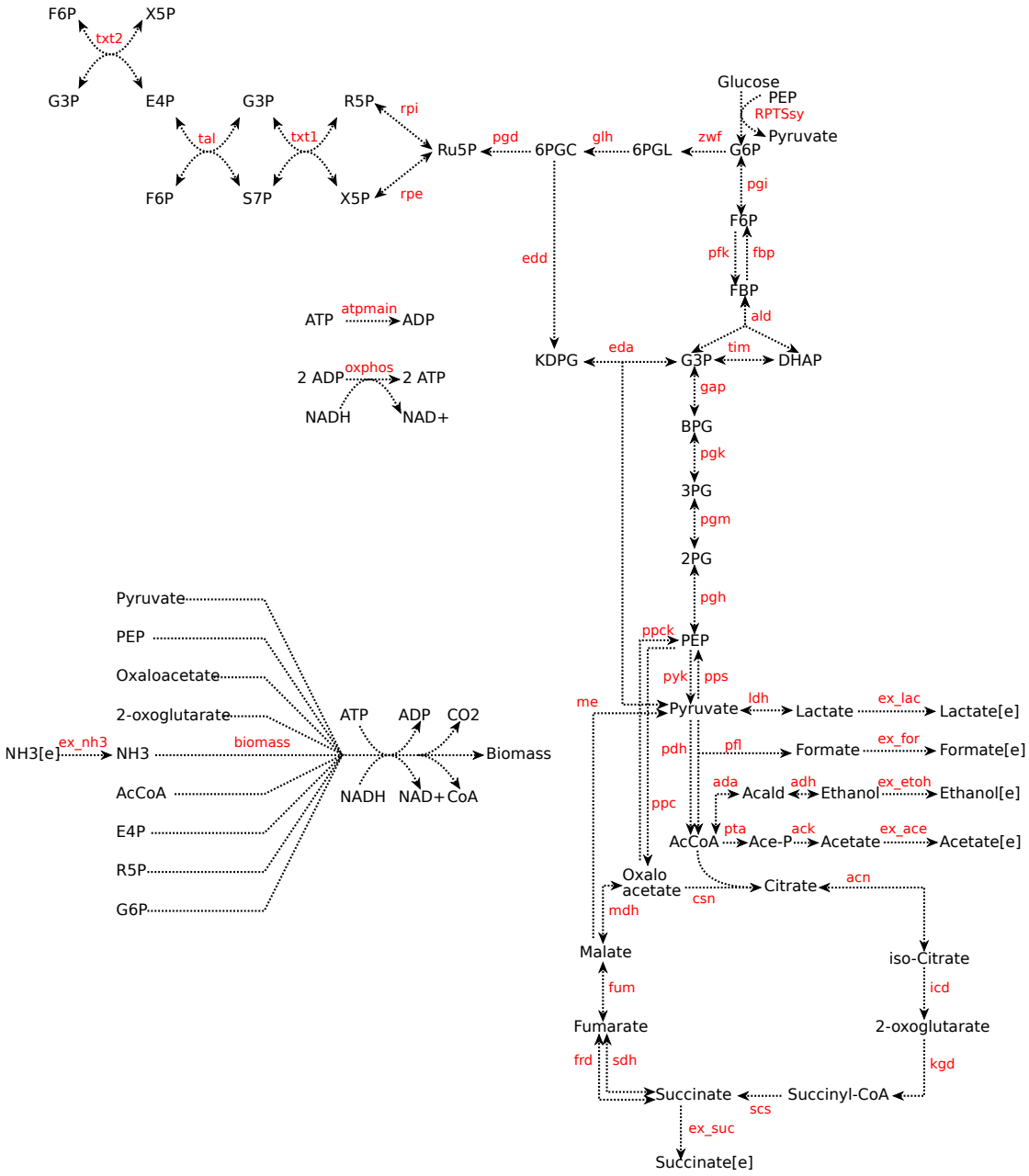

Figure S4: **The reaction network of central carbon metabolism in *E. coli*.**

values the turnover numbers and the K_{eq} values the equilibrium constants. In our model, macromolecule production is quantified by a single biomass production rate v_{BM}. The rate law represents the action of many (not explicitly modeled) cellular processes operating in steady state [9]. The biomass reaction is a lumped reaction of all processes involved in biomass production.

$$v_{\text{biom}} = \frac{e_{\text{biom}} \cdot k_{\text{cat},biom}}{1 + \sum_j \frac{K_{M,s_j,biom}}{s_j}} \tag{S12}$$

Equation (S12) corresponds to the equation in [9] with a single unit in the template ($n = 1$) and e_{biom} as the template concentration. The s_j are the substrates of the biomass reaction and $K_{M,s_j,biom}$ their Michaelis constants. Since this reaction is not a real biochemical reaction, but a lumped reaction that summarizes a

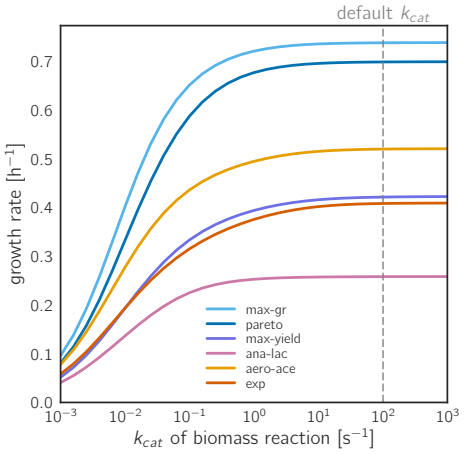

Figure S5: **Growth rates of the focal EFMs, depending on the k_{cat} value of the biomass reaction (R70)**. In all cases, the growth rates increase when the maximal velocity of the biomass reaction is increased. This is caused by the reduction of enzyme cost in the affected reaction, which is then redistributed across the network. The maximal growth rates, determined by metabolism rather than the biomass production itself, differ between EFMs.

wide range of biosynthetic reactions, its catalytic constant does not have a clearcut biochemical meaning. We thus opted not give too much weight to the biomass production, as it is a crude approximation, and the chosen k_{cat} does not affect the rate very much (see Figure S5 and compare with Figure S18). With the catalytic constant chosen, the kinetics of the biomass reaction does not limit growth, i.e., growth control is only exerted by the other, metabolic reactions.

S2.3 Choice of consistent model parameters by parameter balancing

A complete and consistent set of model parameters (k_{cat}, K_{eq}, and K_{M} values) was determined by parameter balancing [1]. Measured kinetic constants from the literature, which were incomplete and may be contradictory, were translated into model parameters that satisfy thermodynamic constraints. Unknown k_{cat} values, for example, would be substituted by values around 100 s^{-1}, but adjusted to satisfy the Haldane relationships, thermodynamics-based laws that link them to K_{M} values and equilibrium constants. Plausible orders of magnitude of kinetic parameters were defined by prior distributions (e.g., mean values and a standard deviation for logarithmic k_{cat} values in general). The median value of 100 s^{-1} was chosen because k_{cat} values in central metabolism tend to be higher than generally in metabolism (typical value around 10 s^{-1}) [10]. The kinetic parameters used as input data for parameter balancing were obtained from the literature (see Supplementary Files). Some k_{cat} values were calculated from specific activities (SA) with the formula $k_{\mathrm{cat}} = \mathrm{SA} \cdot \mathrm{MW(in\ kDalton)}/(60 \cdot \mathrm{nr_{catalyticsites}})$. The input and output files of the parameter balancing can be found in the Supplementary Files. The weights were obtained by calculating the weights for enzyme complexes and dividing by the number of catalytic sites. Whenever the number of catalytic sites was not known, we used the number of subunits as a proxy for the number of catalytic sites. There was no information available for the biomass equation and we set the K_M values to the measured intracellular concentrations when available, and otherwise to a low value as not to influence the results (see Supplementary Files). Overall, we found literature values for the k_{cat} of 26 out of 51 reactions, and K_{M} values for 87 out of the 171 reactant-enzyme pairs (i.e. about 50% coverage in both cases).

S2.4 Calculation of specific growth rate and yield

To calculate the growth rate μ (using formulae from section S1.3), we first translate the biomass flux to actual mass units (e.g. grams) by summing up the molecular masses of the biomass reactants times their stoichiometric coefficient. The ATP/ADP and NADH/NAD$^+$ couples are left out from this calculation because they produce a negligible amount of biomass. The details of the mass calculation of the biomass are given in

Table S5.1. That means that one mole of biomass weighs about 20.7 [kg]. Then, we convert the biomass flux in the model v_{R70} to the biomass production rate in the growth equations v_{BM}:

$$\begin{aligned} v_{\mathrm{BM}} &= v_{R70}\ [\mathrm{mM\ s^{-1}}] \cdot 2.07 \times 10^4\ [\mathrm{mg\ mmol^{-1}}] \cdot 3.6 \times 10^3\ [\mathrm{s\ h^{-1}}] \\ &= v_{R70} \cdot 7.45 \times 10^7\ [\mathrm{mg\ s\ mmol^{-1}\ h^{-1}}] \end{aligned} \tag{S13}$$

Our model calculates the abundance of each enzyme (e_i, in mM) required for this biomass flux. To get the total mass of enzyme requirement we multiply each e_i by the molecular weight per active site of each enzyme (w_i, given in $[\mathrm{mg\ mmol^{-1}}]$). Therefore, the total cost will be $E_{\mathrm{met}} = \sum_i e_i w_i\ [\mathrm{mg\ l^{-1}}]$. Finally, the *enzyme doubling time* would be:

$$\tau_{\mathrm{met}} = \frac{\ln(2) E_{\mathrm{met}}}{v_{\mathrm{BM}}} = v_{R70}^{-1} \cdot 9.3 \times 10^{-9} [\mathrm{mmol\ h\ mg^{-1}\ s^{-1}}] \cdot E_{\mathrm{met}} \tag{S14}$$

And according to Equation S7, the doubling time of the cell would be:

$$T = 7.4\ \tau_{\mathrm{met}} + 0.51[\mathrm{h}] = \frac{1}{v_{R70}} \cdot 6.9 \times 10^{-8}\ [\mathrm{mg\ s\ mmol^{-1}\ h^{-1}}] \cdot E_{\mathrm{met}} + 0.51[\mathrm{h}] \tag{S15}$$

From equations (S1), (S4) and (S13) the growth rate μ can be calculated from the total cost E_{met} and the biomass flux in the model v_{R70} with the following formula:

$$\mu = \frac{v_{R70} \cdot 10^7\ [\mathrm{mg\ s\ mmol^{-1}\ h^{-1}}]}{E_{\mathrm{met}} + v_{R70} \cdot 7.5 \times 10^6\ [\mathrm{mg\ s\ mmol^{-1}}]} \tag{S16}$$

The biomass yield is expressed in milligram biomass per millimole of carbon atoms uptake and, therefore, we need to convert biomass production to grams and substrate uptake rates to mole carbons. Since glucose molecules contain six carbon atoms, and a mole of biomass weights 20666 grams (see Table S5.1) the biomass yield is given by:

$$Y_{\mathrm{g/C}} = \frac{20666\ [\mathrm{mg\ mmol^{-1}}]\, v_{R70}}{6 \cdot v_{\mathrm{pts}}} = \frac{v_{R70}}{v_{\mathrm{pts}}} \cdot 3.4 \times 10^3\ [\mathrm{mg\ mmol^{-1}}], \tag{S17}$$

where v_{pts} is the flux in the PTS glucose uptake system (reaction **R1** in the model).

S2.5 Realistic values for enzyme concentrations

Although the model calculates enzyme abundances in mM, there is a hidden scaling factor which needs to be taken into account in order to use this estimated values for validation. In our implementation of the optimization problem, we assume that v_{R70} is equal to $1\ [\mathrm{mM\ s^{-1}}]$. But is this a realistic value for actual *E. coli* cells growing exponentially?

To answer this question, we must first calculate the value of v_{BM}. An average bacterial cell consists of about 30% dry matter [11], which translates to a dry density of $\rho \approx 3 \times 10^5\ [\mathrm{mg\ l^{-1}}]$. A typical value for the growth rate would be $\mu \approx 1\ [\mathrm{h^{-1}}]$, which would be enough to for a rough estimate of the biomass rate, given by $v_{\mathrm{BM}} = \rho \cdot \mu$. Using equation S13 we can now get:

$$\begin{aligned} v_{R70} &= \frac{v_{\mathrm{BM}}\ [\mathrm{mg\ l^{-1}\ h^{-1}}]}{7.45 \cdot 10^7\ [\mathrm{mg\ s\ mmol^{-1}\ h^{-1}}]} = \frac{\rho\ [\mathrm{mg\ l^{-1}}] \cdot \mu\ [\mathrm{h^{-1}}]}{7.45 \cdot 10^7\ [\mathrm{mg\ s\ mmol^{-1}\ h^{-1}}]} \\ &\approx 4 \times 10^{-3}\ [\mathrm{mM\ s^{-1}}]\,. \end{aligned} \tag{S18}$$

This result means that our standard value for v_{R70} is ∼250 times too high, and so are the estimates for enzyme concentrations. To obtain realistic estimated value should, one must therefore divide each e_i by 250.

It is important to note, that this overestimation of v_{R70} and e_i does not affect the overall model prediction of growth rate and yield. One can see, that in formulas S2.4 and S2.4, this scaling factor affects both the numerators and the denominators and therefore cancels out. This is not a coincidence, but a feature of the way we calculate growth rate by dividing the rate of biomass production by the required enzyme amount for that specific rate. As an outcome of this independence, one can also use the predicted growth rate directly in equation S18 instead of the typical value ($1\ [\mathrm{h}^{-1}]$), even though the value of v_{R70} and the enzyme concentrations were used to calculate the growth rate in the first place.

S2.6 Choice of standard external conditions

As a standard condition for our simulated cells, we chose a high external glucose concentration of 100 mM. For O_2 we chose the concentration from the same paper as we used for the kinetics of the oxygen using reactions [12], namely 0.21 mM. The concentration of other external compound that is taken up, NH_3, was set to 10 times more than the highest K_M to ensure saturation (1.0 mM). The levels of excreted metabolites were assumed to be low and were set to 0.01 mM, except for ethanol and CO_2 which were set to 1 mM (which are actually internal metabolites, since we treated the export reactions as non-enzymatic).

S2.7 Details on the elementary flux modes

Some biological and statistical properties of the EFMs are shown in Figures S6, S7, S8, and S9.

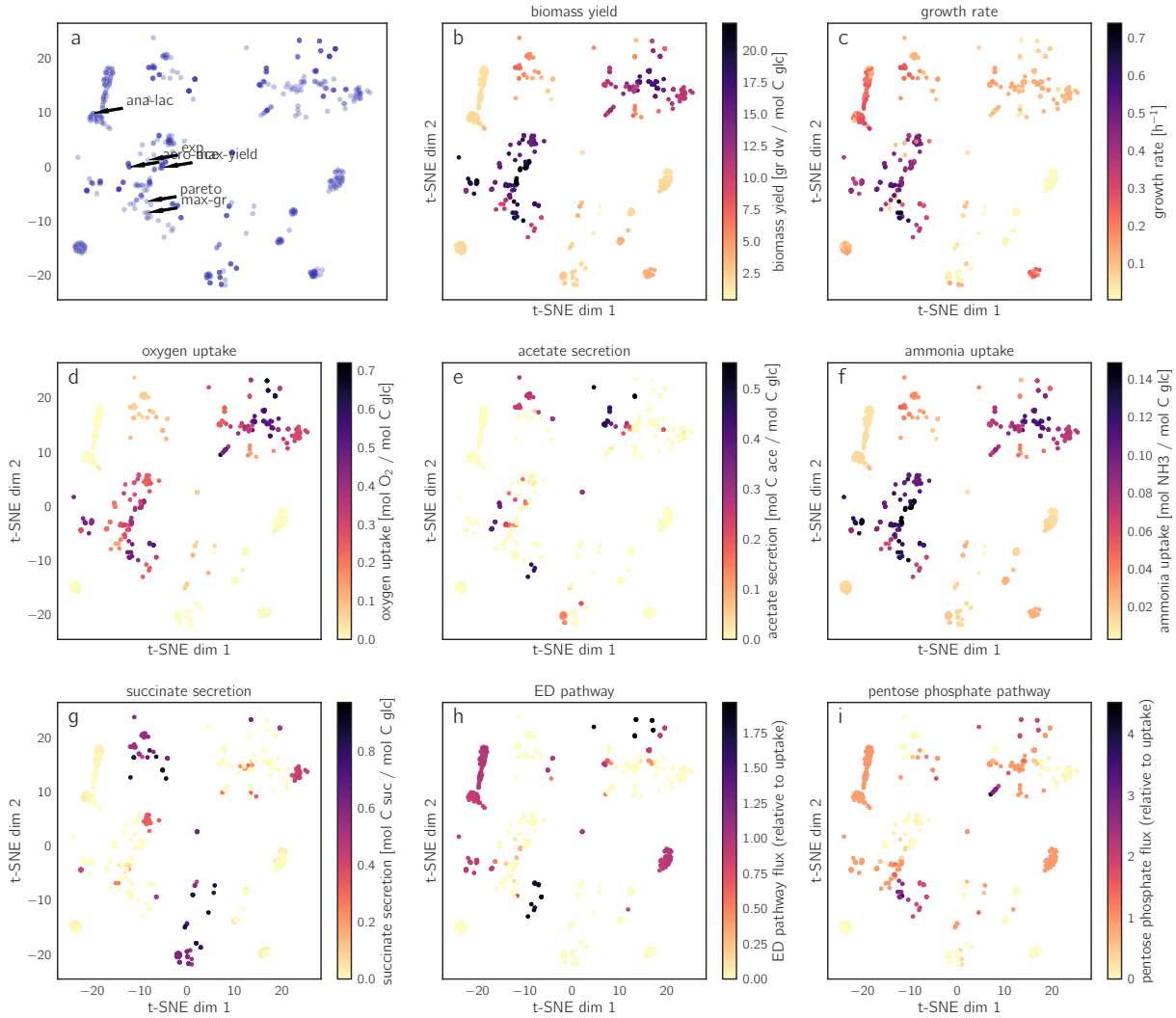

Figure S6: **Elementary flux modes, visualized using the t-SNE algorithm.** (a) Elementary flux modes are vectors in a high-dimensional flux space. The t-SNE algorithm [13] represents the EFMs by points on a two-dimensional plane. It tries to preserve their original distances (i.e., the Euclidean distances in flux space) while spreading the points evenly over the plane. The biomass yield (b) and specific growth rate (c) are represented by a color scale. As expected, similar EFMs have very similar yields (since the yield is one of the dimensions in the flux space) but growth rate can sometime change significantly for neighboring EFMs. In each of the panels (d-i) we show a single feature in color coding on the same EFM map.

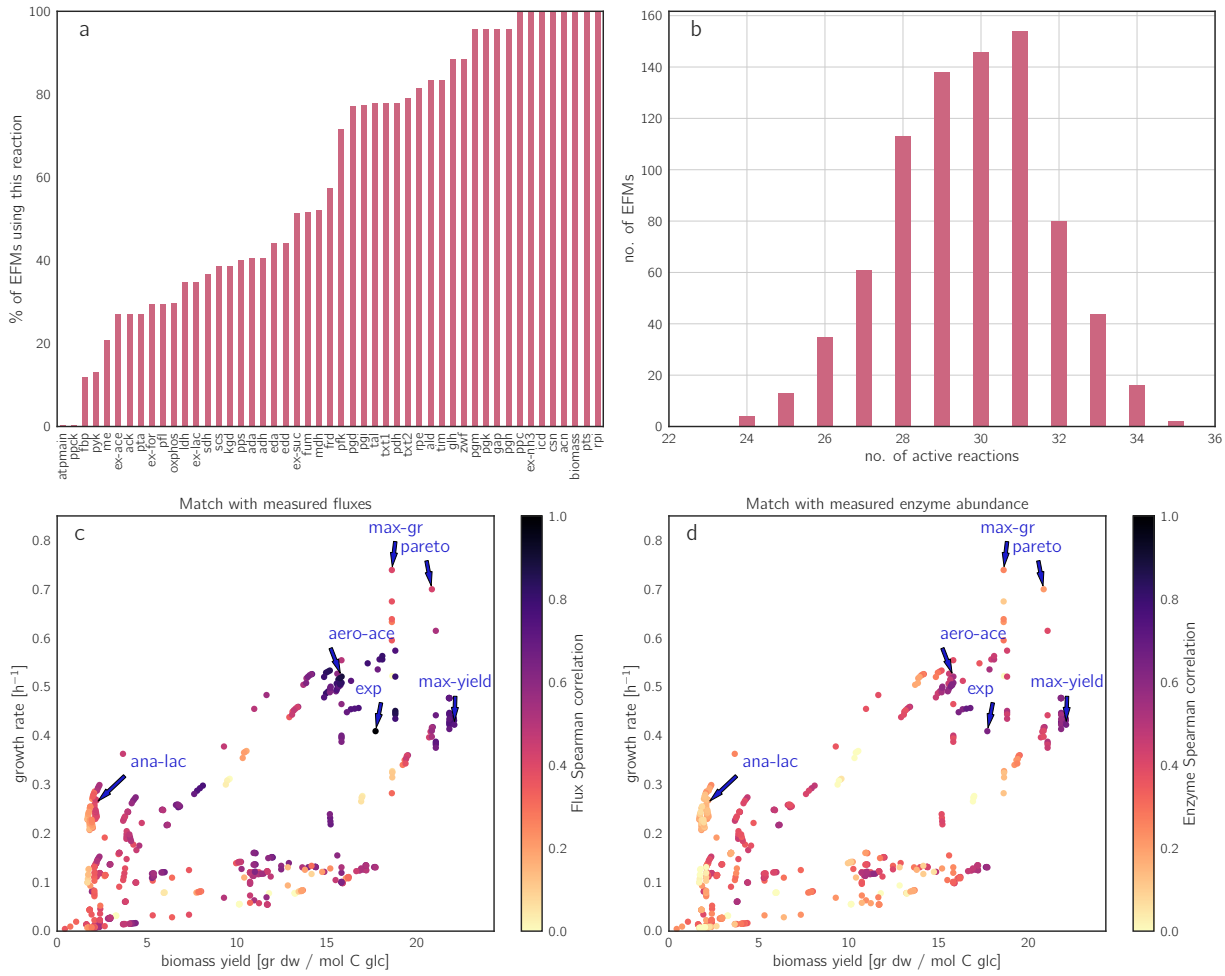

Figure S7: **Statistical properties of biomass producing EFMs and similarity to measured fluxes** (a) Usage of individual reactions by biomass-producing EFMs. For each reaction, a bar shows in what percentage of all EFMs this reaction is active. A core of 8 essential reactions is active in all biomass producing EFMs (**ppc**, **ex-nh3**, **icd**, **csn**, **acn**, **biomass**, **pts**, and **rpi**). (b) Size distribution of biomass producing EFMs (number of active reactions). (c) Correlation of measured flux distribution with each other EFM. The color of each EFM in this Pareto plot corresponds to the Spearman correlation between the fluxes in that EFM and the experimentally measured fluxes from [14]. (d) Similarly, we plot the Spearman correlation between the estimated enzyme levels for each EFM and the measured enzyme levels from [15]. Note that even the point corresponding to *exp* does not have a correlation of 1, since even though the fluxes are taken from experiments, the estimated enzyme abundances are still given by the ECM algorithm. Nevertheless, *exp* is among the EFMs with the highest correlation.

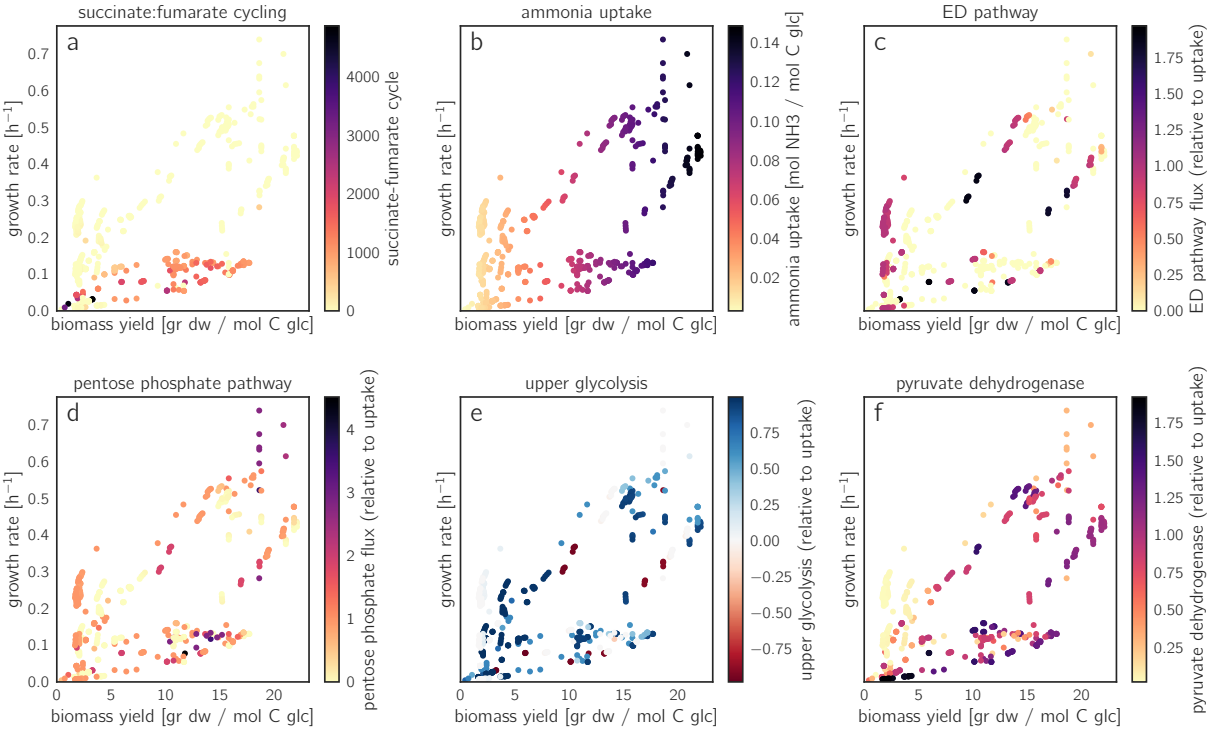

Figure S8: **Some relevant fluxes for all EFMs plotted in a rate/yield diagram under standard conditions.** More fluxes are given in main text Figure 3. (a) Succinate:fumerate cycling is not very beneficial and occurs only in suboptimal EFMs. (b) Ammonia uptake is very correlated with yield, because ammonia enters the biomass equation directly. (c) The ED pathway is used throughout the spectrum but clearly by a whole group of low yield/high growth rate EFMs. (d) The pentose phosphate pathway is also used throughout, but more in the higher growth rate EFMs. (e) The flux in upper glycolysis, which is reversible and therefore is depicted using a red-blue colormap, is zero for the seven EFMs with the highest growth rate and otherwise usually positive. Only a few EFMs with medium-low growth rates use the reverse direction. (f) The flux through pyruvate dehydrogenase, which is a very large enzyme complex, is relatively low for the fastest EFMs.

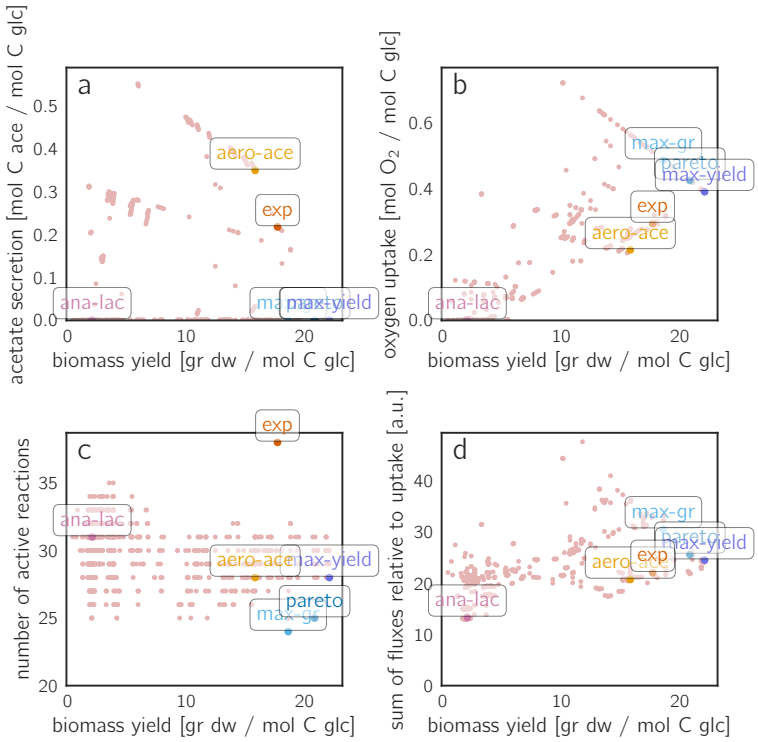

Figure S9: **Correlations of biomass yield with acetate flux, oxygen flux, number of reactions and sum of all fluxes, all scaled with the uptake rate** All quantities shown follow directly from the shapes of EFMs, independently of a specific kinetic model. (a) Among the acetate-secreting EFMs, higher acetate secretion tends to imply lower biomass yields. (b) The oxygen uptake is optimal at around 0.4 [mol O_2 per mol C], as lower uptake forces higher byproduct secretion rates (to make up for the ATP requirements) and higher O_2 uptake rates reduce yield as well, since more carbon is oxidized and released as CO_2. (c) The number of active reactions, a very simplified measure of enzyme demand, shows relatively little correlation with biomass yield. (d) The same holds for the sum of fluxes relative to glucose uptake.

S3 Model results

This section provides additional results on simulated growth rates for different EFMs, under different external conditions (variation of glucose and oxygen levels), different choice of kinetic constants, under a restricted usage of pathways, and under enzyme knockouts. It also provides results on the necessary enzyme investments (for different EFMs, and under varying external conditions) and on the optimal metabolic strategies in chemostats at different growth rates (as determined from predicted Monod curve parameters).

S3.1 Growth rates achieved by elementary flux modes

As shown in Figure S10, the predicted cell growth rates do not only vary widely across EFMs, but their distribution also assumes very different shapes depending on biochemical external conditions. At standard (high oxygen) conditions, they are relatively evenly distributed, while under low-oxygen conditions, a large number of EFMs (the oxygen-dependent ones) show very low growth rates, and only a very small percentage reaches comes close to the maximal growth rate. This has practical consequences for modelling: since the number of EFMs in a model can be large, it might be tempting to sample EFMs instead of enumerating all of them, in order to find at least some EFMs with realistic biological properties. However, this approach would probably fail in the low-oxygen case shown. Almost all EFMs would yield very unfavourable growth rates. In this example, a pragmatic solution would be to sample EFMs in a way that oxygen-dependent EFMs are automatically discarded. However, in general cases, such heuristics may be difficult to find. We conclude that a sampling of EFMs may yield some well-performing EFMs (as here, in high-oxygen conditions), but it may also fail (as here, in low-oxygen conditions), and which of the two is the case may be hard to control or assess.

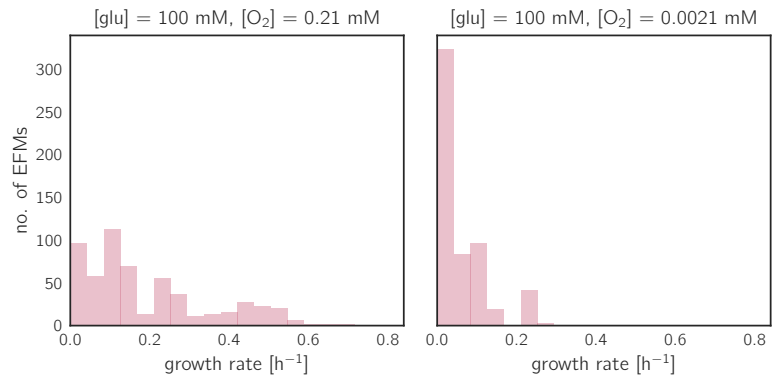

Figure S10: **Distribution of growth rates over EFMs.** (a) Under standard conditions, the growth rates are relatively evenly distributed, and by randomly sampling a small number of EFMs, there would still be a chance to find EFMs that maximize the growth rate. (b) Under low-oxygen conditions, however, the distribution is very far from uniform and skewed to the left. The chances of finding one of the few EFMs with high growth rates (i.e. > 0.2 [h^{-1}]) using random sampling are very small.

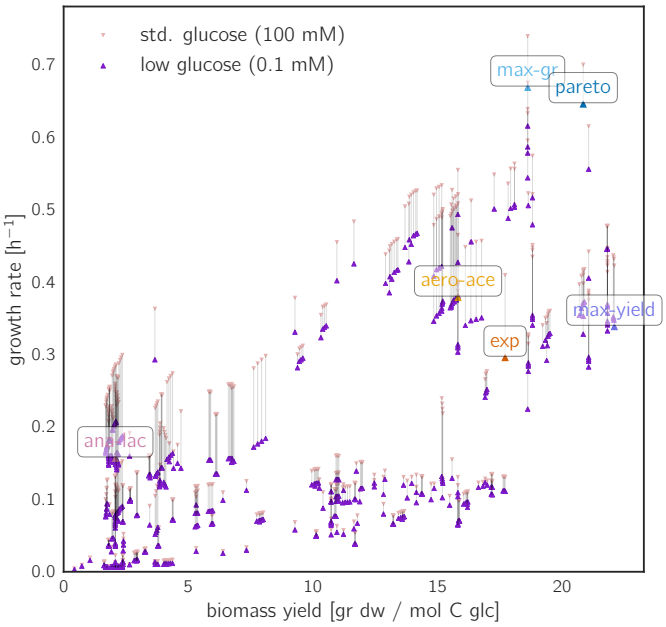

Figure S11: **Growth effects of a decrease in glucose level**. A drop in glucose level (from 100 mM to 0.1 mM) decreases the growth rate for all the EFMs, each to a different extent. The EFMs with low yield are more affected (ana-lac, for example), since they have higher glucose uptake rates, and the enzyme burden of the PTS system is larger. Nevertheless, the Pareto front changes only slightly, and still consists of a few EFMs.

S3.2 How the choice of metabolic strategies depends on glucose and oxygen levels

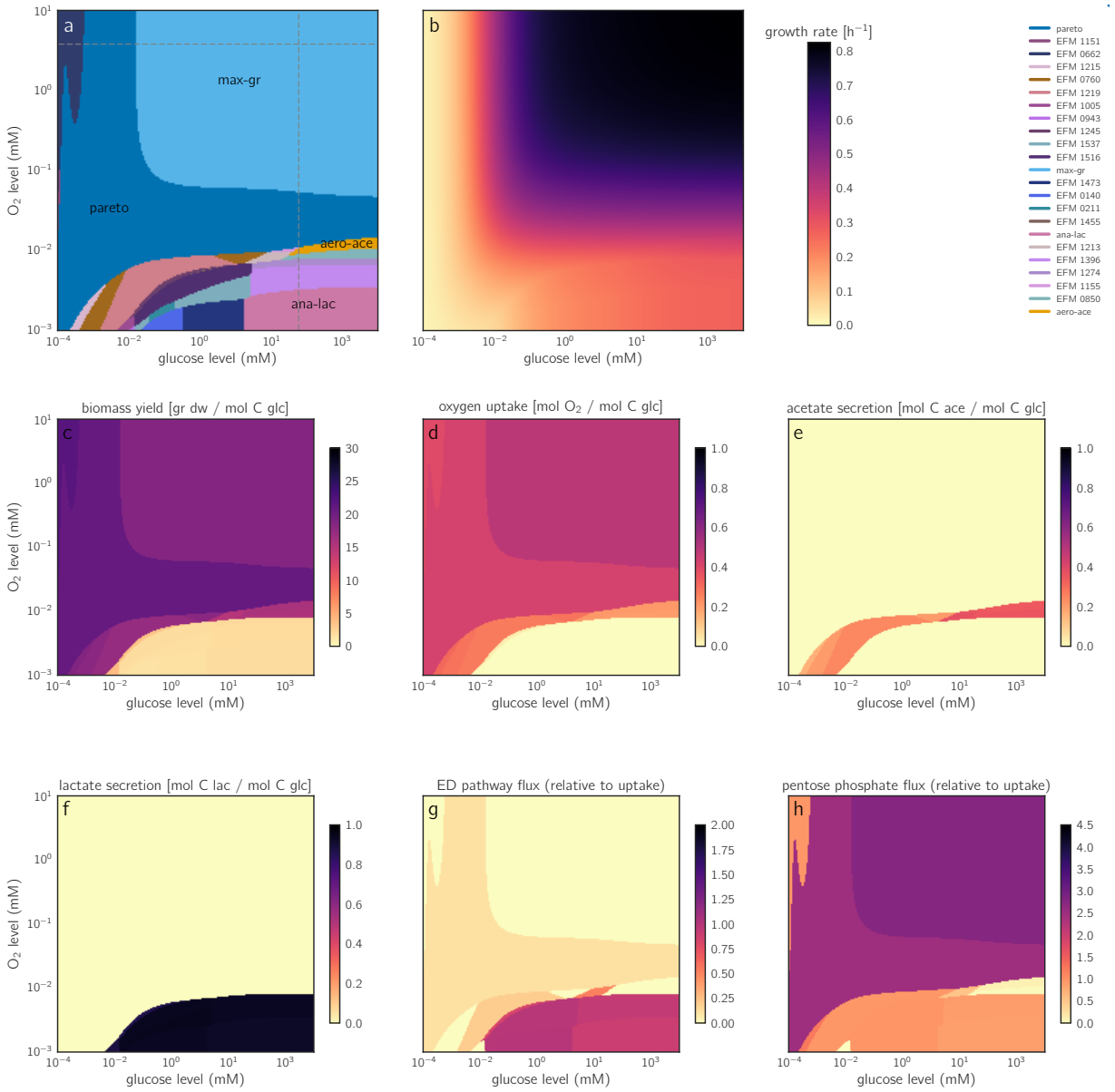

Figure S12: **Cell growth rate as a function of external glucose and oxygen levels.** For each combination of glucose and oxygen levels, maximal growth is realized by one "optimal EFM" (marked by colors). In chemostat experiments, the same fixed growth rate could be realized by a (one-dimensional) continuum of possible states, each entailing a different combination of external glucose and oxygen levels arising in the medium. The specific choice will probably depend on the ratio of glucose and oxygen supplies to the chemostat. (a) 23 different EFMs are optimal (i.e. have the highest growth rate) in different regions in the glucose/oxygen plane. (b) In contrast to the complex map of optimal EFMs, the optimal growth rates change very smoothly and it is difficult to see any transitions except for one boundary that curves up around the lower right corner. From overlaying this plane with different EFM properties (c-h) of the optimal EFM at each region, we can see that this low oxygen/high glucose region consists only of anaerobic EFMs (d). Interestingly, acetate fermentation is only favorable in a narrow band around this anaerobic region and becomes unfavorable again at high O_2 levels (e).

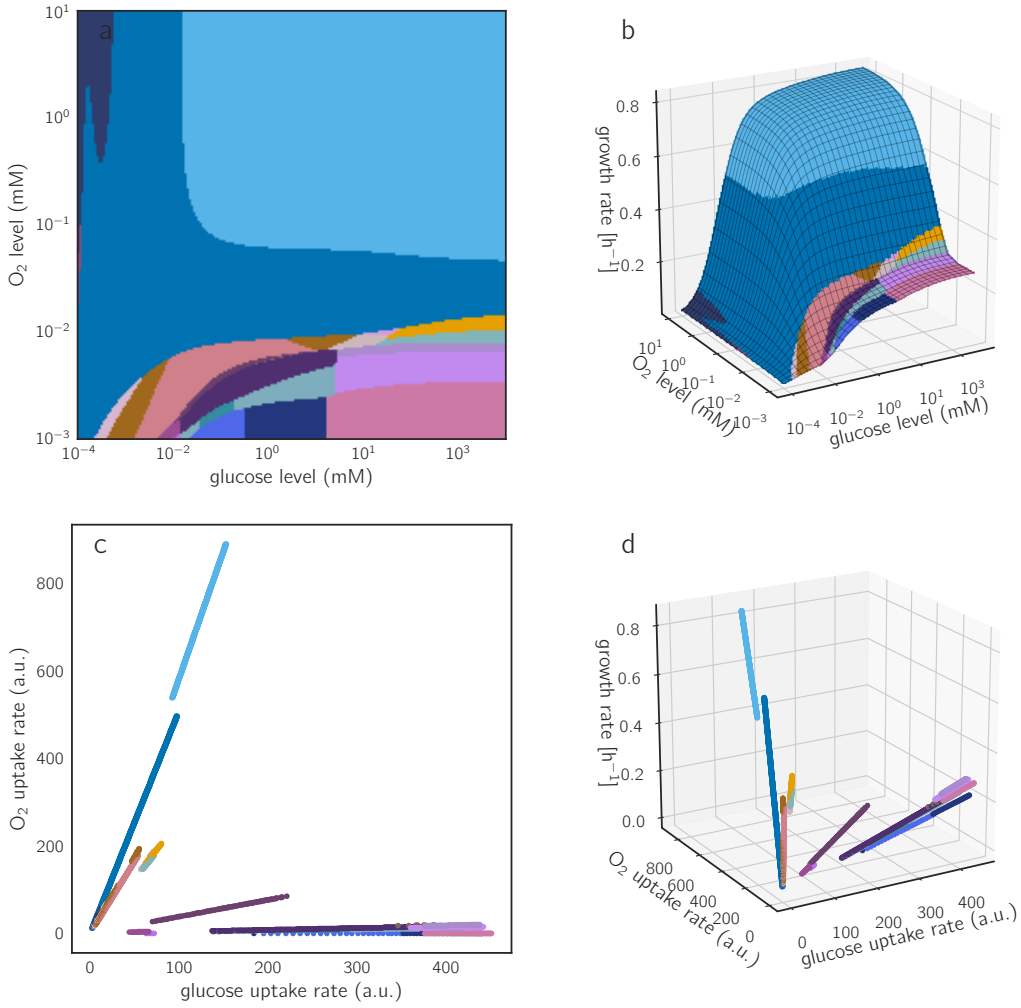

Figure S13: **Optimal EFMs as functions of uptake rates.** (a) The optimal EFMs across the glucose/oxygen plane (same figure as S12a). (b) The optimal EFM colors overlaid on a 3D surface plot, where hight represents the optimal growth rate at each condition. (c) Changing the axes in panel (a) to the glucose and oxygen uptake rates, paints a very different picture. Due to our model assumptions, only EFMs can maximize the growth rate at a given condition, and since each EFM defines a constant ratio of glucose to oxygen uptake, its points are all positioned along a straight line (scaled by the growth rate). Therefore, this scatter plot is very sparse. (d) the same data as in (c), shown as a 3D plot where the z-axis is growth rate. Again, the points of each EFM are aligned since all rates scale linearly with the growth rate.

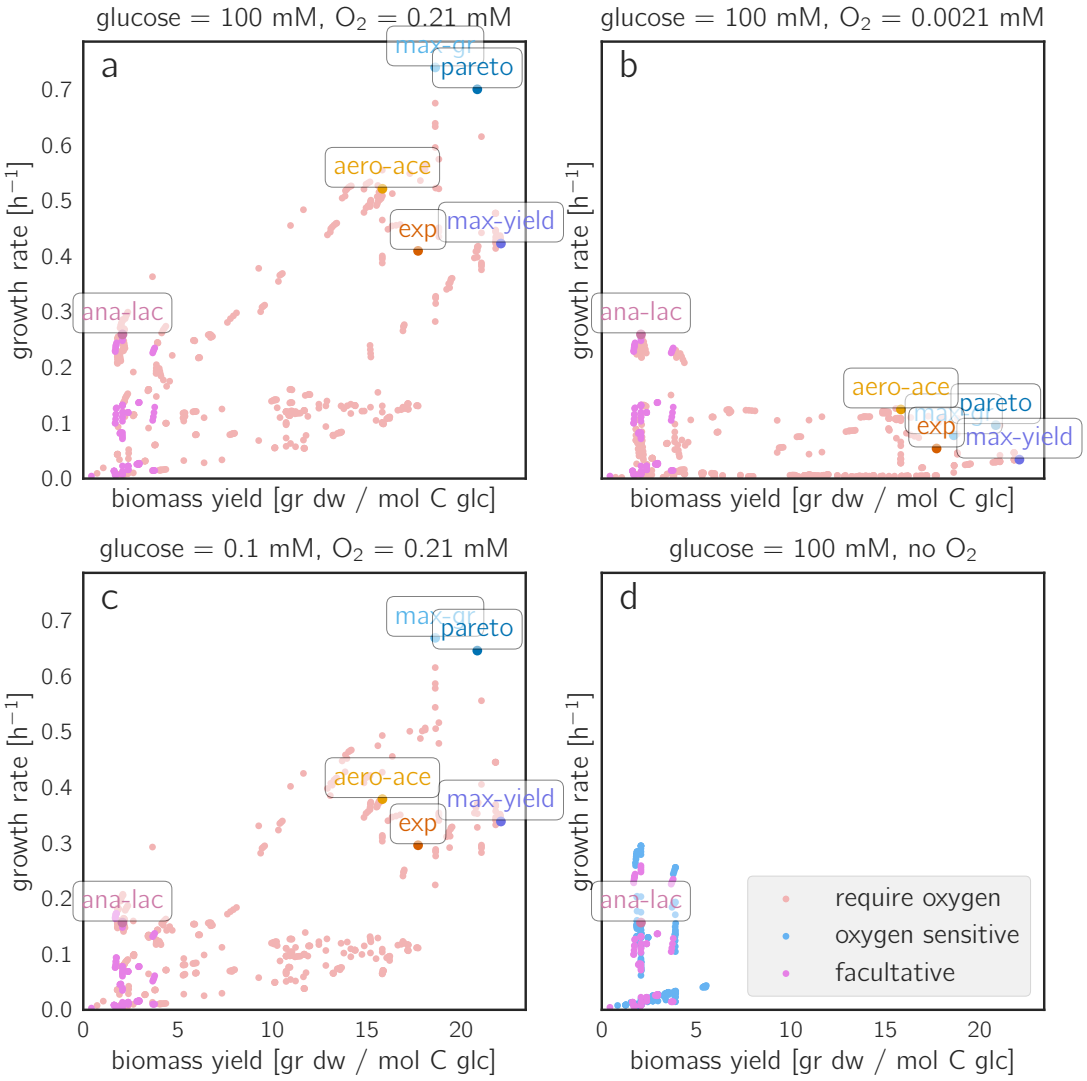

Figure S14: **Growth rate and biomass yield of all EFMs under standard conditions (a), low oxygen conditions (b), low glucose conditions (c), no oxygen (d).** Decreased glucose and oxygen levels lead to a growth decrease in many of the EFMs. At the same time, different EFMs may become optimal, and a wide Pareto front consisting of diverse EFMs may emerge, such as in panel (b).

S3.3 Enzyme investment in selected elementary flux modes

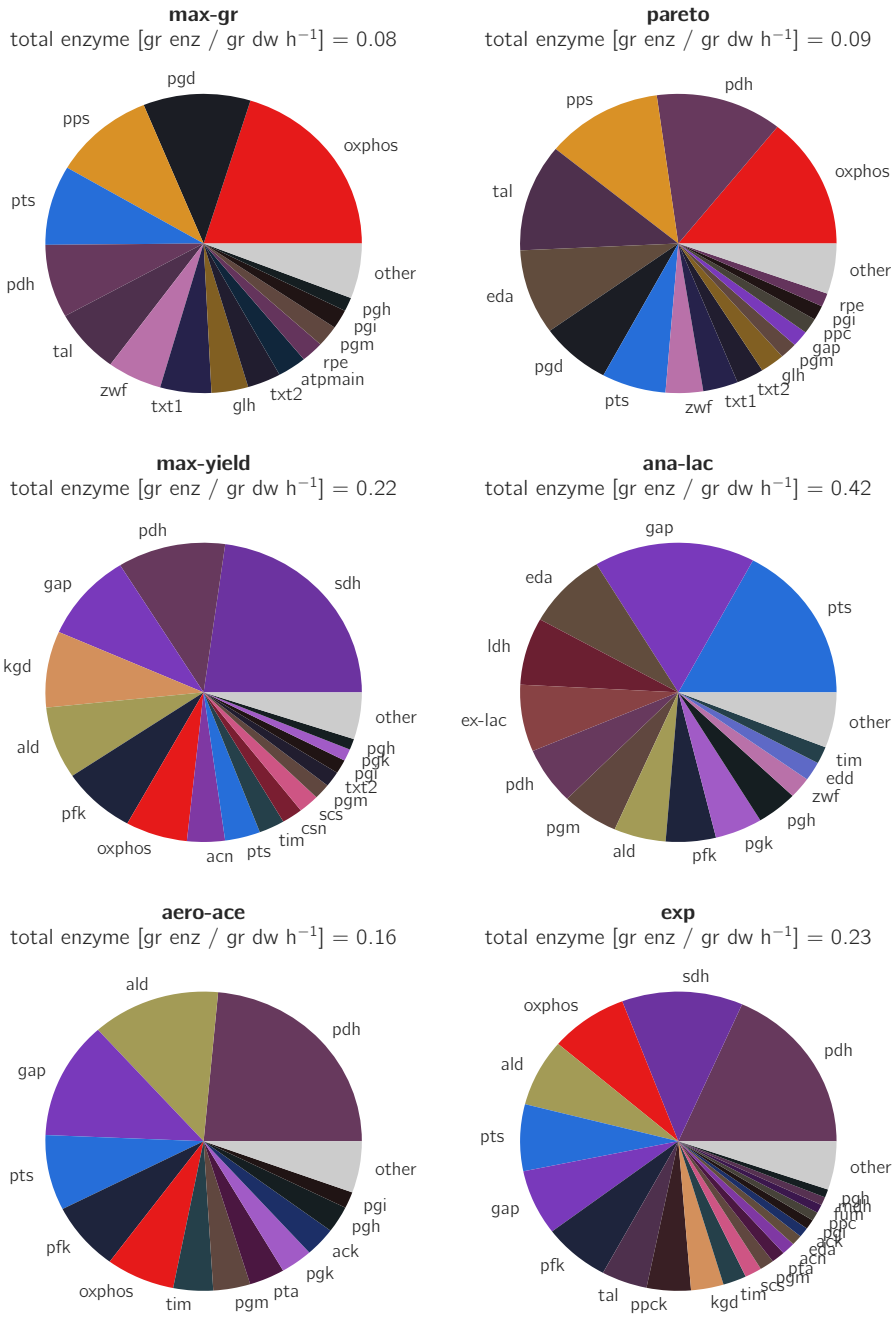

Figure S15: **Breakdown of protein cost for selected EFMs (standard concentration conditions).** The reactions that require a lot of enzyme investment are: oxidative phophorylation (**R80**), citrate synthase (**R21**), glucose uptake PTS system (**R1**) and glyceraldehyde-3P dehydrogenase (**R7ra**).

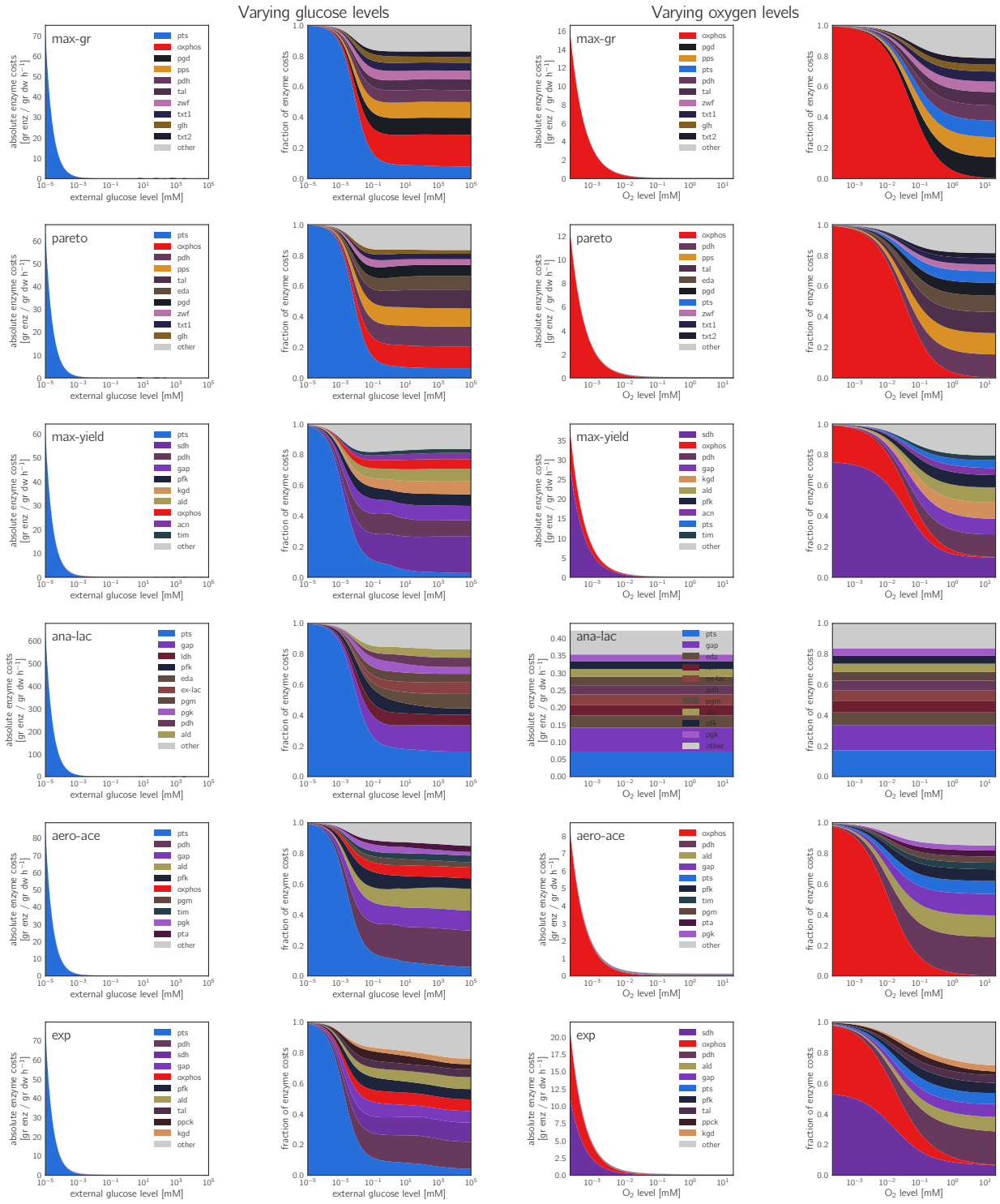

Figure S16: **Breakdown of protein cost for varying external concentrations.** Varying glucose concentration (two left columns). Allocation of resources for reactions that require more than 5% of the total cost, in absolute concentration (left) or as a fraction of the total cost (right). The allocations going to all other reactions are lumped together in the grey area denoted "other". Only the six focal EFMs are shown. The two right columns show the same for varying oxygen concentration.

S3.4 Trade-off between growth under high-glucose or low-glucose conditions

Cells in a chemostat, at different dilution rates, face different metabolic challenges. Can we expect that cells use the same metabolic strategies at a wide range of growth rates, or would we rather expect a trade-off, by which some strategies perform better at low growth rates and others perform better at high growth rates? To answer this question, Monod curve parameters (Monod coefficient and maximal growth rate) were calculated for all EFMs by fitting a Hill function to the estimated growth rates across a wide range of external glucose concentrations. The Monod curve is typically characterized by the formula $\mu = \mu^{\max} \frac{[S]^h}{[S]^h + K_{\mathrm{S}}^h}$, where $\mu^{\max}$ is the maximal growth rate, $[S]$ is the concentration of the limiting nutrient (i.e. glucose) and K_{S} is the substrate saturation constant (or "Monod coefficient"). The Monod coefficient is equal to the concentration of glucose where the growth rate is exactly half of the maximum ($\mu = \frac{1}{2}\mu^{\max}$), and its reciprocal value can be seen as the cell's overall affinity for glucose. In a chemostat at high dilution rates cells must grow fast because they can only survive if their maximal growth rate exceeds the dilution rate; at low dilution rates, in contrast, the higher cell density leads to very low glucose levels, and cells with a high growth rate at low glucose concentrations will be selected for – typically the ones that have a low Monod constant.

To observe possible trade-offs between these two quantities, we used our model to estimate the growth rate of all EFMs in a wide range of glucose concentrations, either in aerobic or anaerobic conditions, and fitted the parameters of a Monod curve for each EFM separately. Plotting $\mu^{\max}$ versus the affinity $1/K_{\mathrm{S}}$ under aerobic conditions we do find a trade-off (Figure S17(a)). Under anaerobic conditions, a more pronounced trade-off develops (Figure S17(e)). The growth at low glucose concentration is a combination of the Monod constant and the maximal growth rate, and as shown in Figure S17(i), our simulations predict almost no trade-off as long as oxygen levels are high, but in anaerobic conditions (Figure S17(m)), the trade-off becomes more pronounced, suggesting different winning strategies depending on the dilution rate.

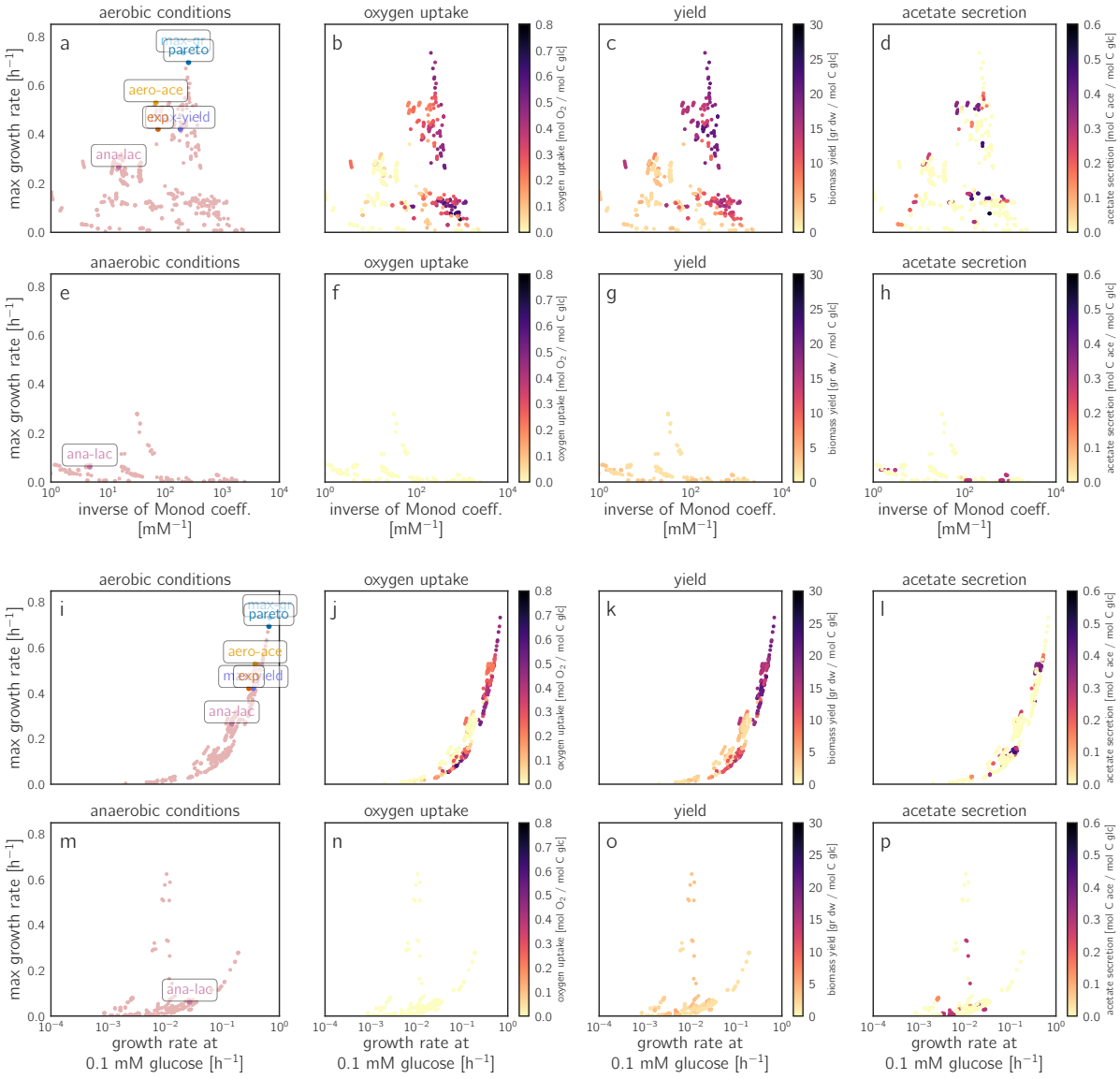

Figure S17: **Trade-off between growth under high-glucose or low-glucose conditions.** The trade-off is presented as a scatter plot of the maximal growth rate (y-axis) either as a function of the inverse Monod coefficient (panels a-h) or as a function of the growth rate at very low glucose levels (1 μM, panels i-p). Odd rows show the distribution of all EFMs in aerobic conditions ($[O_2]$ = 0.21 mM) and even rows are for anaerobic conditions (only EFMs that do not consume O_2 are shown). The first column (left) displays all the EFMs, and the three other columns show the oxygen uptake, biomass yield and acetate secretion rate respectively. One can appreciate, that in aerobic conditions there is almost no trade-off between the maximal growth rate and the growth on low glucose concentrations. In other words, the best EFMs in high glucose levels are often the best also in low glucose.

S3.5 Effect of an individual enzyme parameter on cell growth

If the catalytic constant of an enzyme is decreased, all EFMs using this enzyme will show lower growth rates because in each EFM, in order to maintain the predefined flux mode, the lower catalytic efficiecy must be compensated by higher enzyme levels (see proof in SI section S4.4); as a consequence, the biomass production per enzyme investment decreases, and so does the growth rate. Figure S18 shows the simulated effects of a changing catalytic constant of triose-phosphate isomerase. In contrast to the case in standard conditions (light grey dots, same as in main text Figure 2b), the Pareto front now comprises a few EFMs with very different growth rates, but similar yields. The growth rate of *max-yield*, which uses *tpi*, is greatly reduced, while *max-gr* is not affected since it has no flux through *tpi* at all. The sensitivities shown in Figure S18 (b) correspond to the slope of the graph in Figure S18 (c) at the standard k_{cat} value. The grey shading marks the range of a two-fold increase or decrease around the standard k_{cat}. One can see that the linear approximation is not suitable for much larger changes in k_{cat}, such as the change shown in panel a (i.e. a 1000-fold decrease to "low k_{cat}"). The calculation of cost sensitivities is explained in more detail in sections S4.2 and S4.3. Figure S18c shows the results of a global enzyme optimization after parameter changes. Therefore, focusing on relatively small parameter changes (which allow us to ignore effects higher than first order) and given the optimal enzyme profile for some EFM under some standard conditions, the parameter sensitivities can be computed easily (as described in SI section S4.3 for k_{cat} values, equilbrium constants, and K_{M} values).

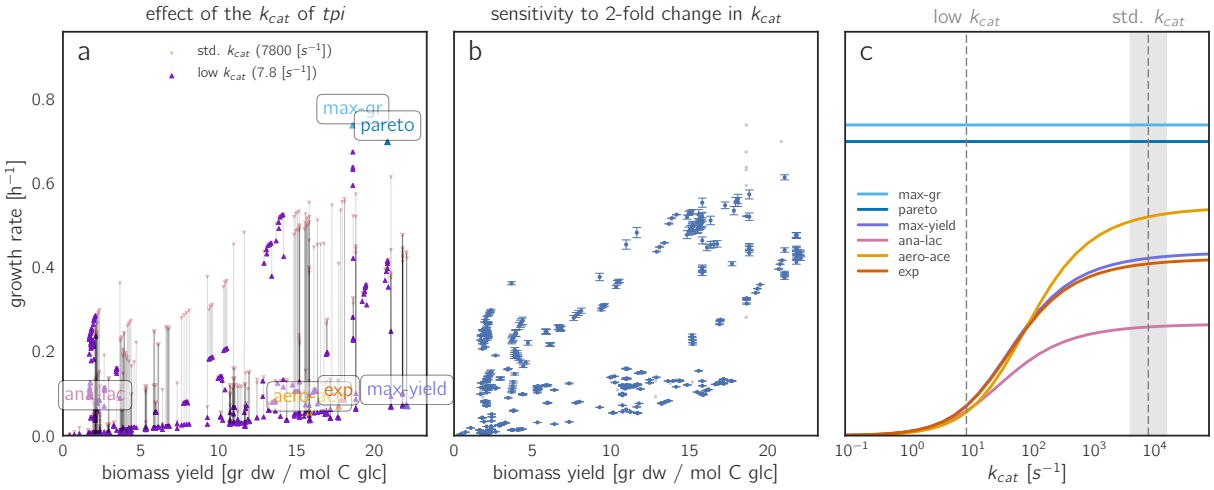

Figure S18: **Growth effects of a varying enzyme parameter.** (a) Growth/yield diagram in which the catalytic constant of triose-phosphate isomerase (*tpi*) is decreased by a factor of 1000 (from its original value 7800 s^{-1} to 7.8 s^{-1}). (b) Sensitivity of the biomass production to the k_{cat} of *tpi* for different EFMs. Error bars depict the change in growth rate for a two-fold increase or decrease in the k_{cat} value. An approximation based on an direct compensation of the affected enzyme was used to compute the sensitivity (see SI section S4.3). (c) Effect of an increased triose-phosphate isomerase k_{cat} value on the growth rate of selected EFMs. The 2-fold range around the "standard k_{cat}" value is marked by grey shading. The top two EFMs *max-gr* and *pareto* do not use *tpi* at all and are not sensitive to changes in its k_{cat}.

S3.6 Growth of strains deficient in EMP glycolysis, ED glycolysis, or respiration

Our analysis of growth-yield trade-offs extends an earlier study of cost/yield trade-offs in ATP regeneration. Flamholz *et al.* had studied cost-yield trade-offs between two versions of glycolysis, the Embden-Meyerhof-Parnas (EMP) and the the Entner-Doudoroff (ED) pathway [16]. Their model comprised only glycolysis, and ATP yield and enzyme cost (both at a fixed glucose uptake rate) were compared. The EMP pathway, which provides a two-fold yield, requires much higher amounts of protein per ATP flux, which entails a cost-yield trade-off. The authors concluded that cells, under a high demand for ATP produced in glycolysis, should use the EMP pathway, while cells that already produce "cheap" ATP (e.g., by photosynthesis), should rather use the ED pathway.

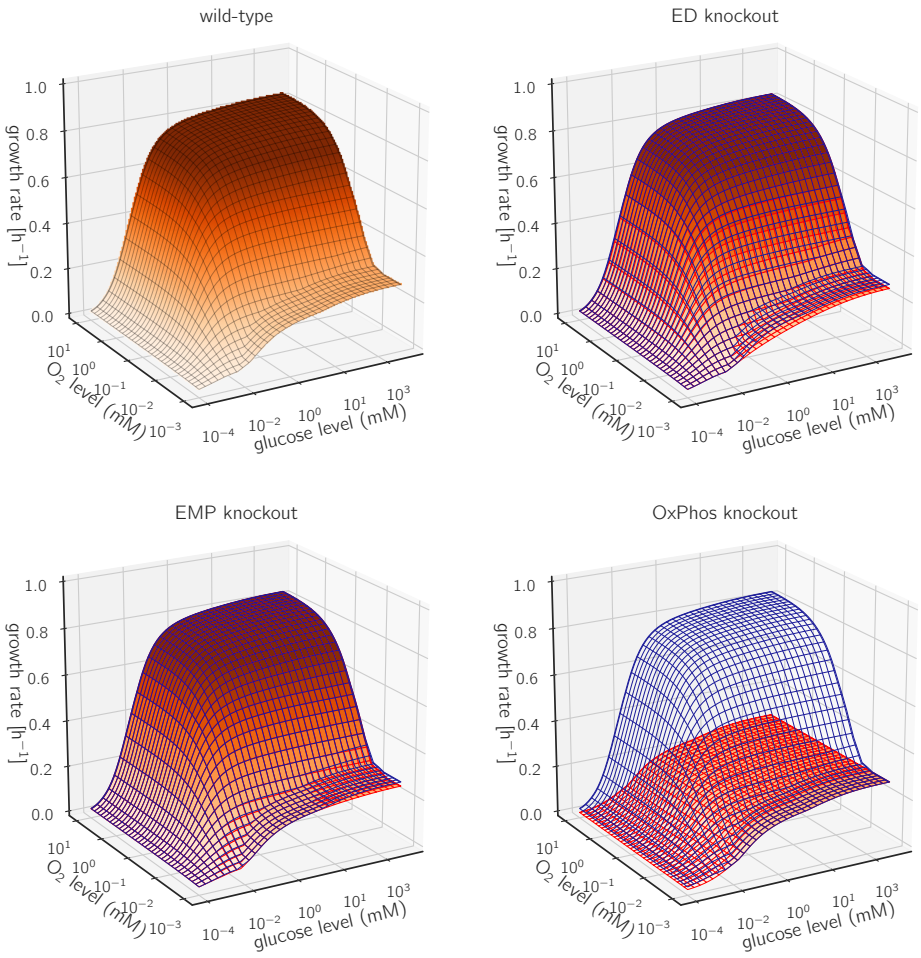

Figure S19: **Growth surface plots for strains deficient in EMP glycolysis, ED glycolysis, or respiration.** (a) Growth surface plot for simulated wild-type *E. coli*. (b) Changes in growth surface after simulated suppression of ED pathway (wild-type surface shown in blue, mutant surface shown in red). (c) Changes in growth surface after simulated suppression of EMP pathway. (d) Changes in growth surface after simulated suppression of respiration.

With our model, we can now revisit this question. Compared to the model in [16], our model has the following additional features: (i) the model includes the TCA cycle and respiration, so choices between fermentation and respiration and choices between EMP and ED pathways, which may be entangled, can be studied at the same time. (ii) Our model predict overall biomass yield and cell growth rate (instead of only ATP yield and ATP production cost in glycolysis) as output variables. (iii) We screen the behavior for various glucose and oxygen levels, each providing different possible enzyme costs and possible growth-yield

trade-offs.

To compare the usages of EMP and ED glycolysis, as well as the usage of respiration, we consider model versions in which some of these pathways are suppressed, i.e., all EFMs that include these pathways will be discarded[3]. If a metabolic pathway cannot be used, it may have to be by-passed by other, less efficient pathways. This will decrease the growth rate under all (glucose and oxygen) conditions in which the original pathway was used, and the quantitative decrease may again depend on the (glucose and oxygen) conditions. Here we compare our simulated wildtype *E. coli* (capable of using EMP glycolysis, ED glycolysis, and respiration) to variants in which one of the three pathways, respectively, is suppressed by simulated gene deletions (see Figures 6 and S19). The suppression of either the EMP or ED pathway have the most marked effects at low oxygen levels, where respiration is not used and cells rely entirely on glycolysis for their ATP production. Suppressing the respiration pathway has almost the same effect as setting the external oxygen concentration to extremely low values (except for the case of extremly low glucose concentrations).

In [16], it was hypothesized that non-respiring cells may be more dependent on ATP generated in glycolysis, which and may therefore employ the high-ATP-yield EMP pathway despite it higher enzyme cost. Here we find the opposite: at (low oxygen and) high glucose levels, EMP and ED pathways yield approximately the same growth rates, and at low to medium glucose levels, the ED pathway even performs better, possibly to ATP cost/yield ratio, which is still better than the cost/yield ratio of the EMP pathway – a conclusion that is in line with the simulation results, but not with the verbal conclusions from [16]. Note that our three model variants could also be seen as simple models of bacterial species that are lacking the genes for EMP glycolysis, ED glycolysis, or respiration, respectively.

S3.7 Epistatic effects between gene knock-outs

Once the growth rates and yields for all EFMs have been computed, simulating single, double, or multiple gene knockouts is really easy – we just need to exclude all EFMs that are affected by a knockout and redo the statistical analysis (e.g., finding the growth-maximising EFM, or determining the Pareto front). By comparing the performance of simulated single and double knockout strains, we can compute epistasis values, which quantify whether a double knockout has a less severe or more severe effect than we would have expected based on the separate single knockouts. Epistasis values can be defined based on growth rate or based on yield as the fitness objective.

Figure S20 show predicted epistasis values related to growth rate, instead of biomass yield as selection objective. Lethal knock-outs are the same as for the yield (see Figure S21). However, there are less cases of positive epistatic interactions. A second knock-out in the PPP (**R11**-**R15**) after a knock-out in lower glycosis (**R7**-**R8**) does give an extra reduction in the growth rate (no positive scaled epistasis), while it does not have a big effect on the yield. There is a positive epistasis in the growth rate for the 2 reactions in the ED pathway (**R60** and **R61r**), while this pathway is not used in high yield pathways (and therefore no epistasis there). Some of the epistatic interactions from the yield come back in the case of low oxygen, because high yield pathways are often also high growth rate pathways at low oxygen (panel d). Interesting is that some epistatic interactions change signs between different conditions, such as the positive epistasis between uptake (**R1**) with the ED pathway (**R60** and **R61r**) (panel b and f) which turns negative at low oxygen (panel d).

The epistatic effects on yield are shown in Figure S21. Panel (a) shows the relative yields of each double knockout as a fraction of the maximal yield (i.e. the yield of the wild-type). Single knock-outs can be seen on the diagonal. Obviously, the uptake reaction (**R1**) and biomass reaction (**R70**) are essential. As can be seen, a knock-out of reactions **R21**-**R24** and **R40** is lethal because there is no way to produce 2-oxogluterate,

[3]This analysis does not require any additional optimization runs. It suffices to analyse the existing simulation results, while discarding some of the EFMs.

Figure S20: **Epistatis effect of double knockouts on growth rate.** Relative growth rate (a,c,e) and "scaled growth rate epistatis" (b,d,f). Epistasis values are computed using Equation S19, but based on growth rates instead of biomass yields. The three rows refer to standard conditions (a-b), low oxygen conditions (c-d), and low glucose conditions (e-f) respectively. The relative growth rate of single reaction knock-outs can be seen on the diagonal (only on the left column).

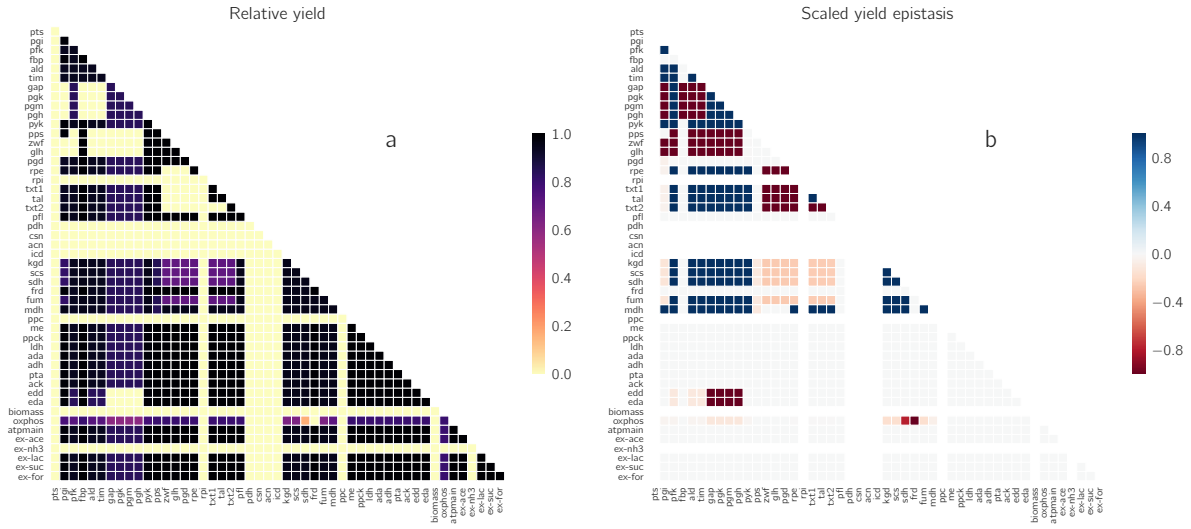

Figure S21: **Epistatic effect of double knockouts on biomass yield.** (a) Relative yields of each double knockout as a fraction of the maximal yield (i.e. the yield of the wild-type). (b) Scaled epistatis for biomass yields.

which is needed for the biomass reaction (although this is not directly obvious for reaction **R20** and **R40**). Other essential reactions are **R93** (to obtain ammonia) and R12r (to obtain ribose-5-P). Single knock-outs in oxidative phosphorylation (**R80**) or lower glycolysis (**R7**/**R8**) decrease the yield, because those are used by the high yield EFMs. Some double knock-outs are synthetically lethal—i.e. only the double knockout is lethal but each of the single ones are viable, mostly combinations of knock-outs in lower glycolysis, the ED-pathway and the pentose phosphate pathway. A few double knock-outs lower the yield dramatically, mostly combinations of the PPP and the TCA cycle. In panel (b), the "scaled yield epistatis" is calculated with the formula

$$\frac{Y_{1,2} - Y_1 \cdot Y_2}{|\widetilde{Y}_{1,2} - Y_1 \cdot Y_2|}\,, \tag{S19}$$

where Y_i is the scaled yield of knock-out i and $\widetilde{Y}_{1,2} = \min(Y_1, Y_2)$ if $\widetilde{Y}_{1,2} > Y_1 \cdot Y_2$ and $\widetilde{Y}_{1,2} = 0$ otherwise (definition from [17]). Here we can clearly see the synthetically lethal double knockouts and the double knockouts that have a dramatically lower yield in red. In blue (i.e. epistasis score = 1) we find combinations or knockouts where one is dominant, while the other doesn't reduce the yield further. For instance, this is the case for sequential reactions in the same pathway (e.g. **R10a** and **R10b**).

S4 Mathematical derivations

S4.1 Computing a stationary flux distribution in a metabolic network from measured fluxes

Assume we obtain absolute flux measurements using ^{13}C metabolic flux analysis (MFA), but the list of fluxes does not cover all the 52 reactions in our CCM model. We now describe a method to estimate the missing reaction fluxes (and maybe adjust the measured ones) in order to get a consistent flux, i.e. one that satisfies the constrains of the system include mass balance. One option is to find the flux mode (v) that minimizes the l_1 distance to the measured exchange fluxes ($\bar{v} \pm \sigma$), and fulfills mass balance constraints:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in I_{\text{meas}}} |v_i - \hat{v}_i| \\ Sv &= 0 \\ v_{min} &\leq v \leq v_{max} \\ \bar{v}_i - \sigma_i &\leq \hat{v}_i \leq \bar{v}_i + \sigma_i \end{aligned} \tag{S20}$$

where $\hat{v}$ is an auxiliary variable that represents the actual flux, and constrained to be within the confidence interval of the absolute measured fluxes. This LP typically has redundant solutions, therefore we add a secondary optimization goal which is to maximize the ATP production (in our model it means maximizing v_{r82}).

$$\begin{aligned} \text{maximize} \quad & v_{r82} \\ \sum_{i \in I_{\text{meas}}} |v_i - \hat{v}_i| &= \epsilon_1 \\ Sv &= 0 \\ v_{min} &\leq v \leq v_{max} \\ \bar{v}_i - \sigma_i &\leq \hat{v}_i \leq \bar{v}_i + \sigma_i \end{aligned} \tag{S21}$$

where ϵ_1 is the value of the minimum in the first optimization.

S4.2 The global cost sensitivities can be approximated by local cost sensitivities

We now compute the sensitivities between enzyme cost and model parameters, e.g., k_{cat} values. In analogy to the local and global flux sensitivities in Metabolic Control Analysis, called elasticities and flux response coefficients respectively, we distinguish local and global sensitivities for enzyme cost at predefined fluxes. From these sensitivities, we can compute local and global sensitivities between parameter perturbations and the cell growth rate. . To define the local sensitivities, we perturb one parameter and adapt only the enzyme of the affected reaction (while all metabolite levels remain unchanged, and the flux must remain the same). To define the global sensitivities, we perturb the same parameter and adapt all enzymes (and accordingly, all metabolite levels, where the flux must remain the same and the cost-optimal adaptation is chosen). Below we show that, for small parameter changes, local and global adaptation lead to the same first-order cost changes, i.e., that local and global sensitivities are identical.

To define the optimal enzyme cost q^{opt} and the optimal metabolite profile s_0 for a given flux vector v, we

solve the enzyme cost minimization problem

$$\begin{aligned} q^{\mathrm{opt}}(v,k) &= \min_s q(v,s,k) \\ s^{\mathrm{opt}}(v,k) &= \operatorname{argmin}_s q(v,s,k) \end{aligned} \tag{S22}$$

with an enzyme-based metabolite cost function $q(v,s,k)$ derived from a kinetic model. The fluxes v are fixed and given, the vector of logarithmic metabolite levels s can be varied within the metabolite polytope, and the vector k contains model parameters (e.g., kinetic constants) that affect individual reaction rates.

To show that local and global sensitivities are identical, we start from the unperturbed reference values k_0 and obtain an optimal metabolite vector s_0; then we expand the cost function quadratically around this point with respect to s. To simplify the notation (and without loss of generality), we consider a one-dimensional problem (with scalar logarithmic concentration s):

$$q(v,s,k) \approx a(s-s_0)^2 + \underbrace{b}_{0}(s-s_0) + c, \tag{S23}$$

with expansion coefficients a, b, and c. Since we expand around an optimum point, the coefficient b vanishes. Now we consider a small parameter change $\mathrm{d}k$, which changes the cost landscape (and thereby the expansion coefficients):

$$q(v,s,k+\mathrm{d}k) \approx [a+\alpha\,\mathrm{d}k]\,(s-s_0)^2 + [\beta\,\mathrm{d}k]\,(s-s_0) + [c+\gamma\,\mathrm{d}k]. \tag{S24}$$

We compute the new optimum point by equating the derivative to zero:

$$\begin{aligned} 0 &= 2[a+\alpha\,\mathrm{d}k]\,(s^{\mathrm{opt}*}-s_0) + [\beta\,\mathrm{d}k] \\ \Rightarrow s^{\mathrm{opt}*} &= s_0 - \frac{\beta\,\mathrm{d}k}{2(a+\alpha\,\mathrm{d}k)} \approx s_0 - \frac{\beta\,\mathrm{d}k}{2\,a} \end{aligned} \tag{S25}$$

where we assume that $a \geq 0$ (i.e., we assume that we started from a well-defined, unique optimum optimal). To compute the resulting cost, we insert $s^{\mathrm{opt}*}$ back into the perturbed cost function:

$$\begin{aligned} q^{\mathrm{opt}}(v,k+\mathrm{d}k) &= q(v,s^{\mathrm{opt}*}(k+\mathrm{d}k),k+\mathrm{d}k) \\ &\approx \underbrace{[a+\alpha\,\mathrm{d}k](-\frac{\beta\,\mathrm{d}k}{2\,a})^2 + [\beta\,\mathrm{d}k]\,(-\frac{\beta\,\mathrm{d}k}{2\,a})}_{\text{second-order terms in } \mathrm{d}k} + \underbrace{[c+\gamma\,\mathrm{d}k]}_{\approx q(v,s_0,k+\mathrm{d}k)} \end{aligned} \tag{S26}$$

This means:

$$\frac{\partial q^{\mathrm{opt}}(v,k)}{\partial k} = \frac{\partial q(v,s^{\mathrm{opt}}(k),k)}{\partial k} = \frac{\partial q(v,s,k)}{\partial k}|_{s=s_0} \tag{S27}$$

or, in other words: the cost change after a parameter perturbation (and with an optimal adaptation of all enzyme levels) is, to first order, given by the cost change that would ensue from adapting only the affected enzyme (and keeping all metabolite levels unchanged). Note that our first-order expansion holds only for small (relative) parameter changes.

S4.3 Cost sensitivities of kinetic constants

Cost sensitivities for k_{cat} values In the case of k_{cat} values, we obtain a simple formula for local cost sensitivities: With the rate law $v = E\ k\ f(s)$, with fixed flux v and metabolite levels s (and thus fixed f), we obtain $E_{\mathrm{adapt}} \cdot k_{\mathrm{adapt}} = E_{\mathrm{ref}} \cdot k_{\mathrm{ref}} = \frac{v}{f} = \text{const.}$. (where ref refers to the reference state, and adapt to

the perturbed state with adapted enzyme level). The adapted enzyme level reads $E_{\text{adapt}} = E_{\text{ref}} \frac{k_{\text{ref}}}{k_{\text{adapt}}}$ and its derivative reads $\frac{\mathrm{d}E_{\text{adapt}}}{\mathrm{d}k_{\text{adapt}}} = -\frac{E_{\text{ref}}\, k_{\text{ref}}}{k_{\text{adapt}}^2}$, and in the original reference state: (where $k_{\text{adapt}} = k_{\text{ref}}$):

$$\frac{\mathrm{d}E_{\text{adapt}}}{\mathrm{d}k_{\text{adapt}}} = -\frac{E_{\text{ref}}}{k_{\text{ref}}}$$

The local cost sensitivity (with enzyme cost weight h) thus reads

$$\frac{\mathrm{d}q}{\mathrm{d}k} = -h\, \frac{E_{\text{ref}}}{k_{\text{ref}}} = -\frac{q_{\text{ref}}}{k_{\text{ref}}}$$

.

As we saw before, these (first-order) local cost sensitivities are identical to the (first-order) global cost sensitivities.

Cost sensitivities for K_{eq} values To quantify the sensitivity of the total enzyme cost to an independent change in one of the K_{eq} value, we first calculate the elasticity

$$q = hv \cdot k_{\text{cat}}^{-1} \cdot \eta^{\text{sat}}(\mathbf{c})^{-1} \cdot \left[1 - Q(\mathbf{c}) K_{\text{eq}}^{-1}\right]^{-1} \tag{S28}$$

$$\begin{aligned} \frac{\partial q}{\partial K_{\text{eq}}} &= \frac{q}{K_{\text{eq}}} \cdot \frac{\partial \log q}{\partial \log K_{\text{eq}}} = -\frac{q}{K_{\text{eq}}} \cdot \frac{Q K_{\text{eq}}^{-1}}{1 - Q K_{\text{eq}}^{-1}} \\ &= -\frac{q}{K_{\text{eq}}} \cdot \frac{1}{K_{\text{eq}} Q^{-1} - 1} \end{aligned} \tag{S29}$$

where $Q(\mathbf{c})$ is the reaction quotient, and $\eta(\mathbf{c})$ represents the saturation and regulation terms (that do not depend on K_{eq}). Note that if we use the definition $\eta^{\text{thr}} = \left(1 - Q(\mathbf{c}) K_{\text{eq}}^{-1}\right)$, we can rewrite this result as

$$\frac{\partial q}{\partial K_{\text{eq}}} = -\frac{q}{K_{\text{eq}}} \cdot \left[\eta^{\text{thr}}(\mathbf{c})^{-1} - 1\right] \tag{S30}$$

Cost sensitivities for K_M values (reaction substrates) The expression for the sensitivity of the total enzyme cost to a certain K_M value is given by:

$$\begin{aligned} q &= h\; v\; k_{\text{cat}}^{-1}\; \eta^{\text{thr}}(\mathbf{c})^{-1} \cdot \left[\frac{\prod_j \frac{c_j}{K_{M,c_j}}}{\prod_j \left(1 + \frac{c_j}{K_{M,c_j}}\right) + \prod_k \left(1 + \frac{c_k}{K_{M,c_k}}\right) - 1}\right]^{-1} \\ &= h\; v\; k_{\text{cat}}^{-1}\; \eta^{\text{thr}}(\mathbf{c})^{-1} \cdot \left[\prod_j \left(\frac{K_{M,c_j}}{c_j} + 1\right) + \prod_j \left(\frac{K_{M,c_j}}{c_j}\right) \cdot \left(\prod_k \left(1 + \frac{c_k}{K_{M,c_k}}\right) - 1\right)\right] \end{aligned} \tag{S31}$$

where j is the index for $\mathbf{c}$ for all substrates and k for all products. We will calculate the sensitivity to changes in the Michaelis constant of a substrate s, which we denote in short $K_s \equiv K_{M,c_s}$. For this cost function, the

K_M sensitivity would be:

$$\begin{aligned}
\frac{\partial q}{\partial K_s} &= h\ v\ k_{\text{cat}}^{-1}\ \eta^{\text{thr}}(\mathbf{c})^{-1} \cdot \left[\frac{1}{c_s}\prod_{j\neq s}\left(\frac{K_{M,c_j}}{c_j}+1\right)+\frac{1}{c_s}\prod_{j\neq s}\frac{K_{M,c_j}}{c_j}\cdot\left(\prod_k\left(1+\frac{c_k}{K_{M,c_k}}\right)-1\right)\right] \\
&= \frac{q\ \eta^{\text{sat}}(\mathbf{c})}{K_s}\left[\frac{K_s}{c_s}\prod_{j\neq s}\left(\frac{K_{M,c_j}}{c_j}+1\right)+\frac{K_s}{c_s}\prod_{j\neq s}\frac{K_{M,c_j}}{c_j}\cdot\left(\prod_k\left(1+\frac{c_k}{K_{M,c_k}}\right)-1\right)\right] \\
&= \frac{q\ \eta^{\text{sat}}(\mathbf{c})}{K_s}\left[\prod_j\left(\frac{K_{M,c_j}}{c_j}+1\right)-\prod_{j\neq s}\left(\frac{K_{M,c_j}}{c_j}+1\right)+\prod_j\frac{K_{M,c_j}}{c_j}\cdot\left(\prod_k\left(1+\frac{c_k}{K_{M,c_k}}\right)-1\right)\right] \\
&= \frac{q\ \eta^{\text{sat}}(\mathbf{c})}{K_s}\left[\eta^{\text{sat}}(\mathbf{c})^{-1}-\prod_{j\neq s}\left(\frac{K_{M,c_j}}{c_j}+1\right)\right] \\
&= \frac{q}{K_s}\left[1-\eta^{\text{sat}}(\mathbf{c})\prod_{j\neq s}\left(\frac{K_{M,c_j}}{c_j}+1\right)\right] \qquad \text{(S32)}
\end{aligned}$$

and if s is the only substrate for this reaction, we are left with

$$\frac{\partial q}{\partial K_s} = \frac{q}{K_s}\left[1-\eta^{\text{sat}}(\mathbf{c})\right]$$

Cost sensitivities for K_M values (reaction products) We again start with equation S31 and derive by $K_p \equiv K_{M,c_p}$ – the Michaelis constant to one of the products. In this case the sensitivity is given by:

$$\begin{aligned}
\frac{\partial q}{\partial K_p} &= -\frac{q}{K_p}\cdot\eta^{\text{sat}}(\mathbf{c})\cdot\prod_j\frac{K_{M,c_j}}{c_j}\cdot\prod_k\frac{c_k}{K_{M,c_k}}\cdot\prod_{k\neq p}\left(\frac{K_{M,c_k}}{c_k}+1\right) \\
&= -\frac{q}{K_p}\cdot\eta^{\text{sat}}(\mathbf{c})\cdot Q(\mathbf{c})\cdot\frac{\prod_j K_{M,c_j}}{\prod_k K_{M,c_k}}\cdot\prod_{k\neq p}\left(\frac{K_{M,c_k}}{c_k}+1\right)\ . \qquad \text{(S33)}
\end{aligned}$$

and in case there is only one substrate and one product, this simplifies to

$$\frac{\partial q}{\partial K_p} = -\frac{q}{K_p}\cdot\frac{\prod_k\frac{c_k}{K_{M,c_k}}}{1+\prod_k\frac{c_j}{K_{M,c_j}}+\prod_k\frac{c_k}{K_{M,c_k}}} \qquad \text{(S34)}$$

S4.4 Increasing a k_{cat} value will never force the cell to invest more enzyme

Theorem 1 *A local increase in one enzyme's k_{cat} can never increase the minimal total enzyme cost required*

Proof 1 *Let $q^{\text{opt}}(v,k) \equiv min_s\ q(v,s,k)$ be the minimal enzyme cost (over all metabolite level profiles s) for a given flux vector v and set of kinetic parameters k. Let $\hat{k}$ be another set of kinetic parameters which is identical to k except for the k_{cat} of a single enzyme i, specifically $\hat{k}_{\text{cat},i} > k_{\text{cat},i}$. Now, for any metabolite profile s, if we compare $q(v,s,\hat{k})$ to $q(v,s,k)$ the only change would be the cost associated with that one enzyme, so*

$$q(v,s,\hat{k}) - q(v,s,k) = q_i(v_i,s,\hat{k}) - q_i(v_i,s,k) = \frac{h_i v_i}{\hat{k}_{\text{cat},i}\ f(s)} - \frac{h_i v_i}{k_{\text{cat},i}\ f(s)} = \frac{h_i v_i}{f(s)}\left(\hat{k}_{\text{cat},i}^{-1} - k_{\text{cat},i}^{-1}\right) \leq 0\ .$$

Therefore, $\forall s\ \ q(v,s,\hat{k}) \leq q(v,s,k)$, and this inequality will also trivially apply to the minima of both functions.

S5 Model details and elementary flux mode statistics

S5.1 Tables with model details

Original reaction	New reaction(s)	Reaction formula
R7r	R7ra	GA˙3P + NAD = DPG + NADH
	R7rb	DPG + ADP = PG3 + ATP
	R7rc	PG3 = PG
R10	R10a	GLU˙6˙P + NAD = GLULAC6P + NADH
	R10b	GLULAC6P = GLUCO6P
	R10c	GLUCO6P + NAD = NADH + CO2 + RUBILOSE˙5˙P
R27r	R27	SUCC + ADP + OXYext = FUMURATE + ATP
R81		merged with R27
R27r	R27b	FUMERATE + NADH = SUCC + NAD
R83		merged with R27b
R54r	R54ra	ACETYL˙CoA + NADH = ACALD + NAD + CoASH
	R54rb	ACALD + NADH = ETOH + NAD
R55r	R55a	ACETYL˙CoA + ADP = ACETYLP + CoASH
	R55b	ACETYLP + ADP = ACETATE + ATP
new	R60	GLUCO6P = KDGP
new	R61r	KDGP = GA3P + PYR

Table S2: Changes in reactions compared to the original model in Carlson and Srienc [7].

Name	KEGG ID	Stoichiometry	Molecular Wt. $[\mathrm{mg\ mmol^{-1}}]$	Weight in biomass $[\mathrm{mg\ mmol^{-1}}]$	#Carbons	#Carbons in biomass
acetyl-CoA	C00024	-41	59	-2419	2	-82
α-ketoglutarate	C00026	-14	146	-2044	5	-70
CO2	C00011	2	44	88	1	2
DHAP	C00236	-5	266	-1330	3	-15
glucose-6P	C00345	-4	180	-720	6	-24
NH_3	C00014	-139	17	-2363	0	0
oxaloacetate	C00007	-24	132	-3168	4	-96
PEP	C00631	-32	88	-2816	3	-96
pyruvate	C00022	-38	88	-3344	3	-11
ribose-5P	C00117	-13	150	-1950	5	-65
erythrose-4P	C00279	-5	120	-600	4	-20
TOTAL				**-20666**		**-580**

Table S3: Breakdown of biomass-producing reaction

S5.2 Selected elementary flux modes

Acronym*	EFM #	Biomass yield [g/C-mol]	Growth rate $[h^{-1}]^{\dagger}$	Oxygen uptake	Acetate secre-tion	Lactate secre-tion	Succinate secre-tion	fraction PPP	# active reactions
max-gr	1565	18.6	0.739	0.49	0	0	0	2.77	24
pareto	1218	20.8	0.699	0.42	0	0	0	2.50	25
max-yield	938	22.1	0.422	0.39	0	0	0	0	28
ana-lac	1295	2.1	0.258	0	0	0.92	0.02	0.90	31
aero-ace	559	15.8	0.520	0.21	0.35	0	0	0.11	28
exp	9999	17.7	0.409	0.29	0.22	0	0	0.39	38

*max-gr: maximum growth rate; max-yield: maximum yield; ana-lac: anaerobic lactate fermentation; aero-suc: aerobic succinate fermentation; aero-ace: aerobic acetate fermentation; exp: experimentally measured flux distribution †Growth rate is given for standard conditions, where [glucose] = 100 mM, and $[O_2]$ = 0.21 mM.

Table S4: Details on selected EFMs representing different growth strategies. Metabolic fluxes are given in carbon moles (or O_2 moles) per carbon mole of glucose taken up. The flux fraction through the pentose phosphate pathway (PPP) is defined as the ratio R1/R10a (see Supplementary Figure S4 for the reaction numbers in the network).

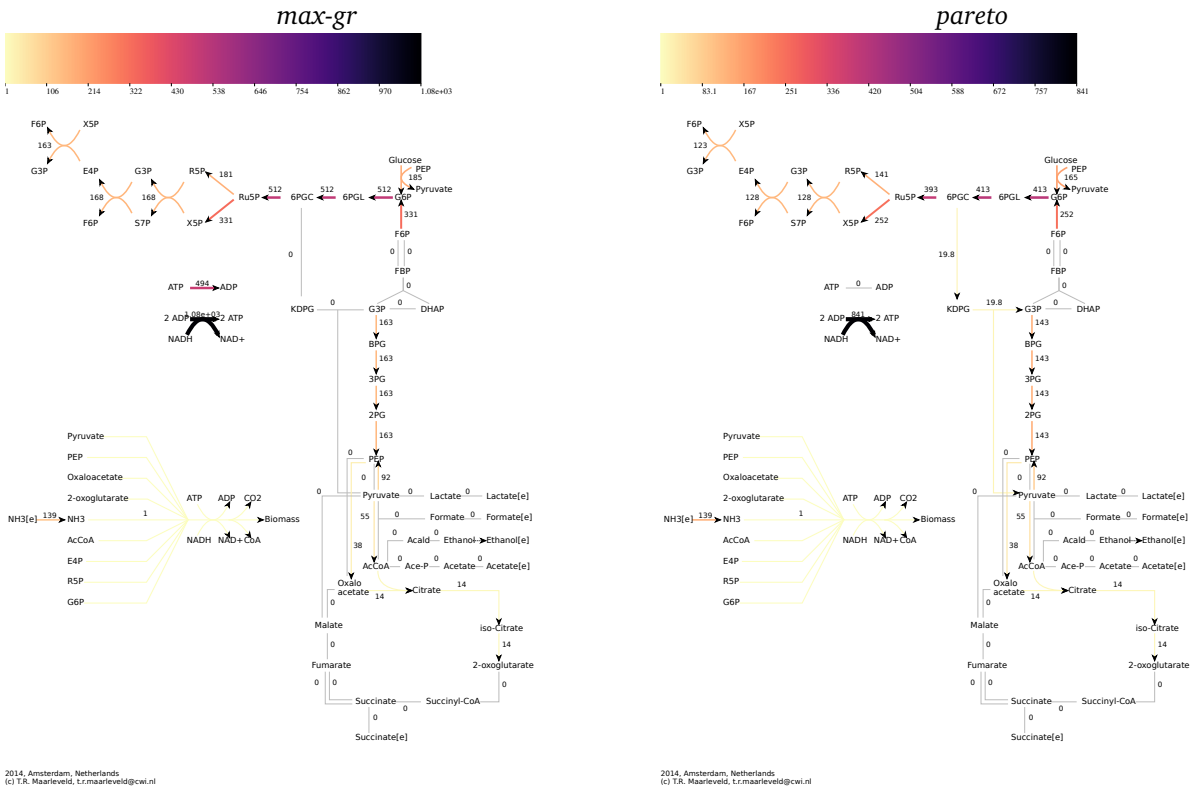

Figure S22: **The flux distributions of *max-gr* (EFM #1565) and *pareto* (EFM #1218).**

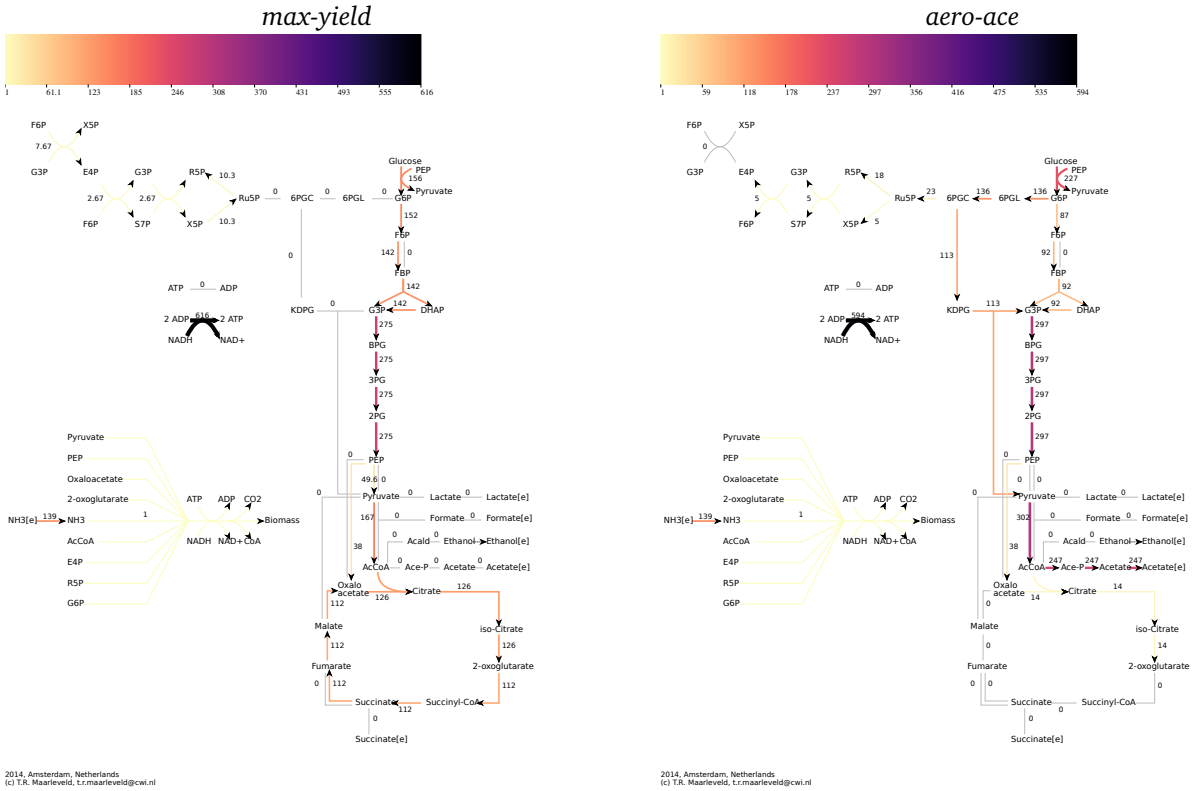

Figure S23: **The flux distributions of *max-yield* (EFM #938) and *aero-ace* (EFM #1155).**

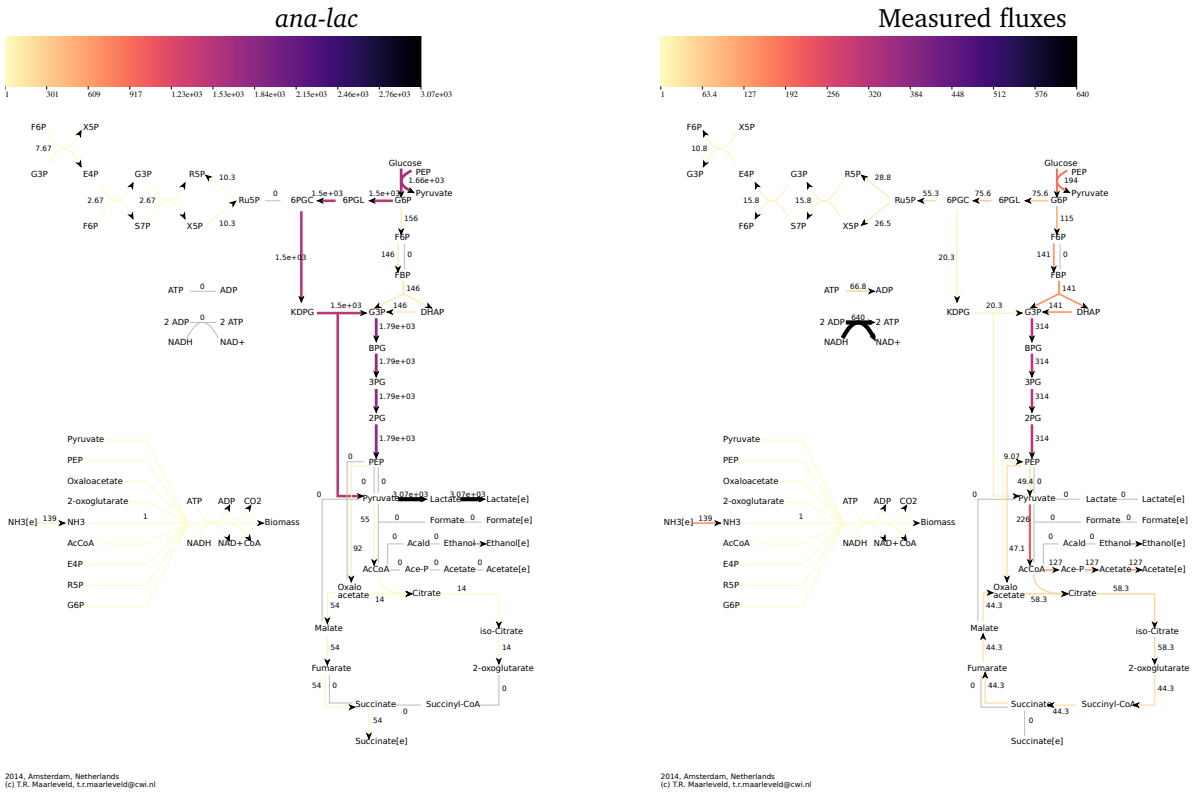

Figure S24: **The flux distributions of *ana-lac* (EFM #1295) and *exp* (flux distribution adjusted to measured fluxes).**

S6 List of supplementary files

Supplementary files describing the *E. coli model* and the analysis of metabolic strategies therein can be downloaded from github: `https://github.com/liebermeister/flux-enzyme-cost-minimization`.

- `Ecoli_Central_Metabolism.xml` – SBML file of the central metabolism network of *E. coli*
- `Model_Structure.xlsx` – reactions, metabolites, stoichiometry matrix and elementary flux modes
- `Kinetic_Parameters.xlsx` – kinetic parameters obtained from literature with references
- `SBtab_Literature_data.tsv` – the (averages of the) literature data in a parameter balancing input file
- `SBtab_Output_Parameter_Balancing.tsv` – kinetic parameter values after parameter balancing
- `Input_File_GAMS.csv` – single file GAMS input for the optimization of enzyme levels
- `Input_Files_GAMS.zip` – multi file GAMS input for the optimization of enzyme levels
- `GAMS_run_files.zip` – GAMS code files
- `Sensitivities_kcat.csv` – sensitivities of the k_{cat} values under standard conditions
- `Sensitivities_Keq.csv` – sensitivities of the k_{eq} values under standard conditions
- `Sensitivities_KM.csv` – sensitivities of the K_M values under standard conditions

References

[1] T. Lubitz, M. Schulz, E. Klipp, and W. Liebermeister. Parameter balancing for kinetic models of cell metabolism. *J. Phys. Chem. B*, 114(49):16298–16303, 2010.

[2] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. The general algebraic modeling system. *GAMS Development Corporation*, 1998.

[3] W. Liebermeister, J. Uhlendorf, and E. Klipp. Modular rate laws for enzymatic reactions: thermodynamics, elasticities, and implementation. *Bioinformatics*, 26(12):1528–1534, 2010.

[4] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. Bionumbers – the database of key numbers in molecular and cell biology. *Nucleic Acids Res.*, 38(Database issue):D750–D753, 2010.

[5] K. Valgepea, K. Adamberg, A. Seiman, and R. Vilu. Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst*, 9(9):2344–23458, 2013.

[6] M. Scott, C.W. Gunderson, E.M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330:1099, 2010.

[7] R. Carlson and F. Srienc. Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions. *Biotechnology and bioengineering*, 85(1):1–19, 2004.

[8] M. Terzer and J. Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19):2229–2235, 2008.

[9] J.S. Hofmeyr, O.P.C. Gqwaka, and J.M. Rohwer. A generic rate equation for catalysed, template-directed polymerisation. *FEBS Letters*, 587:2868–2875, 2013.

[10] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D.S. Tawfik, and R. Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–4410, 2011.

[11] L. R. Bakken and R. A. Olsen. Buoyant densities and dry-matter contents of microorganisms: conversion of a measured biovolume into biomass. *Appl. Environ. Microbiol.*, 45(4):1188–1195", apr 1983.

[12] P. Millard, K. Smallbone, and P. Mendes. Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in escherichia coli. *PLOS Computational Biology*, DOI:10.1371/journal.pcbi.1005396, 2017.

[13] L.J.P. van der Maaten and G.E. Hinton. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[14] L. Gerosa, B.R.B. Haverkorn van Rijsewijk, D. Christodoulou, K. Kochanowski, T.S.B. Schmidt, E. Noor, and U. Sauer. Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Systems*, 1(1):1, 2015.

[15] A. Schmidt, Kochanowski K, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nat. Biotechnol.*, 34(1):104–110, 2016.

[16] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *PNAS*, 110(24):10039–10044, 2013.

[17] D. Segre, A. DeLuna, M.G. Church, and R. Kishony. Modular epistasis in yeast metabolism. *Nature Genetics*, 37:77–83, 2005.