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

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

Abstract

Microbes in fragmented environments profit from yield-efficient metabolic strategies, which allow for a maximal number of cells. In contrast, cells in well-mixed, nutrient-rich environments need to grow and divide fast to out-compete others. Paradoxically, a fast growth can entail wasteful, yield-inefficient modes of metabolism and smaller cell numbers. Therefore, general trade-offs between biomass yield and growth rate have been hypothesized. To study the conditions for such rate-yield trade-offs, we considered a kinetic model of *E. coli* central metabolism and determined flux distributions that provide maximal growth rates or maximal biomass yields. In the model, maximal growth rates or yields are achieved by sparse flux distributions called elementary flux modes (EFMs). We screened all EFMs in the network model and computed the biomass yields and growth rates enabled by these EFMs. Growth rates were computed from the amount of protein required to sustain a given biomass production, computed from the kinetic model by enzyme cost minimization (ECM). In a scatter plot between the growth rates and yields of all EFMs, a trade-off shows up as a Pareto front. At reference glucose and oxygen levels, we find that the rate-yield trade-off is almost negligible. However, in low-oxygen environments, a clear trade-off emerges: low-yield fermentation EFMs allow for a growth 2-3 times faster than the maximal-yield EFM. The trade-off is therefore strongly condition-dependent and should be almost unnoticeable at high oxygen and glucose levels, the typical conditions in laboratory experiments. Our public web service `www.neos-guide.org/content/enzyme-cost-minimization` allows users to run ECM to compute enzyme costs for metabolic models flux distributions of their choice.

Introduction

Metabolic networks, their dynamics, and their regulation are shaped by evolutionary selection. When nutrients are in excess and the environment is well mixed, fast-growing bacterial cells will outcompete others. Under this selection pressure, organisms should evolve to maximize their growth rate. Indeed, microbiologists use the terms *growth rate* and *fitness* almost synonymously. As much as such well-mixed rich environments are common in laboratory settings, natural environments are much more diverse. In fragmented ecological niches with limited resources, the is no direct competition during the growth phase, and fecundity determines the evolutionary success regardless of growth speed. This puts a selection pressure on biomass yield rather than growth because the number of offspring for bacteria in a limited-nutrient environment is directly proportional to the biomass yield of their catabolism (for a given cell size).

At first glance, high growth rate and yield could be expected to go hand in hand: image a cell that can produce more offspring for the same amount of nutrient (i.e., higher yield); then it seems logical that it would reproduce fast (i.e., produce more offspring per hour). However, this is not what we see in experiments. Many fast-growing cells employ low-yield metabolic pathways, e.g., bacteria that, when grown on glucose, display respiro-fermentative metabolism at high growth rates even though a complete respiratory growth would have a higher yield per mole of glucose. Similarly, yeast cells that produce ethanol (Crabtree effect) and cancer cells that product lactate (Warburg effect) in the presence of oxygen seem to waste much of the carbon that they take up (see [1] for a review of these strategies and hypotheses). These yield-inefficient strategies observed in completely unrelated organisms have lead to the suggestion that fast growth and high cell yield may even exclude each other due to physico-chemical reasons (e.g. following thermodynamic principles [2, 3]). This hypothesis has been supported by simple cell models, in which lower thermodynamic forces or higher enzyme costs in the high-yield pathways caused a rate/yield trade-off.

In [4], two versions of the glycolysis, both common among bacteria, have been compared in terms of ATP yield on glucose and by of enzyme demand (or, equivalently, their ATP production rate at a given enzyme investment). At a given glucose influx, the Embden-Meyerhof-Parnas (EMP) pathway yields twice as much ATP, but requires about 4.5 as much enzyme than the Entner-Doudoroff (ED) pathway. Thus, it was hypothesized that cells under yield selection will use the EMP pathway while those under rate selection would use the ED pathway. The economics of other metabolic choices, e.g. respiration *versus* fermentation, and the resulting trade-offs, remain to be better quantified (approximations for yeast [5] and *E. coli* [6]).

Several lab-evolution experiments with fast-growing microorganisms have been conducted to bring the rate/yield hypothesis to the test, with varying levels of success. Growth rate and yield of microbial strains have been compared between different wild-type and evolved strains [8–11]. Most of these studies found poor correlations between growth rate and yield. Novak et al. [9] found a negative correlation within evolved *E. coli* populations, indicating a growth/yield trade-off. One of the few examples of bacteria evolving for high yield in the laboratory was the work of Bachmann et al. [12]. In their protocol, each cell is kept in a separate droplet in a medium-in-oil suspension, simulating a fragmented environment, and the offspring are mixed only after the nutrients in each droplet are depleted. This creates a strong selection pressure for maximizing biomass yield. Indeed, the strains gradually evolved towards higher yields at the expense of their growth rate, again indicating a trade-off between the two objectives. However, most of the evidence relies on laboratory experiments in which microorganisms may behave sub-optimally, and the existence of rate-yield trade-offs remains debatable.

How can we study growth/yield trade-offs by models? A major difficulty is the prediction of cell growth rates and their dependence on the metabolic state. For exponentially growing cells, the rate of biomass synthesis per cell dry weight – typically measured in grams of biomass per gram cell dry weight per hour – is equal to the specific growth rate μ. At balanced growth, the relative amounts of all cellular components

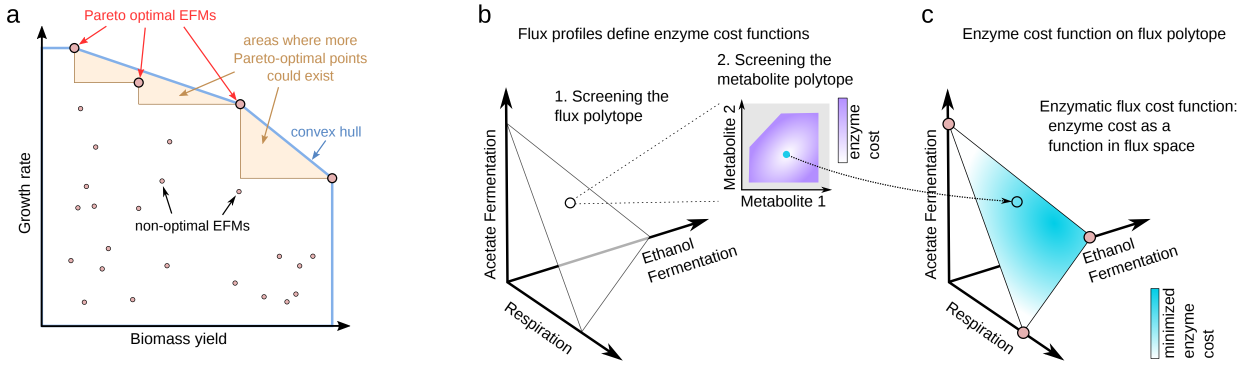
a
Pareto optimal EFMs
areas where more Pareto-optimal points could exist
convex hull
non-optimal EFMs
Growth rate
Biomass yield
b
Flux profiles define enzyme cost functions
1. Screening the flux polytope
2. Screening the metabolite polytope
Metabolite 2
Metabolite 1
enzyme cost
Acetate Fermentation
Ethanol Fermentation
Respiration
c
Enzyme cost function on flux polytope
Enzymatic flux cost function: enzyme cost as a function in flux space
Acetate Fermentation
Ethanol Fermentation
Respiration
minimized enzyme cost

Figure 1: **Growth/yield trade-offs in metabolism and the prediction of growth-optimal fluxes.** (a) Pareto front in the growth/yield scatter plot (schematic drawing). Elementary flux modes (EFMs) are represented by points, indicating their biomass yield and the maximal achievable growth rate (in a given simulation scenario). The Pareto-optimal EFMs define a convex hull within which all fluxes must reside [7] (mock data). Although there can be non-elementary, Pareto-optimal flux modes, they must lie within a narrow area defined by the Pareto-optimal EFMs inside the convex hull (light brown area). (b) Computing the cost of metabolic fluxes and finding optimal flux modes. In a hypothetical example model, the space of metabolic flux distributions is spanned by three EFMs. Each point in this space represents a stationary flux mode. Flux modes with a fixed benefit (predefined biomass production rate) form a triangle. Each flux mode defines an optimality problem for enzyme cost as a function of metabolite levels. The enzyme cost as a function on the metabolite polytope is shown as an inset graphics. Minimizing this enzyme cost is a convex optimality problem. By solving this problem, we obtain the optimal metabolite and enzyme levels and the minimal enzyme cost for the chosen flux distribution. (c) By evaluating the minimal enzyme cost in every point of the flux polytope, we obtain a kinetic flux cost function. Since this function is concave, at least one one of the polytope vertices (i.e., an EFM) must be cost-optimal. The search for optimal flux modes is this reduced to a screening of EFMs.

are preserved over time, including the protein fraction associated with enzymes that catalyze central carbon metabolism. If a metabolic strategy achieved the same biomass synthesis rate, but with a lower cost in terms of total enzyme mass, evolution would have the chance to reallocate the freed protein resources to other cellular processes that contribute to growth, and thus increase the cell's growth rate. Thus, a growth-optimal strategy will be one that minimizes enzyme cost (at a given rate of glucose-to-biomass conversion). This drive for low enzyme or nutrient investments should be reflected in the choice of metabolic strategies: if low-yield pathways provide higher metabolic fluxes per enzyme investment, this leads to a growth advantage [13–15]. Thus, the rate-yield trade-off in cells reflects a trade-off between *enzyme efficiency* and *substrate efficiency* in metabolic pathways. On the contrary, if cells can choose between a substrate-efficient and an enzyme-efficient pathway, they face a growth-yield trade-off. Since the enzyme cost of a pathway depends on substrate concentrations, this trade-off is likely to be condition-dependent.

The relation between protein investments and biomass yield $Y_{X/S}$, which we measure in grams of biomass per mole of carbon source carbons (i.e. per 1/6 mole of glucose), is not immediately clear. If the rate of carbon uptake q_S were known, we could directly relate the yield and growth rate, using this formula: $Y_{X/S} = \frac{\mu}{q_S}$. However, since changing metabolic fluxes affect carbon uptake, yield, and growth rate at the same time, it is difficult to gain insight about the relation between μ and $Y_{X/S}$ without understanding the changes in q_S. Some metabolic models (in particular, classical Flux Balance Analysis (FBA)) assume that q_S is constant, and therefore effectively assume a fixed relation between growth rate and biomass yield. Of course, rate-yield trade-offs cannot be explored in this way. Here, we take a different approach, which combines kinetic modeling with elementary flux mode analysis.

Elementary flux modes (EFMs) describe the fundamental ways in which a metabolic network can operate [16–19]. Among the possible steady-state flux modes, EFMs are minimal in the sense that they do not contain any smaller subnetworks that can support steady-state flux modes [16, 17, 19]. The EFMs of a metabolic network can be enumerated (even though a large number of EFMs may preclude this in practice). Each EFM has a fixed yield, defined as the output flux divided by the input flux, and is easy to compute. The maximal possible yield is always achieved by an EFM. EFMs might be expected to have very simple shapes, but since biomass production requires many different precursors, biomass-producing EFMs can be highly branched (e.g. Figure 2A). All biomass-producing EFMs are thermodynamically feasible, and in a model with predefined flux directions, the set of steady-state flux distribution is a convex polytope spanned by the EFMs.

EFMs have a remarkable property that makes them suited for studying rate/yield trade-offs: in a given metabolic network, the maximal rate of biomass production at a given enzyme investment is always achieved by an EFM [7, 20]. Therefore, to find flux modes that maximize growth, we only need to enumerate the EFMs and assess them one by one. Since biomass yield is a fixed property of EFMs, we can observe rate/yield trade-offs simply by plotting yields versus growth rates of all EFMs (Figure 1(a)).

To study metabolic strategies and trade-offs in microbes, we developed a new method for predicting cellular growth rates achievable by each EFM. To do so, we consider a kinetic model of metabolism and and determine an optimal enzyme allocation pattern in the network, realising the required EFM and a predefined rate of biomass production at a minimal total enzyme investment. This optimization can be efficiently performed using the newly developed method of Enzyme Cost Minimization (ECM) [21]. For each EFM, the optimized enzyme amount at unit biomass flux is then translated into a mass doubling time of the proteins (i.e. the amount of time that metabolism would have to be running just to duplicate all the enzymes). Assuming that the protein fraction of the cell dry weight is a constant, this can further be translated into a cell growth rate by a semi-empirical formula described below. Building on recent developments in the field [7, 20, 21], we can then effectively scan the space of feasible flux modes and define a region of Pareto-optimal strategies – i.e., flux modes that maximize growth at a given yield or maximize yield at a given growth rate. The shape of this Pareto front tells us whether a rate/yield trade-off exists. We focus our efforts on the central metabolism of *E. coli*, because these fast-growing bacteria have often been used for experiments on the rate/yield trade-off and because of their well-studied enzyme kinetics.

Results

Computing the cell growth rate achievable by an elementary flux mode (EFM)

Flux-balance Enzyme Cost Minimization (fECM) is a method for finding optimal metabolic states in kinetic models, i.e., flux distributions that realize a given metabolic objective (e.g., a given biomass production rate) at a minimal enzyme investment. These are the flux distributions that can be expected to allow for maximal growth rates. The fundamental difference between fECM and constraint-based methods is the underlying kinetic model. For scoring a flux profile, rather than using approximations such as the *sum of fluxes* [22] (or other linear/quadratic functions of the flux vector [23]), we directly the total amount of required enzyme predicted by a kinetic model.

To compute the optimal enzyme investments for a given flux mode, we use Enzyme Cost Minimization (ECM), a method that uses metabolite log-concentrations as the variables and can be quickly solved using convex optimization [21]. ECM finds the most favorable enzyme and metabolite profiles that support the provided fluxes – i.e. the profiles that minimize the total enzyme cost at a given biomass formation flux and therefore maximizes the specific biomass formation flux. The minimal enzyme cost of all EFMs can be computed in reasonable time (a few minutes on a shared server).

Given the enzyme demands, we next ask: how fast can a cell grow at the given metabolic fluxes and enzyme abundances? The steady-state cell growth rate is given by the biomass production rate divided by the biomass amount. Focusing on enzymes in central metabolism of *E. coli*, we first try to answer what is the total amount of enzyme required (E_{met}) to produce biomass at a given rate (v_{BM}) in a kinetic model. Then, we define the *enzyme doubling time* in hours as $\tau_{\mathrm{met}} \equiv \frac{\ln(2) \cdot E_{\mathrm{met}}}{v_{\mathrm{BM}}}$. This is the time a cell would need to reproduce all its metabolic enzymes if its didn't have to produce any other biomass. Since *E. coli* cells contain also other proteins and biomass constituents, the real doubling time is longer and depends on the fraction of metabolic enzymes within the total biomass. This fraction, however, decreases with the growth rate as seen in experiments [24] and as expected from trade-offs between metabolic enzymes and ribosome investment [25]. Here, we use the approximation $T = 7.4 \cdot \tau_{\mathrm{met}} + 0.51[\mathrm{h}]$ derived in the Methods section. The resulting growth rate (in h^{-1}), $\mu = \frac{\ln(2)}{T}$, is a decreasing function of the enzyme cost (E_{met}).

We can now compute the minimal total enzyme cost associated with a given flux mode, and we know that flux modes that minimize this cost will also maximize growth. So how can we find an optimal flux mode? Under some reasonable model assumptions, and no matter how the kinetic model parameters are chosen, the enzyme-specific biomass production rate, and therefore growth, is maximized by elementary flux modes [7, 20]. What is the reason? If we consider all feasible flux modes, constrained to predefined flux directions and to a fixed biomass production rate, these flux modes together form a convex polytope in flux space, called benefit-constrained flux polytope (see Figure 1(b)). The vertices of this polytope are EFMs. Since the flux cost function is concave [7], any cost-optimal flux mode must be a vertex of the flux polytope, and therefore be an EFM. Thus, in fECM, we can screen all EFMs, score each of them by the minimal necessary enzyme cost per unit of biomass production flux, and determine the growth-maximizing mode.

To summarize, rate-yield trade-offs in metabolism can be studied by considering a kinetic metabolic model (with rate laws and rate constants, external conditions such as glucose and oxygen levels, and constraints on possible metabolite and enzyme levels), screening all of its metabolic steady states, and computing for each of them growth rate and yield. Finally, the EFM with the maximal growth rate is chosen. However, our method provides not only the growth-optimal flux mode, but the full spectrum of growth rates and yields of all EFMs. The growth-yield diagram, a scatter plot between the two quantities, shows the possible trade-offs between the two objectives. An EFM that is not beaten by any other EFM in terms of both growth rate and yield is called Pareto-optimal. If we connect these Pareto-optimal points by straight lines, we obtain their convex hull (see Figure 1(a)). If we could evaluate the growth rates and yields for *all* metabolic states in the model (including non-elementary flux modes), the resulting growth/yield points would form a compact set, and this entire set would be enclosed in the convex hull of the EFM points. The Pareto-optimal EFMs therefore mark the best compromises between growth rate and yield that are achievable in the model. By inspecting the growth/rate diagram, we can tell, for the set of *all* metabolic states, whether there is an extended Pareto front or a single solution that optimizes both rate and yield. While the yields depend only on shapes of the EFMs, the growth rates are condition dependent. Therefore, the entire picture and the emergence of rate-yield trade-offs can vary between conditions.

Metabolic strategies in *E. coli* central metabolism

To study optimal metabolic strategies in *E. coli*, we applied our method to a model of central carbon metabolism, which provides precursors for biomass production. Our model is a modified version of the model presented in [26] and comprises glycolysis, the Entner-Doudoroff pathway, the TCA cycle, the pentose phosphate pathway and by-product formation (SI Figure S4). Precursors and cofactors, provided by central metabolism, are converted into macromolecules ("biomass") by a variety of processes. These processes are not explicitly covered by our network model, but summarized in a biomass reaction. Reaction

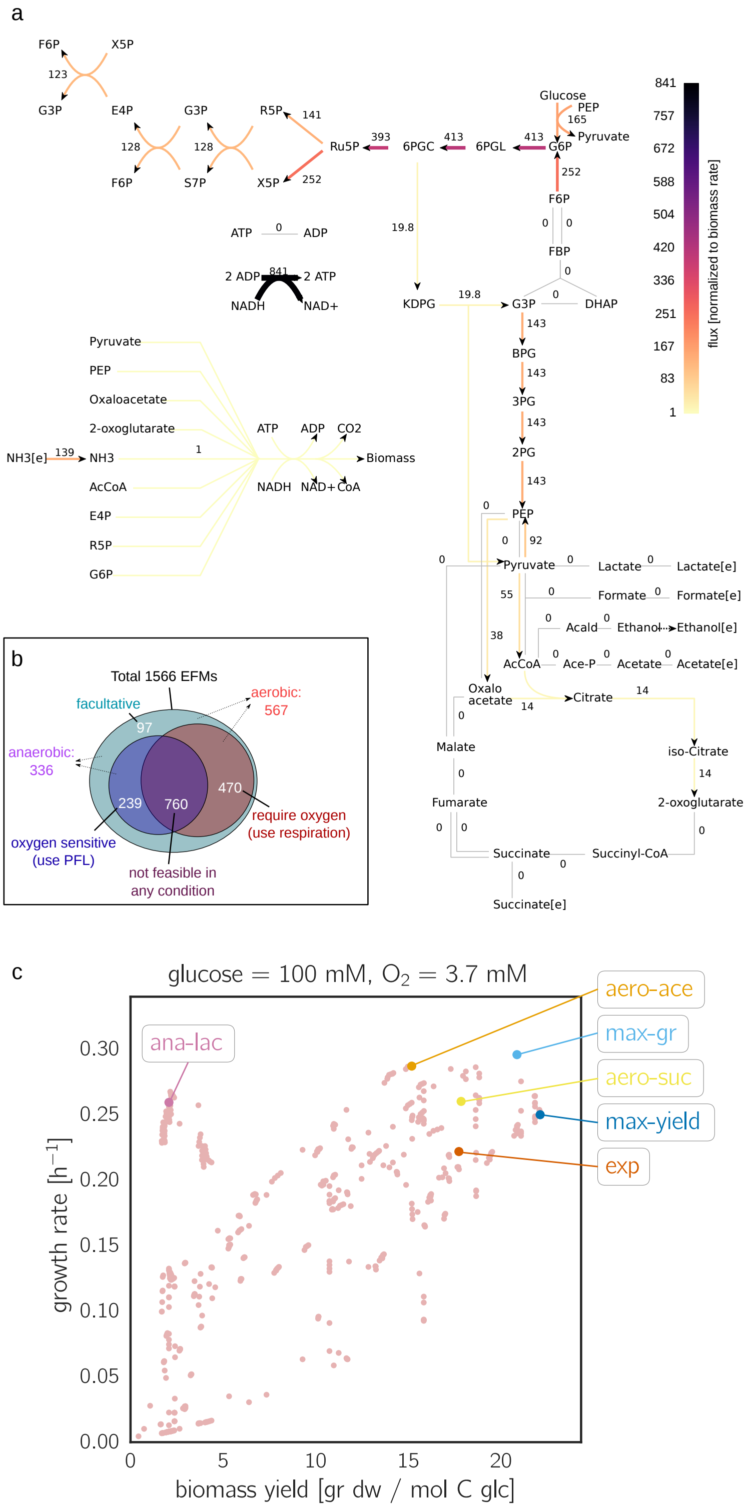
a
F6P
X5P
123
G3P
E4P
G3P
R5P
141
128
128
F6P
S7P
X5P
252
Ru5P
393
6PGC
413
6PGL
413
G6P
Glucose
PEP
165
Pyruvate
252
F6P
0
0
FBP
0
ATP
0
ADP
19.8
2 ADP
841
2 ATP
NADH
NAD+
KDPG
19.8
G3P
0
DHAP
143
BPG
143
3PG
143
2PG
143
PEP
0
0
92
Pyruvate
0
Lactate
0
Lactate[e]
55
0
Formate
0
Formate[e]
0
0
Acald
Ethanol
Ethanol[e]
38
AcCoA
0
Ace-P
0
Acetate
0
Acetate[e]
Oxalo acetate
14
Citrate
14
0
Malate
iso-Citrate
0
14
Fumarate
2-oxoglutarate
0
0
0
Succinate
0
Succinyl-CoA
0
Succinate[e]
Pyruvate
PEP
Oxaloacetate
2-oxoglutarate
NH3[e]
139
NH3
1
ATP
ADP
CO2
Biomass
AcCoA
NADH
NAD+
CoA
E4P
R5P
G6P
841
757
672
588
504
420
336
251
167
83
1
flux [normalized to biomass rate]
b
Total 1566 EFMs
facultative
aerobic: 567
97
anaerobic: 336
239
760
470
require oxygen (use respiration)
oxygen sensitive (use PFL)
not feasible in any condition
c
glucose = 100 mM, O$_2$ = 3.7 mM
ana-lac
aero-ace
max-gr
aero-suc
max-yield
exp
0.30
0.25
0.20
0.15
0.10
0.05
0.00
growth rate [h^{-1}]
0
5
10
15
20
biomass yield [gr dw / mol C glc]

Figure 2: **Metabolic strategies in a model of** ***E. coli*** **central carbon metabolism. (a)** Network model of *E. coli* central metabolism. The reaction fluxes, defined by the Elementary Flux Mode (EFM) *max-gr*, are shown by colors. In standard conditions – i.e. high extracellular glucose and oxygen concentrations – this EFM has the highest growth rate among all EFMs. It uses the TCA cycle at the right amount to satisfy its ATP demand and does not secrete any organic carbon. Cofactors are not shown in the figure, but exist in the model. **(b)** Growth rates and yields of the 1040 biomass-producing EFMs. Each EFM represents a steady metabolic flux mode through the network. **(c)** Spectrum of growth rates and yields achieved by the EFMs. The labeled EFMs are listed in Table 1.

Acronym*	Biomass yield (g/C-mol)	Growth rate† (h^{-1})	Oxygen uptake	Acetate secretion	Lactate secretion
max-gr	20.8	0.295	0.42	0	0
max-yield	22.1	0.249	0.39	0	0
ana-lac	2.1	0.258	0	0	0.92
aero-suc	17.8	0.259	0.32	0	0
aero-ace	15.2	0.286	0.22	0.36	0
exp	17.7	0.221	0.29	0.22	0

*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 the reference conditions, where [glucose] = 100 mM, and [O_2] = 3.7 mM.

Table 1: **Selected EFMs representing different growth strategies.** Metabolic fluxes are given in carbon moles (or O_2 moles) per carbon mole of glucose uptake. For more details, see Supplementary Table S4.

kinetics are described by modular rate laws [27], with enzyme parameters obtained by balancing [28] a large collection of literature values. Parameter balancing yields consistent parameters in realistic ranges, satisfying thermodynamic constraints, and in optimal agreement with measured parameters.

All EFMs were scaled to a standard biomass production of 1, and their yields are defined as grams of biomass produced per mole of carbon atoms taken up in the form of glucose. EFMs that contain oxygen-sensitive enzymes and oxygen-dependent reactions cannot be used by the cell. After removing such EFMs, we obtained 567 EFMs that produce biomass under aerobic conditions and 336 under anaerobic conditions, of which 97 can operate under both conditions (Figure 2). Statistical properties of the EFMs (size distribution, usage of individual reactions, and similarity between EFMs and a measured flux distribution) are shown in Supplementary Figure S7. Some EFMs have very similar shapes, and we used the t-SNE layout algorithm to arrange all EFMs by their similarities (see Supplementary Figure S6). To avoid any biases in our growth predictions, we considered all EFMs that have a non-zero biomass yield, even those that contain physiologically unreasonable fluxes. For example, the futile cycling between PEP and pyruvate (by the combined activity of pyruvate kinase (*R9*) and pyruvate water dikinase (*RR9*)) wastes ATP and is generally expected to be suppressed by strict enzyme regulation [29]. Such EFMs were consistently predicted to show low growth rates and had no effect on the outcomes of our study.

The spectrum of possible growth rates and yields is shown in a growth/yield diagram (see Figure 2c). The growth rates refer to our standard conditions ([glucose] = 100 mM, [O_2] = 3.7 mM). While the yields, as immediate properties of the EFMs, are condition-independent, growth depends on enzyme requirements and therefore on kinetics and external conditions such as nutrient levels. If we naively assumed that all EFMs, scaled to equal glucose uptake, required the same total amount of enzyme, growth rates and yields would be proportional. On the contrary, if we assumed that all EFMs, scaled to equal biomass production, required the same enzyme amount, all EFMs would show identical growth rates. The actual spectrum of growth rates and yields for all EFMs was determined as described above.

To compare some typical metabolic strategies, we focused on six particular flux modes with different characteristics and followed them across varying external conditions and kinetic parameter values. These EFMs and an experimentally determined flux distribution (data from [30]; for calculations see Supplementary Text S4.1) are marked by colors in Figure 2b and listed in Table 1. Their full flux maps (produced using software from [31]) can be found in the SI section S5.2. The map for *max-gr* is also included in Figure 2a. *max-yield*, as mentioned earlier, has the highest yield. This EFM does not produce any by-products nor does it use the pentose-phosphate pathway. *max-gr* has a slightly lower yield, but reaches the highest growth rate (0.33 h^{-1}) in standard conditions and uses the pentose-phosphate pathway with a relatively high flux. We have also selected three by-product forming modes that use the pentose phosphate pathway and have similar growth rates to the *max-yield* EFM: An anaerobic lactate fermenting mode (*ana-lac*) with a very low yield

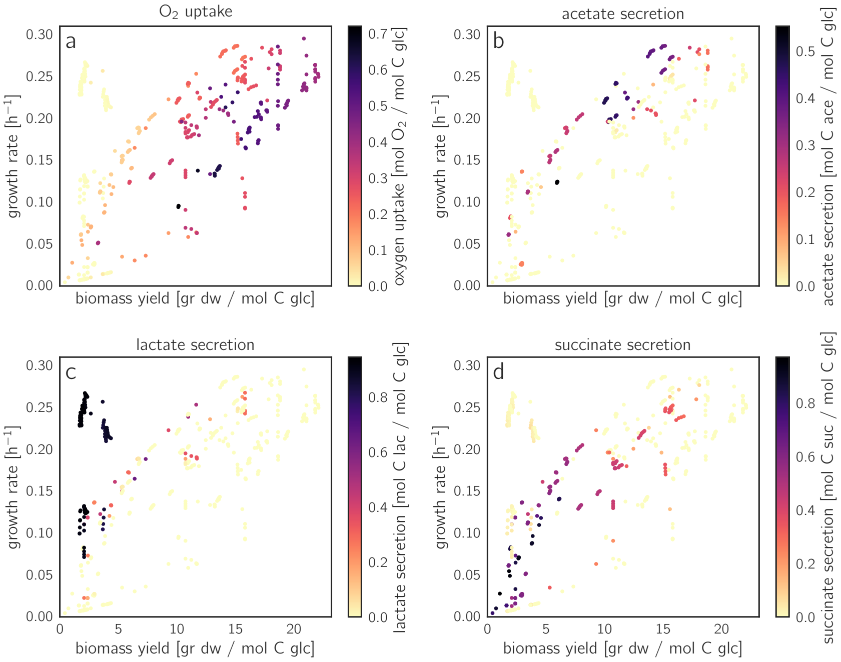

Figure 3: **Fluxes in uptake and secretion reactions, shown for all EFMs.** (a) Oxygen uptake per glucose taken up. Flux values for the different EFMs are shown by colors in the growth/yield diagram (point positions as in Figure 2b). The EFMs with the highest growth rate consume only intermediate levels of oxygen. The following diagrams show **(b)** acetate secretion, **(c)** lactate secretion and **(d)** succinate secretion, each normalized by glucose uptake.

(2.4), an aerobic succinate fermenting mode (*aero-suc*) with a decent yield (20.7), and an aerobic acetate fermenting mode (*aero-ace*) with a slightly lower yield (17.6). Interestingly, *ana-lac* reaches a similar growth rate as the other by-product forming modes with a much lower yield, thanks to the lower enzyme cost of the PPP and lower glycolysis compared to the TCA cycle (per mol of ATP generated). This recapitulates a classic rate-versus-yield problem, associated with overflow metabolism. The acetate producing EFM has the highest growth rate of the by-product producing EFMs, which explains that *E. coli* in fact excretes acetate and not lactate or succinate.

The rate-yield diagram (with enzyme-specific biomass production on the y-axis) and the resulting growth-yield diagram (showing predicted cell growth) have very similar shapes: the y-axis is nonlinearly scaled, but the Pareto front still consists of the same EFMs (Supplementary Fig S1). In our model at standard conditions, we find only a relatively low correspondence between growth and yield (Figure 2b), although both quantities have wide distributions (see SI Figure S10). Many low-yield EFMs (generating less than 8 gram of biomass per carbon mole of glucose) have high growth rates (above 0.3 h^{-1}, which is more than 75% of the maximal growth rate). The reason is that these EFMs use reactions that require a low amount of enzyme. In the standard conditions used in Figure 2c, the EFM max-yield achieves the maximal yield of 25.7 (gram biomass per C-mol glucose). The highest growth rate (0.33 h^{-1}) is achieved by the EFM *max-gr*, which also has a very high yield (24.2). Therefore there is a very narrow Pareto front. This is not a trivial finding, and other choices of parameters or extracellular conditions lead to situations in which the Pareto front is much broader, thus displaying a clearer trade-off between growth rate and yield (for example see Figure 4c).

To associate high yields or high growth rates with specific reaction fluxes or chemical products, we selected four uptake or secretion reactions, computed their fluxes in the different EFMs, and visualized them by colors in the growth/yield diagram (Figure 3). The best-performing EFMs (in the top-right corner) consume intermediate amounts of oxygen and do not secrete any acetate, lactate or succinate. Another group of EFMs (visible in red in Figure 3b) consume slightly less oxygen, but secrete large amounts of acetate. This aerobic fermentation exhibits lower biomass yields compared to pure respiration, but it maintains comparable growth rates, suggesting that a lower demand for enzyme compensates for the lower yield. Other important fluxes are shown in Supplementary Figure S8.

The growth rates associated with metabolic strategies depend on environmental conditions and enzyme parameters

To study how optimal EFMs and the resulting growth rates vary across external conditions, we varied a single model parameter and traced its effects on the growth rate. Figure 4a shows how a decreasing external oxygen concentration affects growth: lower oxygen levels need to be compensated by higher enzyme levels in oxidative phosphorylation, which again lowers the growth rate (Figure 4b). However, EFMs that function anaerobically, such as *ana-lac*, are not affected (see SI Figure S16 for the enzyme allocations). The effects of varying glucose levels can be studied in a similar way (see SI Figure S11): at a lower glucose concentration, the PTS transporter becomes less efficient and to maintain the flux, cells must compensate this by higher expression levels of the transporter. This increases the total enzyme cost, and therefore slows down growth. At very low glucose concentrations, as low as 10^{-5} mM, the cost of the transporter completely dominates the enzyme cost (see Figure 5b and SI Figure S16 for a breakdown of the enzyme allocations).

By varying glucose and oxygen levels, we can screen environmental conditions and see which EFM reaches the highest growth rate. For a slightly lower oxygen concentration than used in the model, *aero-ace* becomes beneficial at high glucose concentrations (see Figure 4b). The switch to aerobic acetate production, predicted for increasing glucose levels, agrees with experimental observations. The growth rate of the selected EFMs in this glucose/oxygen phase diagram shows a different response to glucose and oxygen for the different EFMs (see Supplementary Figure S12). An evaluation of all EFMs over this oxygen and glucose range shows that several EFMs become optimal (see Supplementary Figure S12), and that these EFMs represent four clearly distinct strategies (see Figure 4d-f). The effect of low oxygen levels can also be demonstrated in the growth/yield diagram. Oxygen-dependent EFMs tend to show higher yields (probably due to the high ATP yield in oxidative phosphorylation). The growth rates of these EFMs drop considerably at low oxygen levels (Figure 4b). Therefore, the growth/yield tradeoff becomes much more prominent at low oxygen levels, with a Pareto front spanning a wide range of growth rates and yields (Figure 4a).

Bacterial growth typically increases at higher glucose levels in the growth medium. The quantitative relationship, called Monod curve, can be 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 find a relatively small trade-off (see SI Figure S17a), which would mean that one EFM can remain favorable over a wide range of dilution rates. Under anaerobic conditions, a pronounced trade-off develops (SI Figure S17e) and bacteria are predicted to use different metabolic strategies depending on the dilution rates.

Cell growth rates and choices of metabolic strategies do not only depend on external conditions, but also on enzyme parameters. As an example case, we considered a single k_{cat} value, the one of the enzyme triose-phosphate isomerase (*tpi*), varied its value, and studied its effects on the growth/yield diagram. Not surprisingly, slowing down the enzyme decreases the growth rate (Supplementary Figure S18c). But to what extent? Most of our selected EFMs are strongly affected by this k_{cat} value, except for *max-gr*, which does not use the reaction. At lower k_{cat} values, the resulting Pareto front (Supplementary Figure S18a) becomes broader because the optimal yield EFM now has a much lower growth rate, but *max-gr* is not affected. To study the effect of parameter changes more generally, we predicted the growth effects of *all* enzyme parameters in the model by computing their growth sensitivities, i.e., the first derivatives of the growth rate (or biomass-specific enzyme cost) with respect to the enzyme parameter in question (see Supplementary Files). For *tpi*, EFMs with relatively low enzyme costs (and therefore a high growth rate to yield ratio) are usually sensitive for the k_{cat} of this enzyme (Supplementary Figure S18b). Growth sensitivities are informative for more than one reason. On the one hand, parameters with large sensitivities are likely to be under strong selection pressures (where positive or negative sensitivities indicate a selection for larger or smaller parameter values, respectively). On the other hand, these parameters have a big effect on growth predictions, and precise estimates of these parameters are critical to obtain reliable models. Different parameters in a reaction can have very different growth sensitivities. For example the sensitivity of the k_{cat} and K_M values of *R2r* are low, but the growth rate is very sensitive to the K_{eq} value. Some parameters are equally sensitive in all EFMs, but others are only sensitive in a subgroup of EFMs.

Finally, a selection of EFMs allows us to systematically study the choice between metabolic pathways, for example, the (high ATP yield, high enzyme demand) EMP and (low ATP yield, low enzyme demand) ED versions of glycolysis. Unlike *Flamholz et al.*, [4], we can now study the choice between these pathways as part of a whole-network metabolic strategy, which also involves the choice between respiration and fermentation, or combinations of them. Moreover, we can compare the effect of constraining the model to use only one of these pathways across the different environmental conditions, specifically the external concentrations of glucose and oxygen. We implement these constraints using enzyme knock-outs in the two pathways (which simply make some of the possible EFMs disappear). Since glycolysis is essential for growth in any condition, all EFMs that do not use the ED must use EMP instead, and vice versa. By calculating the (condition-dependent) growth defects of the two knock-outs compared to the wild-type, we can asses the importance of each of the pathways to the fitness in that condition (see SI section S3.6). As shown in Figure 6, at relatively low oxygen levels and medium glucose levels (10 μM – 100 mM) cells profit considerably from employing the ED pathway, therefore knocking it out would decrease growth rate by up to 25%. The EMP pathway has virtually no benefit in these conditions. On the other hand, the cost of not using EMP at low glucose levels (below 10 μM) is much higher, probably since the low yield of the alternative ED pathway requires much higher rates of glucose uptake which incurs a high cost when glucose is scarce.

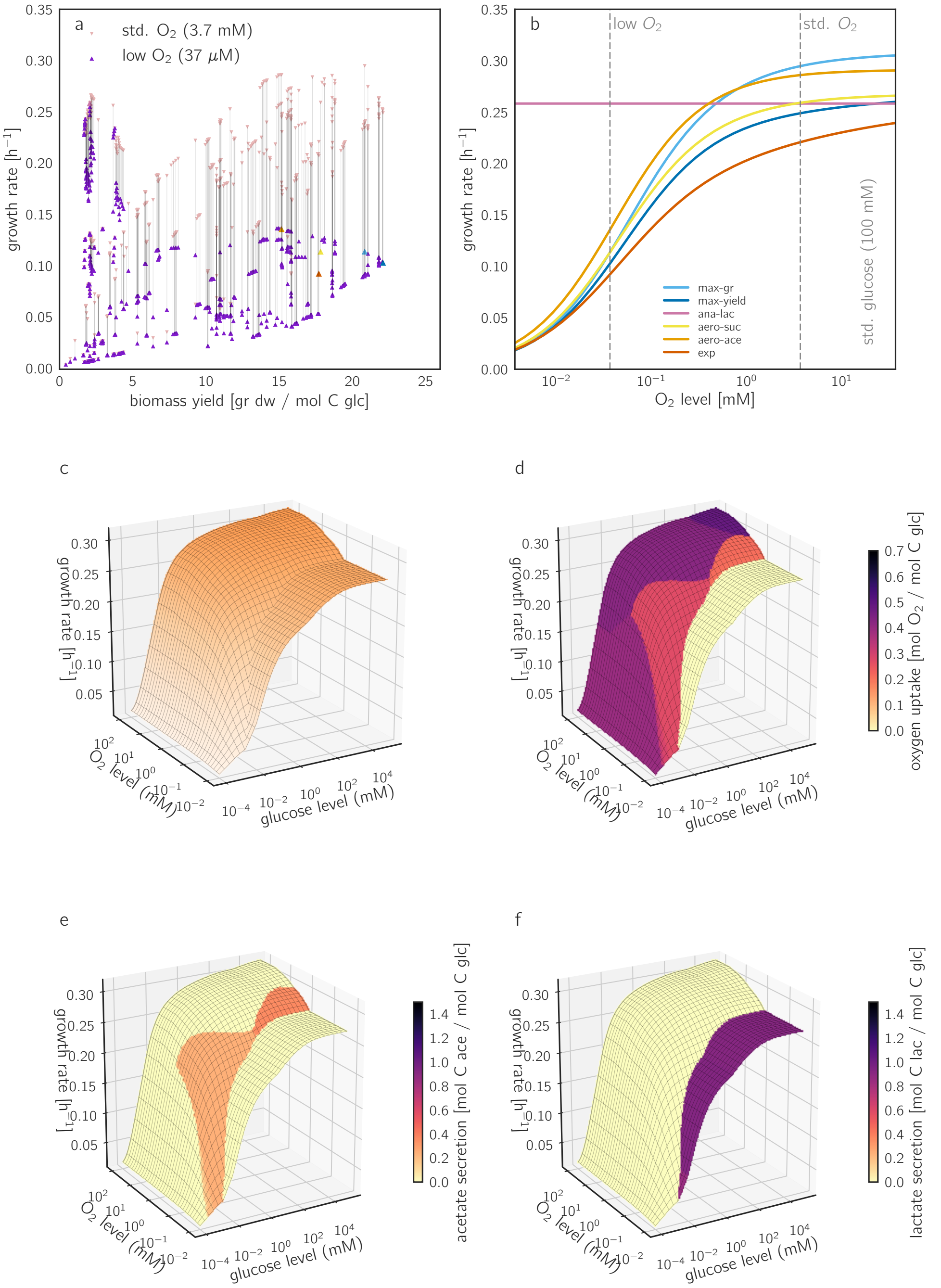

Figure 4: **Growth and growth/yield trade-offs depend on external glucose and oxygen levels.** (a) Predicted growth rates and yields of all aerobic EFMs, at standard oxygen levels (3.7 mM) and at low levels (37 μM). Since only the growth rate is affected, but not the yield, vertical lines connect between the two values. (b) Growth rates as a function of external oxygen concentration for the six selected EFms. The oxygen concentration directly affects the rate of oxidative phosphorylation (reactions *R80* and *R27*), hence changes in oxygen require enzyme changes to keep the fluxes unchanged. The *ana-lac* EFM (non-respiring) does not depend on oxygen and shows an oxygen-independent growth rate. All other selected EFMs show similar responses, except for *max-gr* which uses a higher amount of oxygen and thus has a steeper slope, which causes it to lose its lead when oxygen levels are below 1 mM. In general, growth increases with oxygen level and saturates around 10 mM. The changes in required enzyme levels are shown in SI Figure S16. Colors of EFMs correspond to panel a. (c) Maximal growth rate reached at different external glucose and oxygen levels. (d)-(f) The same plot, with oxygen uptake, acetate secretion, and lactate secretion shown by colors. The distinct areas in this “landscape” correspond to different optimal EFMs, representing different metabolic strategies (compare SI Figure S12).

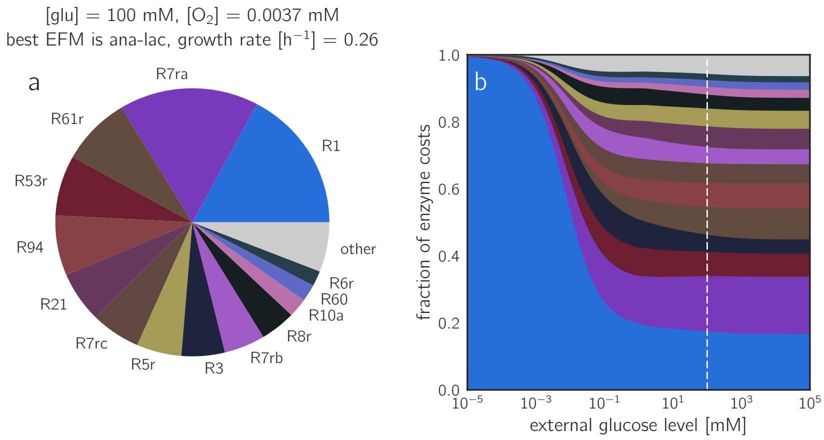

Figure 5: **Predicted protein investments** (a) Breakdown of predicted protein investments at high glucose / low oxygen conditions for the EFM *ana-lac*. (b) Breakdown of predicted protein investments as a function of the glucose level at standard oxygen level for the EFM *ana-lac*

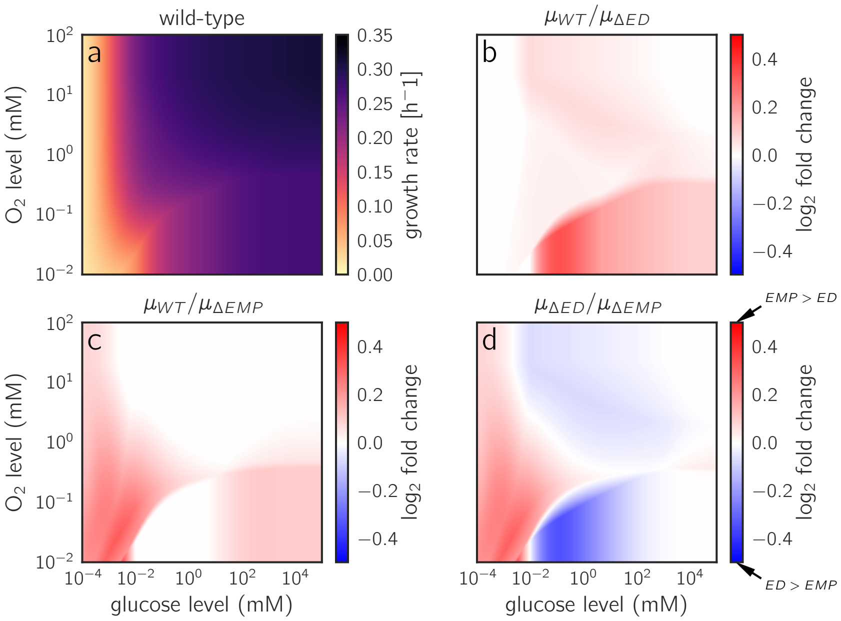

Figure 6: **Growth rates achieved by using EMP glycolysis, ED glycolysis, or both.** (a) Growth at varying external glucose and oxygen levels for simulated wild-type *E. coli*. Same data as in Figure 4c, shown as a heatmap. Bacteria can employ two glycolytic pathways: the Embden-Meyerhof-Parnas (EMP) pathway that is common also to eukaryotes, and the Entner-Doudoroff (ED) pathway, which provides a lower ATP yield, but at a much lower enzyme cost [4]. (b) A simulated ED knockout strain can only rely on the EMP pathway. The heatmap shows the relative increase in growth rate of the wild-type strain (i.e. if the ED pathway is reintroduced to the cell). The ED pathway provides its highest advantage in very low oxygen and medium to low glucose levels. Similarly, (c) shows the increase in growth rate for reintroducing the EMP pathway, which provides its highest advantage glucose concentrations below 10 μM. (d) A direct comparison of the two knockouts, where blue areas mark conditions where shifting from EMP to ED is beneficial and red areas mark conditions where shifting from ED to EMP is favored. The strong blue region of low oxygen and medium glucose levels might represent the environment of bacteria such as *Z. mobilis* that indeed use the ED pathway exclusively [32]. The same data are shown as surface plots in Supplementary Figure S19.

Discussion

Our case study in *E. coli* shows that there is neither a general coupling of growth rate and biomass yield nor a general trade-off between them. Growth/yield trade-offs depend on the circumstances, especially on all factors that affect enzyme cost. This is no surprise. Kinetically, yield influences growth in two contrary ways. On the one hand, high-yield strategies produce biomass at a lower glucose influx, and this lower influx allows for lower enzyme investments and therefore for higher growth rates. On the other hand, high-yield pathways leave a smaller amount of Gibbs free energy to be dissipated in reactions. This, however, needs to be compensated by higher enzyme levels and leads to a lower growth rate. The second relation may be obscured by a second substrate such as oxygen, which provides additional driving force. If the first relation dominates, there may be an EFM that maximizes growth and yield; if the second relation dominates, there will be a trade-off, i.e., a Pareto front formed by several EFMs. In our simulations, the extent of this trade-off strongly depended on conditions and kinetic parameters. At high oxygen levels, our growth-maximizing solution showed almost the maximal yield and the Pareto front was negligible. Under low-oxygen conditions, low-yield strategies showed the highest growth rates and a broad Pareto front emerged.

Experimental results indicating rate-yield trade-offs are difficult to interpret; as shown in [9], the original cell populations may be distant from the trade-off line, and a selection for growth may push the populations and individuals close to the trade-off line. Selection for growth rate and selection for yield would be needed to demonstrate trade-offs experimentally. From our simulation results, we expect the experimental results to be as scattered as they are. It would be interesting to see whether the experimental results are in fact condition-dependent (e.g. dependent on oxygen availability).

Our standard conditions used in this paper describe almost saturating glucose and oxygen concentrations, which are comparable to typical laboratory conditions. However, different conditions are used in different experiments, and actual oxygen concentrations are very hard to estimate (the question of oxygen availability may be as complex as in yeast, where it has been suggested that oxygen may diffuse too slowly to supply the mitochondria with enough oxygen [33]). Since we do also not know realistic values for the affinity of the reactions for oxygen, knowledge of the actual oxygen concentration would not directly solve this either. Under these standard conditions we do not predict a pronounced trade-off, although we know *E. coli* uses a low-yield acetate producing strategy. However, the acetate producing mode would be optimal under a slightly different oxygen or glucose level (see Supplementary Figure S12). Moreover, our strain might not be optimally adapted; recently it has been shown that different strains of *E. coli* show different phenotypes and the strain we used for the data in this paper is not the fastest growing one [34].

To predict metabolic fluxes and growth rates, we developed a new method in which enzyme cost minimization, a numerically efficient method to predict metabolic fluxes and enzyme profiles *ab initio*, is combined with an exhaustive screening of EFMs, i.e., potentially optimal flux modes. To translate enzyme-specific biomass production rates into growth rates, we used a nonlinear formula that accounts for the growth-rate dependent composition of the proteome. The enzyme cost calculations by ECM are only based on a network model, on kinetic enzyme properties, and on a few transparent model assumptions. No flux or proteome measurements are used. The fast optimization of fluxes and enzyme levels is enabled by two mathematical results concerning the convexity and concavity of enzymatic cost functions in kinetic models. In the model, we use common modular rate laws because they yields realistic results [21] and because strict convexity is guaranteed for these rate laws (Joost Hulshof, personal communication). The optimized enzyme cost is a concave function in flux space. Our implementation of ECM allows for other rate laws and can be used by the community via our website. The fast implementation in the optimization platform NEOS allows for a variety of applications, including screens of parameter values and external conditions or knock-out studies.

Our model predicts a much higher maximal biomass yield than the yield measured in batch cultures (25.7

vs 11.8 g dw (mole carbon)$^{-1}$ [35]), while the predicted growth rate is lower (0.33 vs 0.89 h^{-1}). For the experimentally determined flux mode, we overestimate the yield (20.6 vs 12 [30]) and underestimate the growth rate (0.25 vs 0.61) as well. However, the predicted yield under anaerobic conditions is lower (usually below 5 and the maximum is 6.4 see SI Figure S14) and agrees with experimental observations. The overestimation of the yield (which solely depends on the stoichiometric model structure) might be caused by the fact that some waste products or processes that dissipate energy are missing in our model. The low predicted growth rates might result from our simplistic conversion of enzyme costs into growth rates, a part of the method that could be improved. Another cause of the low growth rates could be the high costs of oxidative phosphorylation, which are difficult to estimate. However, we expect that the over- and underestimation occur consistently across EFMs and will not affect the qualitative results of this study.

Our calculations of growth rates rely on a large number of kinetic constants. Uncertainties in these parameters will introduce uncertainties into all our predictions. Stoichiometry-based methods do not require such parameters, but such analyses (such as FBA without any additional flux constraints) would not even be able to address rate/yield trade-offs because their model assumptions force growth rates and yields to be proportional (see Figure S13). More recent FBA methods that bound or minimize the presumable enzyme demand (such as FBA with flux minimization [22] or molecular crowding [23]) can address such trade-offs and predict low-yield flux modes. One could also employ a simplified version of fECM that resembles those methods, relating fluxes and enzyme levels not by rate laws, but by a simple proportionality. For example, assuming that all enzymes work at their maximal speed (as given by their k_{cat} values), the ECM optimization would become obsolete: using the enzyme weights and k_{cat} values, we could directly translate any flux mode into a total required amount of enzyme by a simple linear formula [21]. Such a linear formula has also been used in [23] to put crowding constraints on metabolic fluxes. To account, at least, for the effect of external concentrations on transporter cost, variable cost weights for transporters could be defined by linking external concentration, amount of transporter, and influx by kinetic rate laws [36].

However, a simplified version of fECM employing such a linear formula would lead to an overestimation of growth rates because it would ignore the “unused enzyme fraction” [37], which will always appear, for example, when enzymes are not fully satured with substrate. The term itself may be misleading: as suggested by our ECM results, enzymes may work below their maximal rates not because they are deliberately left unused, but because of the fact that reactions, to be thermodynamically and kinetically efficient, require high substrate and low product levels. This causes contradicting requirements in different reactions, and even in the best possible compromise, many enzymes will be used inefficiently. Other methods, such as ME-Models [38] or Resource Balance Analysis (RBA) [39] yield similar “unused enzyme fractions”, although it is hard to estimate them in these cases. Any “linear” version of fECM calculations will overestimate the growth rate by a factor of about 2.4 (see supplementary figure S3) and, more severely, distort the growth differences between EFMs. This overestimation is purely an artifact and has no biological interpretation, therefore results form these “linear” methods that agree with measurements could actually have wrong assumptions and have to be carefully interpreted. Given the overestimation of the growth rate, it seems quite surprising that these methods can actually be quite predictive (e.g. [6]). In the case of RBA, accuracy is achieved because growth-rate dependent, experimentally measured apparent k_{cat} values are incorporated into the model. These apparent k_{cat} values, in turn, could be explained by efficiency factors defined in ECM [21].

Beyond the high uncertainty in kinetic constants, it seems that the fluxes employed by “true wild-type” *E. coli* are vaguely defined. Recent data shows that even closely-related strains of *E. coli* sometimes use drastically different metabolic strategies, even though we expect their metabolism to be mostly identical [34]. Interestingly, a few of these strains do not display any respiro-fermentative metabolism in aerobic environments, but rather use a fully respiratory strategy without secreting any byproducts. Furthermore, the growth rate of these strains is among the highest. This finding raises questions regarding the universality of the rate/yield trade-off principle and supports our conclusion that it is almost non-existent in highly oxidative conditions.

Being based on enzyme kinetics, fECM is fully quantitative and allows modellers to address a great variety of questions. Unlike other flux prediction methods, our method can account for allosteric regulation and for the quantitative effects of external conditions such as oxygen concentration, kinetic parameters, and enzyme costs (see Supplementary Figure S2). The fact that our model can account for low glucose concentrations also implies that our method can be used to describe chemostat settings, while in flux-only methods everything would just scale linearly with a lower glucose uptake flux. In a chemostat the steady-state growth rate is externally controlled by setting the dilution rate, and the steady-state glucose level reaches a value that supports exactly this growth rate. There are likely trade-offs between growth at low and high oxygen concentrations and our model can be used to estimate these (see SI Figure S17). In standard conditions we do not see a trade-off between the Monod constant and the growth rate, perhaps explaining why there is no significant negative selection on the Monod constant in the long term experimental evolution of *E. coli*, where rate selection could have been expected [40].

Once an fECM analysis has been run, additional analyses require only very little additional effort. For example, parameter sensitivities or uncertainties caused by small parameter variations can be easily computed without re-running any optimizations: all necessary sensitivities can be obtained from the existing results. Moreover, the decomposition into EFMs already provides all information that is needed to study gene knock-outs. To simulate a single or multiple knock-out, we simply need to exclude all affected EFMs from our analysis (Supplementary Figure S2f). The yield of knock-out mutants and the yield-related epistatic interactions between knock-outs have been computed before (see SI Figure S21), but the growth rates of the knock-out mutants and their epistatic effects under different conditions have not been computed so far (see Supplementary Figure S20).

Flux balance ECM can be extended, both to larger network models and to incorporating more detailed kinetic information than we did in this paper. Bigger networks will bring two main challenges: data availability and calculation issues. The EFMs of a large network would greatly increase in number. A subsampling of EFMs can be problematic because depending on model conditions, the high-growth EFMs may easily be missed (see SI section S3.1). However, a promising avenue is to subdivide large networks by setting all strongly connected metabolites external [41] and predefining their concentrations. These concentrations could also be varied to assess their effects on predicted metabolic strategies. The resulting subnetworks can then be analyzed independently, and their EFMs can be combined to yield favorable, global, elementary flux distributions. The convexity of the ECM of individual EFMs allows for large networks to be solved with ECM (as noted by e.g. [42]). Predictions with fECM will be strengthened by more accurate knowledge of enzyme properties. Although kinetic information about enzymes in the central carbon metabolism of *E. coli* is relatively complete, we still had to complete some missing values by parameter balancing. To make the model parameters more precise, it would be possible to take temperature and pH into account [28]. The biomass composition of *E. coli* depends on the growth rate. Considering such a growth-dependent biomass composition will improve the predictions, but requires changes in the algorithm. Ideally, model results should be self-consistent, i.e., the predicted growth rate should match the growth rate for which the biomass composition has been assumed. This challenge can perhaps be solved by an iterative procedure. Another open question concerns models with non-enzymatic reactions, which can effectively render the flux-polytope non-convex, possibly leading to non-elementary growth-optimal flux modes. Finally, we could study optimal EFMs and growth rates at different predefined rates of ATP consumption, implemented by a flux constraint. Since non-EFMs can be optimal with additional flux constraints, the algorithm for finding the possible optimal flux modes will have to be adjusted. Since these new points can be found by interpolating between EFMs we expect that an efficient algorithm can be developed.

Although the importance of efficient protein allocation for reaching high growth rates has often been stressed (reviewed in [43]), real cells may not always minimize enzyme cost. *Lactococcus lactis*, for example, can display a metabolic switch associated with large changes in growth rate, but without any changes in protein investments [44]: these cells could save enzyme resources, but do not do so – possibly because unused enzyme provides other benefits, e.g., being prepared for metabolic changes to come. To capture such behavior in optimality approaches, other optimization criteria could be considered, including an extra ATP production for maintenance or stress or a need for robustness, or further flux constraints could be assumed. If different EFMs provide growth rates close to the optimal one, these EFMs, or maybe mixtures of them, may coexist in a cell population. Mixtures of EFMs would also increase robustness, as EFMs are inherently not robust against a repression of one of the active enzymes. To model such flux patterns in a population, instead of considering only one optimal EFM, we could consider a set of EFMs close to the optimal growth rate (e.g. between 99% and 100% of the optimal growth rate), because selection between these EFMs may be rather weak. Averaging over these EFMs could yield smoother transitions when parameters are varied, e.g., in the condition-dependent enzyme expression or the usage of alternative pathways. Specific fluxes over a range of conditions, such as shown in Supplementary Figure S12, could be averaged over a set of suboptimal EFMs as well, to give a more robust prediction of population behavior. As mentioned before, a sensitivity analysis can indicate selection pressures on kinetic parameters. However, larger changes such as how to evolve from using one EFM to another are still a challenge. Our method can be used to sketch a fitness landscape and predict what mutations would be necessary for such larger transitions.

Although kinetic approaches to flux optimization pose some challenges, they are a necessary and useful addition to existing constraint-based flux analysis methods. Only with relatively assumption-free methods we can address fundamental issues of unicellular growth and cell metabolism, such as the trade-off between growth rate and biomass yield.

Methods

Flux and enzyme profiles for maximal enzyme-specific biomass production

A metabolic state is characterized by its enzyme levels, metabolite levels, and fluxes. The relationships between all these variables are defined by rate laws and are condition- and kinetics-dependent. Our algorithm finds optimal metabolic states in the following way. The elementary flux modes of a network, which constitute the set of potentially growth-optimal flux modes, are enumerated. Now we consider a specific model condition, defined by a choice of kinetic constants and external metabolite levels in the kinetic model. For this condition, we first compute the growth rates for all EFMs. To determine the optimal metabolic strategy – the one expected to evolve in a selection for fast growth – we then choose the EFM with the highest growth rate.

To determine the growth rate of an EFM, we predefine a biomass production rate v_{BM}, scale the EFM to realise this production rate, and compute the enzyme demand by applying ECM. Computing the enzyme demand involves an optimisation of metabolite levels $\mathbf{c}$ and enzyme levels $\mathbf{E}$ by ECM. Thus, in summary, the optimal state $(\mathbf{v}, \mathbf{c}, \mathbf{E})$ can be found efficiently by a nested screening procedure. First, we consider all feasible flux modes $\mathbf{v}$, requiring stationary and a predefined biomass production rate v_{BM}. For each flux mode $\mathbf{v}$, we consider all possible logarithmic metabolite concentration profiles $\ln \mathbf{c}$, where an upper and a lower bound is set for each metabolite. For each such profile, we compute the necessary enzyme levels E_l and obtain the total enzyme cost E_{met}. Since the cost function with respect to logarithmic metabolite concentrations is convex, it can be easily minimized; and since the optimized cost, as a function of fluxes, is concave, we need not screen the flux space exhaustively, but can restrict our search to elementary flux modes. This yields an optimization over all the possible states of our kinetic model.

ECM and NEOS online tool

Enzyme Cost Minimization has been recently applied to a similar kinetic model of *E. coli*'s central carbon metabolism network [21]. It uses a given flux distribution (in our case, given by an EFM) to formulate the enzyme concentrations as explicit functions of their substrate and product levels. We then score each possible enzyme concentration profile by the total enzyme mass concentration $E_{\text{met}} = \sum_l \beta_l \, E_l$ (in mg l^{-1}), where β_l denotes the molar mass of enzyme l in Daltons (mg mmol^{-1}) and enzyme concentrations are measured in mM (i.e., mmol l^{-1}). Written as a function of the logarithmic metabolite levels, E_{met} is a convex function; this greatly facilitates optimization and allows us to find the global minimum efficiently. Our online service for enzyme cost minimization and is freely available to the community. Users can run ECM for their own models. For the model in this paper, the optimization for one flux distribution takes several seconds, and for the complete set of all EFMs several minutes on a shared Dell PowerEdge R430 server with 32 intel xeon cores. Details can be found on the web page describing this case study (`http://www.neos-guide.org/content/enzyme-cost-minimization`).

Computing growth rates from enzyme-specific biomass production rates

The cell growth rate can be approximately computed from the enzyme cost of biomass production. In the spirit of Scott et al. [25], the growth rate of a cell is given by $\mu = v_{\text{BM}}/c_{\text{BM}}$, 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 time). We further define the enzyme-specific biomass production rate $r_{\text{BM}} = v_{\text{BM}}/E_{\text{met}}$, which would be exactly equal to the growth rate if the entire cell biomass were composed of central metabolism enzymes (i.e. the ones directly accounted for in E_{met}). Since that is not the case, we account for the conversion between E_{met} and c_{BM} using the following approximation $E_{\text{met}}/c_{\text{BM}} = \alpha_{\text{prot}}(a - b\,\mu)$, where $\alpha_{\text{prot}} = 0.5$ is the fraction of protein mass out of the cell dry weight, and $a = 0.27$, and $b = 0.2[\text{h}]$ are fitted parameters that approximate the proteomic fraction dedicated to central metabolism (as a linear function of the growth rate [25]). As shown in Supplementary Text S1.3, we obtain the following formula

$$\mu = \frac{\alpha_{\text{prot}} \; a \; v_{\text{BM}}}{E_{\text{met}} + b \; \alpha_{\text{prot}} \; v_{\text{BM}}} \,. \tag{1}$$

Note that in our model, the biomass flux, v_{R70}, is always set to 1 $[\text{mM s}^{-1}]$ and by pure unit conversion we obtain $v_{\text{BM}} = 7.45 \times 10^7 [\text{mg l}^{-1} \text{ h}^{-1}]$. As shown in the previous subsection, the total enzyme mass concentration is given by the formula $E_{\text{met}} = \sum_l \beta_l \, E_l$ in units of $[\text{mg l}^{-1}]$, so it requires no further conversion. The final formula for growth rate is thus

$$\mu = \frac{1.01 \times 10^7 \; [\text{mg l}^{-1} \text{ h}^{-1}]}{\sum_l \beta_l \, E_l + 7.45 \times 10^6 \; [\text{mg l}^{-1}]} \,. \tag{2}$$

Therefore, maximizing the growth rate μ is equivalent to minimizing $\sum_l \beta_l \, E_l$. See Supplementary Text S2.4 for more details.

The connection between biomass rate, the total enzyme mass concentration, and the growth rate can also be understood through the cell doubling time. We first define the enzyme doubling time $\tau_{\text{met}} \equiv \frac{\ln(2)}{r_{\text{BM}}} =$

$\frac{\ln(2)\cdot E_{\mathrm{met}}}{v_{\mathrm{BM}}}$, which represents the doubling time of a hypothetical cell comprised only of central metabolism enzymes. The doubling time of a whole cell would thus be

$$T = \frac{\ln(2)}{\mu} = \frac{\tau_{\mathrm{met}}}{\alpha_{\mathrm{prot}}\ a} + \frac{\ln(2)\ b}{a} = 7.4 \cdot \tau_{\mathrm{met}} + 0.51[\mathrm{h}] \tag{3}$$

Growth sensitivities

The sensitivities between enzyme parameters and growth rate can be approximated in the following way. A parameter change that slows down a specific reaction could to be compensated by increasing the enzyme level in the same reaction, thus keeping all metabolite levels and fluxes unchanged. For example, as a catalytic constant decreases by a factor of 0.5, the enzyme level needs to increase by a factor of 2. More generally, the cost increase for an enzyme follows from the simple formula $\Delta\mathrm{cost} = (\frac{k_{\mathrm{cat,old}}}{k_{\mathrm{cat,new}}} - 1)\,[\mathrm{old\ enzyme\ cost}]$. For other parameters, this local enzyme increase could be simply computed from the reaction's rate law. However, instead of adapting only one enzyme level, the cell may also adjust other enzyme levels, accept a change in metabolite levels, and therefore decrease its total cost even further. The additional cost decrease is only a second-order effect: for small parameter variations, it can be neglected, and the first-order local and global cost sensitivities are therefore identical (proof in SI section S4.2). Sensitivities to external parameters (e.g., extracellular glucose concentration) can be computed similarly. The growth sensitivities for a given EFM can be easily computed by multiplying the enzyme cost sensitivities by the derivative between growth rate and enzyme cost in a reference state.

Acknowledgements

We thank H.-G. Holzhütter, Joost Hulshof, Avi Flamholz, and Bas Teusink for fruitful discussions, and the Lorenz Center of Leiden University for providing a space for developing ideas. This work was funded by the Swiss Initiative in Systems Biology (SystemsX.ch) TPdF fellowship (2014-230) (to EN) and by the German Research Foundation (Ll1676/2-1) (to WL).

References

[1] A. Goel, M.T. Wortel, D. Molenaar, and B. Teusink. Metabolic shifts: a fitness perspective for microbial cell factories. *Biotechnology letters*, 34(12):2147–2160, 2012.

[2] T. Pfeiffer, S. Schuster, and S. Bonhoeffer. Cooperation and competition in the evolution of ATP-producing pathways. *Science*, 292(5516):504–507, 2001.

[3] S. Werner, G. Diekert, and S. Schuster. Revisiting the thermodynamic theory of optimal ATP stoichiometries by analysis of various ATP-producing metabolic pathways. *J. Mol. Evol.*, 71(5-6):346–355, 2010.

[4] 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.

[5] M.T. Wortel, E. Bosdriesz, B. Teusink, and F.J. Bruggeman. Evolutionary pressures on microbial metabolic strategies in the chemostat. *Scientific reports*, 6, 2016.

[6] M. Mori, T. Hwa, O.C. Martin, A. De Martino, and E. Marinari. Constrained allocation flux balance analysis. *PLoS Comput Biol*, 12(6):e1004913, 2016.

[7] M.T. Wortel, H. Peters, J. Hulshof, B. Teusink, and F.J. Bruggeman. Metabolic states with maximal specific rate carry flux through an elementary flux mode. *FEBS Journal*, 281(6):1547–1555, 2014.

[8] G.J. Velicer and R.E. Lenski. Evolutionary trade-offs under conditions of resource abundance and scarcity: experiments with bacteria. *Ecology*, 80(4):1168–1179, 1999.

[9] M. Novak, T. Pfeiffer, R.E. Lenski, U. Sauer, and S. Bonhoeffer. Experimental tests for an evolutionary trade-off between growth rate and yield in E. coli. *The American Naturalist*, 168(2):242–251, 2006.

[10] R. Maharjan Prasad, S. Seeto, and T. Ferenci. Divergence and redundancy of transport and metabolic rate-yield strategies in a single Escherichia coli population. *Journal of Bacteriology*, 189(6):2350–2358, 2007.

[11] J.M. Fitzsimmons, S.E. Schoustra, J.T. Kerr, and R. Kassen. Population consequences of mutational events: effects of antibiotic resistance on the r/K trade-off. *Evolutionary ecology*, 24(1):227–236, 2010.

[12] H. Bachmann, M. Fischlechner, I. Rabbers, N. Barfa, F.B. dos Santos, D. Molenaar, and B. Teusink. Availability of public goods shapes the evolution of competing metabolic strategies. *Proceedings of the National Academy of Sciences*, 110(35):14302–14307, 2013.

[13] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology*, 5(1):323, 2009.

[14] S. Schuster, L.F. de Figueiredo, A. Schroeter, and C. Kaleta. Combining metabolic pathway analysis with evolutionary game theory. explaining the occurrence of low-yield pathways by an analytic optimization approach. *Biosystems*, 105(2):147–153, 2011.

[15] F. Wessely, M. Bart, R. Guthke, P. Li, S. Schuster, and C. Kaleta. Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs. *Molecular Systems Biology*, 7(515), 2011.

[16] S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2):53–60, 1999.

[17] S. Schuster, D. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotech*, 18:326–332, 2000.

[18] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.-D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, 2002.

[19] S.J. Jol, A. Kümmel, M. Terzer, J. Stelling, and M. Heinemann. System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Computational Biology*, 8 (3):e1002415, 2012.

[20] S. Müller, G. Regensburger, and R. Steuer. Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. *Journal of Theoretical Biology*, 347:182–190, 2014.

[21] E. Noor, A. Flamholz, A. Bar-Even, D. Davidi, R. Milo, and W. Liebermeister. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLoS Comput Biol*, 12 (10):e1005167, 2016.

[22] H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, 271(14):2905–2922, 2004.

[23] Q. K. Beg, A. Vazquez, J. Ernst, M.A. de Menezes, Z. Bar-Joseph, A.L. Barabási, and Z.N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, 104(31):12663–12668, 2007.

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] Y.-F. Xu, D. Amador-Noguez, M.L. Reaves, X.-J. Feng, and J.D. Rabinowitz. Ultrasensitive regulation of anaplerosis via allosteric activation of PEP carboxylase. *Nat. Chem. Biol.*, 8(6):562–568, jun 2012.

[30] B.R.B. Haverkorn van Rijsewijk, A. Nanchen, S. Nallet, R.J. Kleijn, and U. Sauer. Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli. *Molecular systems biology*, 7(1):477, 2011.

[31] T.R. Maarleveld, J. Boele, F.J. Bruggeman, and B. Teusink. A data integration and visualization resource for the metabolic network of Synechocystis sp. PCC 6803. *Plant physiology*, pages pp–113, 2014.

[32] T. Fuhrer, E. Fischer, and U. Sauer. Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.*, 187:1581–1590, 2005.

[33] G. López-Torrejón, E. Jiménez-Vicente, J.M. Buesa, J.A. Hernandez, H.K. Verma, and L.M. Rubio. Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nature communications*, 7, 2016.

[34] J.M. Monk, A. Koza, M.A. Campodonico, D. Machado, J.M. Seoane, B.O. Palsson, M.J. Herrgård, and A.M. Feist. Multi-omics quantification of species variation of Escherichia coli links molecular features with strain phenotypes. *Cell Systems*, 3(3):238–251, 2016.

[35] K. B. Andersen and K. von Meyenburg. Are growth rates of Escherichia coli in batch cultures limited by respiration? *Journal of Bacteriology*, 144(1):114–123, 1980.

[36] S. Müller, G. Regensburger, and R. Steuer. Resource allocation in metabolic networks: kinetic optimization and approximations by FBA. *Biochemical Society Transactions*, 43:1195–1200, 2015.

[37] E.J. O'Brien, J. Utrilla, and B.O. Palsson. Quantification and classification of E. coli proteome utilization and unused protein costs across environments. *PLoS Comput Biol*, 12(6):e1004998, 2016.

[38] J.A. Lerman, D.R. Hyduke, H. Latif, V.A. Portnoy, N.E. Lewis, J.D. Orth, A.C. Schrimpe-Rutledge, R.D. Smith, J.N. Adkins, K. Zengler, and B.O. Palsson. In silico method for modelling metabolism and gene product expression at genome scale. *Nature communications*, 3:929, 2012.

[39] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210–1218, 2011.

[40] F. Vasi, M. Travisano, and R.E. Lenski. Long-term experimental evolution in Escherichia coli. ii. changes in life-history traits during adaptation to a seasonal environment. *The American Naturalist*, 144(3): 432–456, 1994.

[41] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae. *Bioinformatics*, 18(2):351–361, 2002.

[42] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1810(10):978–988, 2011.

[43] K. Valgepea, K. Peebo, K. Adamberg, and R. Vilu. Lean-proteome strains–next step in metabolic engineering. *Frontiers in bioengineering and biotechnology*, 3, 2015.

[44] A. Goel, T.H. Eckhardt, P. Puri, A. de Jong, F. Branco dos Santos, M. Giera, F. Fusetti, W.M. de Vos, J. Kok, B. Poolman, D. Molenaar, O.P. Kuipers, and B. Teusink. Protein costs do not explain evolution of metabolic strategies and regulation of ribosomal content: does protein investment explain an anaerobic bacterial Crabtree effect? *Molecular Microbiology*, 97(1):77–92, 2015.