Searching for principles of microbial physiology

Frank J Bruggeman[1,#], Robert Planqué[2], Douwe Molenaar[1] and Bas Teusink[1]

1. Systems Biology Lab, AIMMS, VU University, Amsterdam, The Netherlands
2. Dept of Mathematics, VU University, Amsterdam, The Netherlands

corresponding author: f.j.bruggeman@vu.nl

Abstract

Why do evolutionarily distinct microorganisms display similar physiological behaviours? Why are transitions from high-ATP yield to low(er)-ATP yield metabolism strategies so widespread across species? Why is fast growth generally accompanied with low stress tolerance? Do these phenomena occur because most microbial species are subject to the same selective pressures and physicochemical constraints? If so, a broadly-applicable theory might be developed that predicts common microbiological behaviours. Microbial systems biologists have been working out the contours of this theory for the last two decades, guided by experimental data. At its foundations lie basic principles from evolutionary biology, enzyme biochemistry, cellular metabolism, cellular composition, and steady-state cell growth. The theory makes predictions about fitness costs and benefits of protein expression, characteristics of optimal metabolisms, states of maximal growth rate, and physicochemical constraints on growth rate. Comparing the theory with experimental data suggests that microorganisms aim to express proteins to optimal concentrations in order to maximise growth rate, also in the presence of stresses. This review explains the current status of the theory; its roots, predictions, experimental evidence, and future directions.

Introduction

One goal of microbiology is to explain the differences and commonalities between microbes, given their genetic information, molecular biology, biochemistry, ecology and evolutionary history. For instance, why do so many unicellular microorganisms respire sugars at low growth rate and initiate seemingly wasteful overflow metabolism at higher growth rates? Why are fast growing cell less stress tolerant? Do stress tolerance and growth rate always trade off? Why do microbes often form tiny subpopulations of hardly growing, stress-tolerant persister cells, while the majority is stress-sensitive and fast growing? When is one metabolic pathway better to use than another? And what explains the variation of lag phases of microbial adaptation across conditions? Can a cell robustly steer protein expression to states that support increased, or perhaps even maximal, fitness in the face of environmental dynamics?

Many of these questions touch on fundamentals of evolutionary biology, enzyme biochemistry, cellular metabolism and cellular growth. In this review, we provide predictions and — hopefully — understanding by taking a systems biology perspective on cellular protein economy, growth and fitness, using experiments, models and theory. We will discuss a body of theory, explain its fundaments and provide experimental evidence. The theory predicts common behaviours of microorganisms from first principles, derived from evolutionary, biochemical and molecular-biological considerations. The basic premise is that microorganisms have been selected in evolution to maximise their growth rate; this maximal growth rate may be very low or even zero during stressed conditions. Microbes achieve this by expressing needed proteins to optimal concentrations, within physicochemical constraints that limit the concentrations and activities of proteins. An important aspects of this theory is that it considers the (optimal) expression of all cellular proteins. It aims to be as generic as possible. Like any theory, it needs to stand the test of empirical validation to be able to stand the test of time.

I. Foundational principles

Unity of microbial life?

Our planet is teeming with microbial life [1]. Microbial niches vary greatly and individual species vary in their mechanisms for energy and mass assimilation [2]. Microorganisms are in constant competition to scavenge nutrients for growth and survival. Selection filters out those that perform best, either alone or in communities.

A theory about the physiological consequences of growth-rate maximisation strategies by microorganisms can only be widely applicable if evolutionarily distinct species are alike in their molecular biology and selective pressures. This appears to be the case: microorganisms function very similarly [3, 4]; they all obey the same physicochemical laws and are composed out of the same types of macromolecules. Thus, their basic biochemistry and molecular biology is essentially identical. Moreover, macromolecule synthesis routes are largely conserved. The enzymes all work according to the same kinetic principles [5]. The metabolic reactions that occur follow the same limited set of basic chemical principles [6, 7]. It appears, therefore, that molecular mechanisms, biosynthesis and growth can all be

understood in terms of the same fundamental principles. Some of these are: conservation of chemical elements, reaction stoichiometry, Gibbs-free energy potentials, and energy-equivalent recycling [8]. Thus, as Jacques Monod put it: "What's true for *E. coli* is also true for the elephant."

Selection also has a highly similar influence on different microorganisms [9]. It is much like François Jacob once remarked, "The basic purpose and desire of each cell is to become two cells.". Accordingly, genotypes are selected that produce (the most) offspring, despite dynamic, sometimes harsh, conditions. They either fix, or stably co-exist in communities.

In this light, it is perhaps not a surprise that evolutionarily distinct microorganisms have such similar physiologies, i.e., metabolic behaviours. In 1924, Kluyver, then a Dutch pioneering microbiologist, called this similarity the "Unity of Biochemistry" [10, 11]. For instance, many microorganisms display wasteful metabolisms at high growth rate, while metabolising their nutrients more efficiently at low growth rate. Or, they prefer some carbon sources over others. They can withstand sudden stresses better at low growth rate than at high growth rate, as if they were prepared [12].

If indeed many microorganisms obey the same biochemical and metabolic principles, and they are subject to the same physicochemical constraints and selective forces, then a 'universal' theory might be developed that describes and predicts microbial physiology. This review is about this emerging theory: what its premises and predictions are, its roots, its experimental evidence, and future directions.

We provide and explain the basic equations in the main text rather than relegate them to appendices. We hope that we do not scare away experimentalists. In fact, we aim for the opposite, to provide the reader with concrete food for thought. We are explaining a *developing* quantitative theory after all, with well-defined quantitative, experimentally testable principles and models. We believe that this theory is useful, accessible and insightful to many microbiologists — theoreticians and experimentalists alike.

As is the case for all scientific endeavours, it is hard, if not impossible, to provide a complete overview of the underlying literature. We have, therefore, chosen to provide the main references that we think are minimally needed.

A universal fitness measure

The common concept of the fittest genotype is that it increases most in frequency when competing with others [9, 13]. This competition may occur in a dynamic environment with varying nutrients, periods of famine and stresses. The eventual winner, at some moment in time, left the most offspring: its fold change in abundance was the greatest.

Let's be precise. We denote the number of cells of a genotype at time t by $N(t)$. Its net fold change after (evolutionary) time t_E equals,

$$\frac{N(t_E)}{N(0)} = \frac{N(t_E)}{N(t_{E-1})}\frac{N(t_{E-1})}{N(t_{E-2})}\cdots\frac{N(t_2)}{N(t_1)}\frac{N(t_1)}{N(t_0)},$$

with $N(t_j)/N(t_{j-1})$ as the fold change in the j-th epoch, which we shall denote by w_j from now on, that lasted $\Delta t_j = t_j - t_{j-1}$ time. The net fold change equals the product of the fold changes of the sequence of epochs. Epochs may have different stresses or nutrient levels.

The geometric fitness F over the entire time period is defined as [14, 15],[1]

$$F = \frac{1}{t_E}\ln\frac{N(t_E)}{N(0)} = \frac{1}{t_E}\ln\prod_{j=1}^{N} w_j = \frac{1}{t_E}\sum_{j=1}^{N}\ln w_j = \frac{1}{t_E}\sum_{j=1}^{N}\Delta t_j\frac{\ln w_j}{\Delta t_j} = \left\langle\frac{\ln w}{\Delta t}\right\rangle,$$

with $\langle\,.\,\rangle$ denoting an average value. Note that $\Delta t_j/t_E$ equals the probability of observing epoch j. If only a single environment occurred, during which the micro-organism grew at a constant growth rate μ, then $F = \mu$; this explains why we consider the logarithms of the fold changes in abundances and divide by the total elapsed time.

[1] Note that the original literature defines the geometric mean fitness as $\frac{1}{N}\ln\frac{N(t_N)}{N(0)} = \frac{1}{N}\ln\prod_{j=1}^{N} w_j = \ln\left(\prod_{j=1}^{N} w_j\right)^{1/N}$ with $\left(\prod_{j=1}^{N} w_j\right)^{1/N}$ as the geometric mean of the w_j's.

F is indeed a fitness measure, because the ratio of the abundance of a mutant (N_m) over a wild type (N_m) evolves, in the simplest case as,

$$\frac{d}{dt}\ln\frac{N_m}{N_w} = \frac{d}{dt}\ln N_m - \frac{d}{dt}\ln N_w = \mu_m - \mu_w,$$

the difference between the fitness of the two strains, i.e., $F_m - F_w$. Clearly, the fastest grower becomes most abundant. Thus, one can think of F, when defined for a sequence of environments, as a net growth rate.

Two extreme cases can now be considered. Lewontin and Cohen [16] considered a randomly changing environment (the random limit). They assumed that the fold-change factors $w_j = N(t_j)/N(t_{j-1})$ are independent random variables (independent also of $N(t_{j-1})$). They argued that then

$$\langle(\ln w)/\Delta t\rangle \approx \ln\langle w/\Delta t\rangle - \langle\delta^2 w/\Delta t\rangle/(2\langle w/\Delta t\rangle^2)$$

with $\langle w/\Delta t\rangle$ and $\langle\delta^2 w/\Delta t\rangle$ as the mean and variance of the fold change per unit of time, respectively. This result indicates that, when conditions are very unpredictable the microorganism with the highest average growth rate, $\langle w/\Delta t\rangle$, will not necessarily win. It may lose from a competitor that has a lower fitness variance $\langle\delta^2 w/\Delta t\rangle$ across conditions, for example, because it is better equipped with signalling systems.

The non-random limit is reached when microorganisms always adapt perfectly and instantaneously to new environmental conditions (i.e., in the absence of lag phases and phenotypic heterogeneity); then, exponential growth occurs, and

$$w_j = \frac{N(t_j)}{N(t_{j-1})} = e^{\mu_j(t_j - t_{j-1})}.$$

Geometric fitness is now equal to the average specific growth rate across all conditions,[2]

[2] Note that $\frac{t_j - t_{j-1}}{T}$ is not the probability for environmental state j but for epoch j. Since epochs can have the same environmental state, the probability for an environmental state e is the sum of all epoch durations with e divided by the total time.

$$F = \frac{1}{t_E} \sum_j \ln e^{\mu_j (t_j - t_{j-1})} = \sum_j \frac{t_j - t_{j-1}}{t_E} \mu_j = \sum_j p_j \mu_j = \langle \mu \rangle.$$

Here p_j is the probability of environment j during which the growth rate equals μ_j. Maximal fitness now requires maximisation of the mean growth rate.

In between those two limiting cases lies any realistic case. The analyses then becomes more involved. Lag times, phenotypic heterogeneity and diversification, stresses, and fitness costs can, however, still be introduced. This makes the theory harder to interpret, but still intelligible [17, 18].

Different microorganisms may have different strategies to deal with the same dynamic environment. Accordingly, they vary in their fold change value in the same environment. Who wins depends on their net behaviour, their F value. This might be a microorganism that performs badly in some environments and really well in others. It can also be a microbe that performs reasonably well in all environments. It can be one that makes persisters cells to prepare for future, extinction-threatening conditions, one that senses and aims to adapt quickly, or one that is always stress tolerant, at the expense of its instantaneous growth rate. Who wins depends also on the exact dynamics of the environment. Using fitness theory, the pros and cons of the three fitness-maximisation strategies can be compared [e.g. 14, 15, 19].

The theory we will introduce below assumes that microorganisms always aim to maximise their growth rate. This leads, without major assumptions, to microorganisms that have a mixed growth-rate-maximisation and stress-tolerance strategy that is dependent on growth rate. At low growth rates, they prepare for future adverse conditions. At high growth rate, they invest all resources into growth, at the expense of stress tolerance, and phenotypically diversify into growing and stress-tolerant subpopulations. To us, this makes sense as most environments are unpredictable, certainly from the perspective of the microorganism. In all cases, growth rate is limited by constraints acting on protein concentrations — in addition to physicochemical limits.

Genetic variation

Genetic variation arises continuously in microbial populations, e.g. during DNA replication, due to toxins, etc. The mutation frequency of *E. coli* MG1655 equals 2.2×10^{-10} mutations per nucleotide per generation or, equivalently, 1×10^{-3} mutations per genome per generations [20]. For *E. coli,* with a genome size of 4639675 bp's, this means that a population size of 1000×4639675 cells contains on average all single point mutations. That is about 5×10^{9} cells, which corresponds to 1 ml of a *E. coli* M9 -glucose culture at an OD_{600} of 1. Hence in 5 ml of such a culture, all single point mutations are present with certainty.

Fixation of microbial genotypes is fast: it only takes 530 generations before a single, 5%-faster-growing mutant-cell has reached a frequency of 50% in a population that started with 10^{8} cells. Many laboratory evolution experiments indeed show substantial fitness increases in less than 1000 generations [21, 22]. One cannot exclude, in principle, that microorganisms have evolved optimal circuitry for regulation of protein expression that maximise growth rate. Certainly not on the presumption that their evolution would be slow. On the contrary, we think.

Although evolution is fast, we do not have to worry about it much during growth experiments in the lab. When cells are grown from a stock in batch or chemostat to study cell physiology, roughly 10-20 generations are considered, and fixation of genetic variants does not play an important role. Therefore, during the studies that we provide as evidence below, we will generally not have to consider different genotypes.

Evolution: innovation and pruning

The mean growth rate $\langle \mu \rangle$ of a population of competing genotypes changes in time because faster growing genotypes become relatively more abundant. In 1930 Fisher published a theorem [23], now known as Fisher's theorem, indicating that the rate of change of $\langle \mu \rangle$ equals the genetic variation in the growth rate, i.e., $\frac{d}{dt}\langle \mu \rangle = \langle \delta^2 \mu \rangle$. This corresponds to intuition: If a greater spread (variance) in fitness exists, some

genotypes grow much faster than others, outgrow them quickly and the population growth rate varies greatly.

However, evolution is, for all we know, everlasting; there is at least no reason to believe it is now nearing its end. How can a microorganism then ever behave close to its optimal behaviour if evolution continues indefinitely? We need to address this question, because in the theory we shall be outlining below, we will assume that microorganisms operate close to their optimal behaviour.

The answer is twofold. First, a microorganism can attain a maximal growth rate only *given its genotype*, by expressing the right combinations of proteins at the right concentrations (without its growth rate being *evolutionarily* maximal; it is maximal given its current capacities).

Second, we should distinguish two complementary evolutionary processes, innovation and pruning. Innovation is the evolutionary emergence of new capabilities, such as the evolution of novel metabolic pathways that degrade plastics, via horizontal gene transfer or evolution after gene duplication. Pruning refers to the improved exploitation of existing capacities. It proceeds via mutations too, such as via those that lead to improved protein expression. Pruning improves phenotypic adaptation; innovation facilitates exploration and niche expansion.

Evolutionary pruning improves phenotypic adaptation to growth-supporting conditions, leading to faster adaptation and an increased growth rate. Mutations in key transcription factors or in promoter sequences are an example [21, 24].

One possible outcome of evolution is that microorganisms have evolved protein-expression control strategies that allow them to maximise growth rate in each condition, and that these strategies work optimally regardless of which metabolic proteins are being regulated [25, 26]. Thus, evolution may have stumbled on regulatory motifs of protein expression that are capable of optimal expression across conditions, regardless of conditions.

Thus, after four billion years of evolution, *current* microorganisms may have optimal protein-expression control systems that maximise growth rate, even though

they are still subject to evolution. For something as important and conserved as ribosomes, there is convincing evidence that this is indeed what cells are capable of (see below) [27, 28].

Thus, in our view, a cell may be able to express proteins optimally to give rise to the maximal growth rate. Accordingly, the objective of many, if not all, (gene-expression) control systems in a cell is then ultimately to contribute optimally to this "goal". We will say more about this control strategy below, where we will make it explicit. Whether or not a microorganisms always reaches its maximal growth rate, because its proteins have been optimality expressed, does not mean that its growth rate cannot still increase via mutations, e.g., via those that alter kinetic constants of enzymes.

The pressing question is therefore whether microbes are indeed capable of optimal expression of its proteins to maximise its growth rate, such that, regardless of the environmental conditions they are confronted with, they will adapt their protein expression until growth rate is maximal in that condition. No alternative protein expression state would then have a higher growth rate.

Optimal protein expression can be tested experimentally and compared to theoretical expectations. And, indeed, experiments indicate optimal protein expression. But before we review the evidence, we will shortly discuss the quantitative definition of growth rate in constant conditions.

Balanced growth of cell populations

When a population of (isogenic) microorganisms is cultivated under constant, excess-nutrient conditions it generally settles in a state of steady-state growth. Then, all the properties that are proportional to mass (extensive properties) increase exponentially in time, at a fixed rate (Figure 1). The ratios of extensive properties (intensive properties, such as concentrations;) remain constant. This state is formally called "balanced growth" [29].

Two views exist on the balanced growth state: a population-level, macroscopic view and a single-cell, microscopic view. The microscopic view [e.g. 30] has recently been

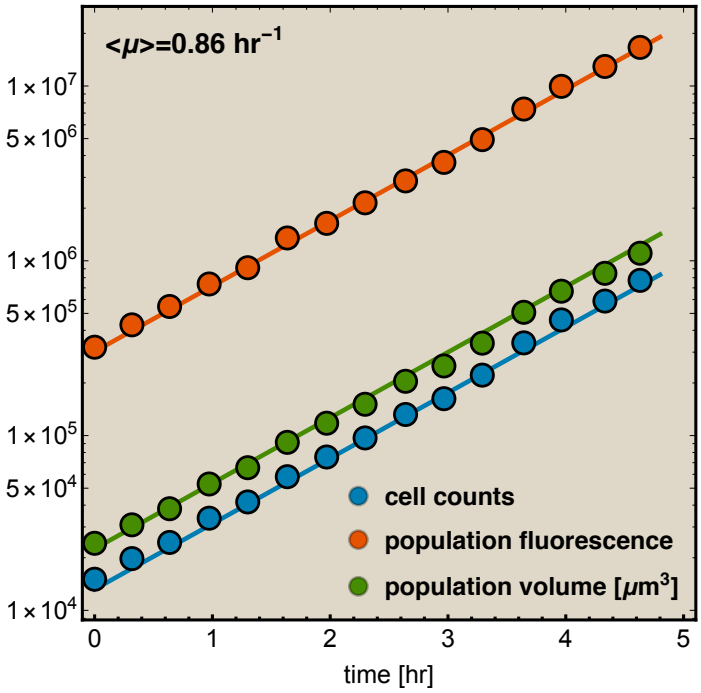

Figure 1. Experimental illustration of balanced growth. A fluorescent-protein expressing *B. subtilis* strain was grown in mineral medium on glucose in shake flask. Samples of it were measured in a flow cytometer. This data was reproduced from Nordholt *et al.* (Nordholt, van Heerden *et al.* 2017) .

reviewed [31]. The theory we discuss in this review is limited to the macroscopic view. In that description, we refer to the average cell, typically at steady state. It is, therefore, relevant to consider to which state of a single cell that corresponds.

In the microscopic perspective on balanced growth, individual cells differ in their birth and division sizes, durations of their cell cycles and molecular content, all due to the inherent stochasticity of molecular processes [31]. At balanced growth, such single-cell properties obey time-invariant probability distributions [31]. Thus, over time, the observed sizes at birth and division all obey, for instance, normal distributions with fixed variances and means. These distributions can be measured with real-time imaging of the growth of single cells [32], using, for instance, also fluorescent reporters of gene expression. Results from single-cell experiments are generally remarkably close to theoretical expectations [33].

The balanced growth state of a cell in the macroscopic theory corresponds to a cell in the microscopic theory that has the mean age $\langle a \rangle$ in the growing population [31].

This mean age is related to the mean generation time $\langle t_g \rangle$ of cells as (see the Appendix)

$$\langle a \rangle = (1/\ln(2) - 1)\langle t_g \rangle \approx 0.44\langle t_g \rangle.$$

Thus, in the macroscopic theory we refer to the average cell that has completed its cell cycle by about 44%. We can also use the microscopic theory to calculate the mean copy number of a molecule, made by a zero-order process with rate constant k, that the average cell contains. Then, we indeed recover the expected macroscopic value, k/μ, with μ as the (balanced) growth rate of the population (see the Appendix). We note that these two results hold when it is assumed that no variation of generation times exists, thus the variance of generation times $\langle \delta^2 t_g \rangle$ equals zero (see Appendix). The assumption that $\langle \delta^2 t_g \rangle \to 0$ can therefore be viewed as a macroscopic limit of the microscopic description. Since real populations have a nonzero variance in their generation times the macroscopic theory is an approximation.

In balanced growth, the extensive properties all grow exponentially in time, at the same specific growth rate μ (Figure 1) [29, 31]:

$$\mu = \frac{1}{M}\frac{dM}{dt} = \frac{1}{V}\frac{dV}{dt} = \frac{1}{N}\frac{dN}{dt} = \frac{1}{n_i}\frac{dn_i}{dt}$$

with M is the (bio)mass of the culture, V the total cell volume of the culture, N the number of cells, and n_i the copy number of the i-th molecule.

We note that the growth rate μ is sometimes called the specific growth rate. It captures the rate of synthesis of new cells per unit cell, or, equivalently, the rate of protein synthesis per unit protein. In the text below, we shall often omit the adjective “specific”.

Finally, individual cells are not always in the average state that the macroscopic description of balanced growth refers to. This becomes clear when we, for instance, consider *E. coli*, which can have generation times that exceed the replication time of its genome [34, 35]. Accordingly, DNA replication is not continuous along the cell cycle. It starts after a certain time after birth. The concentration of DNA changes due to its synthesis, whereas before that time it only decreased because of dilution

by cell-volume growth. The same applies to septum formation, which is happening at later times in the cell cycle [35]. Thus, individual cells do not experience constant concentrations of all molecules during their cell cycle and the metabolism of single cells may adapt during the cell cycle [35].

Thus, the macroscopic theory of balanced growth that we present below describes the average behaviour of a population of single cells.

At balanced growth, metabolism is at steady state

From the relations for extrinsic properties given above, it follows that the concentration $c = n/V$ of all molecules in the cell are constant at balanced growth (see Appendix). The rates of all cellular reactions are then constant too, as they depend on those (constant) concentrations [8].

If we consider the stoichiometry and rates of all cellular reactions then we obtain at balanced growth that [8, 26, 36],

$$\frac{1}{V}\frac{d\mathbf{n}}{dt} = \mathbf{N}\mathbf{j}(\mathbf{c}) = \mu\mathbf{c},$$

with $\mathbf{N}$ as the stoichiometry matrix (containing the reaction stoichiometries), $\mathbf{j}(\mathbf{c})$ as the steady-state flux vector that depends on the concentrations (and kinetic and environmental parameters), and $\mu\mathbf{c}$ capturing the dilution of molecules due to cell-volume growth. This equation tells us that the concentration of molecules stays constant when their net synthesis rate is balanced by an equal dilution rate.

Note that this formalism is not restricted to metabolites only, but also applies to the concentrations of macromolecules at balanced growth, such as proteins, lipids and RNA/DNA. When only metabolism is considered, dilution by growth is neglected based on the (often implicit) assumption that metabolic fluxes (j_i) are much faster than dilution by growth. In steady state, we then arrive at $\mathbf{N}\mathbf{j} \approx \mathbf{0}$. Flux values can now be estimated using linear programming methods, such as Flux Balance Analysis (FBA) [37]. We will come back to this modelling formalism shortly.

The entries of the flux vector are in fact rates of enzymatic reactions whose kinetics depend on the concentrations, e.g., following Michaelis-Menten kinetics. We find it

useful to distinguish the rate (or activity, symbol v) of an enzyme and a particular steady-state flux through that reaction, denoted by j. The rates v can be a function of time. We will further only consider the steady (or balanced) state, to which the dynamics generally settle if the environmental conditions are constant for long enough [38].

Relevant properties of enzyme kinetics

Many different catalytic mechanisms of enzymes exist [5, 39-41]. One reason is that enzymes vary in their number of substrates, products and effectors. Another is that the order of binding of the reactants can vary. This leads to different catalytic mechanisms, e.g., ordered, random, ping pong, etc. However, all enzyme-catalysed reactions, in the absence of metabolite channeling, have something in common: the rate of an enzyme-catalysed reaction is proportional to the concentration of the enzyme. It can be shown that for all enzyme mechanisms, the rate of the enzyme obeys $v(\lambda e_T) = \lambda v(e_T)$ (see Appendix) [42].

Furthermore, enzyme-kinetic rate equations have a common structure [43],

$$v = k_{cat}^{+} \cdot e \cdot f(\mathbf{m}),$$

with k_{cat}^{+} as the forward catalytic rate constant, e as the concentration of enzyme, and $f(\mathbf{m})$ is a nonlinear function of metabolite concentrations ($\mathbf{m}$) and parameters (e.g. kinetic constants and environmental conditions).

The rate of an enzyme has an upper and a lower bound $-k_{cat}^{-}e < v < k_{cat}^{+}e$, i.e., the maximal forward rate of the enzyme equals $V_{max}^{+} = k_{cat}^{+}e$ and the maximal backward rate $V_{max}^{-} = k_{cat}^{-}e$ [5].

The function $f(\mathbf{m})$ is sometimes called the saturation function (although strictly speaking, $f(\mathbf{m})$ corresponds only to enzyme saturation if the reaction is irreversible) and is bounded:

$$-k_{cat}^{-}e < v < k_{cat}^{+}e \quad \Rightarrow \quad -\frac{k_{cat}^{-}}{k_{cat}^{+}} < f(\mathbf{m}) < 1,$$

which we exploit below.

These are the only enzyme kinetic principles used in the theory.

Physicochemical limits

Rates of biosynthetic reactions are bounded by the concentrations of the catalysing enzymes [5], which are, in turn, bounded by the cellular capacity to contain proteins in its compartment [44]. Those bounds of enzyme-catalysed reaction rates are "cellular".

Rate limits also exist that are "physical". They provide an upper bound for reaction rates of bimolecular reactions in the cytosol, in the membrane, and between a membrane-embedded and an extra- or intracellular molecule [45-47]. Physical limits are hit when enzymes convert substrates faster into products than the rate at which diffusion can replenish these, so that concentration gradients occur. This effect is enhanced when substrates are scarce.

Due to the unintuitive, erratic properties of random walks, the number of membrane proteins required for maximal collision rates of extracellular substrates and membrane receptors is, however, much lower than one would expect, far below 1% [45]. Yet, membranes are packed with protein [48].

Although increased concentrations of enzymes enhance reaction rates [5], it also enhances crowding and viscosity, which lowers diffusion rates [49]. It has been suggested that protein synthesis inside cells operates close to its diffusion limit [50]. It has also been proposed that cells have an optimal protein density that maximises reaction rates [51].

Taken together, this suggests that optimal allocation of biosynthetic resources over protein synthesis (and its reallocation when conditions change to prevent growth-rate reducing over- and underexpression of proteins) indeed has an impact on cellular fitness. Below we will provide evidence that microbes indeed appear to display such optimal behaviour.

II. Fitness effects and optimality of protein expression

Biosynthetic resources for protein expression are limited

Cells have a finite rate of protein synthesis, because of finite concentrations of available RNA polymerases, ribosomes, amino-acid loaded tRNAs, etc., that limit enzyme rates [52]. Thus, "genes compete for biosynthetic resources". The biosynthetic resource demand of one gene that suddenly becomes active will come at the expense of the biosynthetic resources used by others.

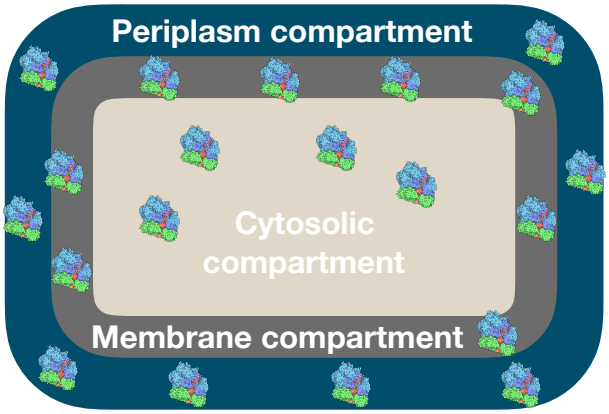

Figure 2. Cellular compartments have finite protein storage capacities.
In the theory, protein compete for biosynthetic resources, like RNA polymerases, sigma factors, nucleic acids, ribosomes, amino acids, etc., and space, as shown in this figure.

During steady-state growth, each active gene has a fixed biosynthetic resource demand, required to keep the concentration of its cognate protein constant in the face of degradation and dilution by growth. Cells have a finite capacity to store proteins (e.g., in their periplasm, plasma membrane or cytoplasm; Figure 2), to insert proteins in membranes, and to export proteins over membranes.

We believe that biosynthetic resource competition has shaped cell physiology [52]. It leads to bounds on cellular growth rate — in addition to physicochemical bounds — that enforce particular metabolic behaviour. In our view, in agreement with that of others [27, 53-55], evolution has led to optimal allocation of limited biosynthetic resources over cellular proteins, such that protein concentrations are optimal and maximise growth rate.

Growth-rate effects of the allocation of limited biosynthesis

It turns out that thinking about the outcomes of the allocation of limited resource allocation is rather straightforward. Consider, for instance, beta-galactosidase, which is required for lactose growth, but a burden during glucose growth. A cell growing on lactose as its sole carbon source does not grow when it does not express beta-galactosidase. Thus, the growth rate increases when a cell starts to express it: the protein is now still underexpressed. Above some threshold expression level (the *optimal* level), the growth rate will reduce again, because beta-galactosidase is overexpressed and its synthesis goes at the cost of other protein concentrations. One should therefore expect some optimum in the protein expression.

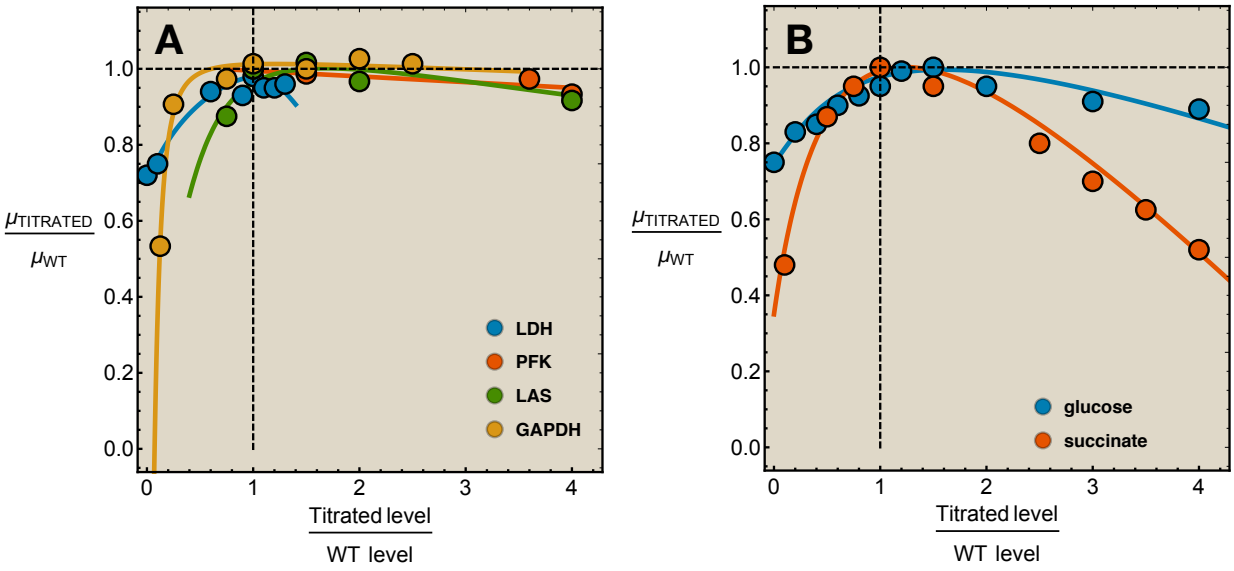

Figure 3. Illustration of enzyme titration and optimal protein-expression by a wild type *L. lactis* (A) and *E. coli* (B) strain. **A.** Three glycolytic enzymes and an operon display optimal expression levels in *L. lactis* (Koebmann et al., 2002, Koebmann et al., 2005, Koebmann et al., 2006, Solem et al., 2007, Solem et al., 2008). **B.** H^+-ATPase of *E. coli* is optimally expressed in two growth environments (Jensen et al., 1993).

This has been tested experimentally with enzyme-titration experiments (Figure 3). The microorganism is then grown under a condition where the titrated protein is needed, and the growth impact is the difference between the benefit and the cost of a protein [e.g. 56, 57]. In the hypothetical case that the enzyme would not cost anything and we only look at the benefit of a protein, one should expect a hyperbolic relationship, with maximal benefit at low enzyme level, and diminishing benefit as the enzyme level increases to higher levels when it becomes progressively more

overexpressed [58]. Protein costs are expected to reduce the growth rate in a linear manner, as we will discuss below.

The relation between growth rate and the (titrated) expression level of a needed protein therefore generally displays an optimum [56]. Remarkably, it is often observed that the titrated optimum *coincides with the wild type growth rate and corresponding protein expression* (Figure 3) [59-63]. (However, exceptions exist too, perhaps hinting at evolutionary trade offs [64, 65]). Such optimal protein expression by the wild type has indeed been found for *Lactococcus lactis*, *Escherichia coli* and *Saccharomyces cerevisiae,* three microorganisms that are evolutionarily unrelated; *L. lactis* is a gram-positive prokaryote, *E. coli* is a gram-negative prokaryote, and *S. cerevisiae* is a eukaryote [59-63, 66, 67].

Thus, optimal protein expression has been found at different growth conditions, for different proteins. We view this as a stunning result that indicates to us that microorganisms are capable of optimising protein concentrations.

The fitness costs of an unneeded enzyme

The cost of a protein can be investigated by artificially titrating the level of an enzyme under conditions when it is not needed. Titration of, for instance, beta-galactosidase under glucose growth then indeed reduces growth rate [68-71] (Figure 4).

In those studies, one result is particularly surprising: the relationship between growth rate and unneeded protein expression is linear. We can explain this behaviour with a simple model if we accept one peculiar assumption, that of “even competition for resources” [58]. All pre-existing protein concentrations then reduce by the same fraction λ after the gene activity of an unneeded protein is changed and a new steady state of growth is established.

We denote the total (cellular) protein concentration (total proteome) by p_T. When the unneeded protein is not expressed, p_T equals the sum of all needed protein concentrations

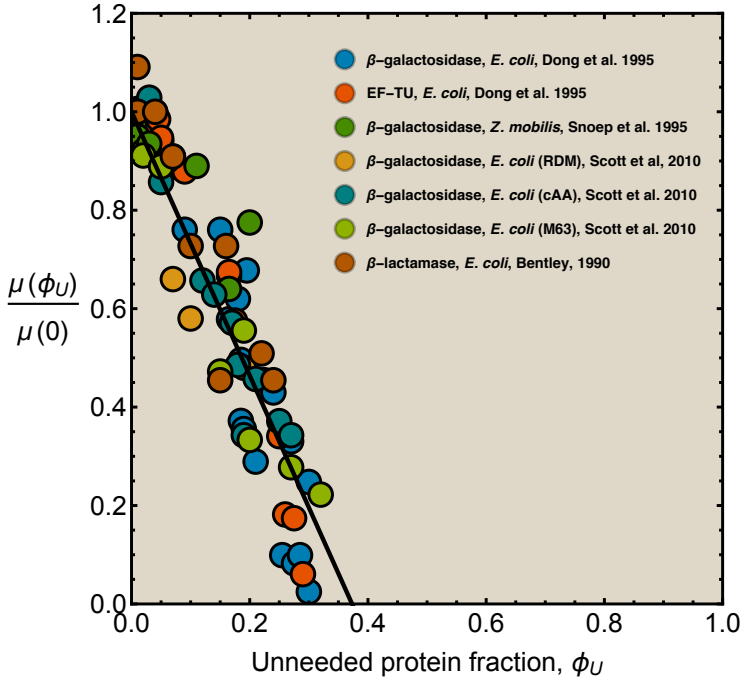

Figure 4. Illustration of the fitness cost (growth-rate reduction) due to the expression of an unneeded enzyme. The black line illustrates a fit with slope -2.7, indicating that growth rate is zero at an unneeded protein expression of 37% (Bentley et al., 1990, Dong et al., 1995, Snoep et al., 1995, Scott et al., 2010).

$$p_T = \sum_i p_{i,N}.$$

When unneeded protein is expressed, the proteome available for needed proteins decreases:

$$p_T - p_U = p_T(1 - \phi_U) = \sum_i p'_{i,N}$$

with $\phi_U = p_U/p_T$ and $p'_{i,N} = (1 - \phi_U)p_{i,N}$. Thus, each needed protein is reduced in concentration by the factor $(1 - \phi_U)$. We already established that the activity of an enzyme, v, is proportional to its concentration such that $v_i(\lambda p_{i,N}) = \lambda v_i(p_{i,N})$; Hence, $\lambda = 1 - \phi_U$.

Since the growth rate μ of a cell equals its overall protein synthesis j_p divided by total cellular protein (derived below), we obtain that,

$$\mu(\lambda p_T) = \frac{j_p(\lambda p_T)}{p_T} = \lambda \frac{j_p(p_T)}{p_T} = \lambda \mu(p_T).$$

Thus, finally we obtain $\mu(p_U)/\mu(0) = 1 - \phi_U$ in the even (or passive) competition model. The growth rate decreases linearly with the unneeded protein fraction ϕ_U.

However, the experimental data of Figure 4 suggests that the slope deviates from -1. One possible explanation is the following: the unneeded protein is always expressed in a particular compartment of the cell, say in the cytosol. Then the unneeded proteins compete for resources with other cytosolic proteins, and its expression fraction amongst all the proteins it competes with equals p_U/p_T. But we always plot the fraction of unneeded protein in the entire protein pool P_T of the cell. Thus, we plot p_U/P_T. Then we have to change the equation we derived above to,

$$\frac{\mu(p_U)}{\mu(0)} = 1 - \frac{P_T}{p_T}\frac{p_U}{P_T},$$

with $P_T/p_T > 1$. The data shown in Figure 4 has a slope of -2.7 indicating that $p_T/P_T \approx 0.37$, suggesting that 37% of all cellular protein is cytosolic. (An improvement of this equation can be obtained by also taking into account that this protein competes only with proteins using the same sigma factor. We do not do this here.)

The fitness potential of an enzyme

To better understand growth-rate versus needed-protein concentration relationships, we derived theory that predicts their slopes [58]. We again made the assumption that the total protein concentration of a cell is constant, and that the rate of an enzyme-catalysed reaction is proportional to the concentration of enzyme. We also assumed that the experimentalist sets the concentration of the titrated protein and that the cell optimally allocates the remaining protein concentrations over its needed reactions. As before, we defined the growth rate as the protein synthesis rate j_p divided by the cellular protein concentration p_T.

Under those assumptions, the normalised slope of the dependence of the growth-rate on the concentration of protein i equals (see Appendix) [58]

$$\Omega_i = \frac{d\ \ln j_p/p_T}{d\ \ln p_i} = \frac{C_{p_i}^{j_p} - \frac{p_i}{p_T}}{1 - \frac{p_i}{p_T}}\ ,\ \text{with}\ C_{p_i}^{j_p} = \frac{\partial \ln j_p}{\partial \ln p_i}.$$

We refer to this coefficient as the fitness potential of an enzyme. This equation expresses the change in the growth rate (fitness) when the concentration p_i of a protein i is changed. The equation gives an exact definition of the terms enzyme underexpression ($\Delta F_i > 0 \rightarrow C_{p_i}^{j_p} > p_i/p_T$) and overexpression ($\Delta F_i < 0 \rightarrow C_{p_i}^{j_p} < p_i/p_T$).

The $C_{p_i}^{j_p}$ coefficient indicates whether a protein has any fitness effect at all. (In metabolic control analysis such terms are called flux control coefficients [72, 73].) $C_p^{J_p}$ quantifies the fractional change in the protein synthesis rate when the protein concentration is slightly perturbed, without the readjustment of the other enzymes to an optimal state. If $C_{p_i}^{j_p}$ is high, the protein has a strong control on protein synthesis. Enzymes that have a $C_p^{j_p}$ that is close to 1 have the greatest influence on fitness, regardless of their protein fraction. Unneeded proteins have a $C_{p_i}^{j_p}$ of zero. For such proteins, integration of Ω_i again reveals the linear relation between growth rate and the unneeded protein fraction (see Appendix).

Abundant enzymes have low $1 - p/p_T$ values, and, therefore, have according to the Ω_i equation, a stronger fitness influence than scarce enzymes. Abundant enzymes, such as ribosomes or the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase, should therefore be more carefully tuned in concentration than transcription factors, which are low in abundance. This also makes sense, as an abundant enzyme with say $p/p_T \approx 0.2$ that is 5% removed from its optimal level corresponds to a significant waste of resources. This suggests that an order exists for the evolutionary need of tuning enzyme concentrations: the order of their abundance. Evolution then proceeds in the direction of diminishing returns.

At the optimum, a change in the enzyme level can only reduce fitness. In this optimal state, we find that the flux control coefficient equals the fraction abundance of the protein: $C_{p_i}^{j_p} = p_i/p_T$ [74]. This indicates that abundant proteins have a high control coefficient on protein synthesis and therefore on the growth rate.

Properties of genetic circuitry that can optimise protein levels

Thus, cells are capable of expressing needed protein to optimal concentrations, across conditions (Figure 3). This implies that their molecular control circuitry, responsible for (tuning of) protein expression, work in such a way that growth rate is maximised [25]. Microorganisms achieve robust growth-rate maximisation, because their regulatory circuitry works optimally and steers protein expression to optimal states. But isn't optimal steering of gene expression a "biochemically hard problem"?

Typically, in microorganisms, regulation of gene expression of (metabolic) proteins is achieved by binding of a metabolic intermediate — the "sensor" — to a transcription factor, affecting its affinity for promoter sequences. For instance, fructose-1,6-bisphopshate binds to the transcription factor Cra and inhibits Cra from binding to promoter sequences [74].

Several theoretical studies have focussed on finding a characterisation of optimal genetic circuits, using numerical methods [25-28, 65, 75, 76]. These papers indicate that numerically parameterising a circuit that achieves optimal steering is rather straightforward (see Appendix). The idea is to first solve the optimal allocation of resources over protein synthesis numerically as function of environmental conditions, using numerical optimisation of only the metabolic model. Then the optimal relationship between the optimal sensor concentration and the metabolic proteins are known. This relationship is the optimal input-output relation of the genetic circuit. This optimal genetic circuit is found by fitting its steady-state behaviour to this input-output relationship. When we did this for galactose metabolism in *S. cerevisiae* it turned out that only a few kinetic parameters in the gene circuit influenced optimality, as if evolution of only those could tune the optimal input-output relationship. The remaining could, for instance, be useful for noise suppression [75].

We investigated the general mathematical formulation of optimal control of gene expression [25]. Its most important finding was that if a metabolic network, with a single total-protein constraint, should display robust optimisation against N environmental parameters, e.g., nutrient concentrations, then minimally N

metabolites should control gene expression. It turns out that simple biochemical binding events between metabolites, transcription factors, and promoters are sufficient to achieve such optimal control of gene expression. This corresponds, e.g., to the binding of a metabolite to a transcription factor of which two, for instance, cooperatively bind to the DNA to enhance transcription. We provide the theoretical analysis of an example in the Appendix.

The observed optimal expression of protein requires that regulatory circuits of gene expression that can optimise. These are, surprisingly, biochemically not hard to construct, suggesting that is also not a huge problem for evolution to have stumbled onto them, independently and many times [25].

Optimal regulation of ribosome concentration in *E. coli*

Ribosomes are abundant proteins and, therefore, if we can use the fitness potential equation as a guide, their synthesis should be carefully tuned to keep them at their optimal levels at all times. Two studies analysed the regulation of the ribosome concentration in *E. coli* from the perspective of optimal resource allocation [27, 28]. Those studies have been inspired by work that dates back decades ago [53-55, 77]. These studies agree on the statement that ribosome concentrations are tuned in such a way *E. coli* to prevent overexpression, about 85% of the ribosomes are actively translating.

The regulation of ribosome concentration with growth rate works as follows [78]. When too many ribosomes are expressed, or when an amino acid is limiting, so that the unloaded-tRNA concentration is high, ribosomes bind more often to unloaded tRNA, leading to the synthesis of ppGpp by RelA. When ppGpp is made, it can bind to ribosomal promoters, in concert with DksA, lowering their affinity for RNA polymerase. RNA polymerase then binds more often to non-ribosomal (catabolic) operons, leading to an enhancement of amino acid synthesis, an increase of loaded tRNA concentration, decrease in ppGpp, and a reestablishment of a steady state at higher growth rate. When ribosomes are underexpressed, for instance when nutrient conditions improve, the converse happens; ppGpp concentration is low, more RNA polymerases allocate to the ribosomal promoter to enhance the ribosome concentration, and establish a novel steady state, again at a higher growth rate.

Using simulations, Bosdriesz *et al.* [28] and Scott *et al.* [27] showed that the known regulatory circuitry, involving ppGpp, is indeed able to maximise cellular growth rate, defined as the protein synthesis flux per unit invested protein. This suggests that indeed *E. coli* controls its ribosome concentrations in a manner that maximises its growth rate. That *E. coli*'s expression of ribosomes might be close to optimal is also shown by Bollenbach *et al.*, who sequentially removed ribosomal promoters (*E. coli* has 7 of them) and found a growth rate maximum at the promoter combination active in the wild type [79].

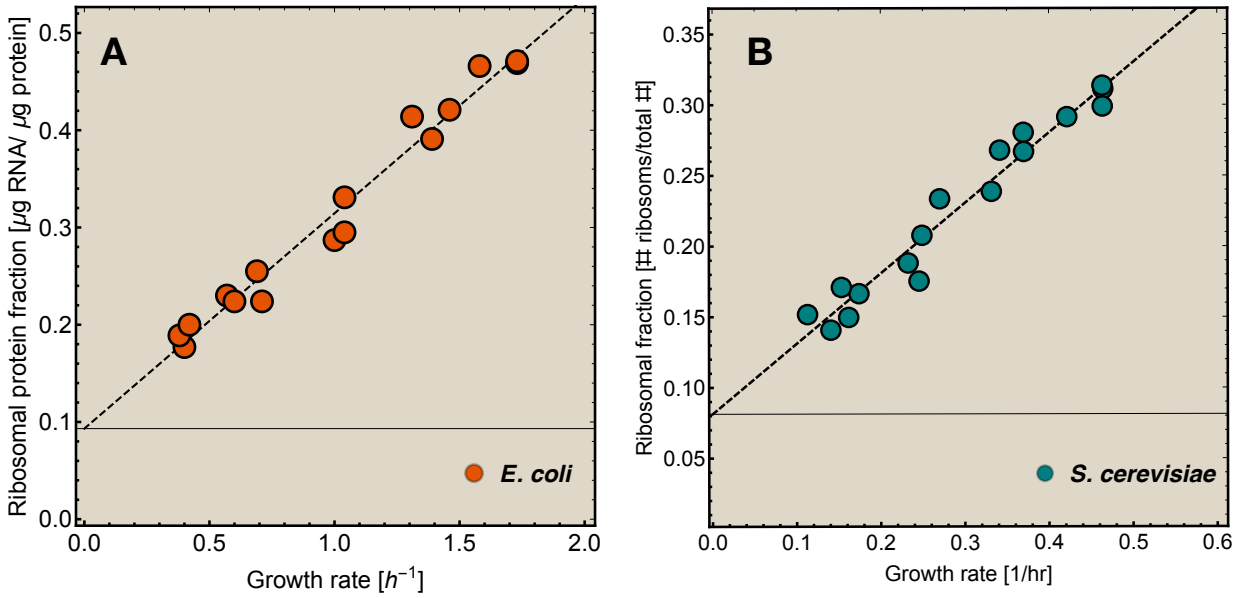

Figure 5. Linear relation between ribosomal activities and growth rate in *E. coli* and *S. cerevisiae*. Data is from Scott *et al.* (2010) and Metzl-Raz *et al.* (2017).

One view on the regulatory circuit for ribosome expression is, therefore, that its control objective is to prevent wasteful under- and overexpression of ribosomes by keeping its saturation function $f_r(\mathbf{m})$ as close to 1 as possible. This had already been suggested decades ago [54, 55]. In fact, Kjeldgaard wrote in 1963 [53], five years after he, together with Maaloe and Schaechter, established the (almost) linear relationship between ribosome concentration and the growth (Figure 4),

"*In other words, the number of protein-synthesising units within the bacteria is regulated in such a way that individual units always function at the same rate. If this fixed rate of protein synthesis per ribosome* [note: $f_r(\mathbf{m})$] *is assumed to be*

optimal, or nearly so, it is clear that RNA synthesis in the cell is regulated in a manner which affords a high degree of economy to the growing cell. This again would mean a selective advantage in the competitive environment in which bacterial species have probably evolved. It would be tempting to assume that such constant efficiency also applies to the synthesis of other macromolecular species of the bacteria."

When the saturation function is (high and) constant, the ribosomal protein fraction in *E. coli* should become a linear function of the growth rate, which we shall show below. This relation was first found experimentally in 1958 [77] and has recently been studied by Scott *et al.* [71] (Figure 4).

We expect that the linear relation between ribosomal protein fraction and growth rate holds for all microorganisms. In fact is has been found for quite a few microorganisms already [71, 80, 81]. Suggesting that they all try to prevent ribosome overexpression that would be a fitness cost. The linearity may have a surprising evolutionary origin. When the maximisation of a flux per unit protein is the objective, it is advantageous that enzymes have high affinities for substrates (to enhance substrate saturation) and low affinities for products (to prevent inhibition). This can therefore be expected to happen in evolution (assuming this introduces no trade-offs) as well. It turns out that this condition is exactly the limit in which the relationship between growth rate and protein fraction becomes linear. (We illustrate this with a simple example in the Appendix.) This might still be an unlikely coincidence, but we doubt that.

That evolution towards maximisation of flux per unit invested protein leads to enhanced affinities for substrates and reduced affinities for products can also shown theoretically. Consider a linear metabolic pathway for which we aim to maximise the steady-state pathway-flux per unit invested protein. At steady state, all enzymes carry the same flux, $J = v_1 = v_2 = \ldots = v_r$, and the total enzyme contraction equals $e_T = \sum_{i=1}^{r} e_i$. Since maximisation of J/e_T is the same as minimisation of e_T/J, we obtain,

$$\min_{\mathbf{e}} \sum_i \frac{e_i}{J} = \min_{\mathbf{e}} \sum_i \frac{e_i}{v_i} = \min_{\mathbf{e}} \sum_i \frac{1}{k^+_{cat,i} f_i(\mathbf{m}(\mathbf{e}))}.$$

We already derived that $-k^-_{cat,i}/k^+_{cat,i} < f_i(\mathbf{m}(\mathbf{e})) < 1$. Thus, evolution pushes all saturation functions to high values, maximally to 1, which means to high affinities for substrates and low affinities for products. We note that this argument applies for metabolic networks with a single degree of freedom in its fluxes (so we can equate overall steady-state flux J with each individual enzyme flux j). Such networks are called elementary flux modes (EFMs) [82-84], and we return to them later as they have a special role to play in optimal metabolic states.

Pushing enzymes to substrate saturation might not always be possible due to a thermodynamic constraint on the values of the kinetic parameters of an enzyme, known as the Haldane relationship. A perspective on the optimisation of activities of single enzymes, given the Haldane constraint, can be found in Cornish-Bowden [85].

Protein synthesis, ribosomes and growth rate

At balanced growth, the concentration p_i of any protein i in a cell is established by the balance between its rates of synthesis, degradation and dilution by (volume) growth,

$$\frac{dp_i}{dt} = \alpha_i \frac{k_r}{N_i} f_r(\mathbf{m}) r - (\mu + k_{d,i}) p_i = 0.$$

The factor α_i specifies the fraction of the total translation rate devoted to metabolic protein i, N_i equals the number of amino acids in this protein, k_r is the catalytic rate constant of the ribosome (unit: number of amino acids per ribosome per minute), r its concentration and $f_r(\mathbf{m})$ its saturation function.

The concentration of the ribosomes is set by the same balance,

$$\frac{dr}{dt} = (\alpha_r \frac{k_r}{N_r} f_r(\mathbf{m}) - \mu) r = 0.$$

(We omitted degradation, because ribosomes are stable proteins.) This last equation indicates that the growth rate of a cell equals the synthesis rate of ribosomes per unit ribosome, $\mu = \alpha_r k_r f_r(\mathbf{m})/N_r$.

Thus, we obtain the following relation between the concentration of a protein and the ribosome,

$$p_i = r\frac{\alpha_i k_r/N_i f_r(\mathbf{m})}{\alpha_r k_r/N_r f_r(\mathbf{m}) + k_{d,i}} \quad \Rightarrow \quad \frac{p_i}{r} \propto \frac{\alpha_i}{\alpha_r},$$

(where the proportionality is valid for stable proteins, i.e., most metabolic proteins.) This relation is remarkable, and leads to the insight that, at balanced growth, the ratio of protein and ribosome concentration is proportional to the ratio of the number of ribosomes respectively allocated to their synthesis. Thus, ribosome profiling results should agree with proteomics experiments at balanced growth. We are not aware of an experimental confirmation of this statement.

The total protein concentration in a cell equals $p_T = r + \sum_i p_i$ and is determined by the balance

$$\frac{dp_T}{dt} = \left(\sum_i \frac{\alpha_i}{N_i} + \frac{\alpha_r}{N_r}\right) k_r f_r(\mathbf{m}) r - \mu p_T = k_r' f_r(\mathbf{m}) r - \mu p_T = 0.$$

This equation leads to a relationship between the ribosomal protein fraction, the growth rate and the saturation function of ribosome $f_r(\mathbf{m})$,

$$\frac{r}{p_T} = \frac{\mu}{k_r' f_r(\mathbf{m})} \quad \Rightarrow \quad \mu = \frac{k_r' f_r(\mathbf{m}) r}{p_T}.$$

This last equation is a very useful one, which we will exploit several times throughout the remaining text. It also allows for an interesting calculation of the growth rate of a cell. We can calculate the maximal growth rate when the cell would only consist of ribosomes and all ribosomes are saturated with their substrates. Then $\mu_{max}^{theor} = k_r/N_r$ which gives rise to a generation time of about 5 min when k_r equals 17 aa/s and the amino acid content of a ribosome is 7536 (bionumbers website).

The derived relation $\mu = k_r' f_r(\mathbf{m}) r/p_T$ shows — perhaps somewhat counterintuitively — that increasing all enzymes by the factor λ such that $r \to \lambda r$ and $p_T \to \lambda p_T$ does not change the growth rate. Only changes in the relative protein concentrations can do that, via reallocation of limited resources; they influence the steady-state metabolite concentrations (entries of $\mathbf{m}$), r and all other enzyme concentrations that sum to $p_T - r$.

That growth rate is proportional to protein synthesis, i.e., $\mu \propto k_r' f_r(\mathbf{m}) r$, does not mean that the ribosome is the most important enzyme in a cell. It is not the "rate-limiting step", as one could (erroneously) conclude. Since, maximisation of $f_r(\mathbf{m})$, i.e., maximisation of the saturation of ribosomes with its loaded-tRNA substrates, requires high concentrations of those substrates, their synthesis rates need to be high enough to replenish those substrates that are binding and converted by the ribosomes. Thus, $f_r(\mathbf{m})$ maximisation requires high enzyme concentrations for sufficient amino-acid production [28]. Keeping high concentrations of substrates also requires that they diffuse fast enough to prevent concentration concentration. It has been suggested that diffusion of loaded-tRNA might limit growth rate in *E. coli*; thus, keeping the concentrations high when the ribosome is very active might be limited by substrate-diffusion rates [50]. Since $f_r(\mathbf{m})$ correspond to the fraction of ribosomes that are elongation, its evolutionary maximisation makes sense. Experimental data indicates that $f_r(\mathbf{m}) \approx 0.85$.

Growth rate also increases when the catalytic rate constant, i.e. k_r, of the ribosome increases. Ehrenberg and Kurland have, however, suggested that an upper bound to this value likely exists [55]; As increasing it reduces the effectiveness of kinetic proofreading. Kinetic proofreading prevents incorporation of the wrong amino acid into the developing protein. It works by pausing the incorporation of the amino acid onto the growing peptide chain, such that the wrong amino-acid-tRNA pair typically dissociates from the ribosome, before the ribosome attached it. (Assuming that the ribosome has a higher affinity for the correct pair that, therefore, spends (on average) a longer time on the ribosome.) The allowable number of errors in proteins then sets a bound to k_r. Thus, understanding ribosome optimisation is a complex problem.

The ribosomal fraction of *E. coli* in μg RNA per μg protein, a measure proportional to r/p_T, as function of growth rate is shown in Figure 5. It is not a linear relation between the total ribosomal protein content and the growth rate of the cell as was derived above ($\mu = k_r' f_r(\mathbf{m}) r/p_T$). It rather obeys,

$$\frac{r_T}{p_T} = \frac{\mu}{a k_r} + b,$$

with a and b as phenomenological constants for now. This relation has been explained in two ways.

One view is that all ribosomes are active (hence $r = r_T$), and that the saturation degree, $f_r(\mathbf{m})$, of the ribosome is growth-rate dependent. Combining the above two expressions, we find

$$f_r(\mathbf{m}) = \frac{\mu p_T}{r k_r} = \frac{a\mu}{abk_r + \mu},$$

which argues against the control objective of the ribosome system described above, as it makes the saturation degree of the ribosome dependent on growth rate, and only at high growth ($\mu \gg abk_r$) does the saturation become growth rate independent and equal to a.

Another view is that a fixed pool of inactive ribosomes exists, with concentration r^U, such that the concentration r is in fact the pool of active enzymes, and $r_T = r + r^U$. Then,

$$\frac{r}{p_T} = \frac{r_T - r^U}{p_T} = \frac{\mu}{k_r f_r(\mathbf{m})} \quad \Rightarrow \quad \frac{r_T}{p_T} = \frac{\mu}{k_r f_r(\mathbf{m})} + \frac{r^U}{p_T}.$$[3]

In this last model, $f_r(\mathbf{m})$ should be a constant for the relation between $\frac{r_T}{p_T}$ versus μ to be linear, and it also defines the constants a and b as the saturation of the ribosome and the fraction of inactive ribosomes, respectively.

---

[3] Note that now $\frac{dr_T}{dt} = \alpha_r k_r f_r(\mathbf{m}) r - \mu r_T = 0$ such that the now growth rate equals $\mu = \alpha_r k_r f_r(\mathbf{m}) \frac{r}{r_T}$ and is no longer equal to the ribosome synthesis rate per ribosome.

We favour the last view, for a reason that is still speculative. The determination of r_T/p_T is done by measurement of the (stable) rRNA and total protein. It is seems very plausible to us that free rRNAs and non-translating, partially assembled ribosomes exist due to the thermodynamic equilibrium of the association and dissociation events involved in ribosome assembly. The sum of the partially-assembled concentrations is then equal to r^U, which may be large since the assembled ribosome complex exists of many subunits such that even more partially-assembled complexes exist. (This argument may also be only partially correct, given the fact that during a nutrient upshift almost instantaneous translation occurs, as if idle, unused, fully-assembled ribosomes exists.)

Overcapacity of ribosomes at slow growth rate have been reported for *E. coli* [86] and *S. cerevisiae* [81], indicating that having a reserve of idle ribosomes, when nutrient are scarce, provides a fitness benefit when nutrients become suddenly available. This, however, introduces the question why those idle ribosomes do not reduce the growth rate when nutrients are scarce. If that would be the case, then prolonged growth at nutrient limitation would lead to a reallocation of biosynthetic resources to other now-needed proteins at the expense of the idle ribosomes. So, either those microorganisms have evolved under dynamic nutrient conditions or idle ribosomes at nutrient limitation do not reduce growth rate. We think that the latter is the case — perhaps, paradoxically so. Below we will show that growth rate is limited at nutrient limitation by the protein-storage capacity of membranes, in particular of nutrient-uptake proteins. And, that under those circumstances, the cytolic protein storage is exploited below its maximum. Thus, then space is free in the cytosol for now-unneeded ribosomes. That in turn poses the question why having unneeded proteins in a non-limited cellular compartment does not reduce growth rate. For that we do not have a good answer, we provide one below on the basis of an argument made by Dill *et al.* [87] that the protein density of microorganisms is perhaps optimal for maximal reaction rates. That would explain why microorganisms keep a constant protein density across conditions, regardless of whether expressed proteins are needed [86, 88].

In any case, the following definition of growth rate is independent of the nature of the inactive ribosomal protein fraction:

$$\mu = \frac{k_r f_r(\mathbf{m}) r}{p_T} = \frac{\text{total translation rate}}{\text{total cellular protein}} = \frac{J_p(\mathbf{m})}{p_T}.$$

This is the definition we used above already.

III. Fitness and microbial physiology

From proteins to metabolic pathways

Obtaining an understanding of growth-rate-maximising protein expression addresses only part of the challenge. Proteins function concertedly in networks, and it is those networks that ultimately set cellular growth rate. They are responsible for the conversion of all nutrients into all the cell material required for growth, including the relief of stresses acting on growth processes in the cell. One goal is, therefore, to understand why a cell chooses for the network it expresses and not another one. Microbial physiology is therefore at its heart a systems biology problem [89, 90].

Remarkably, many evolutionarily distinct microorganisms show the same qualitative metabolic behaviour as function of nutrient availability, and hence, growth rate [74]. First, the vast majority shows catabolite repression and diauxie [91], i.e., they often favour sequential metabolism of substrates over co-consumption [74, 92-94]. Second, they shift from a high yield to a low yield metabolism as function of growth rate and/or carbon excess (Figure 6). For some microorganisms, such as the best studied *E. coli* and *S. cerevisiae*, this means that respiratory metabolism occurs at low growth rates and increases in rate up to a critical growth rate, whereafter metabolism shifts to a respirofermentative or "overflow" mode that progressively shows a higher rate of fermentation or overflow-product formation [95]. In other words, microorganisms spoil nutrients at high growth rate.

What is the evolutionary benefit of recurrent physiological behaviours? It turns out that constraints acting on protein expression and growth-rate maximisation can explain these common findings [52, 96].

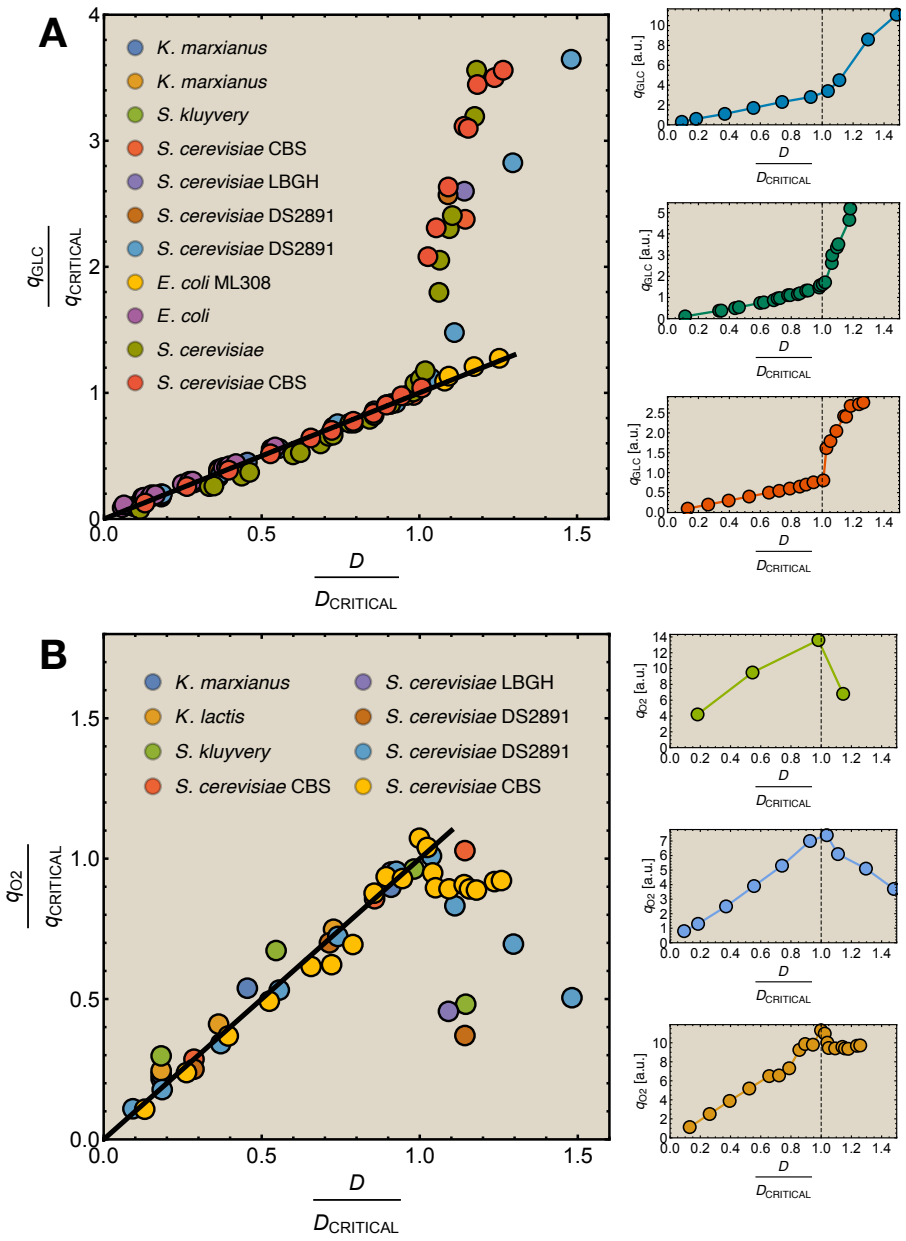

Figure 6. Illustration of common microbial physiology: overflow metabolism occurs after a critical growth rate and linear flux-growth rate relations. Values of the glucose (A) and oxygen (B) uptake fluxes in glucose-limited chemostats from different studies (Postma et al., 1989, Holms, 1996, Van Hoek et al., 1998, Nanchen et al., 2006, Fonseca et al., 2007). The figures shows on the right are examples, shown also in the left figure.

Stoichiometric models of whole-cell metabolism

Shortly after metabolic pathways were experimentally determined in the 1950 and 1960's, stoichiometric modelling of metabolism was started with the work of Umbarger [97] and Stouthamer [98]. Central to those maps is the concept of reactant stoichiometry, how many moles of each of the substrates are needed to

make each of the products in certain molar amounts [8]. These stoichiometric coefficients follow from knowing the reactants, their elemental composition, and balancing the number of each element right and left of the reaction arrow. When all reactants of all enzymes have been identified, the reaction network is known.

At balanced growth, all the concentrations are constant and therefore all the net synthesis and degradation of all molecules balance. Under this condition, pathway maps can be used to figure out, for instance, how much ATP, NAD(P)H, and central precursors of metabolism were required to make each amino acid, nucleic acid and lipid at steady state [3]. This what Umbarger and Stouthamer did. Then, from the rate of the growth and the compositions of cells in terms of those monomers they could calculate their synthesis rates. Given a particular set of nutrients, they could determine how many mole of those nutrients are required to synthesize one mole of ATP, NAD(P)H and precursors [3]. Knowing these demand rates, the uptake of nutrients can be calculated. These models are so-called structured models [99]. (Unstructured models also exist. They are very simple, but lack detail, and just give the overall conversion stoichiometry of nutrients into cells and byproducts [100].)

The most detailed structured model considers the stoichiometry of all metabolic reactions encoded on an organism's genome [101]. These are called genome-scale stoichiometric models, pioneered by Bernard Palsson and colleagues [102-104]. They are analysed using methods from linear programming, pioneered by Fell and Small [105]. Those methods are nowadays better know as flux balance analysis [37]. Recently those methods have been extended to deal with resource-allocation based optimisations [88, 106-108].

Protein-expression constraints

Since the kinetic parameters of enzymes can only change through mutations, cells adapt to new conditions to optimise their growth rate via changes in protein expression. That occurs, in addition, to metabolic regulation of proteins, via post-translational modifications and (allosteric) feedback regulation [74]. A higher growth rate requires for many reactions higher rates and, therefore, also higher enzyme concentrations. A cell then eventually runs into bounds of allowable protein concentrations, since cells have a finite capacity to store proteins. The growth rate it

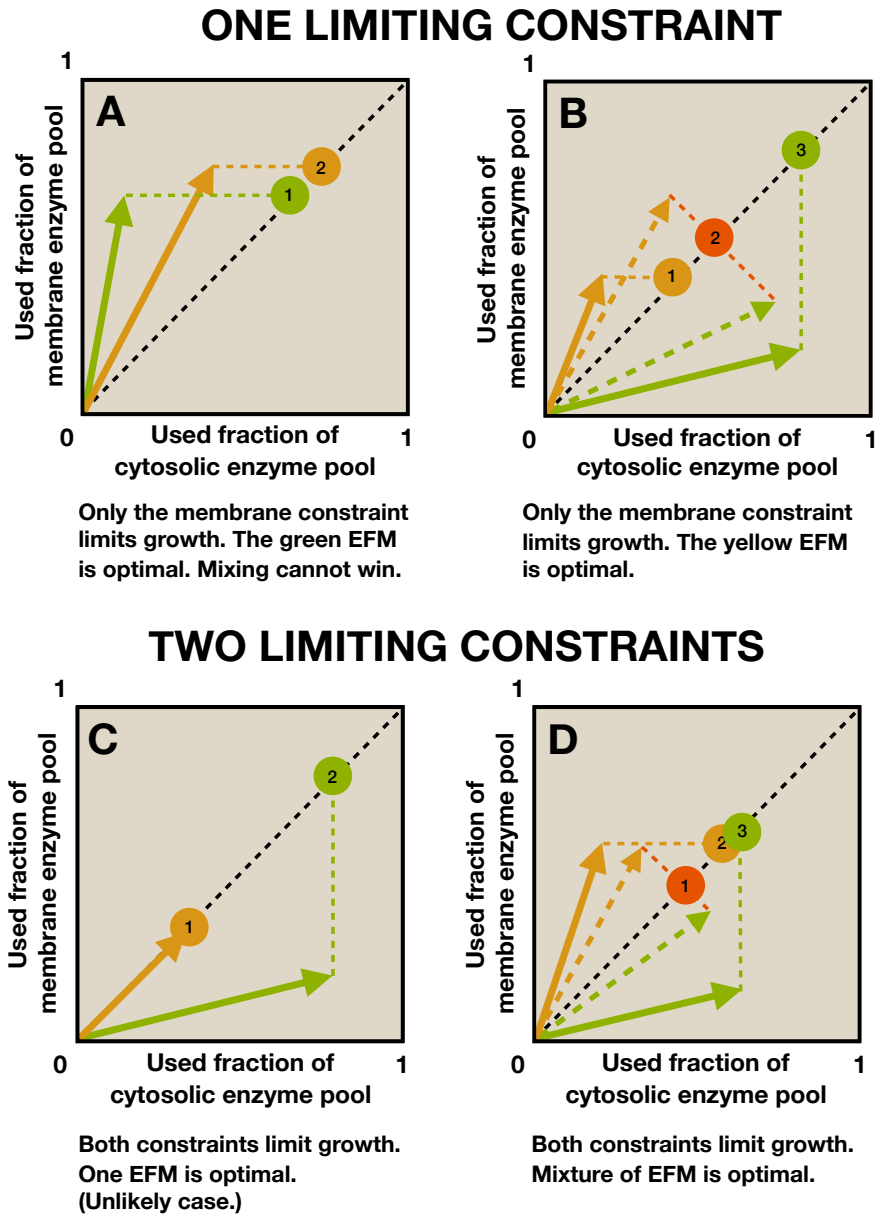

Figure 7. Illustration of the use of cost vectors of EFMs for 1 unit flux. A scenario of with two protein constraints is considered, one for the membrane and another for the cytosol compartment. We also limit the case to two possible EFMs (yellow and green) and their mixture (dotted lines). We always optimise the EFMs and their mixture independently, which leads to protein costs vectors, corresponding to usage fractions of the two pools required for 1 unit objective flux. If all EFM cost vectors lie above the diagonal (Figure 7A; or below, not shown) then one wins. This is the one with the lowest costs, marked 1 on the diagonal. The green EFM is optimal, it requires the least resources for 1 unit flux. Since it lies above the diagonal, it exhausts only one (the membrane) pool, and leaves the other (cytosolic) pool largely unused. Conversely, when both EFMs lie below the diagonal only the other (cytosolic) constraint is limiting (not shown). One of the EFMs can also win if both lie on opposite sides of the diagonal. For instance, in Figure 7B the yellow EFM wins, as it also beats a mixture of EFMs. Again only the membrane constraint is limiting. Both constraints can become limiting too (Figure 7C and 7D). Then one unlikely case exists, when one of the EFMs lies on the diagonal and wins (Figure 7C). The other is that the mixtures win (Figure 7D).

then attains is maximal and limited by at least one of those bounds. (This happens

regardless of the occurrence of post-translational or feedback regulations.)

These bounds may correspond to the maximal protein solvent capacities of the various protein-containing compartment of a microbial cell. For bacteria these are the periplasm, plasma membranes, and cytoplasm, while for eukaryotic microorganisms organelles are also relevant.

It turns out that the interiors [49] and membranes [48] of cells are extremely crowded with proteins. About 20% of the interior of cells is occupied by proteins (with 58% being the theoretical maximum) and they occupy 30-50% of the membrane. This means that protein-to-protein distances are of the order of the diameter of a protein (~5 nm), both in a cell's interiors and in membranes. Cells are thus extremely packed with protein. Making more proteins of one type will therefore influence the space available for others. In fact, increasing the protein concentration of cells, by reducing the water content by increasing the osmotic pressure, reduces growth rate [109]. This suggests that diffusional speeds of proteins and large molecular complexes can limit growth rate [50].

To summarise, each protein-containing compartment in a cell has a limited protein storing capacity; A growth rate increase requires increased rates of biosynthetic reactions, and those rates are proportional to the concentration of their catalysing enzyme. Thus, growth rate can increase until one or more protein containing compartments is filled with needed protein. We will show below that not all protein-containing compartments are limiting growth simultaneously. Which ones do depends on conditions, and the resulting limiting constraints explain changes in metabolism.

Elementary flux modes maximise metabolic flux per unit protein

Again we view growth rate as the flux of protein synthesis per unit protein. Its optimisation requires the optimal allocation of a finite amount of protein over all the metabolic reactions required to make proteins. Equivalently, we can ask what the amount of protein is that is minimally required to obtain 1 unit of protein synthesis flux.

This is a complicated problem. To see this, imagine a linear metabolic pathway and suppose we aim to find the amount of protein that we should invest in it in order to reach 1 unit of pathway flux. All enzyme rates are now equal to $1 = v_1 = v_2 = \ldots = v_r$. Each rate equals $v_i = k_i e_i f_i(\mathbf{m}) = 1$. The required enzyme concentrations cannot be calculated without knowing the steady-state metabolite concentrations, entries of $\mathbf{m}$. These, in turn, depend on the concentrations of the enzymes, and appear nonlinearly in the saturation functions. Thus, this a nonlinear optimisation problem; these are usually difficult to solve for larger systems.

Since we did not want to solve this for a linear metabolic pathway, but for the general case, to identify the general characteristics of optimal metabolic networks, we initially feared the worst. To our surprise, it turned out that this problem *is* solvable. We [76], and Mueller *et al.* [110], found that elementary flux modes (EFMs) are the metabolic networks that maximise the flux given a finite of amount of enzyme that can be allocated to the enzymes in the metabolic network. Also for nearly all enzyme kinetic equation this optimisation problem is convex, meaning that the landscape of the j/e_T as function of all the (bounded) enzyme concentrations has only a single peak [25, 111]. That was an unexpected finding too.

EFMs are minimal pathways through a reaction network [82-84]. All flux directions in an EFM are in accordance with thermodynamics. The EFM's network is minimal, in the sense that no reaction can be removed without violating the steady-state requirement. EFMs (therefore) have only one degree of freedom, meaning that if you know one flux value, you can determine them for all reactions in the EFM. Therefore, an EFM has fixed yields of its products given its substrates, which corresponds to fixed flux ratios. Finally, all possible steady-state flux distributions are expressible in terms of weighted sums of flux distributions of single EFMs. Thus, the smallest units of physiological activities of a cell are its EFMs. Some of these are growth supporting, i.e., they produce new cells from all the required nutrients.

The mathematical proof that EFMs are the constrained flux optimisers can be found in the Appendix [76]. (Mueller *et al.* came to the same conclusion, using a

completely different approach [110].) The intuitive explanation is the following: each metabolic strategy is a weighted sum of EFMs. For each EFM, we can compute the costs of reaching a flux of 1 unit. One EFM will be cheapest; it is very unlikely that two EFMs will have exactly the same costs. If one EFM is the cheapest, it will be the best strategy, no mix of EFMs will do better. This result is completely general and holds for arbitrary networks (not linear, it can have branches, cycles, etc.), with enzyme kinetics ($v_i = k_i e_i f_i(\mathbf{m})$) that can include allosteric feedback regulation.

Whether a flux distribution derives from an EFM can be tested by computing the rank of the associated stoichiometric matrix, which equals the number of reactions minus 1 in the case of EFMs [82]. Moreover, due to its fixed yield, the net reaction stoichiometry of an EFM is constant too. Therefore, the substrates of the EFM and the products are always consumed and produced in the same proportions. Consequently, if a single EFM is used across conditions, at varying rates, the relation between each of these rates and growth rate are linear. We will come back to that later, after we have dealt with the situation where more than one proteome constraint is active.

A maximal bound on the number of used EFMs in the optimum

So far, we have limited ourselves to the situation that a single enzyme constraint exists that limits the allocation of enzyme concentrations over metabolic reactions. In that situation, the optimal protein investment is the usage of a single EFM. But what happens if two, or more, constraints exist that each limit protein concentrations? For example, when one limits the protein concentrations in the membrane, another in the cytosol and a third in the periplasm? Will all compartments always be filled with protein when growth rate is maximised? How many EFMs will generally be used? Recently, we have solved this problem [112].

We illustrate the solution with a case that can be understood visually (Figure 7). The general case can be found in de Groot *et al.* [112]. We consider two protein-concentration constraints, one for the membrane and another for the cytosolic compartment. For each EFM we compute the minimal protein requirement for 1 unit of flux. This leads to protein-cost vectors, whose coordinates correspond to the usage fractions of the two protein pools. These are the outcomes of nonlinear

optimisations given the enzymes kinetics of all the reactions in the EFM and thus employ a steady-state kinetic model — not a stoichiometric model. One can see, that depending on the conditions, one or two EFMs maximise total flux, but never more than two.

If all EFM cost vectors lie above or below the diagonal, only one constraint is limiting, and only one EFM can be optimal. As conditions change (for example, the concentrations of nutrients increase), the length and direction of the cost vectors change, and another constraint may become limiting as soon as the diagonal is crossed by one or more EFMs. Then, a combination of two EFMs may make optimal use of the available, limiting, resources and deplete both protein pools.

In general, the number of EFMs that can carry flux in the optimum is bounded by the number of constraints on protein pools that actively limit the growth rate; when the size of those pools would increase, then the growth rate would also increase.

When nutrient conditions change, the cost vectors of EFMs change along. This is because the optimal activity of the EFM depends on the nutrient concentration. For each new concentration, the EFM optimisations have to be repeated, the protein-costs change and the picture changes. Thus in response to a changing condition, a metabolic network can show a qualitative change in behaviour, from the use of a single EFM to a mixture of two, or from two to three, etc. We think that this mechanism underlies the shift from respiratory to respirofermentative growth.

To summarise, we obtained the following insight: Microorganisms are continuously selected for optimal expression and usage of proteins to grow fast enough and to prevent being outcompeted. This means that they inevitably run into protein constraints that limit growth rate, since rates of enzymes are proportional to their concentrations, but not all constraints are ‘hit’ at once. When conditions change, other constraints may become limiting (in addition) and cause a shift in the qualitative behaviour of metabolism, e.g., from pure respiratory to mixed respirofermentative growth. The number of mixed EFMs will nonetheless remain small, because the number or protein concentrations are small too; we cannot think of more than a handful. Note that all this assumes optimal protein expression

leading to maximisation of growth rate under constant conditions. Let us now turn to the evidence that cells indeed behave in this way.

Experimental evidence that only a few EFMs are used

EFMs have only a single independent flux, so that all the flux ratios for an EFM are constant. Therefore, a change of one flux in the EFM is accompanied by a proportional change in all the other fluxes. This is characteristic of usage of a single EFM. Figure 8 shows some of the nutrient-uptake and byproduct-formation fluxes of a glucose-limited chemostat culture of *S. cerevisiae* as function of the dilution rate [95].

Below the critical dilution rate, at $0.28\ hr^{-1}$, only respiratory growth occurs (no by-products are formed and oxygen is consumed). The linear fits indicate a proportional relationship between the fluxes and the growth rate, indicating the activity of a single EFM. We confirmed this computationally, using FBA, by fixing the growth rate and minimisation the glucose-uptake flux, and determining the rank of the used metabolic network (P. Grigaitis, Teusink, Bruggeman, unpublished; this optimisation corresponds to a linear program with a single constraint that always has a single EFM as its solution.) Famili *et al.* [113] also confirmed that FBA results agree with the data, providing additional evidence for the usage of a single EFM.

An interesting property of EFMs is that the maximal yield of any metabolic network is always achieved by one of its EFMs. Thus, growth-rate maximisation in the presence of one nutrient-uptake flux constraint with FBA also maximises the biomass yield on that nutrient. Growth rate and yield do not necessarily trade-off: it depends on the condition-dependent active constraints. This suggests that below the critical growth rate in the chemostat cells maximise growth rate, as they always do, but only hit a single constraint, with an EFM is the optimal solution. The EFM that maximises growth rate now also has the maximal biomass yield on glucose, because cells are now limited by the glucose-uptake capacity (membrane is full) and cannot afford to spoil.

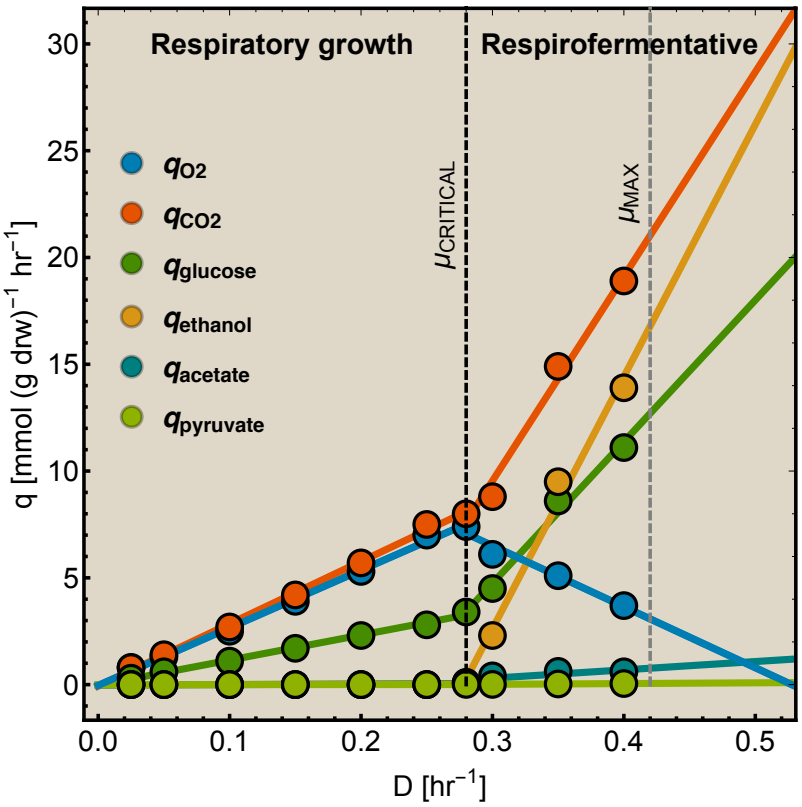

Figure 8. Occurrence of linear relations between flux and growth rate and a critical dilution after which fermentation starts in a glucose-limited chemostat of *S. cerevisiae* (Van Hoek *et al.*, 1998). The dots are the experimental data and the lines are linear fits.

When the dilution rate rises, the glucose concentration rises, growth rate increases until the critical dilution rate occurs. Then we expect that a new constraint becomes active, and that now two EFMs, respiration and fermentation in the case of yeast, are being used together. Which constraint becomes active is likely to be condition specific, and many options have been proposed [114], but the underlying extremum principle, based on constrained growth-rate optimisation, is always the same. Along these lines is also how Basan *et al.* [115] rationalised the onset of acetate overflow metabolism by *E. coli* as function of its growth rate. They found strong evidence that the cytosolic protein pool becomes limiting when acetate formation starts. By over-expressing an unneeded enzyme in the cytosol they could change the critical dilution rate.

The data of Van Hoek *et al.* [95] also show that, beyond the critical dilution rate, the respiratory fluxes decrease, while the fermentation flux increases, in a linear fashion. Fermentative growth has a lower yield and hence the glucose uptake rate increases. In this way, a higher-yield EFM (respiration) is exchanged for a lower-

yield EFM. In quantitative terms, the flux distributions of the two EFMs are mixed with a coefficient $\alpha(\mu)$ that is growth-rate dependent, such that

$$J_R\alpha(\mu) + (1 - \alpha(\mu))J_F = J(\mu), \qquad (0 \leq \alpha(\mu) \leq 1)$$

with $J(\mu)$ as the entire metabolic flux distribution of the cell. (Note that the ribosome concentration increases linearly with growth rate too (Figure 5).) Fitting to the data of Van Hoek *et al.* shows that $\alpha(\mu) = 1.9\mu$. At the critical dilution rate, $\alpha(\mu_c) = 1$ and $J_R = J(\mu_c)$. At $\mu = 0.53\ hr^{-1}$ the cellular growth would be entirely fermentative; this is, however, above the maximal growth rate ($0.42\ hr^{-1}$). Using FBA, it can be shown that the fluxes value after the critical dilution rate can be predicted using a single new flux constraint (P. Grigaitis, Teusink, Bruggeman, unpublished).

Basan *et al.* [115] found a similar linearity in *E. coli* in the overflow regime. They showed with proteomics that flux changes coincided with protein expression changes. They suggested an even allocation of protein, i.e., if the respiratory system reduces in protein content by 1% then this amount of protein is allocated to the ribosome and the overflow EFM. In the overflow EFM, the increased protein resource would then be divided in proportion to the current protein concentrations, so that all protein ratios remain the same. Since the relative fluxes were already the same, the saturation degrees (the *f*'s) of all the protein in the fermentative EFM remain the same. This is still speculative at this moment, and more experimental work on protein and metabolite levels is needed, but if it is true then metabolic regulation directly follows from growth-rate maximisation principles. We think that is a worthwhile endeavour.

The trade off between growth and stress demands

The fitness effect of unneeded protein expression, which we derived earlier, indicates that preparatory protein expression reduces instantaneous fitness. It reduces the growth rate, because it occurs at the expense of growth-promoting protein expression. The derivation above, however, only applies if the preparatory protein occupied a protein pool that is limiting growth rate. It does not say anything about preparatory protein expression in pools that are not limiting. Before we consider preparatory protein expression under such conditions, we have to convince ourselves that preparatory protein expression is not always growth-rate

reducing. It does require resources after all. Isn't it better to leave non-limiting protein compartments partially empty? Thus, why would a cell fill the cytosol with protein if it is only membrane limited?

Dill *et al.* [51] may have provided an answer: They showed that the protein density of a cell (~20% volume fraction [44]) is remarkably close to the optimal volume fraction of macromolecules that maximises biosynthetic rates. That such an optimum must exist follows from the insight that at low-volume fractions macromolecule collision rates are limited by their collision time, while, at high volume-fractions, protein crowding increases viscosity, decreasing the diffusion coefficient of proteins. Thus at intermediate protein concentrations rates be higher. Thus, perhaps, under- *and* overfilling of compartments with proteins reduces fitness, regardless of whether those proteins are needed or not. The fact is that experimental data show that protein content of cells (g per g dry weight) is very similar across conditions and microorganisms.

Given the experimental evidence that cells do fill protein compartments, then under which circumstances do we expect preparatory protein expression? We have seen that at low growth rates, under nutrient limitation, the membrane is the likely only limiting constraint. This implies that the entire capacity of the cytosol to store growth-promoting proteins is not required in the optimal state. O'Brien and Palsson [116] came to the same conclusion using a detailed model of metabolism and expression of *E. coli*.[4] They also found that below the critical dilution in the chemostat, before acetate is made, the cell is nutrient limited and that the cytosol has excess protein-storage capacity. After the critical growth rate, the cytosolic protein compartment has no longer excess capacity and is completely filled with needed protein. O'Brien and Palsson's computations are therefore fully in line with our theory.

[4] In this model all the cellular macromolecules are made out of their building blocks, incl. mRNAs, DNA and proteins, and mRNA needs to be made so that translation can proceed. Classical genome-scale stoichiometric models do not do this, they only consider metabolism and the formation of macromolecules is not made explicit. In the latter, simpler, case, metabolic fluxes of enzymes are not limited by their concentration, enzyme concentrations are not limited by their mRNA concentration, and enzyme concentrations are not set by the balance between their rates of synthesis, dilution by growth and degradation.

A direct consequence is, that at low growth rates caused by nutrient limitations, proteins can be made in the cytosol that have no impact on the current growth rate by forcing growth-promoting, needed, cytosolic proteins to reduce in concentration. If that is indeed so, then an increase in geometric fitness can be achieved by expressing anticipatory proteins (proteins that are currently not used) but would prepare the cell for future environments. Experiments of Egli's lab [117, 118] confirmed this: *E. coli* was found to be more stress tolerant and capable of instantaneous growth on alternative carbon sources when harvested at a dilution rate below its critical value than above. This also agrees with our understanding of ppGpp, which increases with decreasing growth rates, activating alternative stress-related sigma factors, priming the cell for future stressful conditions [78]. In addition, catabolite repression is relieved when growth rate reduces, increasing the (leaky) expression rate of alternative carbon utilisation systems [117]. Thus, perhaps, growth rate does not always trade-off with stress [119].

Thus, the emerging picture is that, as growth rate increases, cellular protein compartments are progressively filled with needed proteins, reducing preparatory protein expression. It may even explain the need for phenotypic diversification, using stochastic mechanisms, leading to persister formation, for example, at high growth rate. Although circumstantial evidence exists, many aspects need to be more firmly established experimentally.

The intrinsic nonlinearity of cellular self-fabrication

Most genome-scale models of metabolism and growth circumvent the heart of the matter. They fail to truly capture a defining essence of cell growth, namely that a cell uses its own molecules to synthesise these same molecules and double them in number, from its birth to its division. It has to choose the right molecules, the set that can make itself from the available nutrients. Natural selection forces cells to choose components that run this process of *self-fabrication* at the highest possible rate. This means that growth rate depends on the identity and concentrations of all those components, as these determine reaction rates.

Classical analysis of genome-scale models uses flux balance analysis, linear programs, in which the flux values are the unknowns [37]. Linear (in)equality

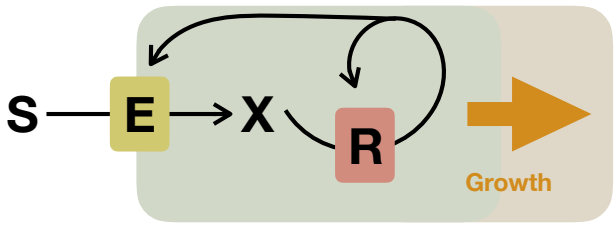

Figure 9. The simplest, biological self-fabricator. For a cell to grow, it needs to make its own components out of nutrients, which it takes up from its environment. For it to grow at a constant rate, it needs to make all its components in constant proportions, such that their concentrations remain constant and, therefore all biosynthetic fluxes. Then, self-fabrication occurs as balanced growth. The nutrient S is the growth substrate, X is a metabolite, E is an membrane-embedded uptake enzyme, and R is the ribosome that makes E and itself. The regulation of the synthesis of the E and R are regulated by the concentration of X [26].

relations between fluxes are given as constraints, corresponding to steady-state mass balances and flux bounds. Then an objective function, also linear in the fluxes, is optimised. Linear equations are solved and concentrations of molecules do not play any role. Consequently, the fact that an enzyme with a certain concentration catalyses its reaction at a certain rate and, hereby, indirectly contributes to its own synthesis, to balance its dilution due to cell-volume growth, is not considered. Thus, by not focusing on concentrations the essence of self-fabrication is missing. The ME model by O'Brien and Palsson [116] was the next step in the direction of self-fabrication models. Its optimisation problem is nonlinear in the growth rate of the microorganism. It considers protein concentrations, but not metabolite concentrations. As a consequence, the saturation functions of enzymes (the f's) are constants. ME models are the simplest variants of models of self-fabrication [26].

To truly model self-fabrication, we need to consider all cellular machinery, so including all the kinetics of all the processes in a cell, and the concentrations of all the molecules it contains. We have recently proposed such a general formalism [26]. In it, the ribosome makes itself and all the enzymes that are required to supply the substrates for protein synthesis. While this happens, the rates of metabolic reactions and translation increases, but the volume of the cell increases too, causing dilution, and these rates balance at balanced growth. To see why this problem is nonlinear in the growth rate, we have added an example to the Appendix.

Elementary growth modes

We have found that even in the full-blown nonlinear case, i.e., with enzyme kinetics, dilution by growth of all the components and the ribosome making all the enzymes and itself, one can still define networks that are minimally required for growth [26]. They are minimal self-replicating metabolic systems, called elementary growth modes (EGMs), that allow self-fabrication of all their components at a fixed steady-state rate from nutrients. They truly describe balanced growth in all biochemical detail. EFMs refer to metabolic networks, in the absence of volume dilution of metabolites; EGMs refer to metabolic networks catalysed by proteins synthesised by ribosomes, and all these components are diluted by growth. Thus, metabolic networks are subnetworks of EGMs.

Some of the results obtained with EFMs turn out to be valid for EGMs too [26]. We could show that the number of EGMs is maximally equal to the number of active constraints on protein pools. With only one active constraint (which is the condition that cellular volume is directly tied to cellular contents), a single EGM is the growth rate maximiser. A remarkable insight is that if the amino-acid composition of all proteins in a cell is constant across conditions, the metabolic subnetwork of any EGM is a corresponding EFM. Experimental evidence supports this constancy of amino acid composition [26]. To summarise, the vast majority of the interpretations of experiments discussed above are preserved when the EGM framework is used. EGMs are the most general molecular descriptions of the states underlying balanced growth of microorganisms. Therefore, any general balanced-growth theory should start from them. This is the challenge ahead of us in the coming years.

IV. Microbial optimality: fact or fiction?

A perspective on microorganisms from the allocation of finite biosynthetic resources has proven to be extremely fruitful in the last decade. Without this perspective it is hard to take a systemic perspective on cells and realise how all molecules in it form a functional entity.

The resource allocation perspective is in itself not a hypothesis about the optimality of microorganisms. Optimality only comes into play when one starts to wonder

about *why* microorganisms express certain proteins and networks and not others. Then it is very natural to think about natural selection and evolution leading to microorganisms that allocate resource optimally to maximise their fitness. Fitness then equals growth rate, in presence and absence of stresses. It thus also incorporates survival investments to avoid a negative growth rate, i.e., death. (Fitness in dynamic conditions can then be equated in terms of that growth rate, like we did above.)

Without the optimality hypothesis, it is hard to hypothesise how a cell should direct its protein economy, in which proteins it should invest. For us, only when we think about maximisation of the return of those investments, i.e., making offspring the fastest, do hypotheses emerge about protein and network expression.

We started the growth-rate maximisation viewpoint as a null hypothesis, a starting point for the design of experiments and interpretation of the microbial physiology literature. What has been mind-boggling is how many puzzle pieces have fallen into place in the last decade of research. So much makes sense now, that we start thinking that a general optimisation-based theory about microorganisms can really be developed. The next step is to continue to quantitatively test the theory against experiments, preferentially by multiple labs using highly standardised protocols. We also need to improve the theory, in particular concerning the competition for protein involved in different types of constrained commodities.

V. Closing remarks

Richard Feynman once wrote (on his blackboard): “What I cannot create, I do not understand.”, referring to the fact that if he couldn’t write down a set of equations describing some phenomenon he did not understand that phenomenon. We hope that microbiology will reach this high standard too. It would give microbiologists a common language that would facilitate comparative microbiology and offer a method to come to grips with the enormous microbial biodiversity on our planet. It would also improve the communication of evolutionary biologists, microbiologists and biophysicists, leading to a more unified biology.

Thus, is there a unity in microbiology? We certainly hope there is.

Acknowledgements

FJB acknowledges funding from NWO-VIDI grant 864.11.011 and BT from NWO-VICI grant 865.14.005. We thank the following researchers for their contributions and discussions: Jan Berkhout, Evert Bosdriesz, Daan de Groot, Johan van Heerden, Matthias Heinemann, Joost Hulshof, Terry Hwa, Hidde de Jong, Stefan Klumpp, Wolfram Liebermeister, Timo Maarleveld, Steffen Klamt, Stefan Müller, Erik van Nimwegen, Elad Noor, Niclas Nordholt, Gosse Overal, Han Peters, Jack Pronk, Iraes Rabbers, Matt Scott, Peter Swain, Pieter Rein ten Wolde and Meike Wortel.

Appendix

Derivation of Fisher's theorem

We define at time t the mean growth rate $\langle \mu \,|\, t \rangle$ of a population of N genetically different microorganisms as,

$$\langle \mu \,|\, t \rangle = \sum_{i=1}^{N} p_i(t)\mu_i.$$

Its rate of change of equals

$$\frac{d}{dt}\langle \mu \,|\, t \rangle = \sum_{i=1}^{N} \mu_i \frac{d}{dt} p_i(t).$$

The probability of occurrence of genotype i is given by

$$p_j(t) = \frac{n_j(t)}{\sum_{i=1}^{N} n_i(t)},$$

with $n_i(t)$ denoting the number of organisms with genotype i growing at rate μ_i, and,

$$\frac{d}{dt} p_j(t) = \frac{\sum_{i=1}^{N} n_i(t) \frac{d}{dt} n_j(t) - n_j(t) \frac{d}{dt} \sum_{i=1}^{N} n_i(t)}{\left(\sum_{i=1}^{N} n_i(t) \right)^2} = p_j(t)(\mu_j - \langle \mu \,|\, t \rangle).$$

This last equation shows that the abundance of a genotype that grows faster than average increases while that of a slower-than-average grower decreases. Finally, the rate of change in the mean growth rate equals,

$$\frac{d}{dt}\langle \mu \,|\, t \rangle = \sum_{i=1}^{N} \mu_i p_i(t)(\mu_i - \langle \mu \,|\, t \rangle) = \langle \mu^2 \,|\, t \rangle - \langle \mu \,|\, t \rangle^2 = \langle \delta^2 \mu \,|\, t \rangle,$$

which is Fisher's theorem [23],

$$\frac{d}{dt}\langle \mu \,|\, t \rangle = \langle \delta^2 \mu \,|\, t \rangle \qquad \text{(Fisher's Theorem).}$$

This theorem indicates that variation in growth rate causes changes in the mean growth rate.

Derivation of properties of the average cell in an isogenic population

A key aspect of the balanced growth theory is that it applies to the average cell in a population of cells that grows at a logarithmic growth rate. If the total number of cells equals n then its rate of change equals

$$\frac{dn(t)}{dt} = \mu n(t).$$

The number of cells at time t with age a, defined as the time elapsed since birth, equals [30],

$$n(a|t) = n(t)u(a)da = 2\mu n(t-a)da \int_a^\infty f(t)dt = 2\mu n(t)e^{-\mu a}da \int_a^\infty f(t)dt,$$

with:

1. $u(a)da$ the probability to have age a and $u(a)$ equals the probability density function for the cell age a,
2. $2\mu n(t-a)da$ is the number of daughter cells formed at $t-a$ in the time period of length da,
3. $n(t) = n(t-a)e^{\mu a}$,
4. $\int_a^\infty f(t)dt$ equals the probability that a cell has an age greater than a, and thus a generation time greater than a. Therefore, $f(t)$ is the probability density function for the generation time (or the interdivision time).

The probability density function for the cell age (the "age distribution") is hence given by

$$u(a) = 2\mu e^{-\mu a} \int_a^\infty f(t)dt. \qquad \text{(Cell-age Distribution)}$$

Since, $\int_0^\infty u(a)da = 1$ the growth rate can be solved from the following characteristic equation,

$$\int_0^\infty 2\mu e^{-\mu a} \int_a^\infty f(t)dtda = 1. \quad \text{(Population growth rate)}$$

Case of gamma distributed generation times

Experimental generation-time distributions, $f(t)$'s, are often well described by gamma distributions,

$$t_g \sim f(t_g) = \frac{e^{-\frac{t_g}{\beta}} t_g^{\alpha-1} \beta^{-\alpha}}{\Gamma(\alpha)},$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the Gamma function and α and β are positive parameters. The mean generation time then equals $\langle t_g \rangle = \alpha\beta$ and the variance of the generation time $\langle \delta^2 t_g \rangle = \alpha\beta^2$. For this case, the growth rate of the population is given by

$$\mu = \frac{\langle t_g \rangle}{\langle \delta^2 t_g \rangle} \left(2^{\frac{\langle \delta^2 t_g \rangle}{\langle t_g \rangle^2}} - 1 \right).$$

Note that (by L'Hopital's Rule)

$$\lim_{\langle \delta^2 t_g \rangle \to 0} \mu = \frac{\ln 2}{\langle t_g \rangle},$$

which corresponds to the generation time definition for the population level description

$$\frac{dn(t)}{dt} = \mu n(t).$$

Thus, generation time, measured at the level of the population in a culture of isogenic cells, does not equal the mean generation time when measured at the level of single cells. This discrepancy has been shown to be in agreement with experimental data.

We find that

$$\lim_{\langle \delta^2 t_g \rangle \to 0} \langle a \rangle = (\ln(2)^{-1} - 1)\langle t_g \rangle \approx 0.44 \langle t_g \rangle, \qquad \text{(Mean cell age)}$$

indicating that the average cell is almost at half its cell cycle.

The balanced growth theory applies to the state of the average cell. To show that this leads to correct predictions, consider a population level model of a molecule X that is produced at a constant rate and is diluted only by growth. The rate of change of its concentration (in copy number per cell) then satisfies

$$\frac{d\langle x\rangle}{dt} = k - \mu\langle x\rangle,$$

so that $\langle x\rangle_s = \frac{k}{\mu}$. Now we consider the single cell scenario. The number of molecules of X that a cell contains at time a equals,

$$n_x|a = n_x|0 + N_x|a,$$

with $n_x|0$ as the number of molecules at cell birth and the number of molecules produced since birth as $N_x|a$. The latter quantity is distributed according to a Poisson distribution (can be derived from the chemical master equation),

$$N_x|a \sim \textbf{Poisson}(ka) \quad \text{(with mean } k\langle a\rangle\text{)}.$$

Since $\langle n_x|\langle t_g\rangle\rangle = 2\langle n_x|0\rangle$ we know that $\langle N_x|\langle t_g\rangle\rangle = \langle n_x|0\rangle = k\langle t_g\rangle$; thus, in the limit $\langle\delta^2 t_g\rangle \to 0$,

$$\langle n_x|\langle t_g\rangle\rangle = k\langle t_g\rangle + k\langle a\rangle \approx k(\langle t_g\rangle + (\ln(2)^{-1} - 1)\langle t_g\rangle) = k\langle t_g\rangle\ln(2)^{-1} = \frac{k}{\mu}.$$

This equals the population level model result. Thus, the population level model hence assumes that $\langle\delta^2 t_g\rangle \to 0$ or that

$$\frac{\langle\delta^2 t_g\rangle}{\langle t_g\rangle^2} = CV(t_g)^2 \approx 0$$

This is therefore also one of the key assumptions in the balanced growth theory. According to measurements, the coefficient of variation of the generation time equals,

$$CV(t_g)^2 \approx 0.1,$$

indicating that the balanced growth theory can be an oversimplification.

Constancy of the concentrations at balanced growth

Since $c = n/V$ we obtain for its rate of change,

$$\frac{dc}{dt} = \frac{d}{dt}\frac{n}{V} = \frac{1}{V}\frac{dn}{dt} - \frac{1}{V}\frac{dV}{dt}\frac{n}{V} = c\left(\frac{1}{n}\frac{dn}{dt} - \frac{1}{V}\frac{dV}{dt}\right),$$

and, since

$$\frac{1}{n}\frac{dn}{dt} = \frac{1}{V}\frac{dV}{dt},$$

during balanced growth, concentrations are constant:

$$\frac{dc}{dt} = 0. \qquad \text{(Steady-state concentrations at Balanced Growth.)}$$

The rate of an enzyme is proportional to its concentration

That the rate of an enzyme is proportional to its concentration follows from a universal property of enzyme mechanisms [5]. When we consider those in terms of their elementary reactions, i.e., of reactant association, dissociation and catalysis, using mass action kinetics, we observe that all those rates are proportional to the concentration of an enzyme species.

Consider, for instance, the following mechanism of an enzyme that catalyses the reaction $A + B \rightleftarrows C$, in an ordered mechanism of 4 sequential elementary reactions,

$$E + A \overset{1}{\rightleftarrows} EA + B \overset{2}{\rightleftarrows} EAB \overset{3}{\rightleftarrows} EC \overset{4}{\rightleftarrows} E + C\,.$$

When we multiply all enzymes species with a factor λ then the total enzyme concentration, $e_T = e + ea + eab + ec$, increases with this factor too. The rate of the enzyme will do so too because the rates of all four elementary reactions ($v_1 = k_1^+ \cdot e \cdot a - k_1^- \cdot ea$, $v_2 = k_2^+ \cdot ea \cdot b - k_2^- \cdot eab$, $v_3 = k_3^+ \cdot eab - k_3^- \cdot ec$ and $v_4 = k_4^+ \cdot eb - k_4^- \cdot e \cdot b$) then increase by a factor of λ as well. Thus, the rate of the enzyme obeys

$$v(\lambda e_T) = \lambda v(e_T). \text{ (Proportionality of enzyme rate and concentration)}$$

It can be shown that this holds for all enzyme mechanisms, and is preserved if nonlinear saturation functions are derived using Quasi-Steady-State techniques [5].

Derivation of the fitness potential equation

Say we change by titration the concentration of protein i with concentration p_i. The cell responds by changing the concentration of all other proteins. Thus, the change in the objective flux equals,

$$dJ = \left(\frac{\partial J}{\partial p_i} + \sum_{j \neq i} \frac{\partial J}{\partial p_j} \frac{\partial p_j}{\partial p_i} \right) dp_i \,.$$

We assume that the cell responded by maximising the flux given the new value of p_i by optimally allocating $p_T - p_i$ over all the remaining enzymes. This mean that we can define the following objective with a Lagrange multiplier λ,

$$\mathscr{L}(\mathbf{p}) = J(\mathbf{p}) - \lambda \left(\sum_{j \neq i} p_j - (p_T - p_i) \right).$$

For all proteins, except i, optimality requires that,

$$\frac{\partial \mathscr{L}}{\partial p_j} = \frac{\partial J}{\partial p_i} - \lambda = 0 \qquad (\text{for all } j, \text{ except } j{=}i).$$

Thus, $\dfrac{\partial \ln J}{\partial \ln p_j} = \dfrac{p_j}{J} \dfrac{\partial J}{\partial p_j} = \dfrac{p_j}{J} \lambda$. From the flux summation theorem from metabolic control analysis, we know that,

$$\sum_{j \neq i} \frac{\partial \ln J}{\partial \ln p_j} = 1 - \frac{\partial \ln J}{\partial \ln p_i}.$$

Using the last two relations, we obtain

$$\sum_{j \neq i} \frac{\partial \ln J}{\partial \ln p_j} = \frac{\lambda}{J} \sum_{j \neq i} p_j = \frac{\lambda}{J} \left(p_T - p_i \right)$$

so that

$$\lambda = J \frac{1 - \frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}$$

and, therefore,

$$\frac{\partial \ln J}{\partial \ln p_j} = \lambda \frac{p_j}{J} = p_j \frac{1 - \frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}.$$

Now we substitute this last equation in the one we started with,

$$
\begin{aligned}
d\ln J &= \left(\frac{\partial \ln J}{\partial \ln p_i} + \sum_{j\neq i} \frac{\partial \ln J}{\partial \ln p_j}\frac{\partial \ln p_j}{\partial \ln p_i}\right) d\ln p_i \\
&= \left(\frac{\partial \ln J}{\partial \ln p_i} + \sum_{j\neq i} p_j \frac{1-\frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}\frac{\partial \ln p_j}{\partial \ln p_i}\right) d\ln p_i \\
&= \left(\frac{\partial \ln J}{\partial \ln p_i} + \sum_{j\neq i} p_i \frac{1-\frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}\frac{\partial p_j}{\partial p_i}\right) d\ln p_i \\
&= \left(\frac{\partial \ln J}{\partial \ln p_i} + p_i \frac{1-\frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}\frac{\partial \sum_{j\neq i} p_j}{\partial p_i}\right) d\ln p_i \\
&= \left(\frac{\partial \ln J}{\partial \ln p_i} + p_i \frac{1-\frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}\frac{\partial (p_T - p_i)}{\partial p_i}\right) d\ln p_i \\
&= \left(\frac{\partial \ln J}{\partial \ln p_i} + p_i \frac{1-\frac{\partial \ln J}{\partial \ln p_i}}{p_T - p_i}\right) d\ln p_i \\
&= \left(\frac{\frac{\partial \ln J}{\partial \ln p_i} - \frac{p_i}{p_T}}{1-\frac{p_i}{p_T}}\right) d\ln p_i .
\end{aligned}
$$

Thus,

$$\frac{d\ln J}{d\ln p_i} = \frac{C^J_{p_i} - \frac{p_i}{p_T}}{1-\frac{p_i}{p_T}}, \text{ with } C^J_{p_i} = \frac{\partial \ln J}{\partial \ln p_i}.$$

Derivation of unneeded protein cost equation from the fitness potential

An unneeded enzyme has a $C_p^{J_p}$ equal to zero. Thus its fitness potential equals

$$\Delta F_U = \frac{d \ln J_p / p_U}{d \ln p_U} = \frac{-\frac{p_U}{p_T}}{1 - \frac{p_U}{p_T}},$$

so that

$$\frac{1}{J_P} dJ_P = \frac{-1}{p_T - p_U} dp_U,$$

and integrating

$$\int_{J_p(0)}^{J_p(p_U)} \frac{dJ_P}{J_P} = \int_0^{p_U} \frac{-1}{p_T - p_U} dp_U$$

gives

$$\frac{J_p(p_U)}{J_p(0)} = 1 - \frac{p_U}{p_T}. \qquad \text{(Fitness cost of unneeded protein expression.)}$$

When the unneeded protein pool is a subset of the protein pool of the cell, e.g., the unneeded protein is cytosolic and the cell has other protein compartments like the periplasm and the membrane, then

$$\frac{J_p(p_U)}{J_p(0)} = 1 - \frac{P_T}{p_T} \frac{p_U}{P_T},$$

(Fitness cost of unneeded protein expression in a pool $p_T \in P_T$)

with $p_T < P_T$ and P_T as the entire protein pool of the cell. Experiments indicate that $P_T / p_T \approx 3$.

Optimal steering of metabolic protein expression by a genetic circuit

Our aim is to maximise the steady state flux through the simplest metabolic network (see [25] for the general case) and to illustrate how a genetic network can be found that is able to change enzyme concentrations such that the steady state flux is maximal with respect to the nutrient condition. The network we consider is

$$\underline{S} \xrightarrow{1} X \xrightarrow{2} \underline{P}.$$

Underlined metabolites indicates that their concentrations are kept constant. The enzyme kinetics of two enzymes is chosen to be

$$v_1 = k_1 e_1 \frac{\frac{s}{K_{1S}}}{1 + \frac{s}{K_{1S}} + \frac{x}{K_{1X}}} \text{ and } v_2 = k_2 e_2 \frac{\frac{x}{K_{2X}}}{1 + \frac{x}{K_{2X}} + \frac{p}{K_{2P}}}.$$

We aim to maximise the steady state flux $j = v_1(x_s, e_1) = v_2(x_s, e_2)$, with x_s as the steady state concentration, under the constraint $e_T = e_1 + e_2$. The value of x_s is determined by s and p. First, we express e_2 in terms of x_s, using $v_1 - v_2 = 0$ and $e_T = e_1 + e_2$, to yield $e_{2,s} = e_{2,s}(x_s)$. Next, we substitute this expression into v_2. This is now the expression for the steady state flux j and only depends on x_s, and is denoted $j(x_s)$. In the optimum, we require that

$$\left.\frac{\partial j}{\partial x_s}\right|_{x_s = x_s^o} = 0,$$

with x_s^o as the flux maximising value of x_s. A direct computation gives

$$x_s^0 = \sqrt{\frac{k_1 K_{1X} K_{2X} (K_{2P} + p) s}{k_2 K_{2P} K_S}}.$$

The optimal enzyme concentration now equals $e_{2,s}^o = e_{2,s}(x_s^o)$, and the optimal flux equals $j^o = j(x_s^o)$. We note that

$$e_{2,s}^o = \frac{e_T k_1 s \sqrt{K_{1X} K_{2P}} \sqrt{K_{2P} + p} + e_T \sqrt{k_1 k_2 K_{2X} K_S s} (K_{2P} + p)}{\sqrt{K_{1X} K_{2P}} \sqrt{K_{2P} + p} (k_1 s + k_2 (K_S + s)) + 2\sqrt{k_1 k_2 K_{2X} K_S s} (K_{2P} + p)}$$

Although this is a complicated expression, it is nothing but an explicit function in terms of the substrate and product concentrations s and p.

Next, we want to construct the optimal enzyme allocation, as a function of the metabolite X. The idea is that $e_{2,s}^o = e_{2,s}(x_s^o)$ is the optimal input-output relationship of the gene network with x_s^o as its input and $e_{2,s}^o$ as its output. The problem is that the relation $e_{2,s}^o = e_{2,s}(x_s^o)$ we have derived still depends on s and p. In this example, we keep p fixed, but suppose that s may vary. The steady state x_s then varies with it. But we can invert the relation between s and x_s^o,

$$s = \frac{k_2 K_{2P} K_S (x_s^0)^2}{k_1 K_{1X} K_{2X} (K_{2P} + p)}$$

we can substitute this into $e_{2,s}^o$. The interpretation of this s is that is the value of s at which x_s^o is the optimal steady state value of x. If we now substitute this into $e_{2,s}^o$, we have found an expression for $e_{2,s}^o$ in terms of x. This is the input-output relation between $e_{2,s}^o$ and x_s^o, denoted by $e_{2,s}^o = \tilde{e}_{2,s}^o(x_s^o)$, that we require for finding the optimal genetic circuit.

To implement this optimal allocation, we are going to model the genetic circuit really simply, using synthesis and dilution by growth, as

$$\frac{de_2}{dt} = f(x) - e_2$$

and $e_1 = e_T - e_2$. This means that the steady state concentration of e_2 equals $e_{2,s} = f(x_s)$. Thus, for optimal behaviour, we need to require $f(x_s) := \tilde{e}_{2,s}^o(x_s^o)$. In other words, we have written down an explicit synthesis rate, as a function of the metabolite concentration x, with the property that if x is at steady state, the optimal synthesis rate for enzyme 2 is induced (and automatically also for enzyme 1). The surprising thing is that this function generally resembles a hyperbolic or a Hill-type relationship. Such a relationship is simple to achieve from the biochemical interactions between the sensor metabolite X, a transcription factor and a promoter. When you make such a model it is then easy to show that dynamics of the coupled system (metabolic pathway plus enzyme synthesis) track the environment in an optimal manner, without using the environment to control enzyme synthesis directly.

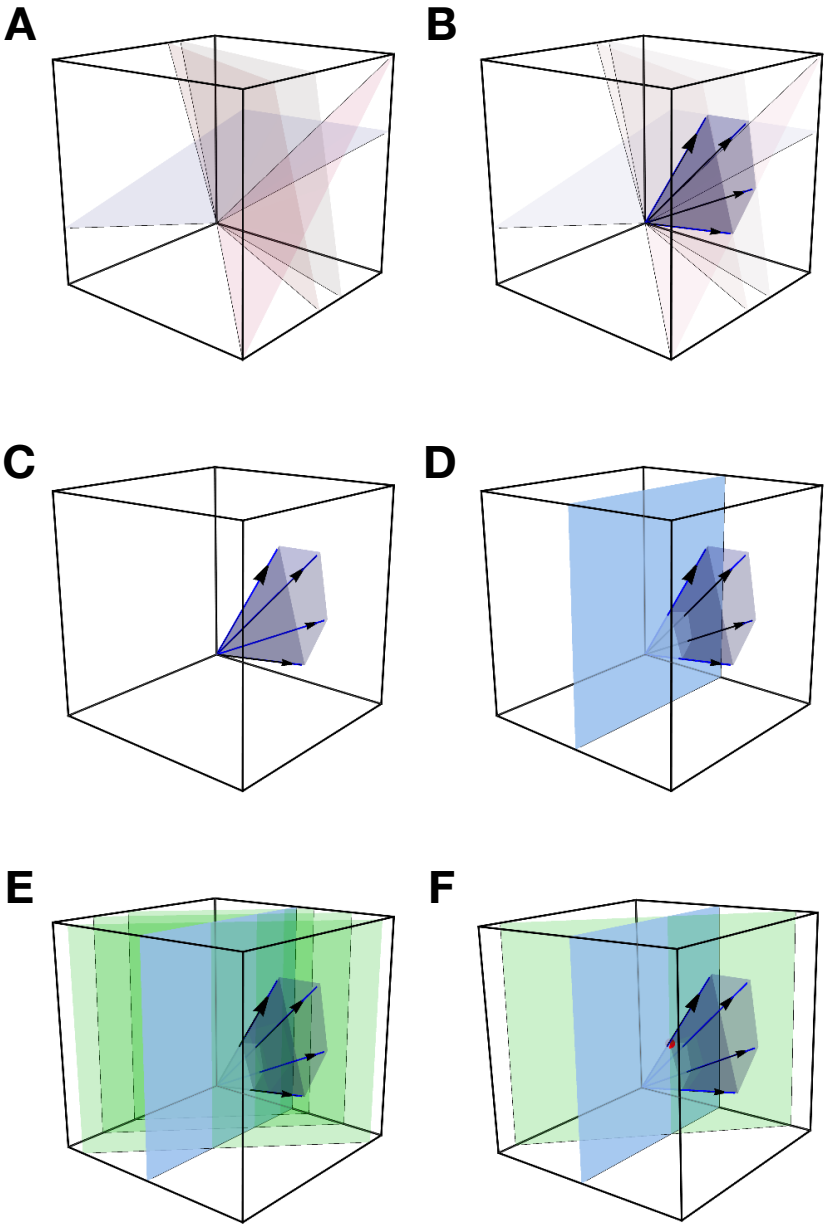

Figure S1: Illustration of the proof that EFMs are the maximisers of linear programs with a single constraint. A. The flux planes that result from setting the mass balances to zero are shown. B. The intersection of the flux planes define a cone. C. The cone EFMs at is corners, the vectors. D. The intersection of the cone with the constraint plane. E. Movement of the objective plane through the cone until the objective plane leaves the cone. F. The optimal value lies on the border of the cone, on an EFM, on the constraint plane.

Even though the behaviour of the network is now optimal, mutations can still

increase the optimal growth rate. This can be understood by studying the enzyme kinetics once more. It shows that if x_s/K_{1X} would be made smaller (reduced product inhibition) and x_s/K_{2X} would be made higher (increased substrate saturation), the optimal flux would increase at the same optimal enzyme investment. To investigate this we define a new parameter $\alpha = K_{2X}/K_{1X}$ that we substitute into the optimal relationships $e^o_{2,s} = e_{2,s}(x^o_s)$ and $j^o = j(x^o_s)$ by eliminating K_{1X}, and then we take the limit of α to infinity. We obtain,

$$\lim_{\alpha\to\infty} e^o_2 = \frac{e_T k_1 s}{k_1 s + k_2(K_S + s)} \text{ and } \lim_{\alpha\to\infty} j^o = k_2 \frac{e_T k_1 s}{k_1 s + k_2(K_S + s)}.$$

Thus, in this limit

$$\frac{j^o(s)}{e_T} = k_2 \frac{e^0_{2,s}(s)}{e_T} \propto \frac{e^o_{2,s}}{e_T}.$$

This recovers the linearity of the growth law of Maaloe [77] and Hwa [71],

$$\frac{e^0_{2,s}(s)}{e_T} = \frac{1}{K_2} \frac{j^o(s)}{e_T},$$

with $j^o(s)/e_T$ as a growth rate measure. The linearity of the growth law is perhaps due to optimisation of the amino-acid affinities of amino-acid synthesis and protein synthesis by the ribosome to reduce product inhibition and enhance ribosome saturation with substrate.

Elementary flux modes maximise J/p_T

We start from an arbitrary network containing enzymes with arbitrary enzyme kinetics, and the objective is to minimise the amount of total enzyme concentration, e_T, needed for 1 unit flux to a target reaction T in the network, which can be stated mathematically as [76]

$$\min_{\mathbf{m},\mathbf{e}} \left\{ \sum_{i=1}^{r} e_i \middle| v_T = 1, \mathbf{Nv} = \mathbf{0}, \forall i : v_i = k_i e_i f_i(\mathbf{m}), \forall i : \ e_i \geq 0 \right\}.$$

Next, we reformulate the optimisation and remove the enzyme concentrations and obtain rates as optimisation variables,

$$\min_{\mathbf{m},\mathbf{v}} \left\{ \sum_{i=1}^{r} \frac{v_i}{k_i f_i(\mathbf{m})} \middle| v_T = 1, \mathbf{Nv} = \mathbf{0} \right\}.$$

Now we split all the reversible reactions into two (we hereby increase the number of columns of the stoichiometric matrix) and force both those new rates to have positive fluxes, by multiplying the corresponding entries in the stoichiometric matrix by -1,

$$\min_{\mathbf{m},\mathbf{v}} \left\{ \sum_{i=1}^{r} \frac{v_i}{k_i f_i(\mathbf{m})} \middle| v_T = 1, \mathbf{N}'\mathbf{v}' = 0, \mathbf{v}' \geq \mathbf{0} \right\}.$$

Next, we fix metabolite concentrations $\mathbf{m}$ and set $\alpha_i = 1/f_i(\mathbf{m})$, so

$$\min_{\mathbf{v}'} \left\{ \sum_{i=1}^{r} \alpha_i v_i' \middle| v_T = 1, \mathbf{N}'\mathbf{v}' = 0, \mathbf{v}' \geq \mathbf{0} \right\}.$$

The saturation functions have now become scalars, gone is the nonlinearity and we are in the realm of linear algebra. The (convex) set $\left\{ \mathbf{N}'\mathbf{v}' = 0, \mathbf{v}' \geq \mathbf{0} \right\}$ is a cone (Figure S1A). It is spanned by its rays (vectors) that are in metabolic theory exactly the *elementary flux mode* (EFM) of the metabolic network [82, 84]. The cone is intersected with plane $v_T = 1$ (Figure S1B), defining the so-called feasible solution space (a polyhedron). A standard result in linear convex optimization is that the optimum must be one of the cornerpoints of this feasible solution space, and hence this is one of the spanning vectors of the cone. (Or in a very unlikely case a convex combination of them, when the objective plane is parallel to a facet of the cone.) This argument is valid for any choice of metabolite concentrations, so in particular, for any optimal choice of these.

The flux through the network that leads to minimal protein usage must be contained in an EFM.

A self-fabrication models leads to a nonlinear balanced-growth equation

We consider the simplest self-fabricating model (see [26] for the general exposition), with one metabolite X that is taken up by enzyme E and used by the ribosome R to synthesise E and itself. We have the following rate of change equations for the concentrations, x, e, and r,

$$\begin{aligned}\frac{dx}{dt} &= ef(x) - r\alpha_E g_E(x) - r\alpha_R g_R(x) - \mu x \\ \frac{de}{dt} &= r\alpha_E g_E(x) - \mu e \\ \frac{dr}{dt} &= r\alpha_R g_R(x) - \mu r\end{aligned} \quad ,$$

with α_E and α_R as the fraction of ribosome allocated to E or R, the functions $f(x)$, $g_E(x)$ and $g_R(x)$ are nonlinear saturation function (enzyme kinetics), μ is growth rate and concentrations are defined as molecule numbers (n) divided by volume V.

We assume, for simplicity, that the volume of a cell is the sum of its protein volumes,

$$V = v_E n_E + v_R n_R,$$

with v_E and v_R as the protein volumes of E and R, respectively. Thus, the growth rate equals,

$$\begin{aligned}\mu &= \frac{1}{V}\frac{dV}{dt} \\ &= v_E \frac{1}{V}\frac{dn_E}{dt} + v_R \frac{1}{V}\frac{dn_R}{dt} \\ &= r(v_E \alpha_E g_E(x) + v_R \alpha_R g_R(x))\end{aligned} \quad .$$

This leads to the equation

$$\frac{\mu}{r} = v_E \alpha_E g_E(x) + v_R \alpha_R g_R(x).$$

Next, we solve part of the balanced growth equations

$$\frac{dx}{dt} = 0 \text{ and } \frac{de}{dt} = 0,$$

so that

$$\begin{aligned}\mu x &= ef(x) - r\alpha_E g_E(x) - r\alpha_R g_R(x) \\ &= r\frac{\alpha_E g_E(x)}{\mu} f(x) - r\alpha_E g_E(x) - r\alpha_R g_R(x)\end{aligned}$$

or

$$\frac{\mu x}{r} = \frac{\alpha_E g_E(x)}{\mu} f(x) - \alpha_E g_E(x) - \alpha_R g_R(x).$$

The two equations in the boxes can be combined to

$$x(v_E \alpha_E g_E(x) + v_R \alpha_R g_R(x)) = \frac{\alpha_E g_E(x)}{\mu} f(x) - \alpha_E g_E(x) - \alpha_R g_R(x).$$

Now using from the final balanced growth requirement, $\frac{dr}{dt} = 0$, from which we deduce $\mu = \alpha_R g_R(x)$, we finally find

$$(1 + x v_R)\mu^2 + (\alpha_E g_E(x) + x v_E \alpha_E g_E(x))\mu - \alpha_E g_E(x) f(x) = 0.$$

This a nonlinear equation in the growth rate, μ, and the metabolite concentrations. The quadratic dependence in the growth rate can be directly attributed to the ribosome, which needs to make both itself and the enzyme. (If the metabolite would in turn catalyse a fourth kind of molecule, the relation would be third order in the growth rate.) When we set the concentration of x fixed we obtain a ME-model, in the spirit of O'Brien and Palsson. The reader is invited to read [26] for the mathematical definition of elementary growth modes. It is built directly on top of systems of nonlinear balanced growth equations such as the one we derived here, but then for whole cells.

Bibliography

1. Gibbons, S.M. and J.A. Gilbert, Microbial diversity--exploration of natural ecosystems and microbiomes. Curr Opin Genet Dev, 2015. **35**: p. 66-72.
2. Falkowski, P.G., T. Fenchel, and E.F. Delong, The microbial engines that drive Earth's biogeochemical cycles. Science, 2008. **320**(5879): p. 1034-9.
3. Neidhart, F.C., J.L. Ingraham, and M. Schaechter, Physiology of the bacterial cell: a molecular approach. 1990: Sinauer Associates, Inc.
4. Peregrin-Alvarez, J.M., C. Sanford, and J. Parkinson, The conservation and evolutionary modularity of metabolism. Genome Biol, 2009. **10**(6): p. R63.
5. Cornish-Bowden, A., Fundamentals of enzyme kinetics. 2012: Wiley-Blackwell.
6. Noor, E., et al., Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. Mol Cell, 2010. **39**(5): p. 809-20.
7. Rabinowitz, J.D. and L. Vastag, Teaching the design principles of metabolism. Nat Chem Biol, 2012. **8**(6): p. 497-501.
8. Schuster, S.C. and R. Heinrich, *The regulation of cellular systems*. 1996: Chapman & Hall.
9. Orr, H.A., Fitness and its role in evolutionary genetics. Nat Rev Genet, 2009. **10**(8): p. 531-9.
10. Kluyver, A.J., Eenheid en verscheidenheid in de stofwisseling der microben. Chemisch weekblad, 1924. **2**(1): p. 266.
11. Singleton, R., Jr. and D.R. Singleton, Remembering Our Forebears: Albert Jan Kluyver and the Unity of Life. J Hist Biol, 2017. **50**(1): p. 169-218.
12. Franchini, A.G. and T. Egli, Global gene expression in Escherichia coli K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. Microbiology, 2006. **152**(Pt 7): p. 2111-2127.
13. Saether, B.E. and S. Engen, The concept of fitness in fluctuating environments. Trends Ecol Evol, 2015. **30**(5): p. 273-81.
14. Bull, J., Evolution of phenotypic variance. Evolution, 1987. **41**: p. 303-315.
15. Haccou, P. and Y. Iwasa, Optimal mixes strategies in stochastic environments. Theoretical Population Biology, 1995. **47**(2): p. 212-243.
16. Lewontin, R.C. and D. Cohen, On population growth in a randomly varying environment. Proc Natl Acad Sci U S A, 1969. **62**(4): p. 1056-60.

17. Kussell, E. and S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments. Science, 2005. **309**(5743): p. 2075-8.
18. Rivoire, O. and S. Leibler, The value of information for populations in varying environments. Journal of Statistical Physics, 2011. **142**: p. 1124-1166.
19. Wolf, L., O.K. Silander, and E. van Nimwegen, Expression noise facilitates the evolution of gene regulation. Elife, 2015. **4**.
20. Foster, P.L., et al., Determinants of spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing. Proc Natl Acad Sci U S A, 2015. **112**(44): p. E5990-9.
21. Kussell, E., *Evolution in microbes*. Annu Rev Biophys, 2013. **42**: p. 493-514.
22. Lenski, R.E., Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J, 2017. **11**(10): p. 2181-2194.
23. Fisher, R.A., The genetical theory of natural selection. 1930, Oxford: Clarendon Press.
24. Price, C.E., et al., Adaption to glucose limitation is modulated by the pleotropic regulator CcpA, independent of selection pressure strength. BMC Evol Biol, 2019. **19**(1): p. 15.
25. Planque, R., et al., Maintaining maximal metabolic flux by gene expression control. PLoS Comput Biol, 2018. **14**(9): p. e1006412.
26. de Groot, D.H., et al., Elementary Growth Modes provide a molecular description of cellular self-fabrication. PLoS Comput Biol, 2020. **16**(1): p. e1007559.
27. Scott, M., et al., Emergence of robust growth laws from optimal regulation of ribosome synthesis. Mol Syst Biol, 2014. **10**: p. 747.
28. Bosdriesz, E., et al., How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. FEBS J, 2015. **282**(10): p. 2029-44.
29. Campbell, A., Synchronization of cell division. Bacteriological reviews, 1957. **21**: p. 263-272.
30. Painter, P.R. and A.G. Marr, Mathematics of microbial populations. Annu Rev Microbiol, 1968. **22**: p. 519-48.
31. Jun, S., et al., Fundamental principles in bacterial physiology-history, recent progress, and the future with focus on cell size control: a review. Rep Prog Phys, 2018. **81**(5): p. 056601.
32. Young, J.W., et al., Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. Nat Protoc, 2011. **7**(1): p. 80-8.

33. van Heerden, J.H., et al., Statistics and simulation of growth of single bacterial cells: illustrations with B. subtilis and E. coli. Sci Rep, 2017. **7**(1): p. 16094.
34. Cooper, S. and C.E. Helmstetter, Chromosome replication and the division cycle of Escherichia coli B/r. J Mol Biol, 1968. **31**(3): p. 519-40.
35. Cooper, S., Bacterial growth and division: biochemistry and regulation of prokaryotic and eukaryotic division cycles. 1991: Academic Press, Inc.
36. de Jong, H., et al., Mathematical modelling of microbes: metabolism, gene expression and growth. J R Soc Interface, 2017. **14**(136).
37. Orth, J.D., I. Thiele, and B.O. Palsson, *What is flux balance analysis?* Nat Biotechnol, 2010. **28**(3): p. 245-8.
38. Heinrich, R., S.M. Rapoport, and T.A. Rapoport, *Metabolic regulation and mathematical models*. Prog Biophys Mol Biol, 1977. **32**(1): p. 1-82.
39. Cleland, W.W., The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. Biochim Biophys Acta, 1963. **67**: p. 104-37.
40. Cleland, W.W., The kinetics of enzyme-catalyzed reactions with two or more substrates or products. II. Inhibition: nomenclature and theory. Biochim Biophys Acta, 1963. **67**: p. 173-87.
41. Cleland, W.W., The kinetics of enzyme-catalyzed reactions with two or more substrates or products. III. Prediction of initial velocity and inhibition patterns by inspection. Biochim Biophys Acta, 1963. **67**: p. 188-96.
42. Giersch, C., Control analysis of metabolic networks. 1. Homogeneous functions and the summation theorems for control coefficients. Eur J Biochem, 1988. **174**(3): p. 509-13.
43. Hofmeyr, J.H., Metabolic regulation: a control analytic perspective. J Bioenerg Biomembr, 1995. **27**(5): p. 479-90.
44. Zimmerman, S.B. and S.O. Trach, Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. J Mol Biol, 1991. **222**(3): p. 599-620.
45. Berg, H.C. and E.M. Purcell, *Physics of chemoreception*. Biophys J, 1977. **20**(2): p. 193-219.
46. Calef, D.F. and J.M. Deutch, *Diffusion-controlled reactions*. Annual Review of Physical Chemistry, 1983. **34**: p. 493-523.
47. Wiegel, F.W., Diffusion and the physics of chemoreception. Physics Reports, 1983. **95**(5): p. 283-319.

48. Guigas, G. and M. Weiss, Effects of protein crowding on membrane systems. Biochim Biophys Acta, 2016. **1858**(10): p. 2441-2450.
49. Zimmerman, S.B. and A.P. Minton, Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct, 1993. **22**: p. 27-65.
50. Klumpp, S., et al., Molecular crowding limits translation and cell growth. Proc Natl Acad Sci U S A, 2013. **110**(42): p. 16754-9.
51. Vazquez, A., Optimal macromolecular density in the cell. Proc Natl Acad Sci U S A, 2012. **109**(9): p. E533; author reply E534.
52. Molenaar, D., et al., Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol, 2009. **5**: p. 323.
53. Kjeldgaard, N.O., *Dynamics of bacterial growth*. 1963, Copenhagen: Nyt Nordisk Forlag Arnold Busck.
54. Maaloe, O., *An analysis of bacterial growth*. Development Biology Supplement, 1969. **3**: p. 33-58.
55. Ehrenberg, M. and C.G. Kurland, Costs of accuracy determined by a maximal growth rate constraint. Q Rev Biophys, 1984. **17**(1): p. 45-82.
56. Dekel, E. and U. Alon, Optimality and evolutionary tuning of the expression level of a protein. Nature, 2005. **436**(7050): p. 588-92.
57. Kalisky, T., E. Dekel, and U. Alon, Cost-benefit theory and optimal design of gene regulation functions. Phys Biol, 2007. **4**(4): p. 229-45.
58. Berkhout, J., et al., How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. Genetics, 2013. **194**(2): p. 505-12.
59. Jensen, P.R., O. Michelsen, and H.V. Westerhoff, Control analysis of the dependence of Escherichia coli physiology on the H(+)-ATPase. Proc Natl Acad Sci U S A, 1993. **90**(17): p. 8068-72.
60. Koebmann, B.J., et al., Experimental determination of control of glycolysis in Lactococcus lactis. Antonie Van Leeuwenhoek, 2002. **82**(1-4): p. 237-48.
61. Koebmann, B., C. Solem, and P.R. Jensen, Control analysis as a tool to understand the formation of the las operon in Lactococcus lactis. FEBS J, 2005. **272**(9): p. 2292-303.
62. Solem, C., et al., The las enzymes control pyruvate metabolism in Lactococcus lactis during growth on maltose. J Bacteriol, 2007. **189**(18): p. 6727-30.

63. Solem, C., B. Koebmann, and P.R. Jensen, Control analysis of the role of triosephosphate isomerase in glucose metabolism in Lactococcus lactis. IET Syst Biol, 2008. **2**(2): p. 64-72.
64. Bren, A., et al., Glucose becomes one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal levels of cAMP. Sci Rep, 2016. **6**: p. 24834.
65. Towbin, B.D., et al., Optimality and sub-optimality in a bacterial growth law. Nat Commun, 2017. **8**: p. 14123.
66. van der Vlag, J., et al., Control of glucose metabolism by the enzymes of the glucose phosphotransferase system in Salmonella typhimurium. Eur J Biochem, 1995. **230**(1): p. 170-82.
67. Keren, L., et al., Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell, 2016. **166**(5): p. 1282-1294 e18.
68. Bentley, W.E., et al., Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng, 1990. **35**(7): p. 668-81.
69. Dong, H., L. Nilsson, and C.G. Kurland, Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. J Bacteriol, 1995. **177**(6): p. 1497-504.
70. Snoep, J.L., et al., Protein burden in Zymomonas mobilis: negative flux and growth control due to overproduction of glycolytic enzymes. Microbiology, 1995. **141**(9): p. 2329-2337.
71. Scott, M., et al., Interdependence of cell growth and gene expression: origins and consequences. Science, 2010. **330**(6007): p. 1099-102.
72. Kacser, H. and J.A. Burns, *The control of flux*. Symp Soc Exp Biol, 1973. **27**: p. 65-104.
73. Fell, D.A., Understanding the control of metabolism. 1997: Ashgate publishing.
74. Chubukov, V., et al., Coordination of microbial metabolism. Nat Rev Microbiol, 2014. **12**(5): p. 327-40.
75. Berkhout, J., B. Teusink, and F.J. Bruggeman, Gene network requirements for regulation of metabolic gene expression to a desired state. Sci Rep, 2013. **3**: p. 1417.
76. Wortel, M.T., et al., Metabolic states with maximal specific rate carry flux through an elementary flux mode. FEBS J, 2014. **281**(6): p. 1547-55.

77. Schaechter, M., O. Maaloe, and N.O. Kjeldgaard, Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. J Gen Microbiol, 1958. **19**(3): p. 592-606.
78. Potrykus, K. and M. Cashel, *(p)ppGpp: still magical?* Annu Rev Microbiol, 2008. **62**: p. 35-51.
79. Bollenbach, T., et al., Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. Cell, 2009. **139**(4): p. 707-18.
80. Cox, R.A., Quantitative relationships for specific growth rates and macromolecular compositions of Mycobacterium tuberculosis, Streptomyces coelicolor A3(2) and Escherichia coli B/r: an integrative theoretical approach. Microbiology, 2004. **150**(Pt 5): p. 1413-1426.
81. Metzl-Raz, E., et al., Principles of cellular resource allocation revealed by condition-dependent proteome profiling. Elife, 2017. **6**.
82. Hilgetag, C. and S. Schuster, On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems, 1994. **2**(2): p. 165-182.
83. Schuster, S., D.A. Fell, and T. Dandekar, A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol, 2000. **18**(3): p. 326-32.
84. Gagneur, J. and S. Klamt, Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics, 2004. **5**: p. 175.
85. Cornish-Bowden, A., The effect of natural selection on enzyme catalysis. Journal of Molecular Biology, 1976. **101**(1): p. 1-9.
86. Mori, M., et al., Quantifying the benefit of a proteome reserve in fluctuating environments. Nat Commun, 2017. **8**(1): p. 1225.
87. Dill, K.A., K. Ghosh, and J.D. Schmit, *Physical limits of cells and proteomes*. Proc Natl Acad Sci U S A, 2011. **108**(44): p. 17876-82.
88. Nilsson, A. and J. Nielsen, Metabolic Trade-offs in Yeast are Caused by F1F0-ATP synthase. Sci Rep, 2016. **6**: p. 22264.
89. Neidhardt, F.C., Bacterial growth: constant obsession with dN/dt. J Bacteriol, 1999. **181**(24): p. 7405-8.
90. Schaechter, M., From growth physiology to systems biology. Int Microbiol, 2006. **9**(3): p. 157-61.
91. Monod, J., *The growth of bacterial cultures*. Annual Review of Microbiology, 1949. **3**: p. 371-394.

92. Gorke, B. and J. Stulke, Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol, 2008. **6**(8): p. 613-24.
93. Hermsen, R., et al., A growth-rate composition formula for the growth of E.coli on co-utilized carbon substrates. Mol Syst Biol, 2015. **11**(4): p. 801.
94. Okano, H., et al., Regulation underlying hierarchical and simultaneous utilization of carbon substrates by flux sensors in Escherichia coli. Nat Microbiol, 2020. **5**(1): p. 206-215.
95. Van Hoek, P., J.P. Van Dijken, and J.T. Pronk, Effect of specific growth rate on fermentative capacity of baker's yeast. Appl Environ Microbiol, 1998. **64**(11): p. 4226-33.
96. Hui, S., et al., Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. Mol Syst Biol, 2015. **11**(1): p. 784.
97. Umbarger, H.E., A one-semester project for the emersion of graduate students in metabolic pathways. Biochemical Education, 1977. **5**(4): p. 67-71.
98. Stouthamer, A.H., A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Van Leeuwenhoek, 1973. **39**(3): p. 545-65.
99. De Hollander, J.A., The use of stoichiometric relations for the description and analysis of microbial cultures. Antonie Van Leeuwenhoek, 1991. **60**: p. 257-273.
100. Roels, J.A., Energetics and kinetics in biotechnology. 1983: Elsevier Biomedical Press.
101. Price, N.D., J.L. Reed, and B.O. Palsson, Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol, 2004. **2**(11): p. 886-97.
102. Varma, A., B.W. Boesch, and B.O. Palsson, Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl Environ Microbiol, 1993. **59**(8): p. 2465-73.
103. Varma, A. and B.O. Palsson, Metabolic capabilities of Escherichia coli: I. synthesis of biosynthetic precursors and cofactors. J Theor Biol, 1993. **165**(4): p. 477-502.
104. Varma, A. and B.O. Palsson, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol, 1994. **60**(10): p. 3724-31.
105. Fell, D.A. and J.R. Small, Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochem J, 1986. **238**(3): p. 781-6.

106. Goelzer, A., et al., Quantitative prediction of genome-wide resource allocation in bacteria. Metab Eng, 2015. **32**: p. 232-243.
107. O'Brien, E.J. and B.O. Palsson, Computing the functional proteome: recent progress and future prospects for genome-scale models. Curr Opin Biotechnol, 2015. **34**: p. 125-34.
108. Mori, M., et al., Constrained Allocation Flux Balance Analysis. PLoS Comput Biol, 2016. **12**(6): p. e1004913.
109. Cayley, S. and M.T. Record, Jr., Large changes in cytoplasmic biopolymer concentration with osmolality indicate that macromolecular crowding may regulate protein-DNA interactions and growth rate in osmotically stressed Escherichia coli K-12. J Mol Recognit, 2004. **17**(5): p. 488-96.
110. Muller, S., G. Regensburger, and R. Steuer, Enzyme allocation problems in kinetic metabolic networks: optimal solutions are elementary flux modes. J Theor Biol, 2014. **347**: p. 182-90.
111. Noor, E., et al., The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. PLoS Comput Biol, 2016. **12**(11): p. e1005167.
112. de Groot, D.H., et al., The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. PLoS Comput Biol, 2019. **15**(3): p. e1006858.
113. Famili, I., et al., Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13134-9.
114. de Groot, D.H., et al., The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. Cell Mol Life Sci, 2020. **77**(3): p. 441-453.
115. Basan, M., et al., Overflow metabolism in Escherichia coli results from efficient proteome allocation. Nature, 2015. **528**(7580): p. 99-104.
116. O'Brien, E.J., et al., Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol, 2013. **9**: p. 693.
117. Ihssen, J. and T. Egli, Global physiological analysis of carbon- and energy-limited growing Escherichia coli confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. Environ Microbiol, 2005. **7**(10): p. 1568-81.

118. Berney, M., et al., Specific growth rate determines the sensitivity of Escherichia coli to thermal, UVA, and solar disinfection. Appl Environ Microbiol, 2006. **72**(4): p. 2586-93.
119. Nystrom, T., Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? Mol Microbiol, 2004. **54**(4): p. 855-62.
120. Klipp, E. and R. Heinrich, Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. Biosystems, 1999. **54**(1-2): p. 1-14.