UNIVERSITY OF AMSTERDAM

# UvA-DARE (Digital Academic Repository)

## Whole-cell metabolic control analysis

Bruggeman, F.J.; Remeijer, M.; Droste, M.; Salinas, L.; Wortel, M.; Planqué, R.; Sauro, H.M.; Teusink, B.; Westerhoff, H.V.

**DOI**
10.1016/j.biosystems.2023.105067

**Publication date**
2023

**Document Version**
Final published version

**Published in**
Biosystems

**License**
CC BY

Link to publication

**Citation for published version (APA):**
Bruggeman, F. J., Remeijer, M., Droste, M., Salinas, L., Wortel, M., Planqué, R., Sauro, H. M., Teusink, B., & Westerhoff, H. V. (2023). Whole-cell metabolic control analysis. *Biosystems*, *234*, Article 105067. https://doi.org/10.1016/j.biosystems.2023.105067

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)*

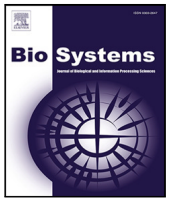
Download date:27 May 2024

BioSystems 234 (2023) 105067

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems



# Whole-cell metabolic control analysis

Frank J. Bruggeman [a,*], Maaike Remeijer [a], Maarten Droste [a,b], Luis Salinas [a], Meike Wortel [c], Robert Planqué [b], Herbert M. Sauro [d], Bas Teusink [a], Hans V. Westerhoff [a]

[a] Systems Biology Lab, A-LIFE, AIMMS, VU University, Amsterdam, Netherlands
[b] Department of Mathematics, VU University, Amsterdam, Netherlands
[c] Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands
[d] Department of Bioengineering, University of Washington, Seattle, WA, 98195-5061, USA

ARTICLE INFO

*Keywords:*
Whole cell models
Elementary flux modes
Metabolic control analysis
Enzyme kinetics

ABSTRACT

Since its conception some fifty years ago, metabolic control analysis (MCA) aims to understand how cells control their metabolism by adjusting the activity of their enzymes. Here we extend its scope to a whole-cell context. We consider metabolism in the evolutionary context of growth-rate maximisation by optimisation of protein concentrations. This framework allows for the prediction of flux control coefficients from proteomics data or stoichiometric modelling. Since genes compete for finite biosynthetic resources, we treat all protein concentrations as interdependent. We show that elementary flux modes (EFMs) emerge naturally as the optimal metabolic networks in the whole-cell context and we derive their control properties. In the evolutionary optimum, the number of expressed EFMs is determined by the number of protein-concentration constraints that limit growth rate. We use published glucose-limited chemostat data of *S. cerevisiae* to illustrate that it uses only two EFMs prior to the onset of fermentation and that it uses four EFMs during fermentation. We discuss published enzyme-titration data to show that *S. cerevisiae* and *E. coli* indeed can express proteins at growth-rate maximising concentrations. Accordingly, we extend MCA to elementary flux modes operating at an optimal state. We find that the expression of growth-unassociated proteins changes results from classical metabolic control analysis. Finally, we show how flux control coefficients can be estimated from proteomics and ribosome-profiling data. We analyse published proteomics data of *E. coli* to provide a whole-cell perspective of the control of metabolic enzymes on growth rate. We hope that this paper stimulates a renewed interest in metabolic control analysis, so that it can serve again the purpose it once had: to identify general principles that emerge from the biochemistry of the cell and are conserved across biological species.

## 1. Introduction

Microbial systems biology has made enormous advances in the last two decades (see for instance, Scott et al., 2014; Price et al., 2004; Kochanowski et al., 2021; Scott et al., 2010; Bruggeman et al., 2020; Scott and Hwa, 2022). A first major insight has been that protein expression is bounded by constraints that emerge from limited biosynthetic resources (Scott et al., 2010; Molenaar et al., 2009). Competition for such resources by active genes implies that different cellular functions can trade off (Basan et al., 2020; Berney et al., 2006; Ihssen and Egli, 2004; Utrilla et al., 2016; Bruggeman et al., 2023; Nyström, 2004). The second main advance has been that cellular growth rate (equal to the protein synthesis rate per unit cellular protein) can be related to the activities of its entire metabolic network (Scott et al., 2010; Molenaar et al., 2009). Together these two insights have allowed for a whole-cell perspective on protein expression into protein sectors (Scott et al., 2010; Hui et al., 2015), and its implications under the assumption that natural selection maximises growth rate (O'Brien et al., 2013; Elsemman et al., 2022). These views allow for the appreciation of the functions of individual molecular networks in the context of the entire cell (Mori et al., 2021; Hui et al., 2015). Mathematical models of whole-cell metabolism exist nowadays that can be used to predict and interpret whole-cell proteomics and fluxomics data (Elsemman et al., 2022; O'Brien et al., 2013). This framework, which integrates quantitative molecular data, computational modelling and microbial physiology, has proven extremely rewarding (Bruggeman et al., 2020; Scott and Hwa, 2022).

These recent developments followed after the introduction of genome-scale stoichiometric models in the 1990s (Varma and Palsson, 1993a,b; Varma et al., 1993; Price et al., 2004) and, before that, a

* Corresponding author.
*E-mail address:* f.j.bruggeman@vu.nl (F.J. Bruggeman).

https://doi.org/10.1016/j.biosystems.2023.105067
Received 9 May 2023; Received in revised form 23 October 2023; Accepted 25 October 2023
Available online 30 October 2023
0303-2647/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

general understanding of enzyme biochemistry in the 1960s (Cornish-Bowden, 2013b; Monod et al., 1963; Cleland, 1963a,b,c). Both of these rely on system-level concepts and quantitative approaches to the biochemistry of the cell. Such system ideas can already be found in the pioneering works of Monod and Maaløe in the 1950s; both have had an enormous influence on current systems biology (Maaløe, 2012; Monod, 1974; Schaechter, 2006; Monod et al., 1963). Maaløe in particular has had an important influence on our current thinking on the protein-expression constraints that limit microbial growth (Maaløe, 2012; Maaløe and Kjeldgaard, 1966).

The introduction of 'systems thinking' in the study of the biochemistry of metabolism has been key to these developments (Kacser, 1986). This involved the understanding that metabolic enzymes should be understood in their 'systemic context' of metabolic pathways and that, since their rates depend on the concentrations of all of their reactants and modifiers, no enzyme can change in rate without affecting the rate of others as they are all jointly active in a network (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). Therefore, there is a priori no reason to believe that a single enzyme in a metabolic network dictates the rates of all others (Kacser and Burns, 1979). All enzymatic rates are set by their concerted actions, and ultimately by their kinetic properties (encoded in genetic information) and the prevailing physicochemical conditions, both in and outside the cell. The cell behaves as an 'organised whole' (Kacser, 1986).

This concept of 'systems biochemistry' has its roots in metabolic control analysis, with its first papers appearing 50 years ago (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). This theory was a logical continuation of quantitative enzymology (Cleland, 1963a,b,c; Koshland Jr. et al., 1966; Monod et al., 1965; Hill, 1977), the elucidation of the key metabolic pathways of catabolism and anabolism (Newsholme, 2009) and the application of non-equilibrium thermodynamics to (catabolic) metabolism (Westerhoff and Van Dam, 1987). At that time, the biochemistry of enzyme catalysis was understood in quantitative terms. Enzymatic rate equations were deduced from underlying models of enzyme mechanisms and associated experimental methods had been developed for the inference of these mechanisms and their kinetic parameters (Cleland, 1963a,b,c). Together with the basic knowledge of metabolic pathways (which was completed around the 1960-70s Newsholme, 2009), the question naturally arose how enzymes operate together and define key features of life. Metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974) and biochemical systems theory (Savageau, 1969a,b, 1970) emerged as two ways to address this scientific challenge. These two theories were always more complementary than competing, as they differed in their approach.[1]

The identification and kinetic understanding of cooperative enzymes (Koshland Jr. et al., 1966; Monod et al., 1965), allosteric feedback and feedforward control (Umbarger, 1956), and the occurrence of (thermodynamically) irreversible enzymes (Goldbeter, 2018) led to the hypothesis that such enzymes act as a single 'rate-limiting', 'pacemaker', 'flux-generating' or 'key' enzyme in metabolism — all qualifications of which the equivalence was uncertain (Krebs, 1957; Bücher and Rüssmann, 1963; Newsholme, 1980). These enzymes were, after all, the locations in cellular metabolism where regulation occurred and where Gibbs free energy was either harvested or invested. What was overlooked in these (overly reductive) deductions is that all enzymes are linked via the concentrations of their reactants and modifiers, which gives all of them in principle the capacity to influence and limit the metabolic flux to varying degrees (Fell, 1997). Although mathematical descriptions of enzyme kinetics were common in the field at the time, mathematical models of metabolic pathways were not yet mainstream[2] and most discussions about pathways were therefore based on intuitions. The inherent nonlinearity of metabolic network dynamics made it hard to make correct predictions without mathematical models (Heinrich et al., 1978).

Another confounding factor was the lack of a strict terminology (Westerhoff et al., 1984). What was precisely meant by 'metabolic flux'? Under which conditions can it be measured? What is precisely meant by 'rate limiting'? Is 'rate limitation' a continuous or a discrete measure? How can it be measured? How do we describe the interconnectedness of enzymes in networks precisely enough to make predictions? What the field needed was a way to integrate the quantitative enzyme kinetics description of single reactions to be able to predict properties of whole metabolic pathways such as flux and regulation. Since not in all cases the exact enzyme kinetics (and its parameterisation) was known, this needed to come from a 'parameter-independent' approach. This is what metabolic control analysis eventually achieved. Two seminal, pioneering papers addressed that issue for the first time (Kacser and Burns, 1973; Heinrich and Rapoport, 1974), now fifty years ago. Soon after, experimental illustrations indicated the experimental applicability of the theory and proved that control of metabolic fluxes indeed need not be confined to single 'rate-limiting' enzymes (Groen et al., 1982). Eventually, the mathematics of MCA was generalised in 1988 by Reder (1988), effectively marrying stoichiometric analysis (Cornish-Bowden and Hofmeyr, 2002) and metabolic control analysis. After this breakthrough, extensions followed that dealt with modular, regulatory and metabolic networks (Bruggeman et al., 2002; Schuster et al., 1993; Kahn and Westerhoff, 1991; Hofmeyr and Westerhoff, 2001). Many of these extensions of MCA, which go beyond metabolism and involve signalling cascades and gene regulation, can be found in a book written by Schuster & Heinrich (Heinrich and Schuster, 2012). A book with a more biochemical and experimental focus is 'Understanding the control of metabolism' by Fell (1997). Both outline quantitative approaches to understand metabolism.

The aim of this work is to consider metabolic control analysis in the contemporary context of whole-cell models that describe the metabolism and self-replication of cells (Fig. 1). The order of this paper is as follows. We start from a definition of evolutionary fitness in terms of cellular growth rate. We present a stoichiometric description of steady-state metabolism in the context of a cell that grows in a state of balanced growth and synthesises both growth-associated and growth-unassociated proteins. This leads to a definition of the cellular growth rate as the cellular protein synthesis rate per unit cellular protein. We consider the maximisation of this quantity and show that a convex combination of elementary flux modes (EFMs) characterises the optimal state. After having explained an example of an EFM, we discuss experimental chemostat data to illustrate that only a few elementary flux modes (EFMs) are used. We proceed by showing how all the optimal concentrations of metabolites and proteins may be calculated in such an elementary mode. This analysis leads to the conclusion that a control coefficient of a metabolic enzyme on metabolic flux is proportional (but not equal) to its protein fraction. We also derive the control coefficient of a metabolic enzyme on cellular growth rate. We discuss how these results are affected by the expression of growth-unassociated proteins. After illustrating with experimental evidence that metabolic enzymes are often expressed optimally, we use proteomics data to provide a whole-cell perspective on the control of growth rate by metabolic enzymes.

[1] Metabolic control analysis was less approximate and more rooted in experimental data than biochemical system theory. Since biochemical systems theory exploited more approximate mechanistic mathematical models, it could focus on general principles. By focusing on derivatives, metabolic control analysis neither needed nor chose to make such simplifying assumptions, which made it therefore also more descriptive than explanatory and predictive. Thus, metabolic control analysis addressed the proximate, mechanistic cause of values of flux control coefficients, in terms of the values of elasticity coefficients, while not addressing the ultimate (teleological) cause: how it favours the organism. Biochemical systems theory always tried to marry both views.

[2] For a pioneering, seminal review on mathematical modelling of biochemical systems see Heinrich & Rapoport (Heinrich et al., 1978).

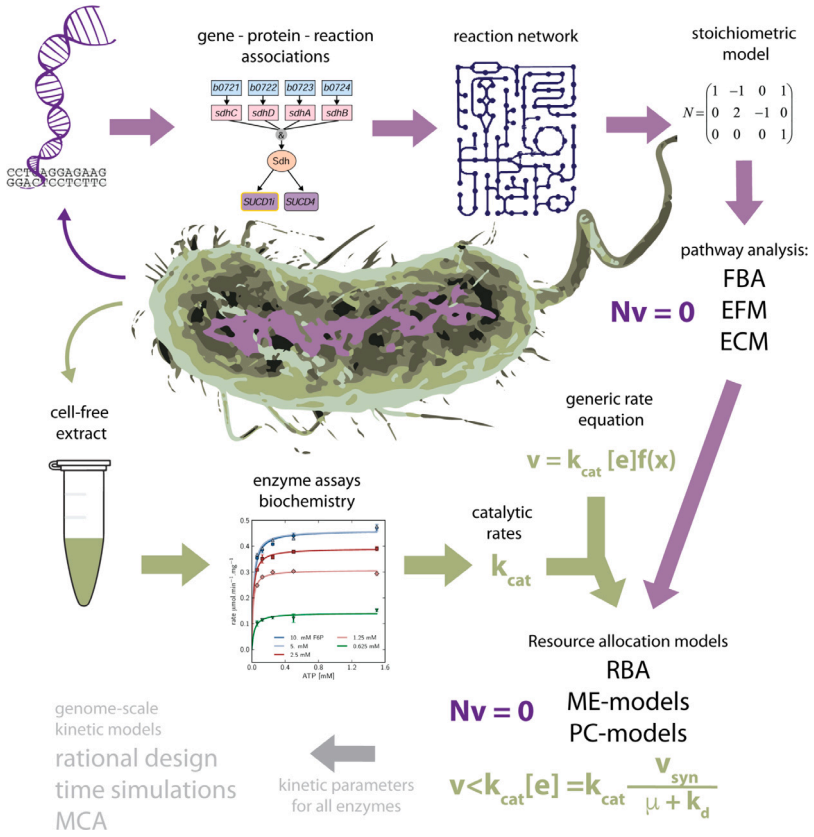



**Fig. 1. Whole-cell, resource-allocation based modelling of microbial physiology.** Contemporary whole-cell, resource-allocation based models of microbial physiology rely on genome-scale metabolic networks and physical biochemistry of microbial cells (O'Brien et al., 2013). This requires, amongst other information, the inference of all the metabolic reactions encoded on the genome of a focal microbe and a basic characterisation of the biochemistry of its enzymes. For an explicit example of such a model and its potential for predicting physiology, see Elsemman et al. (2022).

## 2. Long-term fitness and the instantaneous growth rate

Since we aim to consider the metabolic control analysis of a microbial cell that grows at an evolutionarily-maximal rate – one that has evolved from the sequential fixation of ever-faster-growing ancestral genotypes by natural selection – we require an understanding of evolutionary fitness and its relation to the growth rate.

We define the fitness $F(T)$ of a genotype (at time $t = T$, since time $t = 0$) in accordance with evolutionary theory (Orr, 2009). Let $N(t)$ be the numerical abundance of the genotype at time $t$ and $\mu(t)$ its time-dependent, instantaneous (specific or per-capita) growth rate at time $t$. Then $dN(t)/dt = \mu(t)N(t)$, with solution: $N(T) = N(0)e^{\int_0^T \mu(t)dt}$, so that the fitness $F(T)$ is given by (Bruggeman et al., 2020):

$$F(T) = \frac{1}{T}\ln\frac{N(T)}{N(0)} = \frac{1}{T}\int_0^T \mu(t)dt. \quad (1)$$

Thus, the long-term fitness over a period $T$ equals the time-averaged growth rate during that period (Bruggeman et al., 2020). The 'instantaneous fitness' is taken to be the instantaneous (or specific) growth rate $\mu(t)$. We note that growth rate should be understood here as the difference between the birth and the death rate.

In constant conditions, therefore, maximal long-term fitness is attained by a constant maximal value of the instantaneous growth rate, so that $F(T)$ is then equal to $\mu$.

We note that the long-term fitness of a genotype is not necessarily maximised if it always aims to grow maximally fast, since such behaviour may reduce its adaptive potential in varying and sometimes stressful environments; in fact, this would reduce the long-term fitness (average growth rate). We will see below that expression of proteins that are currently not contributing to growth, but may do so in the future, reduces the instantaneous growth rate. Expression of such growth-unassociated proteins is a common behaviour of, for instance, *E. coli* (Mori et al., 2021; Hui et al., 2015) and is more pronounced at slow growth than at fast growth (O'Brien et al., 2013). Those proteins can benefit the long-term fitness when they increase adaptive and survival potential. Thus, the maximisation of the instantaneous growth rate at each moment in time may not be a fit strategy for a microbe that needs to adapt quickly and prepare for future conditions. Despite that, the metabolic behaviour of *E. coli* and *S. cerevisiae* in constant conditions can be predicted from models that maximise the instantaneous growth rate as long as we assume that some (a priori set) proportion of the protein content of cell is not associated with growth (O'Brien et al., 2013; Elsemman et al., 2022; Hui et al., 2015; Mori et al., 2021; Scott et al., 2010).

The instantaneous or specific growth rate depends on the concentration and kinetics of the expressed proteins, the concentrations of growth-rate-influencing chemical compounds in the microbe's environment, and the prevailing physicochemical conditions (Bruggeman et al., 2020), as we shall see below. A cell therefore requires a regulatory strategy to express the proteins it currently needs and 'expects' to need to meet future challenges. This regulatory strategy has evolved over time and is a reflection of its evolutionary history, during which the competition between genetic variants caused the increase in frequency of those that were fittest (Orr, 2009).

We will limit ourselves to the evolutionary maximisation of the instantaneous growth rate and allow for the synthesis of proteins for future, adaptive purposes. We note that the hypothesis that microorganisms maximise their growth rate by optimal expression of catabolic and anabolic proteins (given a protein fraction that is growth-unassociated), is used in contemporary microbial systems biology to predict metabolic fluxes and enzyme concentrations (O'Brien et al., 2013; Elsemman et al., 2022). We stress that this hypothesis may not hold for all

microorganisms. It appears to hold for a class of microorganisms, including *E. coli* and *S. cerevisiae*, which we refer to growth-rate prioritising microorganisms. A more comprehensive analysis of this issue is provided in a recent perspective paper (Bruggeman et al., 2023).

In order to understand the influence of protein concentrations on the instantaneous growth rate (which is required for MCA) we will first need to express the growth rate in terms of the biochemical activities occurring inside cells.

## 3. Instantaneous growth rate, balanced growth and steady-state metabolism

We will limit ourselves to the derivation of the instantaneous growth rate under conditions of balanced growth (Campbell, 1957), in which it is constant. This state of growth is a property of a population of cells and is attainable in the lab using continuous cultures (such as turbostats, chemostats or (pH-)auxostats) (Kuenen, 2019), and can be approximated with batch cultures. It is characterised by a constant exponential increase of all the extensive properties of a cell culture with time (Schaechter et al., 1958). This, as we will show, implies that the physicochemical state of the average cell in the population remains constant; it has a steady state metabolism, and can therefore be reproducibly studied with quantitative methods. Balanced growth experiments are not limited to growth studies, but can of course also be done in the presence of stressful conditions, forcing the cell to express proteins associated with stress tolerance, in addition to those allocated to metabolism. Alternatively, shifts between balanced-growth states can be studied. Since the exact physiological state of balanced growth in given conditions can be faithfully replicated in different labs, this growth condition underlies most studies in quantitative microbial physiology and biotechnology (Schaechter, 2006; Egli, 2015).

In this section, we will show that a constant growth rate of a cell culture implies that: i. the metabolism of the average cell in the population is at steady-state, and ii. the instantaneous growth rate of the average cell equals its protein synthesis rate per unit cellular protein. These results will allow us to link the growth rate of the average cell to its steady-state enzyme biochemistry, which is the starting point of a metabolic control analysis of a growing cell.

Although the focus is on growth rate and the underlying metabolism, we shall also have to consider the expression of proteins that are not associated with growth, but, for instance, with signalling, stress management and adaptation to new conditions (Hui et al., 2015; Mori et al., 2021; Scott et al., 2010). It is, for instance, customary in microbial systems biology to think of the protein expression by a particular microbe in terms of protein sectors, of which only some serve a metabolic and growth function (Hui et al., 2015; Mori et al., 2021; Scott et al., 2010; Elsemman et al., 2022; O'Brien et al., 2013,?).

When the specific growth rate $\mu$ is constant in time, it can be expressed in terms of different extensive properties of the cell population (e.g., the total cell volume $V$, the number of cells $N_C$, the total copy number of a molecule $i$ $N_i$ or the total cell mass $M$), because they all increase exponentially in time with the same fixed (time-independent) rate,

$$\mu = \frac{1}{V(t)}\frac{dV(t)}{dt} = \frac{1}{M(t)}\frac{dM(t)}{dt} = \frac{1}{N_C(t)}\frac{dN_C(t)}{dt} = \frac{1}{N_i(t)}\frac{dN_i(t)}{dt}. \tag{2}$$

From these relations it directly follows that for the average cell, its volume $v = \frac{V(t)}{N_C(t)}$, molecule content of type $i$ $n_i = \frac{N_i(t)}{N_C(t)}$, and mass $m = \frac{M(t)}{N_C(t)}$ is time independent.

Since the number of molecules $n_i = N_i(t)/N_C(t)$ of type $i$ in the average cell, as well as its volume $v = V(t)/N_C(t)$, are constant, the concentration of that molecule in the average cell $n_i/v$ is constant. The average cell has, therefore, a steady state metabolism: the net rates of synthesis and degradation of all its molecules balance. We note that steady state applies to the average cell of a population and not necessarily also to all the single cells in it; in fact, experimental data suggests that along the cell cycle metabolism is dynamic (Nordholt et al., 2020; van Heerden et al., 2023; Papagiannakis et al., 2017). It turns out that the average cell as considered in population level studies is approximately halfway along the cell cycle (Bruggeman et al., 2020).

Thus, in balanced growth, the concentrations of all individual molecule types in the growing culture are independent of time. An additional way to show this is by defining the concentration $c_i = N_i(t)/V(t)$ of molecule type $i$ and considering its rate of change,

$$0 = \frac{dc_i(t)}{dt} = \frac{1}{V(t)}\frac{dN_i(t)}{dt} - \frac{1}{V(t)}\frac{dV(t)}{dt}\frac{N_i}{V(t)} \tag{3}$$

$$= \frac{1}{V(t)}\frac{dN_i(t)}{dt} - \mu c_i(t). \tag{4}$$

The first line indicates that $d\ln N_i(t)/dt = d\ln V(t)/dt = \mu$ is the necessary for condition $dc_i(t)/dt = 0$ (which agrees with what we concluded above).

The total volume and mass of the cell population may be expressed in terms of the individual molecules through

$$V(t) = \sum_i \bar{v}_i N_i(t) \text{ and } M(t) = \sum_i \bar{m}_i N_i(t), \tag{5}$$

with $\bar{v}_i$ and $\bar{m}_i$ as the molar volume and mass, respectively, of molecular compound $i$. The extensive properties of the cell culture can thus all be attributed to a rise in the number of molecules that make up cells.

The term $\frac{1}{V}\frac{dN_i}{dt}$ in Eq. (4) denotes the rate of change in the concentration of molecular species $i$ due to biochemical and biophysical reactions (i.e., synthesis, degradation or import and export from cells). The rates of these reactions obey,

$$\frac{1}{V(t)}\frac{dN_i(t)}{dt} = \sum_{k=1}^{R} \bar{s}_{ik} j_k(\boldsymbol{c}) = \bar{\boldsymbol{S}}\boldsymbol{j}(\boldsymbol{c}), \tag{6}$$

with $\bar{s}_{ik}$ as the stoichiometric coefficient of molecule $i$ in reaction $k$, $j_k(\boldsymbol{c})$ is the steady-state rate of reaction $k$, and $R$ is the total number of reactions. A stoichiometric coefficient $\bar{s}_{ik}$ is the $i,k$-th entry of the stoichiometric matrix $\bar{\boldsymbol{S}}$ and specifies the number of molecules of type $i$ consumed ($\bar{s}_{ik} < 0$) or produced ($\bar{s}_{ik} > 0$) during a single occurrence of reaction $k$.

The term $-\frac{1}{V}\frac{dV}{dt}c_i = -\mu c_i$ in Eq. (4) denotes the reduction in the concentration due to an increase in cell volume, which is often referred to as 'dilution by growth' or 'flux to expansion' (Flint et al., 1981).

The total cell volume of the cell culture $V(t)$ increases because molecules are taken up by the cell (incl. water) and new molecules are made from them that differ in volume from their substrates. Thus, combining (2), (5) and (6), we conclude that the cellular growth rate is equal to the net increase of volume (per unit volume) due to the import, export and interconversion of the different molecules by the cell (de Groot et al., 2020),

$$\mu = \frac{1}{V(t)}\frac{dV(t)}{dt} = \sum_i \bar{v}_i \sum_{k=1}^{R} \bar{s}_{ik} j_k(\boldsymbol{c}). \tag{7}$$

This relation expresses the growth rate of a culture of cells at a state of balanced growth in terms of the rates of the underlying biochemical processes.

## 4. Instantaneous growth rate equals the ribosomal protein synthesis rate per unit protein

The discussion above did not distinguish between the concentrations of the reactants of metabolism ('metabolites') and proteins. We now focus only on the stoichiometric matrix of metabolism, and denote it with $\boldsymbol{S}$, and denote the vector with metabolite concentrations by $\boldsymbol{c}$ and the vector with metabolic-enzyme concentrations by $\boldsymbol{e}$. We note that, in addition to metabolic proteins, any cell contains also other proteins, having for instance structural or stress-associated tasks.

During balanced growth, it is customary to assume for metabolites that their rate of dilution by growth is negligible relative to the rates of

metabolic reactions. We therefore assume that the steady state reaction rates or 'fluxes' $\boldsymbol{j}$ depend on steady state concentrations of metabolites $\boldsymbol{c}_s$ and metabolic-reaction-catalysing enzymes $\boldsymbol{e}$ in the following manner,

$$\boldsymbol{S}\boldsymbol{j}(\boldsymbol{c}_s;\boldsymbol{e}) \approx \boldsymbol{0} \tag{8}$$

This assumption is a common practice in metabolic modelling (but is not required, see e.g., de Groot et al., 2020). It applies to the reactants of central metabolism in a growing cell and not to its proteins. Since proteins consist of several hundreds of amino acids, say 350–400, a change in their concentration due to the synthesis of a new copy occurs on a time scale that is about 350 to 400 times slower than a change in metabolite concentrations. The steady state of protein concentrations is therefore the outcome of the balance between their synthesis rate (by the ribosome) and their dilution by (volume) growth. In other words, for proteins, dilution by growth cannot be neglected.

Let us now turn our attention to protein synthesis. For the $j$th protein, with copy number $N_{p_j}(t)$ and concentration $p_j(t)$, its rate of concentration change equals the difference between its rates of synthesis, degradation and dilution by growth,

$$\frac{dp_j}{dt} = \frac{1}{V(t)}\frac{dN_{p_j}(t)}{dt} - (k_j + \mu)p_j(t). \tag{9}$$

This equation applies to all proteins. The first term on the right-hand side equals the net synthesis rate of this protein by the fraction of ribosomes $\alpha_j$ that are allocated to its synthesis (Scott et al., 2010),

$$\frac{1}{V(t)}\frac{dN_{p_j}(t)}{dt} = j_{R,j}(\boldsymbol{c}_s;e_r) = \alpha_j\frac{\hat{k}_r}{a_j}e_r f_r(\boldsymbol{c}_s), \tag{10}$$

with $j_{R,j}(\boldsymbol{c}_s;e_r)$ as the net protein synthesis rate of protein $j$, containing the factor $\hat{k}_r/a_j$ (with $a_j$ as the amino-acid length of protein $j$ and $\hat{k}_r$ as the catalytic amino-acid-elongation rate constant of the ribosome (~20 aa/s) Wu et al., 2022; Young and Bremer, 1976), $e_r$ the concentration of ribosomes and $\alpha_j$ as the fraction of ribosomes allocated to the synthesis of protein $j$.

We obtain after summing over all proteins in a cell, and assuming that the total concentration of protein in the cell, $p_T$, is independent of time, the following relation between the cellular protein synthesis rate $j_R(\boldsymbol{c}_s;e_r)$ and the protein dilution into new cellular volume (for simplicity we assume $k_j = 0$, so we assume 'stable' proteins, as most catabolic and metabolic proteins are),

$$j_R(\boldsymbol{c}_s;e_r) = k_r e_r f_r(\boldsymbol{c}_s) = \mu p_T, \tag{11}$$

containing $k_r = \sum_j \alpha_j \hat{k}_r/a_j$ as the average catalytic rate constant of protein synthesis. We assumed that all ribosomes are actively translating proteins. We exclude the hibernating ribosome fraction (Wu et al., 2022; Scott et al., 2010) (we do not require this assumption, but make it here for simplicity). We note that a ribosome is partially composed out of proteins itself and the concentrations of its proteins are included in $p_T$.

The total protein concentration of the cell $p_T$ is the sum of the total growth-associated protein concentration $e_T$ and the total concentration of growth-unassociated proteins $p_{NG}$,

$$p_T = e_T + p_{NG}. \tag{12}$$

Growth-associated proteins are all the enzymes involved in catabolism and anabolism, so carrying fluxes in the metabolic and biosynthetic network of the cell. The reactions they catalyse feature as entries in the stoichiometric matrix of a genome-scale stoichiometric model of a whole cell. Growth-unassociated proteins are metabolically inactive and have, for instance, structural functions. Examples of the latter type of protein include cell-septum proteins (FtsZ), structural proteins (MreB, flagellum), and DNA-binding proteins (H-NS, transcription factors) and proteins that are currently not needed for growth but, potentially, for future conditions (O'Brien et al., 2016).

From (11) we conclude that the specific growth rate of a population at balanced growth equals the protein synthesis rate divided by the cellular protein concentration,

$$\mu = \frac{j_R(\boldsymbol{c}_s;e_r)}{p_T}, \tag{13}$$

which, using (12), shows that the growth rate decreases with the protein fraction allocated to growth-unassociated processes,

$$\mu = \frac{j_R(\boldsymbol{c}_s;e_r)}{p_T}\underbrace{\frac{p_T - p_{NG}}{e_T}}_{=1} = \frac{j_R(\boldsymbol{c}_s;e_r)}{e_T}\left(1 - \frac{p_{NG}}{p_T}\right). \tag{14}$$

This equation is valid during balanced growth (see also Scott et al., 2010). It plays an important role in what follows next when we maximise the specific growth rate, and study the characteristics of the resulting optimal metabolic network and the control coefficients of its growth-associated proteins, at optimal, growth-rate maximising concentrations. It indicates that maximisation of the growth rate at a fixed growth-unassociated protein fraction amounts to maximising $j_R/e_T$, by optimally allocating biosynthetic resources over catabolic and anabolic reactions (Wortel et al., 2014).

From Eqs. (13) and (14) we can deduce several insightful relations between the instantaneous balanced growth rate and the concentrations of protein pools (sectors) in a cell — relations associated with the analysis of metabolic control. First, enhancing the concentration of all proteins equally in a cell, which as we will see in a later section, speeds up all the rates of reactions catalysed by enzymes (Eq. (22)), does not necessarily increase the growth rate, as both the numerator and denominator increase. Second, when we add more proteins to the cell that do not contribute to metabolism or protein synthesis, the denominator of (13) increases without an increase in the numerator, indicating a reduction in growth rate. Thus, the response of the growth rate to an increase of the protein concentration in the cell depends on the cellular function of that protein. Third, (14) indicates that at a constant total cellular protein concentration $p_T$, which is for instance a property of *E. coli* across conditions (deviation of less than ~ 15% Oldewurtel et al., 2021), growth rate is maximised by maximising the protein synthesis rate of a cell. This is to be achieved by gene regulation (Planqué et al., 2018) that optimises the concentrations of metabolic enzymes, including ribosomes, an illustration of 'optimal allocation of finite biosynthetic resources' (Molenaar et al., 2009).

## 5. Steady-state metabolism, the flux cone, and elementary flux modes

At steady-state metabolism, the steady-state flux vector $\boldsymbol{j} = \boldsymbol{j}(\boldsymbol{c}_s;\boldsymbol{e})$ of all the metabolic reactions (growth-associated processes) obeys

$$\boldsymbol{S}\boldsymbol{j} = \boldsymbol{0}. \tag{15}$$

This equation indicates that $\boldsymbol{j}$ lies in the right nullspace of $\boldsymbol{S}$. A basis for this nullspace are the columns of a so-called kernel matrix $\boldsymbol{K}$ (Reder, 1988; Heinrich and Schuster, 2012).

For a given $\boldsymbol{S}$, there exist many kernel matrices. Each column of a kernel matrix represents a set of flux values that together describe a steady-state, mass-flow route through a metabolic network that has the properties that each metabolite in it is synthesised and consumed at an equal rate. The non-uniqueness of the kernel matrix unfortunately leads to an unclear definition of metabolic pathways.

To overcome this problem, we set out the define a set of unique steady-state flux vectors for a given stoichiometric matrix $\boldsymbol{S}$, so that all steady state flux vectors can be constructed from these. We split all the reversible reactions – those that would be described with reversible kinetics in a kinetic model – into an irreversible forward and backward rate such that all reaction rates are again positive. Thus, we introduce

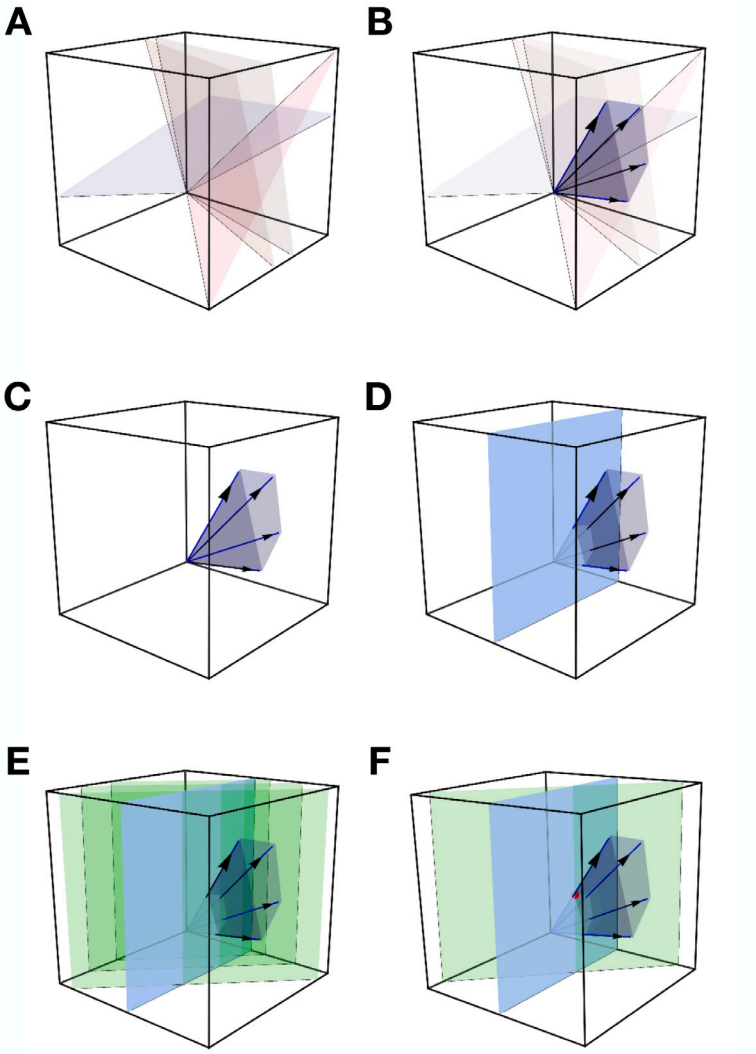



**Fig. 2. The flux cone, a flux-equality constraint and the optimisation of a fluxobjective with flux balance analysis. A.** A 3-D flux space is shown, each axis corresponds to a different flux. All fluxes are positive — reversible reactions have been split, as described in the main text. The planes correspond to the steady-state condition of a single metabolite concentration, which leads to a linear combination of fluxes. **B.** At the intersection lines of the planes, several metabolite concentrations are at steady state simultaneously, giving rise to steady-state subnetworks. These intersection lines together form the edges of the flux cone (shown in **C.**). The edges are the EFMs. The planes between two adjacent EFMs are called facets, which together bound the flux cone. The flux cone corresponds to the feasible steady-state solution space, corresponding to the steady-state and flux-(ir)reversibility constraints. **D.** A new plane is added in blue, which corresponds to a flux-equality constraint. The segment that lies in the flux cone is the feasible steady-state solution space, forming the polytope. **E.** An objective plane is shown, corresponding to a linear combination of flux values, for three different values of the objective (different shades of green). The aim is to maximise the objective value, while meeting all the flux constraints (steady state, (ir)reversibility and (in)equality constraints). This implies that the maximum objective is attained in a corner point of the polytope (red point in **F**), which is an EFM. This is the essence of flux balance analysis.

a new augmented matrix $\boldsymbol{S}'$ and extend the vector of reaction fluxes to obtain $\boldsymbol{j}'$, so that Eq. (15) is replaced by

$$\boldsymbol{S}'\boldsymbol{j}' = \boldsymbol{0}, \quad \boldsymbol{j}' \geq \boldsymbol{0}. \tag{16}$$

All one needs to do is to add to $\boldsymbol{S}$ new columns that are the reversible reaction columns of $\boldsymbol{S}$ with a minus sign and add the new irreversible fluxes to $\boldsymbol{j}$.

The set of flux vectors that satisfies Eqs. (16) form the so-called flux cone $C$ (Fig. 2A–C),

$$C = \left\{ \boldsymbol{j}' \mid \boldsymbol{S}'\boldsymbol{j}' = \boldsymbol{0}, \boldsymbol{j}' \geq \boldsymbol{0} \right\}. \tag{17}$$

The vectors that "span" this flux cone are called elementary flux modes (EFMs, or more generally extreme rays) (Gagneur and Klamt, 2004; Schuster et al., 2000; Schuster and Hilgetag, 1994; Papin et al., 2004). (Here "span" does not refer to the usual linear algebra notion of taking arbitrary linear combinations; instead, we need to restrict to 'conical linear combinations', involving only positive coefficients.) Contrary to the basis of the null space of $\boldsymbol{S}$, the elementary flux modes are unique (up to a scalar multiple, just as, e.g., eigenvectors are; we take the convention that the last element is always set to 1). Together they give rise to an unambiguous set of steady-state metabolic pathways corresponding to a stoichiometric matrix – e.g., the entire set 'encoded' by an entire genome (Schuster et al., 2000) or feasible in one condition (Kelk et al., 2012).

Thus, the set of elementary flux modes represents a unique set of feasible, steady-state flux vectors. From them all the steady-state flux vectors of $\boldsymbol{S}'$ can be constructed as a conical sum of flux vectors of elementary flux modes $\boldsymbol{e}_i$,

$$\boldsymbol{j}' = \sum_{i=1}^{E} \lambda_i \boldsymbol{e}_i, \ \forall i : \lambda_i \geq 0, \tag{18}$$

where $E$ is the number of elementary flux modes. In Fig. 2A–C, we show an example of an idealised cone.

The entire metabolic network encoded by the genome of a cell has an astronomical number of elementary flux modes, which prevents their complete enumeration at genome scale (Gagneur and Klamt,

2004). Although the set of elementary flux modes that are feasible under particular experimental conditions is of course considerably less, it is still a huge number (Kelk et al., 2012). Below we will see, however, that experimental data and theoretical expectations indicate that cells likely use only a handful of EFMs. On top of that, we will also show that evolutionary maximisation of the instantaneous growth rate implies that cells use a minimal number of EFMs in their evolutionarily optimal state. Finally, we will show that the control coefficient of proteins, occurring in this minimal set of EFMs, on the growth rate and metabolic flux can be related to their protein-concentration fraction and growth-unassociated protein fraction.

## 6. Properties and examples of elementary flux modes

An example of the pieces of a single elementary flux mode are shown in Fig. 3. Elementary flux modes have a number of useful properties for the analysis of whole-cell metabolism (Gagneur and Klamt, 2004; Schuster et al., 2000; Schuster and Hilgetag, 1994; Papin et al., 2004; De Groot et al., 2019; Müller et al., 2014; Wortel et al., 2014),

1. They are *minimal*: none of their reactions can be removed without violating the steady-state requirement. Another way of saying this is that they have minimal support (EFM vectors have a maximal number of zero entries),
2. They are *nondecomposable*: no elementary flux mode contains another elementary flux mode,
3. They are *elementary*: all steady-state flux distributions can be retrieved from conical combinations of elementary flux modes,
4. EFMs are *maximal yield solutions*. The metabolic pathway with a maximum value of a ratio of two flux values (generally called a yield) is always an elementary mode,
5. An EFM has one reaction more than the rank of its associated stoichiometric matrix (so *1 independent flux value*, from which all others can be inferred),
6. An EFM is a *specific flux maximiser* of a kinetic model. Thus, the maximal value of a metabolic flux per unit total protein needed to sustain that flux, achieved by optimising the enzyme concentrations under a single total concentration constraint, is achieved by an EFM in a kinetic model of metabolism. The maximal number of EFMs that together maximise any specific flux in a kinetic model with $C$ protein concentration constraints is equal to or smaller than $C$, and
7. A single EFM is a solution of a flux balance analysis computation with one flux equality constraint (i.e., one flux set to a specific value) satisfied in the optimal state. An FBA model with $C$ flux equality constraints satisfied in the optimal state has as an optimal solution a flux vector that is a conical combination of maximally $C$ EFMs.

These amazing properties of EFMs make them ideally suited to study the metabolism of cells, regardless of whether they are modelled with kinetic or stoichiometric models.

Next, we will discuss the role of EFMs in genome-scale stoichiometric models before we shift to the kinetic models of metabolism and a characterisation of the optimal state of cells in terms of metabolic control analysis.

## 7. Elementary flux modes are solutions of yield maximisations with flux balance analysis

An example of a flux balance analysis (Varma and Palsson, 1993a,b; Varma et al., 1993) computation is the following linear program with optimal solution $\boldsymbol{j}^o$ (assuming a medium with glucose as the sole carbon source),

$$\boldsymbol{j}^o = \arg\max_{\boldsymbol{j}} \left\{ j_{biomass} \mid \boldsymbol{j} \in C \text{ and } j_{glc} = 1 \right\}. \tag{19}$$

The flux constraints derive from an intersection of the flux cone (specifying the steady state and flux positivity constraints; Eq. (16)) with the hyperplane that results from the flux-equality constraint that the glucose uptake rate $j_{glc}$ equals 1 (Fig. 2D). The maximisation objective is the flux through the 'biomass reaction', which generally specifies the macromolecular composition of cells, i.e., the number of moles of DNA, RNA, proteins, lipid and ATP (growth-associated maintenance requirement) needed per gram biomass as an elementally-balanced reaction (Orth et al., 2010). These macromolecules need to be made from the chemical building blocks by the anabolic reactions of the cells, given the Gibbs free energy supplied by the catabolic reactions of the cell. Catabolism and anabolism both 'feed' on extracellular nutrients.

When we intersect the flux cone with one flux-equality constraint (Fig. 2D), the space of feasible solutions shrinks from the entire cone to a set for which $j_{glc}$ is always 1. The resulting object is called a polytope, spanned by the EFMs that can accommodate $j_{glc} = 1$ (all the others are now no longer feasible given the constraints).

The objective $j_{biomass}$ corresponds to another hyperplane, which we move through the polytope in the direction of increasing $j_{biomass}$ until we leave the polytope (after which we no longer meet the constraints) (Fig. 2E–F). The maximiser is then generally attained in a cornerpoint (vertex) of the polytope, due to Bauer's maximum principle, and is therefore an EFM. This effectively proves that the maximal yield solution is always attained by an EFM, as maximising the rate of biomass formation given the glucose uptake rate maximises their ratio, which equals the maximal biomass yield on glucose. Note that the same number for the yield is obtained for any value of the glucose uptake rate, as the EFM describing the solution has a constant (maximal) yield (and the cone is otherwise unbounded). Thus in this usual FBA setting, FBA predicts maximal yield solutions and not maximal growth rate solutions.

When an additional constraint is added to the previous linear program and would be hit in addition to $j_{glc} = 1$, e.g., $j_{atpmaintenance} = 4$ as in the next example, then $j_{biomass}$ would generally be lower than the current value (e.g., because glucose is also needed to satisfy the growth rate-independent (ATP) maintenance constraint) and attained in a conical combination of two EFMs as we shall see next. In this case, less biomass is made per unit glucose and, accordingly, the biomass yield on glucose is lower.

Now we focus on a linear program (20) with one additional flux-equality constraint, associated with the growth rate-independent maintenance requirement (Pirt, 1982), to arrive at a minimal, realistic description of a genome-scale model. For instance, we might consider aerobic respiratory growth of *E. coli* or *S. cerevisiae* in a chemostat below the critical growth rate (when overflow metabolism is absent). Consider

$$\boldsymbol{j}^o = \arg\max_{\boldsymbol{j}} \left\{ j_{biomass} \mid \boldsymbol{j} \in C \text{ and } j_{glc} = 1,\ j_{atpmaintenance} = 4 \right\}. \tag{20}$$

The cone is now the feasible solution space, spanned by EFMs, and is intersected with two (orthogonal) hyperplanes, each associated with a single flux-equality constraint. The optimal solution $\boldsymbol{j}^o$ is in this case generally a conical combination of two EFMs, one giving rise to the maximal ATP yield on glucose to satisfy the maintenance requirement (and generally a zero growth rate) and the other giving rise to the maximal biomass yield on glucose. The net biomass yield on glucose, given by the conical combination of those two EFMs, is now lower than its maximum. A useful insight is therefore that, given the active stoichiometric matrix associated with the optimal outcome of (20), the optimal solution can also be obtained by doing two independent FBAs, each giving rise to one EFM and associated with a different equality constraint (one for ATP maintenance and the other for growth). These together make up the optimal solution of (20).

Finally, we consider this example

$$\boldsymbol{j}^o = \arg\max_{\boldsymbol{j}} \left\{ j_{biomass} \mid \boldsymbol{j} \in C \text{ and } j_{glc} = 1,\ j_{atpmaintenance} \geq 4 \right\}. \tag{21}$$

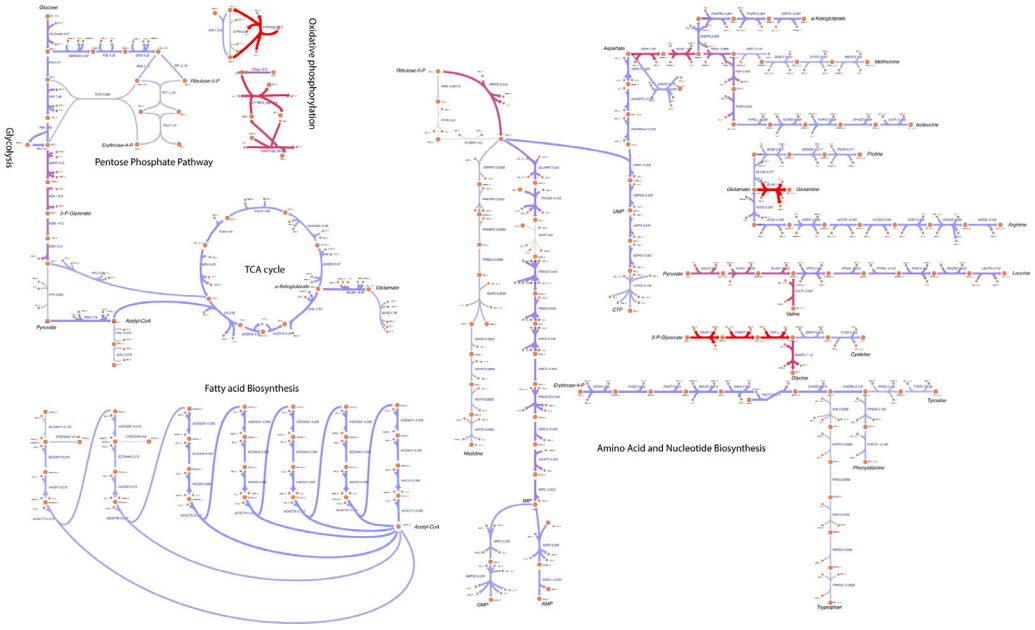



**Fig. 3. Illustration of the complexity of a single biomass-producing EFM, computed with a genome-scale model iJO1366 of *E. coli*** (Orth et al., 2011). A flux balance analysis calculation with two inequality constraints (glucose uptake rate and growth-independent ATP maintenance rate) and with biomass production as optimisation target yielded an optimal flux vector. The EFM associated with the maximal biomass yield was subsequently calculated and some of its metabolic networks are shown here. This figure illustrates the complexity of single EFMs: they can accommodate all the reactions needed to make all cellular macromolecules and byproducts from a supplied (mineral) growth medium. We choose to draw the metabolic network in modules, but this does not mean that the network is disconnected (the same intermediates occur in different modules and are explicitly mentioned, in addition to end products). The shown metabolic networks maps were made with 'Escher' (King et al., 2015).

In this case, we do not know whether $j_{atpmaintenance} = 4$ in the optimal solution (as this flux can exceed 4); the optimal solution might therefore be a single EFM or a combination of two. This may (generally) be determined by observing whether $j_{atpmaintenance}$ equals 4 or exceeds 4 in the optimal solution, i.e., whether the flux equality is satisfied or not. Since most constraints are given as inequalities in FBA, this usage is very common.

Thinking in terms of FBA calculations and conical combination of EFMs – with their intuitive properties – leads to many insights that are not readily apparent when one merely performs FBAs without ever thinking about the associated mathematical theory. For us, this has been particularly instrumental in understanding how to think about growth and metabolic strategies of the cell (de Groot et al., 2020; De Groot et al., 2019; Bruggeman et al., 2020). For instance, when 1 EFM is active, the number of active fluxes in the EFM (with nonzero values) is exactly one more than the rank of the stoichiometry matrix of this EFM (obtained by removing all the columns from the stoichiometry matrix that corresponds to reactions with a zero flux in the EFM). When $C$ constraints are active then the number of nonzero fluxes equals $C$ plus the rank of the relevant stoichiometric matrix.

This basic understanding of FBA will be exploited below to show that EFMs are also the optimal solutions of kinetic models of whole-cell metabolism. After that we can study the flux control properties of metabolic enzymes on fluxes and growth rate in a whole-cell context.

## 8. Universal relation between the rate and concentration of an enzyme

For a general mathematical description of growth-rate maximisation by constrained optimisation of protein concentrations, we need to relate the rate of an enzyme-catalysed reaction to the concentrations of the associated reactants, modifiers and the catalysing enzyme: these are the optimisation variables for growth rate maximisation; for yield maximisations only fluxes are needed as variables. This is achieved by the theory of enzyme kinetics, which was given a firm foundation in the late 1960s by the seminal works of Cleland for single subunit enzymes (Cleland, 1963b,c,a) and later also for cooperative enzymes (Monod et al., 1965; Koshland Jr. et al., 1966; Hofmeyr and Cornish-Bowden, 1997).

Enzyme kinetics theories indicate that the catalytic rate $v$ of an enzyme equals a product of four factors (Heinrich et al., 1978; Cornish-Bowden, 2013a): a catalytic rate constant defined in the forward direction $k^+$, the enzyme concentration $e$, the substrate saturation function of the enzyme $f_j^+(\boldsymbol{c})$, the vector of (steady-state or non-steady-state) metabolite concentrations $\boldsymbol{c}$, and the displacement from thermodynamic equilibrium $1 - e^{\frac{\Delta\mu_j(\boldsymbol{c})}{RT}}$ (with $\Delta\mu_j(\boldsymbol{c})$ as the Gibbs free energy of the reaction),

$$v_j(\boldsymbol{c}; e_j) = k_j^+ e_j \underbrace{f_j^+(\boldsymbol{c}) \left(1 - e^{\frac{\Delta\mu_j(\boldsymbol{c})}{RT}}\right)}_{f_j(\boldsymbol{c})}. \tag{22}$$

Note the convention that a steady state enzymatic rate is denoted by $j$ and otherwise by $v$.[3]

[3] The steady-state flux $j$ applies when all the metabolite concentrations are constant, because then net rates of synthesis and consumption balance — so during balanced growth. The resulting steady-state rate of each enzyme we call a flux. A flux is a function of all the enzyme concentrations, the kinetic parameters of all the expressed enzymes (catalysing the synthesis and consumption reactions occurring in metabolism), and the concentrations of fixed extracellular nutrients and products of growth. The rate $v$ of an enzyme only depends on the kinetic properties and concentration of the associated enzyme and the concentrations of the associated reactants and modifiers. The rate of an enzyme does not require a steady state metabolic network. MCA distinguishes between fluxes and rates, because fluxes are systemic properties of steady-state metabolic networks, while rates are properties also defined for enzyme in isolation (e.g., in a test tube) and defined for all values of its reactants and modifier concentrations.

For our purposes, an important property that follows from this relation is that the rate of the enzyme $v_j$ is directly proportional to the enzyme concentration invested in it, so that for any $\lambda$,

$$v_j(\boldsymbol{c}; \lambda e_j) = \lambda v_j(\boldsymbol{c}; e_j).$$

This relationship results from the derivation of enzyme kinetics regardless of the method (e.g., quasi-steady state or rapid-equilibrium approximation) and applies to mono- as well as multimeric (allosteric or cooperative) enzymes (Cornish-Bowden, 2013a; Segel, 1975). It does not apply when different enzymes form complexes and directly 'channel' reactants to each other (Sauro and Kacser, 1990; Kholodenko and Westerhoff, 1993).

In MCA (Heinrich and Rapoport, 1974; Kacser and Burns, 1973), enzymes enter with their elasticity coefficients, which equal the derivative of the logarithm of their rate with respect to any of the concentrations that influence it (i.e., a reactant, modifier (e.g., feedback metabolite) or enzyme catalyst),

$$\epsilon_{c_k}^{v_j} = \frac{\partial \ln v_j}{\partial \ln c_k}.$$

Note that this derivative is taken with all other concentrations held constant (hence, it is a partial derivative). This implies that all changes made to concentrations that do not occur in the rate equation do not directly affect the rate and lead to a zero elasticity coefficient. Since $v_j \propto e_j$, we also conclude that $\epsilon_{e_j}^{v_j} = 1$, indicating that the response of metabolism to changes in enzymatic rates can be experimentally studied by changing the concentration of enzymes, using, for instance, titratable promoters of metabolic genes (Jensen et al., 1993c; Walsh and Koshland Jr., 1985).

## 9. Flux and growth rate scaling with cellular protein content of growth-associated and unassociated proteins

Before we consider the maximisation of the growth rate by the constrained optimisation of protein concentrations, we derive a few relationships that address the scaling of steady-state properties of metabolic networks, i.e., fluxes and concentrations, with enzyme concentrations. We require these to understand how the specific growth rate, defined in equation (13) as $j_R/p_T$, the protein synthesis rate divided by the protein concentration of a cell, scales with the total protein content allocated to growth-associated and growth-unassociated processes. These relationships also lead to the celebrated summation theorems of MCA.

The rate of an enzyme is usually proportional to its concentration (Eq. (22)). With the notion that a function $f(\boldsymbol{x})$ is called $p$-homogeneous (or $p$-order homogeneous) if, for $\lambda \in \mathbb{R}$, $f(\lambda \boldsymbol{x}) = \lambda^p f(\boldsymbol{x})$, we conclude that the rate equations of enzymes are 1-homogeneous (or first-order homogeneous) functions of enzyme concentrations. Given a steady-state metabolism, $\boldsymbol{Sj} = \boldsymbol{0}$, we conclude also that multiplication of all enzyme concentrations by the factor $\lambda$ maintains steady state at unaltered metabolite concentrations (Giersch, 1988a,b; Westerhoff and Van Dam, 1987):

$$\boldsymbol{Sj}(\boldsymbol{c}_s; \lambda \boldsymbol{e}) = \lambda \boldsymbol{Sj}(\boldsymbol{c}_s; \boldsymbol{e}) = \boldsymbol{0}. \tag{23}$$

This relation indicates that (steady-state) metabolic fluxes and steady-state metabolite concentrations are respectively 1-homogeneous functions and 0-homogeneous functions of enzyme concentrations,

$$\boldsymbol{j}(\lambda \boldsymbol{e}) = \lambda \boldsymbol{j}(\boldsymbol{e}), \text{ and } \boldsymbol{c}_s(\lambda \boldsymbol{e}) = \lambda^0 \boldsymbol{c}_s(\boldsymbol{e}) = \boldsymbol{c}_s(\boldsymbol{e}). \tag{24}$$

Euler's theorem for a $p$-homogeneous function $f : \mathbb{R}^R \to \mathbb{R}$ with variables $x_1, \ldots, x_R$ states that

$$\sum_{k=1}^{R} \frac{\partial \ln f}{\partial \ln x_k} = p.$$

When we apply this to our setting, we obtain

$$\sum_{k=1}^{R} \frac{\partial \ln j_l}{\partial \ln e_k} = 1, \quad \sum_{k=1}^{R} \frac{\partial \ln c_l}{\partial \ln e_k} = 0,$$

When we introduce the notations of the control coefficients from metabolic control analysis,[4]

$$C_k^{j_l} = \frac{e_k}{j_l}\frac{\partial j_l}{\partial e_k} = \frac{\partial \ln j_l}{\partial \ln e_k}, \quad C_k^{c_l} = \frac{e_k}{c_l}\frac{\partial c_l}{\partial e_k} = \frac{\partial \ln c_l}{\partial \ln e_k},$$

we recover, respectively, the summation theorems of flux and concentration control,

$$\sum_{k=1}^{R} C_k^{j_l} = 1, \quad \sum_{k=1}^{R} C_k^{c_l} = 0. \tag{25}$$

The implications of these two last equations relate to a known principle of metabolic control theory, which is either referred to as 'multisite modulation' (Thomas and Fell, 2000) or the 'Universal Method' (Kacser and Acerenza, 1993), with implications for biotechnology. This principle states that changes in metabolic flux can be attained by changes in enzyme concentrations without accompanying changes in metabolite concentrations as long as all enzyme concentration are scaled by the same factor, making (24) valid. How can a cell implement such a method? Consider the situation that a cell expresses a stress gene such that more RNA polymerases transcribe that gene. Let us assume that 0.1% of all polymerases are allocated to that gene, and that therefore 99.9% are left for the genes that were already being transcribed. Then about 0.1% of the mRNA pool, assuming the same life times, etc., becomes 'stress mRNA', leading to about 0.1% usage of ribosomes; the other 99.9% of ribosomes are actively transcribing the remaining protein. Thus, passive and unbiased (same kinetics, same affinity) competition for transcriptional and translation resources may lead naturally to the 'Universal Method'. Plotting proteomics data of metabolic proteins as function of growth rate indeed often shows a linear relation; behaviour in line with the Universal Method (Elsemman et al., 2022; Mori et al., 2021). (A cell might in the end be its own best biotechnologist!) Experiments with titrated over-expression of 'unneeded' proteins also indicate a linear relation between the growth rate and the unneeded protein fraction, in agreement with Eq. (14) (Scott et al., 2010).

We can exploit the scaling relationship (24) between steady-state flux and enzyme concentrations to understand how growth-unassociated proteins influence the growth rate. Consider again the whole cell and assign all the growth-associated enzymes to the enzyme pool $e_T$. These relate to the total protein content $p_T$ as

$$e_T = p_T - p_{NG}. \tag{26}$$

This means that we can write, using equation (24),

$$\boldsymbol{j}(e_T) = \boldsymbol{j}(p_T(1-\phi_{NG})) = (1-\phi_{NG})\boldsymbol{j}(p_T), \tag{27}$$

with $\phi_{NG} = \frac{p_{NG}}{p_T}$ as the non-growth associated protein fraction. Eq. (14) is thus seen to be valid not only for the protein synthesis flux, but for all fluxes. We may also write equation (14) as

$$\mu(\phi_{NG}) = \mu(0)(1-\phi_{NG}) \tag{28}$$

where $\mu(0)$ is the (theoretical) maximal growth rate at which all proteins are associated to growth. Maximisation of growth rate therefore implies that $\phi_{NG} \to 0$. The cell, however, cannot achieve this, because some of those growth-unassociated proteins are essential (such as proteins for cell structure, gene control, etc.) and some are needed to allow adaptation to new conditions, such as stress responses (Hui et al., 2015; Mori et al., 2021; O'Brien et al., 2013) or the rapid uptake of alternative nutrients (Ihssen and Egli, 2005).

[4] Since $\epsilon_{e_k}^{v_k} = 1$, we also deduce that $C_k^{j_l} = \frac{\partial \ln j_l}{\partial \ln e_k} = \frac{\partial \ln j_l}{\partial \ln v_k}$ and $C_k^{c_l} = \frac{\partial \ln c_l}{\partial \ln e_k} = \frac{\partial \ln c_l}{\partial \ln v_k}$. Thus, control coefficients can be understood as steady-state, systemic responses to terms of enzyme-rate or enzyme-concentration perturbations.

## 10. Maximisation of instantaneous growth rate in whole-cell kinetic models is also achieved by elementary flux modes

EFMs do not only have a relevance for purely stoichiometric models, they also play a role in the constrained optimisation of the steady-state fluxes of kinetic models (Müller et al., 2014; Wortel et al., 2014). This is somewhat surprising, because EFMs are purely stoichiometric objects, defined solely on the basis of the stoichiometric matrix. They are therefore independent of enzyme kinetics. Moreover, optimisation of kinetic models does not concern fluxes as optimisation variables, but rather the concentrations of reactants, modifiers and enzymes (from which fluxes are then calculated). Still, EFMs surface as the maximisers of the flux of kinetic models, given a constraint on the maximal concentrations of proteins in a cell and the kinetic parameters of all its enzymes. We will next illustrate that the maximisation of the specific growth rate (Eq. (13)) by the constrained optimisation of protein concentrations is achieved by an elementary flux mode, when we take the kinetics of catabolic and anabolic reactions in a cell into account, in addition to their stoichiometry (Müller et al., 2014; Wortel et al., 2014).

We consider the steady state of a kinetic model of a whole cell (Fig. 1). The dependence of reaction rates on the concentrations of reactants, effectors and catalysts is given by the universal enzyme kinetic Eq. (22). Our aim is to maximise the specific growth rate of the cell at a fixed growth-unassociated protein fraction. Eq. (14) indicates that this amounts to the maximisation of the protein synthesis flux $j_R$ by optimising the concentration of growth-associated proteins under the constraint that they sum to $e_T$, in steady state (cf. equation (15)). The optimisation variables in this problem are the concentrations of all metabolites and the enzymes. This is a nonlinear optimisation problem because the metabolite concentrations enter the rate equations in a nonlinear manner. For a simple example we refer to Wortel et al. (2014).

The maximisation problem of the specific growth rate of a whole-cell kinetic model, assuming a fixed growth-unassociated protein fraction $p_{NG}/p_T$, by optimal allocation of biosynthetic resources for growth-associated proteins, and under the constraints that: i. metabolism is at steady state and ii. the enzyme concentrations sum to a constant ($e_T$), has the following mathematical description,[5]

$$\max_{\boldsymbol{c},\boldsymbol{e}} \left\{ \frac{j_R(\boldsymbol{c},\boldsymbol{e})}{e_T} \;\middle|\; \boldsymbol{S}\boldsymbol{j}(\boldsymbol{c},\boldsymbol{e};\boldsymbol{c}_0)=\boldsymbol{0},\ \forall k:\ j_k(\boldsymbol{c},e_k;\boldsymbol{c}_0)=k_k e_k f_k(\boldsymbol{c};\boldsymbol{c}_0),\ e_T=\boldsymbol{1}^T\boldsymbol{e},\boldsymbol{c}\geq\boldsymbol{0},\boldsymbol{e}\geq\boldsymbol{0} \right\}. \tag{29}$$

where the vector $\boldsymbol{c}_0$ denotes the fixed environmental nutrient and product concentrations. These are included explicitly in the description, because the environmental conditions influence the optimal state.

To show that the solution of this optimisation problem is indeed an elementary flux mode, we first recast the optimisation problem. Eqs. (24) indicate that the optimal metabolite concentrations are independent of the total concentration of growth-associated enzymes ($e_T$) and that the optimal flux values scale in proportion to the total concentration of growth-associated enzyme. This implies that the previous optimisation problem can be rewritten as (for a step-by-step exposition see Wortel et al., 2014):

$$\min_{\boldsymbol{c},\boldsymbol{e}} \left\{ e_T \;\middle|\; \boldsymbol{S}\boldsymbol{j}(\boldsymbol{c},\boldsymbol{e};\boldsymbol{c}_0)=\boldsymbol{0},\forall k:\ j_k(\boldsymbol{c};\boldsymbol{c}_0)=k_k e_k f_k(\boldsymbol{c};\boldsymbol{c}_0),\ e_T=\boldsymbol{1}^T\boldsymbol{e},\boldsymbol{c}\geq\boldsymbol{0},\boldsymbol{e}\geq\boldsymbol{0},j_R=1 \right\}. \tag{30}$$

Thus, when we know the actual amount of available protein for growth, say $e'_T$, we can calculate the optimal protein concentrations from the optimal solution $\boldsymbol{e}^o$ found by solving (30), by scaling them with a factor $e'_T/e^o_T$. (The specific growth rate is in both cases the same, by (24)).

We can simplify the mathematical description further by substituting for each enzyme concentration $e_k = j_k/(k_k f_k(\boldsymbol{c};\boldsymbol{c}_0))$, using Eq. (22),

$$\min_{\boldsymbol{c},\boldsymbol{j}} \left\{ \sum_k \frac{j_k}{k_k f_k(\boldsymbol{c};\boldsymbol{c}_0)} \;\middle|\; \boldsymbol{S}\boldsymbol{j}=\boldsymbol{0},\boldsymbol{c}\geq\boldsymbol{0},\forall k:\ \frac{j_k}{k_k f_k(\boldsymbol{c})}\geq 0,\ j_R=1 \right\}. \tag{31}$$

We proceed by splitting all the reversible reactions into their irreversible parts (as we did above). We consider the resulting optimisation problem for any fixed, steady-state metabolite concentration vector $\boldsymbol{c}^*$ satisfying $f_k(\boldsymbol{c}^*;\boldsymbol{c}_0)\geq 0$ for all $k$,[6]

$$\min_{\boldsymbol{j}'} \left\{ \sum_k \frac{j'_k}{k'_k f'_k(\boldsymbol{c}^*;\boldsymbol{c}_0)} \;\middle|\; \boldsymbol{S}'\boldsymbol{j}'=\boldsymbol{0},\forall k:\ \boldsymbol{j}'\geq\boldsymbol{0},\ j'_R=1 \right\}. \tag{32}$$

This problem is *linear* in the remaining optimisation variables, the irreversible fluxes $\boldsymbol{j}'$, which are all positive. The set of feasible solutions is therefore a flux cone intersected with a single hyperplane $j_R = 1$ and the objective function is a linear function of the fluxes, since $\forall k:\ \frac{1}{k_k f_k(\boldsymbol{c}^*;\boldsymbol{c}_0)} = d_k$ is constant. Therefore, according to the theory above (Fig. 2), the optimal solution has to be a (properly chosen multiple of an) elementary flux mode for any given $\boldsymbol{c}^*$. Therefore, for all choices of $\boldsymbol{c}^*$, including the optimal vector $\boldsymbol{c}^o$ that minimises the objective, the optimiser will be an EFM (Müller et al., 2014; Wortel et al., 2014).

Thus, the growth rate of a cell is maximised by an EFM in the case that a single total protein-expression constraint limits growth, given the kinetics of all the growth-associated proteins encoded on its genome (Müller et al., 2014; Wortel et al., 2014) and assuming a fixed protein fraction of growth-unassociated proteins.

When several linear protein-concentration constraints are considered, it can be shown that the optimal solution can consist of a conical combination of elementary flux modes (De Groot et al., 2019). The number of EFMs appearing in the optimal conical combination is then always smaller than or equal to the number of protein expression constraints that are satisfied (as equalities) in the optimal state.

This theory can be used to rationalise the shift in metabolism from pure respiratory growth to mixed respiratory and overflow-metabolism growth above a critical growth rate, as observed for many microbial species (De Groot et al., 2019). That protein expression constraints play a role at this shift has been experimentally illustrated (Basan et al., 2015), in which the overexpression of an unneeded protein in the cytosol was shown to cause a lower critical and maximal growth rate, in agreement with theoretical expectations. For a critical analysis of the limitations of FBA for explaining the origins of overflow metabolism we refer the reader to De Groot et al. (2020).

Thus, according to the theory laid out in this section (Müller et al., 2014; Wortel et al., 2014; De Groot et al., 2019), the evolutionary outcome of fitness maximisation under constant conditions of balanced growth leads to a cell that exploits a minimal number of elementary flux modes. The exploited number of EFMs equals the number of protein concentration constraints that are 'hit' or, to put it differently, the number of cellular compartments that are completely filled with needed proteins and are, therefore, limiting the growth rate (Bruggeman et al., 2020).

[5] Note that the flux exchanges between the network of growth-associated and growth-unassociated processes are neglected here. The consumption of, for instance, ATP by growth-unassociated processes (e.g., when a signalling protein is phosphorylated) is neglected in the consideration of the growth-associated metabolic network. It is also generally believed that such fluxes are negligible in genome-scale stoichiometric models, as those costs are considered in terms of the maintenance requirements.

[6] We sum again overall reactions, which are now all irreversible. Their enzyme-kinetic rate equations have now all been split in a forward and backward rate.

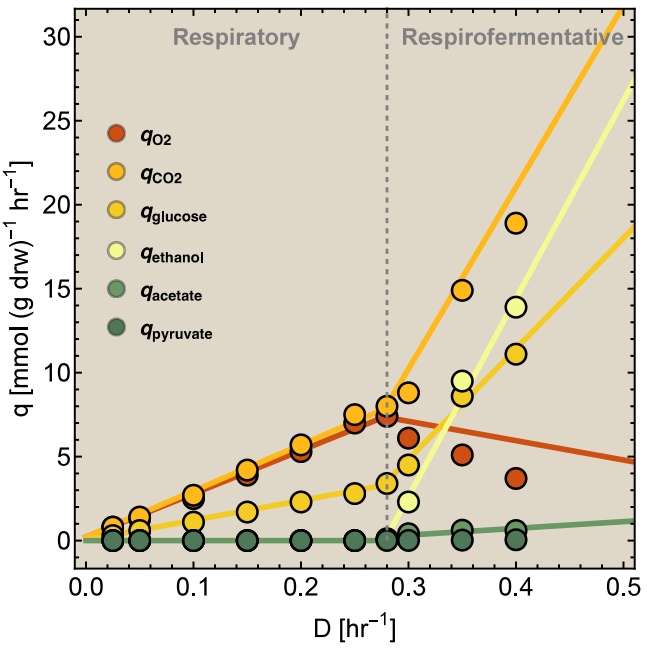



**Fig. 4. *S. cerevisiae* exploits only a few EFMs when growing in a glucose-limited chemostat.** The theoretical expectation is that any microbe minimises the number of exploited EFMs when it evolved to maximise its instantaneous growth rate under constant conditions. We do not know the evolutionary history of *S. cerevisiae*, but in this figure we show that maximally 4 EFMs are needed to fit experimental data on metabolic fluxes in a glucose-limited chemostat. This figure shows that experimental flux data of a glucose-limited chemostat (Van Hoek et al., 1998) (as dots) can be fit (the lines) by a genome-scale model of *S. cerevisiae* (Lu et al., 2019) with either 2 or 4 EFMs, before and after the critical growth rate, respectively.The non-growth associated maintenance supplied by the model, 0.7 mmolATP/(gDW h), was used. That this microbe may indeed optimise protein expression to attain a maximal growth rate is also indicated by protein expression data (Keren et al., 2016).

## 11. Experimental evidence of minimal usage of elementary flux modes in a glucose-limited chemostat

Since we do not know how many protein expression constraints limit the growth rate of a cell, one could argue that inferring elementary flux modes from experimental data is not a meaningful exercise: knowing their number does not inform us about how close cells are to an optimal state as long as we do not know the number of active protein expression constraints. This is particularly relevant when a cell is expected to exploit dozens of EFMs. Knowing just their number then does not indicate much. This changes, however, when a cell expresses only a few EFMs (say 2–5), as then it is feasible to test experimentally whether indeed as many protein expression constraints exist that limit the observed growth rate and whether metabolic enzymes are indeed expressed at optimal concentrations. The experiment to be carried out is, for instance, the expression of an unneeded protein (i.e., a protein the overexpression of which has no direct effect on the specific growth for instance because it is catalytically inactive) in a cellular compartment that is believed to limit growth because it is full with needed proteins. Such an experiment has been carried out by Basan et al. (2015). The protein constraints considered, for instance, derive from the limited protein-solvation capacity of all the protein-containing cellular compartments such as the cytoplasm, periplasm, membranes, and organelles (Bruggeman et al., 2020). In particular for prokaryotes, their number is very small (<4). We therefore also expect that the number of EFMs used by microbes is correspondingly small.

In Fig. 4, the specific uptake rates of key nutrients and products of *S. cerevisiae* grown in a glucose-limited chemostat (Van Hoek et al., 1998) are shown. Two growth rate regions are distinguished, each characterised by a linear relation between the uptake and production rates and the growth rate: a respiratory region below a critical growth rate and a respiratory-fermentative region above it. We note that the slopes of these linear relations signify a fixed ratio of either an uptake or a production rate over the growth rate. These ratios are 'yields' (introduced above; the quantities predicted with FBA). These linear segments suggest that the cell does not change its metabolic strategy over a range of growth rates, it only changes its protein expression and, possibly also, its metabolite concentrations to change growth rate and maintain a balanced metabolism (Elsemman et al., 2022).

We used flux balance analysis and a genome-scale stoichiometric model of *S. cerevisiae* (Lu et al., 2019) to estimate the number of elementary flux modes active in both regions. In the respiratory regions, we minimised the glucose uptake rate for a given growth rate. We found that an acceptable fit can be found with just two EFMs. A single EFM, describing respiratory metabolism and cell growth, is mixed with another that takes care of the growth-rate unassociated maintenance requirement. Above the critical growth rate, fermentation starts and the fermentation flux increases with growth rate while respiratory flux decreases. In this region, we obtained an acceptable fit with four EFMs. These fits are shown as lines in Fig. 4.

Although we can confirm that the linear relations between uptake and production rates as function of growth rates can be shown to be the result of 2 to 4 EFMs, this does not mean that we can always predict which reactions are used intracellularly, as flux variability in the optimal state is a general aspect of flux balance analysis computations (Mahadevan and Schilling, 2003). To address this one can for instance exploit whole-cell, protein-constraint models (Elsemman et al., 2022).

Thus, we conclude that a handful of EFMs can describe the metabolic behaviour of *S. cerevisiae* in a glucose-limited chemostat as a function of growth rate. A more contemporary approach would be to consider so-called ME models (O'Brien et al., 2013), which in addition to fluxes also predict the concentrations of the enzymes in the various EFMs and consider protein-expression constraints. They are also more predictive in terms of the used internal fluxes. Also this more advanced approach leads to the conclusion that only a handful EFMs can explain the experimental data (Elsemman et al., 2022).

That the flux data can be fitted with only a few EFMs (2 or 4) is only indirect proof of the optimisation of metabolism for a maximal growth rate. A more directly supporting piece of evidence – to which we return below – that protein expression maximises growth rate comes from experimentally titrating about 80 different protein levels in *S. cerevisiae* (Keren et al., 2016). The great majority (approx. 80%) of the proteins that were assayed were expressed at growth-rate maximising concentrations. This behaviour was also found for PTS proteins in *Salmonella typhimurium* (van der Vlag et al., 1994), citrate synthase (Walsh and Koshland Jr., 1985) and ATP synthase (Jensen et al., 1993a,b, 1995; Rabbers and Bruggeman, 2022) in *E. coli* and several glycolysis enzymes in *Lactococcus lactis* (Solem et al., 2008, 2003, 2010). On theoretical grounds it has also been suggested that ribosomal proteins are expressed at levels that maximise growth rate in *E. coli* (Bosdriesz et al., 2015; Scott et al., 2014). Thus, many microbes appear to express metabolic enzymes for maximising growth rate. A theoretical study also indicates that this can be achieved by surprisingly simple genetic circuits, involving only a single transcription factor (Planqué et al., 2018).

## 12. The optimal protein expression state of an elementary flux mode

We need to consider briefly how the optimal metabolite and protein concentrations can be calculated for an elementary flux mode, in order to understand, in the two next sections, how individual enzymes contribute to growth-rate maximisation and to analyse the flux and growth-rate control coefficients of the proteins expressed in the optimal state.

Consider again optimisation problem (31) and recall that an elementary flux mode has a single independent flux so that the ratio between any two of its flux values is constant and can be considered a stoichiometric property of the EFM. Thus, we can relate all the flux values in an EFM to a single (independent flux) value, e.g., to the protein synthesis flux $j_R(\boldsymbol{c})$. We define these flux ratios as $\beta_k = j_k/j_R$ so that the objective function (defined in (32)) can be written as (Noor et al., 2016; Planqué et al., 2018),

$$\mathcal{O}(\boldsymbol{c};\boldsymbol{c}_0) = \sum_{k=1}^{R} \frac{\beta_k}{k_k f_k(\boldsymbol{c};\boldsymbol{c}_0)}. \tag{33}$$

The optimal concentration $\boldsymbol{c}^o$ is the vector of metabolite concentrations at which $\mathcal{O}(\boldsymbol{c};\boldsymbol{c}_0)$ is minimised. This objective corresponds to the minimal enzyme concentration that is needed to reach $j_R(\boldsymbol{c}) = 1$ and the optimisation variables are now only metabolite concentrations.[7]

From the optimal metabolite concentrations $\boldsymbol{c}^o$, we can calculate the optimal enzyme concentrations when we know the desired flux value $j_R$. (Alternatively, we can also determine the flux when we know the total concentration of enzymes expressed in the EFM.) For example, in case we know the desired value of the objective flux $j_R(\boldsymbol{c}^o) = j_R^d$ then the optimal enzyme concentrations equal $e_k = j_R^d \beta_k/(k_k f_k(\boldsymbol{c}^o;\boldsymbol{c}_0))$.

We conclude that we can determine the optimal state of an EFM when we have a complete kinetic model of it, containing all the enzyme-kinetic rate equations, their parameter values and a characterisation of the environmental conditions. When we lack the complete kinetic information we can still perform insightful calculations, which is what ME models achieve (O'Brien et al., 2013; Elsemman et al., 2022).

## 13. Kinetic interpretation of the optimal state of an EFM

An EFM in an optimal state has a maximised steady-state flux per unit invested biosynthetic resources (total growth-associated protein concentration) or, alternatively, it requires the smallest total concentration of proteins to achieve a desired flux. What does this optimal state of the system imply for the activity of the metabolic enzymes?

At the level of a single enzyme, the catalytic rate (flux) per unit invested enzyme is maximised when the enzyme is completely saturated with its substrate, so that $f(\boldsymbol{c})$ in Eq. (22) approaches 1. This is however not in agreement with the optimal state of any metabolic network, because: i. a maximal substrate saturation can only be achieved when the product concentrations are zero, implying that the next reaction will carry no flux and thus violating the steady-state requirement of the entire EFM, and ii. a high concentration of the substrate inhibits the previous enzyme, for which it serves as a product, so the substrate concentration cannot be excessively high. Thus, the fact that the catalysis rate of an enzyme always depends on concentrations of substrates, products and effectors implies that all those rates are interdependent and that maximisation of a metabolic flux will result in an optimal comprise between the extents of substrate activation and product inhibition.

Minimisation of the objective function (33), with the metabolite concentrations as optimisation variables, amounts to maximising a harmonic mean of the saturation levels (each multiplied with a constant).[8] The interpretation of each term in the sum (33),

$$\tau_j = \frac{\beta_k}{k_k f_k(\boldsymbol{c};\boldsymbol{c}_0)}, \tag{34}$$

is the expected duration of a catalytic event of reaction $k$. Hence, the objective function measures the mean time for a molecule to pass through all reactions in the EFM. A priori, we do not see how we can say anything about the distribution, or the variance, of those values in the optimal state, without having to make simplifying assumptions regarding the network structure and the enzyme kinetics — as, for instance, done in Klipp and Heinrich (1999).

It has been speculated that evolutionary optimisation may lead (under particular conditions) to reactant concentrations that are close to their $K_M$ constant (Cornish-Bowden, 1976). This appears, however, not to be in agreement with data, obtained with *E. coli*, that indicates that substrate concentrations exceed their affinity constant ($K_M$) (Bennett et al., 2009), indicating closer to optimal enzyme usage. Moreover, since nearly all metabolic intermediates act as both substrates and products (and therefore have multiple affinity constants), and occur in reactions with varying equilibrium constants, we doubt whether any fundamental statement can be made about the ratio of reactant concentrations over affinity constants and enzyme saturation in optimal conditions. We expect that this may be highly system-specific. At this stage, we can only conclude that the average effective turnover time (overall enzymes) is minimised by evolution in constant conditions.

## 14. Derivation of the flux control coefficients at the optimal state of an elementary flux mode

The minimiser $\boldsymbol{c}^o$ of the objective function (33) is a critical point, and so satisfies

$$\forall c_j \in \boldsymbol{c}^o : \frac{\partial \mathcal{O}(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial c_j} = 0. \tag{35}$$

From this relation, we can derive two equations that concern the control coefficients of enzymes in cells with a maximal growth rate. These were previously derived by Klipp and Heinrich (1999), when asking which relations hold when a given flux distribution is attained at a minimal investment of enzymatic resources. They considered any flux distribution, so any conical sum of EFMs. We only consider single EFMs, because we know now that these are the solutions of the maximisation of the flux per unit enzyme. By limiting ourselves to EFMs, we can also extend a previous control analysis result derived for linear pathways (Heinrich and Klipp, 1996) to EFMs that can contain all the stoichiometric complexities of metabolic networks and contain branches, cycles, and moiety conservation (such as the example shown in Fig. 3).

We start from equation (35), which applies to a single EFM, and rewrite it in a few steps (for notational simplification we define $g_k(\boldsymbol{c}) = k_k f_k(\boldsymbol{c})$) to obtain an expression first obtained by Klipp and Heinrich (1999) in the more general setting of any flux distribution,

$$\begin{aligned}
0 = \frac{\partial \mathcal{O}(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial c_j} &= \sum_{k=1}^{R} \beta_k \frac{\partial g_k^{-1}(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial c_j} \\
&= \sum_{k=1}^{R} \beta_k \frac{e_k}{c_j} \frac{\partial (e_k g_k)^{-1}(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial \ln c_j} \\
&= \sum_{k=1}^{R} \beta_k \frac{e_k}{c_j} j_k^{-1} \frac{\partial \ln j_k^{-1}(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial \ln c_j} \\
&= -\sum_{k=1}^{R} \beta_k \frac{e_k}{c_j} j_k^{-1} \frac{\partial \ln j_k(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial \ln c_j} \\
&= -\frac{1}{c_j j_R} \sum_{k=1}^{R} e_k \frac{\partial \ln j_k(\boldsymbol{c}^o;\boldsymbol{c}_0)}{\partial \ln c_j}.
\end{aligned} \tag{36}$$

Since the partial derivative is taken at fixed concentrations of all other metabolites, we deduce that

$$\partial \ln j_k / \ln c_j = \partial \ln v_k / \ln c_j = \epsilon_{c_j}^{j_k}$$

[7] Minimising this objective function is a convex optimisation problem on a convex bounded domain in the logarithms of the metabolite concentrations (Noor et al., 2016) (in fact, it is strictly convex in many cases (Planqué et al., 2018)). The existence of a unique optimiser is thus guaranteed. When we would know the exact convex combination of EFMs, e.g., from data, then this result is still true (Klipp and Heinrich, 1999).

[8] For a vector $\boldsymbol{x}$ of length $n$, the harmonic mean is defined as $n/(\sum_k \frac{1}{x_k})$.

and therefore

$$\sum_{k=1}^{R} e_k \epsilon_{c_j}^{v_k} = 0. \tag{37}$$

Thus, the last equation holds in the optimal state of an EFM, when it has achieved a maximal flux per unit protein investment.

We know from the flux control connectivity theorem of metabolic control analysis that the following relationship holds in all (locally stable) steady states (Reder, 1988):

$$\sum_{k=1}^{R} C_k^j \epsilon_{c_j}^{v_k} = 0, \tag{38}$$

with $C_k^j = \frac{d \ln j}{d \ln e_k}$ as the flux control coefficient of reaction $k$ on the steady-state flux $j$. Since we are considering an EFM there is only one independent flux, so we can choose any flux of the EFM as independent. (Therefore, the flux control coefficient of enzyme $i$ on flux 1 is the same as its control of flux 2, 3, etc.) In our context, the flux we consider as independent is the protein synthesis rate $j_R$.

The summation theorem of flux control coefficients (25) specifies that

$$\sum_{k=1}^{R} C_k^j = 1. \tag{39}$$

From the last three equations, we deduce that in the optimal state of an EFM the flux control coefficients of enzymes equal their protein concentration fraction, i.e.,

$${}^{o}C_k^j = \frac{e_k}{e_T}. \tag{40}$$

An alternative derivation of (40) can be found in a footnote.[9] In the appendix, we revisit the analysis of Burns and Kacser on flux control in linear pathways at a state of maximal flux per unit invested protein, which is remarkably elegant.

We note that (40) was first derived by Heinrich and Klipp (1996), for linear metabolic pathways (using the Lagrange multiplier method for optimisation under constraints). The problem with that approach however is that it is unclear under which conditions it precisely applies; the 'Lagrange multiplier method' does for instance not apply when more than one independent flux exists in the metabolic network.

It is important to realise that we divided by $e_T$ in (40) – the total enzymatic resource available for the EFM – and not by $p_T = e_T + p_{NG}$. When we consider the total protein concentration of a cell then the control of a metabolic protein on the metabolic flux through an EFM in an optimal state equals

$${}^{o}C_k^j = \frac{e_k}{e_T} = \frac{e_k}{p_T} \frac{1}{1-\phi_{NG}}. \tag{41}$$

The quantity $e_k/p_T$ is the protein-concentration fraction of protein $k$. Below, we will analyse published proteomics data from this control perspective.

In the context of growth-rate maximisation with the protein synthesis rate considered as the independent flux the last expression changes into

$${}^{o}C_k^{j_R} = \frac{e_k}{e_T} = \frac{e_k}{p_T} \frac{1}{1-\phi_{NG}}. \tag{42}$$

[9] Let $q(\boldsymbol{e}) = j(\boldsymbol{e})/e_T = j(\boldsymbol{e})/\sum_{k=1}^{n} e_k$ be the specific flux. In an EFM, all reactions are essential to sustain a flux. Hence $j(\boldsymbol{e}) = 0$ if $e_k = 0$ for at least one $e_k$. Therefore, $q(\boldsymbol{e})$ can only have an interior maximum, i.e., a point $\boldsymbol{e} > \boldsymbol{0}$ where $\frac{\partial q}{\partial e_k} = 0$ for all $k = 1, \ldots, n$. At such a point,

$$0 = \frac{\partial q(\boldsymbol{e})}{\partial e_k} = \frac{e_T \frac{\partial j}{\partial e_k} - j(\boldsymbol{e})}{e_T^2}$$

so that

$$\frac{\partial j}{\partial e_k} = \frac{j}{e_T} \quad \Longleftrightarrow \quad \frac{e_k}{j}\frac{\partial j}{\partial e_k} = \frac{e_k}{e_T}.$$

Thus, when we consider the maximisation of the growth rate at a fixed growth-unassociated protein fraction $\phi_{NG}$ then we obtain the last expression at the optimal state. We note that $e_k/e_T$ also equals the protein-number fraction, $n_k/\sum_{j\in G} n_j$ (with $G$ as the set of growth-associated proteins and $e_j = n_j/v$ with $n_j$ as the number of proteins $j$ and $v$ as the volume of the average cell). This means that we can estimate the flux control coefficients on the protein synthesis rate from proteomics data, which we do below.

## 15. Derivation of the growth-rate control coefficients at the optimal state of a growth-supporting elementary flux mode

In Eq. (13) we deduced that the specific growth rate of a cell equals its protein synthesis rate divided by its total protein concentration, i.e., $j_R/p_T$. This implies that the growth rate control coefficients of enzymes are related to their control coefficients on the protein synthesis rate and their protein fraction (see also footnote 9),

$$C_k^{\mu} = \frac{d \ln \mu}{d \ln e_k} = C_k^{j_R} - \frac{e_k}{p_T} = C_k^{j_R} - \frac{e_k}{e_T}\frac{e_T}{p_T}. \tag{43}$$

This equation shows that all proteins without flux control on the protein synthesis rate have a negative growth rate control coefficient equal to minus their protein fraction. Lowering their concentration thus enhances the flux. These proteins are therefore a burden to the cell (Snoep et al., 1995).

At the optimal state of an EFM, we deduced above (Eq. (42)) that ${}^{o}C_k^{j_R} = \frac{e_k}{e_T}$. This means that when $\mu$ is maximal, given a fixed growth-unassociated protein fraction $\phi_{NG}$, that ${}^{o}C_k^{j_R/e_T} = 0$, but ${}^{o}C_k^{\mu} \neq 0$, since

$${}^{o}C_k^{\mu} = \frac{e_k}{e_T}\left(1 - \frac{e_T}{p_T}\right) = \frac{e_k}{p_T}\frac{\phi_{NG}}{1-\phi_{NG}} > 0, \tag{44}$$

indicating that the growth-rate control coefficients are zero when the cell does not express any proteins not associated to growth in the optimal state. (Growth-unassociated proteins we defined above as proteins that do not catalyse any of the reactions in the EFM(-s) making all the biosynthetic components and free energy carriers of the cell from extracellular nutrients.) This happens when $\phi_{NG} = 0$, a scenario that can be most certainly excluded as cells do also express proteins that carry different functions than catalytic functions in the growth-supporting EFM such as structural proteins (H-NS, FtsZ), transcription factors, flagellar proteins, etc..

## 16. Summation theorems of flux and growth-rate control coefficients

So far, we considered two kinds of protein pools: a growth-associated protein pool of enzymes active in the growth-associated metabolic network (catabolism and anabolism; an EFM), responsible for making all the cellular components, and another pool of proteins, which are not associated to that EFM and carry out other growth-unassociated cellular functions. Since these different types of proteins have different control properties on metabolic flux and growth rate, we derive the summation theorems of control coefficients by summing over all the proteins in the cell, pool by pool.

First we show that the flux control coefficient in the optimal state sum to 1 as they have to:

$$\begin{aligned}\sum_j {}^{o}C_j^J &= \sum_{k\in G} {}^{o}C_k^J + \sum_{l\in NG} {}^{o}C_l^J \\ &= \sum_{k\in G} \frac{e_j}{e_T} + \sum_{l\in NG} 0 = 1.\end{aligned} \tag{45}$$

The growth-rate control coefficients sum to 0 (in the optimal and any non-optimal state),

$$\sum_j C_j^{\mu} = \sum_{k\in G} C_k^{\mu} + \sum_{l\in NG} C_l^{\mu}$$

$$= \sum_{k\in G}\left(C_k^{j_R} - \frac{e_k}{p_T}\right) + \sum_{l\in NG} -\frac{p_l}{p_T}$$
$$= \sum_{k\in G} C_k^{j_R} - \sum_{k\in G}\frac{e_k}{p_T} - \phi_{NG}$$
$$= 1 - \phi_G - \phi_{NG}$$
$$= 0. \quad (46)$$

Thus, the growth-unassociated proteins each have a growth rate control coefficient that equals minus their protein fraction (since they increase $p_T$ without affecting the protein synthesis rate).

Note that we took sums here also of proteins that do not have a catalytic activity (unlike a metabolic enzyme). Such growth-unassociated proteins therefore do not affect metabolic flux when changed in concentration but they do affect the specific growth rate via its denominator $p_T$. We note that we deviate here from classical metabolic control analysis which generally does not consider control of non-catalytic proteins.

The last summation theorem was to be expected, of course: since the flux and the total protein concentration are each 1-homogeneous in protein concentrations, any specific flux $j/p_T$ such as the growth rate $\mu = j_R/p_T$ is 0-homogeneous,

$$\mu(\lambda p_T) = \lambda^0 \mu(p_T), \quad \text{for } \lambda > 0, \quad (47)$$

so its control coefficients should add to 0, as explained in Section 9. A cell therefore does not grow faster when it has a higher total protein concentration as it also needs to make all of them.

## 17. Interpreting optimal flux control coefficients

The expressions for the optimal flux control coefficient can be further interpreted, using equations we have already derived.

First, we can exploit the general enzymatic rate equation $v_k = k_k e_k f_k(\boldsymbol{c})$ (22) and the fact that an EFM has a single independent flux to rewrite the metabolic protein fraction as

$${}^{o}C_k^j = \frac{e_k}{e_T} = \frac{\frac{v_k}{k_k f_k(c^o)}}{\sum_j \frac{v_j}{k_j f_j(c^o)}} = \frac{\frac{\beta_k}{k_k f_k(c^o)}}{\sum_j \frac{\beta_j}{k_j f_j(c^o)}} = \frac{\tau_k^o}{\sum_j \tau_j^o}. \quad (48)$$

The optimal flux control coefficient is now expressed in terms of the optimal effective turnover times of all the enzymes $\tau_k^o = \frac{\beta_k}{k_k f_k(c^o)}$. This last equation indicates that the flux control coefficient of an enzyme, say on the protein synthesis flux $j_R$, equals the time that this enzyme needs to convert its substrate molecule(-s) into product molecule(-s) (note the role of $\beta$ and $f(\boldsymbol{c})$ here) divided by the total time needed to run all enzymatic reactions once.

The specific growth rate can also be written in terms of the turnover time (starting from Eq. (13) and using a part of the previous derivation),

$$\mu = \frac{j_R}{e_T}(1-\phi_{NG}) = \frac{1}{\sum_j \frac{\beta_j}{k_j f_j(\boldsymbol{c})}}(1-\phi_{NG}) = \frac{1}{\sum_j \tau_j}(1-\phi_{NG}).$$

Therefore, maximisation of specific growth rate can therefore also be seen as an optimal scheduling task, which was realised earlier by Pugatch (Pugatch, 2015).

Secondly, we can use the steady-state relation between the rates of translation and dilution of a protein to rewrite the optimal flux control coefficient expression in terms of its protein fraction. Recall first that

$$0 = \dot{p}_k = \alpha_k j_R(\boldsymbol{c}) - \mu p_k.$$

Since $\mu = j_R(\boldsymbol{c})/p_T$, this becomes

$$0 = \mu(\alpha_k p_T - p_k)$$

so that $\alpha_k = p_k/p_T$ in steady state. Hence, the fraction of ribosomes translating growth-supporting and non-growth supporting proteins, $\alpha_G$ and $\alpha_{NG}$, are equal to $\phi_G$ and $\phi_{NG}$, respectively,

$$\alpha_G = \sum_{k\in G}\alpha_k = \sum_{k\in N} p_k/p_T = \phi_G,$$
$$\alpha_{NG} = \sum_{k\in NG}\alpha_k = \sum_{k\in NG} p_k/p_T = \phi_{NG}.$$

In terms of these quantities,

$${}^{o}C_j^J = \frac{e_j}{p_T}\frac{1}{1-\phi_{NG}} = \frac{\alpha_j}{1-\phi_{NG}} = \frac{\alpha_j}{\alpha_G}. \quad (49)$$

This relation shows that the flux control coefficients of enzymes can be inferred from experimental ribosome profiling data, in addition to number fraction data of proteomics (assuming the cell is optimal and uses one EFM).

## 18. Experimental evidence of optimal expression of metabolic enzymes and the fitness potential of an enzyme expressed in terms of control coefficients

Whether cells express metabolic enzymes to growth-rate maximising levels can be experimentally verified by titrating the protein concentrations in a mutant strain, with for instance an IPTG-titratable promoter (Jensen et al., 1993c; Walsh and Koshland Jr., 1985; Dekel and Alon, 2005; Keren et al., 2016; Rabbers and Bruggeman, 2022). A comparison of the growth rate of the wild type with the maximal growth rate obtained in the titration experiment then indicates whether the wild type displays an optimal concentration of the protein under investigation. Such studies have been done for metabolic enzymes in *E. coli*, *L. lactis*, *S. typhimurium* and *S. cerevisiae* and these nearly always show optimal expression (Fig. 5) (Jensen et al., 1993a,b, 1995; Keren et al., 2016; Rabbers and Bruggeman, 2022; Solem et al., 2008, 2003, 2010; van der Vlag et al., 1994; Walsh and Koshland Jr., 1985; Dekel and Alon, 2005). We note that suboptimal preparatory protein overexpression also occurs (Grigaitis and Teusink, 2022; Berney et al., 2006; Ihssen and Egli, 2004, 2005; Mashego et al., 2005; Mori et al., 2017), but predominantly at nutrient-limited conditions. All the experimental evidence of optimal protein expression were however obtained at nutrient excess, when protein overexpression is indeed less (O'Brien et al., 2016). It is currently not known how the specific growth rate depends on the expression level of a needed protein under nutrient-limited conditions, for instance, achieved in a chemostat.

How can protein-expression-optimality studies be viewed in the light of metabolic control coefficients? What we have concluded so far is that the control coefficient of an enzyme on the protein synthesis flux equals its protein fraction $e_k/e_T$ (Eq. (40)) under optimality conditions (assuming a fixed growth-unassociated protein fraction), while its growth rate control coefficient is then unequal to zero (Eq. (44); it only equals zero when $e_T = p_T$). But are these results not contradictory to the plots shown in Fig. 5 that show a close-to-zero slope at the point of optimality? To understand this, we have to define a new type of MCA sensitivity coefficient that considers that increases in a (titrated) enzyme concentration occur at the expense of a reduction in concentration of other enzymes — since they compete for biosynthetic resources. So, we need to define a sensitivity coefficient under a constraint of a fixed total protein concentration (Berkhout et al., 2013); one that has not been considered before in MCA.

Consider the scenario that an experimentalist titrates a protein $i$, using, for instance, an IPTG-titratable promoter (Jensen et al., 1993c), to some specific concentration $e_i$ and that the cell then allocates the remaining resources $e_T - e_i$ to optimally express the remaining growth-associated enzymes to maximise the growth rate (assuming again that $p_T$ and $\phi_{NG}$ are fixed). This is one way to interpret the experiments shown in Fig. 5. Under these conditions, the normalised slopes in these figures can be shown to equal (see Appendix)

$$\mathcal{F}_i^{\mu}(e_i) = \frac{d\ln\mu(e_i)}{d\ln e_i} = \frac{C_{e_i}^{j_R}(e_i) - \frac{e_i}{e_T}}{1-\frac{e_i}{e_T}}. \quad (50)$$

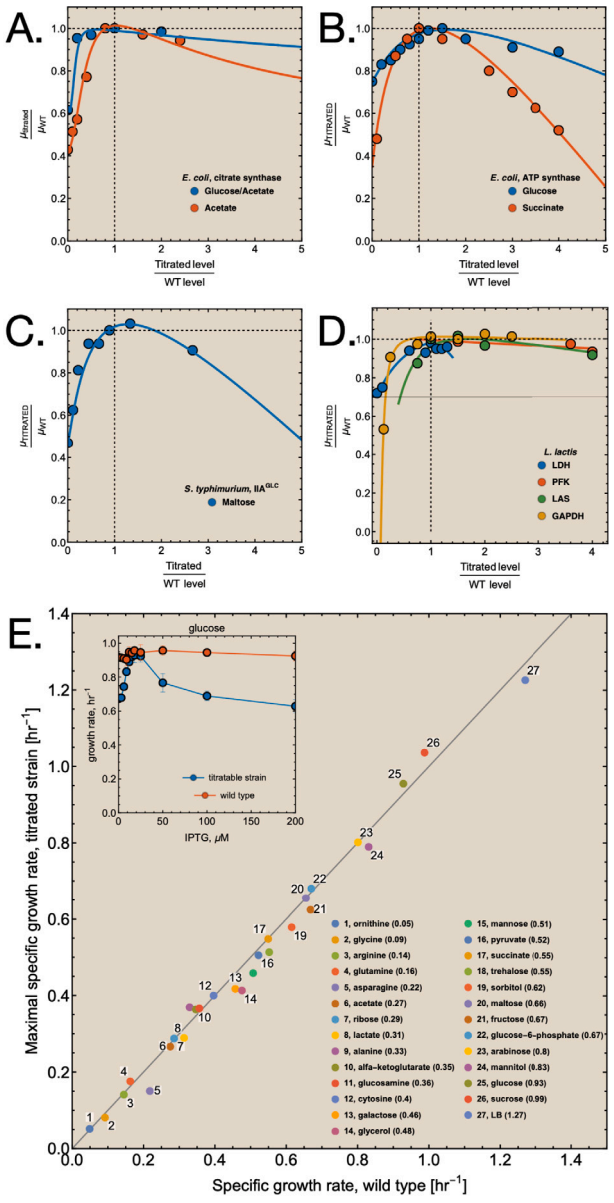



**Fig. 5. Examples of optimal, growth-rate maximising expression of different metabolic enzymes across three different microbial species.** Several examples of optimal expression enzymes are shown: citrate synthase in *E. coli* (Walsh and Koshland Jr., 1985) (A.), ATP synthase in *E. coli* (Jensen et al., 1993a,b, 1995) (B.), C. PTS IIA$^{GLC}$ in *S. typhimurium* (van der Vlag et al., 1994) (C.), several glycolytic enzymes in *L. lactis* (Solem et al., 2008, 2003, 2010) (D.), and ATP synthase in *E. coli* across 27 conditions (Rabbers and Bruggeman, 2022) (E.). Four of the 27 conditions shown in E. had a deviation from optimality of more than 10%.

(For clarity we indicate the functional dependencies on the concentration $e_i$.) When the protein concentration is below its optimal level, at which the growth rate is maximal, $C^{j_R}_{e_i} - \frac{e_i}{e_T} > 0$, and above it $C^{j_R}_{e_i} - \frac{e_i}{e_T} < 0$. At the optimal level, we obtain ${}^{o}C^{j_R}_{e_i} = \frac{e^o_i}{e_T}$ and $\mathcal{F}^{\mu}_i(e^o_i) = 0$, which is in agreement with earlier findings. Note also that generally, $\frac{e_i}{e_T} \ll 1$ and therefore $\mathcal{F}^{\mu}_i(e_i) \approx C^{j_R}_{e_i}(e_i) - \frac{e_i}{e_T}$, which agrees with the earlier conclusion that, at fixed $\phi_{NG}$, optimisation of growth rate amounts to optimisation of $j_R/e_T$.

One can think of $\mathcal{F}^{\mu}_i(e_i)$ as the 'fitness potential' of an enzyme. If a cell aims to optimise its growth rate by optimal allocation of resources then $\mathcal{F}^{\mu}_i(e_i)$ indicates the return of investment when the cell increases $e_i \to e_i + \delta e_i$ by gene expression regulation.

Eq. (50) indicates that two enzymes each with the same absolute distance to optimality $C^{j_R}_{e_i}(e_i) - \frac{e_i}{e_T}$ differ in their fitness potential when they are expressed at different relative abundances. The one with the highest relatively abundance has the highest fitness potential. This indicates that abundant enzymes can be expected to be more tightly controlled in their expression level then low abundance enzymes – because suboptimal expression of abundant enzymes is a significant waste of resources.

## 19. Estimating growth-rate control coefficient from a comprehensive proteomics dataset reporting protein-number fraction data

So far, we have considered the optimal expression of all the growth-associated proteins that together catalyse the reaction in the optimal EFM that maximises the specific growth rate of the cell given a single protein-expression constraint. Eq. (44) indicates that the optimal control coefficient of growth-associated protein (e.g., a metabolic enzyme) on growth rate is proportional to its protein fraction within the EFM it is part of. This result applies however to the case that the cell expresses a single EFM in its optimal state and this is likely not always the case in all growth conditions (e.g., Fig. 4). In the case where several EFMs are active, the $l$th EFM gets a fraction of $e_{T,l}/e_T$ of the growth-associated protein pool and the optimal flux control coefficient of enzyme $k$ on $j_{R,l}$ of EFM $l$ equals $e_{k,l}/e_{T,l}$. Assuming a fixed conical sum of EFMs, the optimal control coefficient of enzyme $k$ on the biosynthetic flux $j_R$ can then be *approximated*[10] by:

$$\frac{d\ln j_R}{d\ln e_k} \approx \sum_l \frac{\partial \ln j_R}{\partial \ln j_{R,l}} \frac{\partial \ln j_{R,l}}{\partial \ln e_{k,l}} \frac{\partial \ln e_{k,l}}{\partial \ln e_k} = \sum_l \frac{\partial \ln j_R}{\partial \ln j_{R,l}} \frac{e_{k,l}}{e_{T,l}} \frac{\partial \ln e_{k,l}}{\partial \ln e_k}$$
$$= \sum_l \frac{e_{k,l}}{e_{T,l}} = \frac{e_k}{e_T} \underbrace{\sum_l \frac{e_{k,l}}{e_k} \frac{e_T}{e_{T,l}}}_{\alpha} \propto \frac{e_k}{e_T}$$

where $j_R = \sum_l j_{R,l}$, $e_k = \sum_l e_{k,l}$, $e_T = \sum_l e_{T,l}$ and $\alpha$ as a proportionality constant. Thus, the optimal control coefficient of enzyme $k$ on the biosynthetic flux is proportional to its concentration fraction of the growth-associated proteins. Its control coefficient on cellular growth rate then becomes ${}^{o}C^{\mu}_{k} \approx \alpha \frac{e_k}{e_T} - \frac{e_k}{p_T} = \frac{e_k}{p_T}\left(\alpha \frac{p_T}{e_T} - 1\right) \propto \frac{e_k}{p_T}$. Thus, the growth-rate control coefficients of the growth-associated proteins of a cell, using several EFMs in its state of maximal growth rate, is approximately proportional to its protein concentration fraction and, therefore also to its number fraction.

Fig. 6 shows experimental data of normalised protein expression, as number fractions, in *E. coli* under a single condition. These numbers are proportional to their growth-rate control coefficients, assuming that the corresponding concentrations of the proteins are not too far removed from optimal concentrations. Enzymes with high control (>0.04) are found in the glucose uptake system (ptsH), glycolysis (D-glyceraldehyde-3-phosphate dehydrogenase: gapA, enolase: eno), TCA (isocitrate dehydrogenase: icd), amino acid metabolism ((cobalamin-independent) methionine synthase: metE, cysteine synthase A: cysK, ketol-acid reductoisomerase; ilvC, serine hydroxymethyltransferase: glyA) and several ribosomal subunits. This indicates growth-rate control is likely distributed over a relatively small number of catabolic and anabolic steps.

[10] We write approximated here because since enzyme $k$ can be active in different EFMs its optimal reactant concentrations may be different in each EFM it participates in. Here we assumed that in all these EFMs, its reactant concentration has the same optimal value.

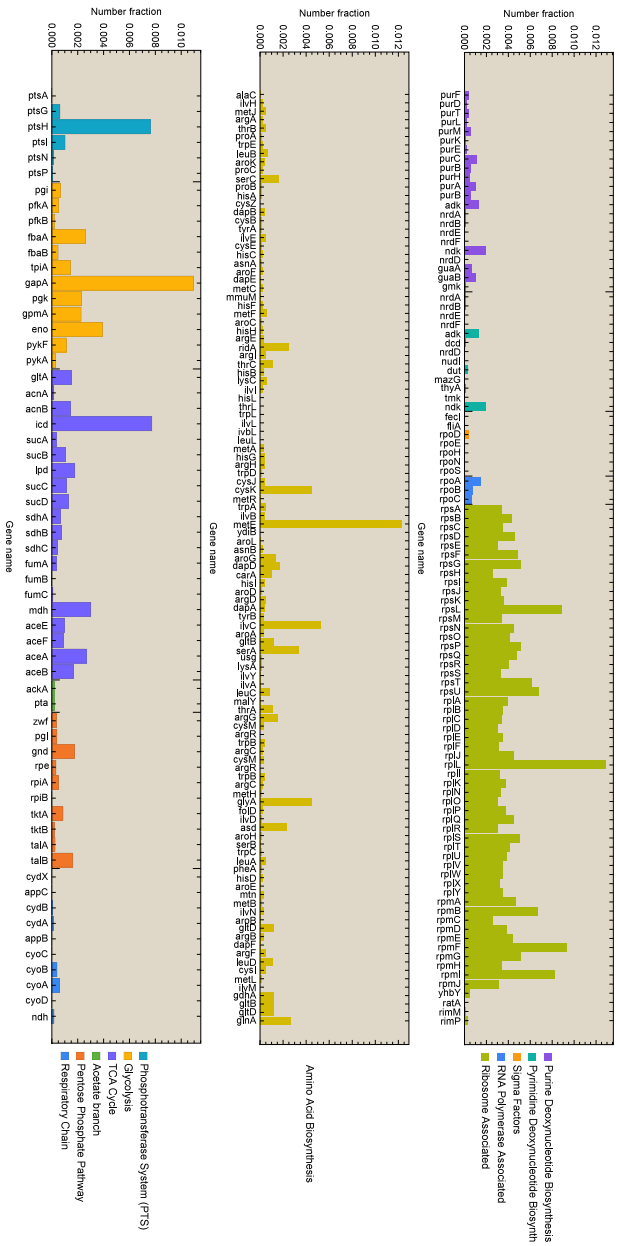



**Fig. 6. Number fractions (proxies for $e_j/p_T$) of metabolic and biosynthetic proteins in *E. coli* categorised according to their metabolic network or function.** We took the proteomics data from Mori et al. (2021) under the reference growth condition (MOPS, glucose, $NH_4Cl$) and selected the metabolic proteins of key metabolic pathways (in addition to sigma factors and RNA polymerase, which are shown for reference). The mean and standard deviation of all the plotted number fractions (273 in total) equal $1.47 \times 10^{-3}$ and $2.14 \times 10^{-3}$.

## 20. Closing remarks

Nowadays it is clear for system biologists and biophysicists: you have to study the 'whole' to understand it (think of 'omics'), the 'whole' can be more than the sum of its 'parts' (emergence is omnipresent in cell biology), and the 'parts' should be understood it terms of the 'whole' (e.g., feedbacks responsible for cellular energy homoeostasis). This was not the case when MCA arose (Kacser, 1986). Reductionistic concepts such as 'rate-limiting steps' reigned and were impeding progress into quantitative understanding of metabolic pathways in terms of enzyme kinetics and their underlying genes.

MCA initiated quantitative studies on a particular type of biochemical system, i.e., metabolic pathways, after enough fundamental understanding of free energy transduction and catalysis by enzymes was attained. MCA subsequently provided the understanding that changes in system properties depend nonlinearly on (all) enzyme concentrations and that regulation emerges from a subtle interplay between pathway topology and enzyme kinetics. The pioneering papers of Burns & Kacser and Heinrich & Rapoport had this as their main aim. Afterwards the MCA community developed mathematical modelling software (e.g., Jarnac by Herbert Sauro and Gepasi by Pedro Mendes and Douglas Kell) for simulation and analysis of mathematical models of biochemical systems. These developments eventually contributed to the emergence of systems biology, about two decades ago (Kitano, 2002; Westerhoff and Palsson, 2004).

In addition to metabolic pathways, one can however define many biochemical systems in cell biology: cooperative enzymes, motor proteins, signalling networks, DNA replication and segregation, gene-regulatory circuits, etc. These are all unintuitive nonlinear systems giving rise to emergent behaviours that correspond to 'tasks' in the living cell. Although this list is not endless, it is huge. We are therefore also interested in principles that apply across a wide range of systems. This is an approach that has been very successful in physics and has been influential in systems biology and biophysics. We exploited many principles throughout this review: the concept of fitness, the universal rate equation, the stoichiometric matrix, the flux cone, elementary flux modes, balanced growth, and the expression of growth rate in terms of all enzyme activities in a cell, etc. Each of these are applicable across living systems.

A clear strength of MCA has always been its applicability to all enzymatic systems, regardless of the precise enzyme kinetics and system structure (reaction stoichiometry and regulatory circuitry), via its generally applicable definitions of elasticity and control coefficients (and their products response coefficients) (Reder, 1988). A downside is, however, that as long as the enzyme kinetics is not known, insightful deductions can only be made of small systems, analytically-solvable simplified cases, or from analysis of experimental data. And, although kinetic models of metabolic and signalling systems can be made (Bakker et al., 1997; Rohwer et al., 2000; Teusink et al., 2000; Bruggeman et al., 2005; Hornberg et al., 2005) and subsequently analysed for control features, they are very laborious to develop (enzyme by enzyme) and often limited by parameter-identifiability problems. Thus, prediction of elasticity and control coefficient values is hard. This, perhaps, may have prevented widespread usage of MCA. In this review, we hope to have shown that this limitation disappears when evolutionary optimality of the growth rate is assumed, for which quite some experimental data also exists. Some nuisances about this viewpoint can be found in Bruggeman et al. (2023).

MCA can also be used to understand the design and function of particular recurrent biochemical systems' designs. Examples are flux control by demand and homeostasis by negative feedback (Hofmeyr and Cornish-Bowden, 2000), zero-order ultrasensitivity, sequestration and sensitivity amplification in signalling circuits (Kholodenko et al., 1997; Blüthgen et al., 2006; Bruggeman et al., 2002), hierarchical control and regulation of metabolism (ter Kuile and Westerhoff, 2001), and moiety-conversed cycles in metabolic pathways (Hofmeyr et al., 1986). In these cases, the elasticity and control coefficients consequences are studied given a 'functional specification' of a system — for example, for an ultrasensitive response of a signalling cascade it can then be shown that the kinases and phosphatases have to operate in their zero-order regime and require low enough $K_M$'s that allow for this (Kholodenko et al., 1997; Blüthgen et al., 2006). Since we lack such insights for whole cell metabolism, applications of MCA to whole cells are limited. Thus, as long as we lack a principle that specifies a desired behaviour of metabolism, given a design or functional requirement from the whole cell level acting top down, we can only model metabolism in the cell bottom up, using mathematical models in terms of the kinetics of the

enzymes (Bakker et al., 1997; Rohwer et al., 2000; Teusink et al., 2000; Bruggeman et al., 2005; Hornberg et al., 2005).

In this review, we hope to have shown that at least for those microbial species that operate close-enough to a state of maximal growth rate, such as *E. coli* and *S. cerevisiae*, sufficient theory has been developed in the last few decades to obtain a near comprehensive understanding of their metabolic stoichiometry, enzymatic activities and control features. This extremum principle has enabled predictions of enzyme activities and protein concentrations at the whole cell level without having to know all the kinetic details of the enzymes. Although growth-rate maximisation may not apply to all microbes or cancer cells, a search for an extremum principle that specifies their behaviour may open up opportunities for whole-cell physiology prediction and MCA — which may, for instance, for cancer cells contribute to the identification of drug targets, such as enzymes controlling key cancer-cell functions.


## Funding information

BP and MD acknowledge funding by NWO grant 613.009.131 'Control of maximal growth rate by single-celled organisms'. FJB and MR acknowledge funding by NWO-XL grant OCENW.XL21.XL21.007 'Taking Control of Metabolism in Microbial Cell Factories by Applying Non-canonical Redox Cofactors'. HMS acknowledges funding by DOE awards BER DE-SC0023091 and DE-EE0008927. LS acknowledges funding by CONACYT.


## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Derivation of the fitness potential

Consider the titration of a protein $i$ to a fixed concentration $e_i$ and the strategy of the cell to subsequently optimally express all the other enzymes, given the total enzyme concentration constraint $e_T - e_i$ (given that $p_T = e_T + p_{NG}$). The change in the growth rate ($\mu = j_R/p_T$) then equals

$$d\ln\mu = \left(\frac{\partial\ln\mu}{\partial\ln e_i} + \sum_{j\neq i}\frac{\partial\ln\mu}{\partial\ln e_j}\frac{\partial\ln e_j}{\partial\ln e_i}\right) d\ln e_i.$$

We consider the following Lagrangian,

$$\mathcal{L}(\mathbf{e}) = \mu(\mathbf{e}) - \lambda\left(\sum_{j\neq i} e_j - (e_T - e_i)\right).$$

Now

$$\forall j\neq i: \frac{\partial\mu}{\partial e_j} = \lambda \quad\Rightarrow\quad \frac{\partial\ln\mu}{\partial\ln e_j} = \frac{e_j}{\mu}\lambda$$

from the summation theorem of growth-rate control coefficients (over only the growth-associated enzymes), which sum to $1 - e_T/p_T$, we deduce that,

$$\begin{aligned}\sum_{j\neq i}\frac{\partial\ln\mu}{\partial\ln e_j} &= \sum_{j\neq i}\frac{\partial\ln j_R}{\partial\ln e_j} - \sum_{j\neq i}\frac{\partial\ln p_T}{\partial\ln e_j} = 1 - \frac{\partial\ln j_R}{\partial\ln e_i} - \left(\frac{e_T}{p_T} - \frac{e_i}{p_T}\right)\\ &= 1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i}.\end{aligned}$$

Also given the above we deduce that

$$\sum_{j\neq i}\frac{\partial\ln\mu}{\partial\ln e_j} = \sum_{j\neq i}\frac{e_j}{\mu}\lambda = \frac{\lambda}{\mu}(e_T - e_i).$$

When we combine the last two results we can determine $\lambda$

$$\frac{\lambda}{\mu}(e_T - e_i) = 1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i} \quad\Rightarrow\quad \lambda = \frac{\mu\left(1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i}\right)}{e_T - e_i}.$$

Now we return to the equation we started with and substitute our results,

$$\begin{aligned}
d\ln\mu &= \left(\frac{\partial\ln\mu}{\partial\ln e_i} + \sum_{j\neq i}\frac{\partial\ln\mu}{\partial\ln e_j}\frac{\partial\ln e_j}{\partial\ln e_i}\right) d\ln e_i\\
&= \left(\frac{\partial\ln\mu}{\partial\ln e_i} + \sum_{j\neq i}\frac{e_j}{\mu}\lambda\frac{\partial\ln e_j}{\partial\ln e_i}\right) d\ln e_i\\
&= \left(\frac{\partial\ln\mu}{\partial\ln e_i} + \sum_{j\neq i}\frac{e_j}{\mu}\frac{\mu\left(1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i}\right)}{e_T - e_i}\frac{\partial\ln e_j}{\partial\ln e_i}\right) d\ln e_i\\
&= \left(\frac{\partial\ln\mu}{\partial\ln e_i} + e_i\frac{1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i}}{e_T - e_i}\sum_{j\neq i}\frac{\partial e_j}{\partial e_i}\right) d\ln e_i\\
&= \left(\frac{\partial\ln\mu}{\partial\ln e_i} + e_i\frac{1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i}}{e_T - e_i}\frac{\partial(e_T - e_i)}{\partial e_i}\right) d\ln e_i\\
&= \left(\frac{\partial\ln\mu}{\partial\ln e_i} - e_i\frac{1 - \frac{e_T}{p_T} - \frac{\partial\ln\mu}{\partial\ln e_i}}{e_T - e_i}\right) d\ln e_i\\
&= \left(\frac{\frac{\partial\ln\mu}{\partial\ln e_i}e_T - \frac{\partial\ln\mu}{\partial\ln e_i}e_i - e_i + e_i\frac{e_T}{p_T} + e_i\frac{\partial\ln\mu}{\partial\ln e_i}}{e_T - e_i}\right) d\ln e_i\\
&= \left(\frac{\frac{\partial\ln\mu}{\partial\ln e_i} - \frac{e_i}{e_T} + \frac{e_i}{e_T}\frac{e_T}{p_T}}{1 - \frac{e_i}{e_T}}\right) d\ln e_i\\
&= \left(\frac{\frac{\partial\ln j_R}{\partial\ln e_i} - \frac{e_i}{p_T} - \frac{e_i}{e_T} + \frac{e_i}{p_T}}{1 - \frac{e_i}{e_T}}\right) d\ln e_i\\
&= \left(\frac{\frac{\partial\ln j_R}{\partial\ln e_i} - \frac{e_i}{e_T}}{1 - \frac{e_i}{e_T}}\right) d\ln e_i.
\end{aligned} \tag{51}$$

Now if we consider the titration of an enzyme in the growth-associated pool that has no flux control, then

$$d\ln\mu = \frac{-\frac{e_i}{e_T}}{1 - \frac{e_i}{e_T}}\frac{1}{e_i}de_i \;\Rightarrow\; \int_{\mu(0)}^{\mu(e_i')} d\ln\mu = \int_0^{e_i'}\frac{-\frac{e_i}{e_T}}{1 - \frac{e_i}{e_T}}\frac{1}{e_i}de_i, \tag{52}$$

which leads to

$$\frac{\mu(e_i')}{\mu(0)} = 1 - \frac{e_i'}{e_T} = 1 - \frac{p_T}{e_T}\frac{e_i'}{p_T}, \tag{53}$$

which is consistent with the other derivations.

## Appendix B. Jim Burns' and Henrik Kacser's view on optimal allocation of resources for the maximisation of growth rate

While writing this review, we read through the metabolic control theory literature again. One classic piece is the thesis of Jim Burns who co-authored the first paper on metabolic control theory with Henrik Kacser (A LaTeX version of the Thesis is available at https://github.com/hsauro/JumBurnsThesis). Surprisingly, this thesis contains a section on optimal metabolic control theory. In this section, we reproduce the main approach and their understanding of it in terms of evolution — since this thesis is not widely available. It is a great addition to the related work by Klipp and Heinrich.

Burns considered a simple two-step pathway at steady-state. He then described the following operational experiment of moving an $\Delta e$ amount of protein from one step to another. If the total amount of enzyme is fixed and the distribution of amounts of the two enzymes is optimal such that we are at maximum flux for a given total amount of protein, then it must be true that $\delta e_1 + \delta e_2 = 0$ and importantly the change in flux is zero. The latter condition is true because as the ratio of enzymes is changed, the flux will increase above the optimum and

then decrease. At the maximum, the rate of change is zero; hence, any change will result in no change in flux.

Guy Brown, in his 1990s paper (Brown, 1991), independently came to the same conclusion. He described the concept better than Burns did in his thesis, but the result is the same.

From Burns, at the optimal allocation, we can state that $\delta e_1 + \delta e_2 = 0$ and $\delta J = 0$. Expressed using flux control coefficients, we can write:

$$\frac{\delta J}{J} = 0 = C^J_{e_1}\frac{\delta e_1}{e_1} + C^J_{e_2}\frac{\delta e_2}{e_2}$$

but since $\delta e_1 = -\delta e_2$, it must be the case that

$$0 = C^J_{e_1}\frac{\delta e_1}{e_1} - C^J_{e_2}\frac{\delta e_1}{e_2} \;\Rightarrow\; C^J_{E_1}\frac{1}{e_1} = C^J_{E_2}\frac{1}{e_2}$$

, which finally allowed Burns to state $e_1 : e_2 = C^J_{E_1} : C^J_{E_2}$. Burns also generalised to pathways of arbitrary size by considering pairs of enzymes.

It is interesting to note that Waley in 1964 (Waley, 1964) published a small note on multi-enzyme systems where he used Lagrange multipliers to show that the optimal distribution of enzymes with a fixed total amount was proportional to the square root of kinetic parameters associated with each enzyme. For a system with simple mass-action kinetics, Brown showed that the square terms were a simple ratio of the forward and reverse rate constants. Burns also came to the same conclusion in his thesis and generalised this result by assuming the mass-action rate laws were modified to include the equilibrium constant: $v_i = k_i e_i (S_i - S_{i+1}/K_{eq,i})$. In this situation, the flux through a linear pathway has a simple form (when the last reaction is irreversible and all $K_{eq,i} = 1$)[11]:

$$J = \frac{S_0}{\sum_i \frac{1}{k_i e_i}}.$$

Burns then went on to show that for a linear pathway, the ratio of enzymes at the optimal state was given by:

$$e_1 : e_2 : \ldots e_n = \frac{1}{\sqrt{k_1}} : \frac{1}{\sqrt{k_2}} : \ldots \frac{1}{\sqrt{k_n}}.$$

From the last equation he deduced that

$$e_i = e_T \frac{\frac{1}{\sqrt{k_1}}}{\sum_i \frac{1}{\sqrt{k_i}}}$$

and therefore

$$\frac{1}{k_i e_i} = \frac{\sum_i \frac{1}{\sqrt{k_i}}}{e_T \sqrt{k_i}}$$

such that

$$J = \frac{S_0 e_T}{\left(\sum_i \frac{1}{\sqrt{k_i}}\right)^2} \;\Rightarrow\; \frac{J}{e_T} = \frac{S_0}{\left(\sum_i \frac{1}{\sqrt{k_i}}\right)^2} \approx \mu.$$

Thus, Burns found an optimal expression for enzyme concentrations and for the specific flux of a pathway in terms of its kinetic parameters, given a finite amount of total enzyme. He even argued that when a cell aims to maximise specific flux of a product when it has two biosynthetic routes, by constrained optimisation of enzyme levels, the optimal choice is always to use one of them — akin to our EFM argument in the presence of a single constraint.

Thus, Burns & Kacser, the initiators of MCA and authors of its pioneering work (Kacser and Burns, 1973), together with Heinrich & Rapoport (Heinrich and Rapoport, 1974), were way ahead of their time. It continually appears as if we are still catching up.

[11] In addition to Burns' thesis, you can also find this as a limiting case of the derivations in Heinrich and Rapoport (1974) and Heinrich and Klipp (1996), Klipp and Heinrich (1999).

## References


Bakker, B.M., Michels, P.A., Opperdoes, F.R., Westerhoff, H.V., 1997. Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. J. Biol. Chem. 272 (6), 3207–3215.

Basan, M., Honda, T., Christodoulou, D., Hörl, M., Chang, Y.-F., Leoncini, E., Mukherjee, A., Okano, H., Taylor, B.R., Silverman, J.M., et al., 2020. A universal trade-off between growth and lag in fluctuating environments. Nature 584 (7821), 470–474.

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J.R., Hwa, T., 2015. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature 528 (7580), 99–104.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat. Chem. Biol. 5 (8), 593–599.

Berkhout, J., Bosdriesz, E., Nikerel, E., Molenaar, D., de Ridder, D., Teusink, B., Bruggeman, F.J., 2013. How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. Genetics 194 (2), 505–512.

Berney, M., Weilenmann, H.-U., Ihssen, J., Bassin, C., Egli, T., 2006. Specific growth rate determines the sensitivity of *Escherichia coli* to thermal, UVA, and solar disinfection. Appl. Environ. Microbiol. 72 (4), 2586–2593.

Blüthgen, N., Bruggeman, F.J., Legewie, S., Herzel, H., Westerhoff, H.V., Kholodenko, B.N., 2006. Effects of sequestration on signal transduction cascades. FEBS J. 273 (5), 895–906.

Bosdriesz, E., Molenaar, D., Teusink, B., Bruggeman, F.J., 2015. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. FEBS J. 282 (10), 2029–2044.

Brown, G.C., 1991. Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. J. Theoret. Biol. 153 (2), 195–203.

Bruggeman, F.J., Boogerd, F.C., Westerhoff, H.V., 2005. The multifarious short-term regulation of ammonium assimilation of *Escherichia coli*: dissection using an in silico replica. FEBS J. 272 (8), 1965–1985.

Bruggeman, F.J., Planqué, R., Molenaar, D., Teusink, B., 2020. Searching for principles of microbial physiology. FEMS Microbiol. Rev. 44 (6), 821–844.

Bruggeman, F.J., Teusink, B., Steuer, R., 2023. Trade-offs between the instantaneous growth rate and long-term fitness: Consequences for microbial physiology and predictive computational models. BioEssays 2300015.

Bruggeman, F.J., Westerhoff, H.V., Hoek, J.B., Kholodenko, B.N., 2002. Modular response analysis of cellular regulatory networks. J. Theoret. Biol. 218 (4), 507–520.

Bücher, T., Rüssmann, W., 1963. Gleichgewicht und ungleichgewicht im system der glykolyse. Angew. Chem. 75 (19), 881–893.

Campbell, A., 1957. Synchronization of cell division. Bacteriol. Rev. 21 (4), 263–272.

Cleland, W.W., 1963a. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations. Biochim. Biophys. Acta 67, 104–137.

Cleland, W., 1963b. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: II. Inhibition: Nomenclature and theory. Biochim. Biophys. Acta 67, 173–187.

Cleland, W., 1963c. 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 67, 188–196.

Cornish-Bowden, A., 1976. The effect of natural selection on enzymic catalysis. J. Mol. Biol. 101 (1), 1–9.

Cornish-Bowden, A., 2013a. Fundamentals of Enzyme Kinetics. John Wiley & Sons.

Cornish-Bowden, A., 2013b. The origins of enzyme kinetics. FEBS Lett. 587 (17), 2725–2730.

Cornish-Bowden, A., Hofmeyr, J.-H.S., 2002. The role of stoichiometric analysis in studies of metabolism: an example. J. Theoret. Biol. 216 (2), 179–191.

de Groot, D.H., Hulshof, J., Teusink, B., Bruggeman, F.J., Planqué, R., 2020. Elementary Growth Modes provide a molecular description of cellular self-fabrication. PLoS Comput. Biol. 16 (1), e1007559.

De Groot, D.H., Lischke, J., Muolo, R., Planqué, R., Bruggeman, F.J., Teusink, B., 2020. The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. Cell. Mol. Life Sci. 77, 441–453.

De Groot, D.H., Van Boxtel, C., Planqué, R., Bruggeman, F.J., Teusink, B., 2019. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. PLoS Comput. Biol. 15 (3), e1006858.

Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436 (7050), 588–592.

Egli, T., 2015. Microbial growth and physiology: a call for better craftsmanship. Front. Microbiol. 6, 287.

Elsemman, I.E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S.W., van Heerden, J., Bruggeman, F.J., Bisschops, M.M., Sonnenschein, N., et al., 2022. Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. Nature Commun. 13 (1), 801.

Fell, D., 1997. Understanding the Control of Metabolism. Vol. 2, Portland Press London.

Flint, H.J., Tateson, R.W., Barthelmess, I.B., Porteous, D., Donachie, W.D., Kacser, H., 1981. Control of the flux in the arginine pathway of *Neurospora crassa*. Modulations of enzyme activity and concentration. Biochem. J. 200 (2), 231–246.

Gagneur, J., Klamt, S., 2004. Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics 5 (1), 1–21.
Giersch, C., 1988a. Control analysis of metabolic networks: 1. Homogeneous functions and the summation theorems for control coefficients. Eur. J. Biochem. 174 (3), 509–513.
Giersch, C., 1988b. Control analysis of metabolic networks: 2. Total differentials and general formulation of the connectivity relations. Eur. J. Biochem. 174 (3), 515–519.
Goldbeter, A., 2018. Dissipative structures in biological systems: bistability, oscillations, spatial patterns and waves. Phil. Trans. R. Soc. A 376 (2124), 20170376.
Grigaitis, P., Teusink, B., 2022. An excess of glycolytic enzymes under glucose-limited conditions may enable *Saccharomyces cerevisiae* to adapt to nutrient availability. FEBS Lett. 596 (24), 3203–3210.
Groen, A.K., Wanders, R., Westerhoff, H., Van der Meer, R., Tager, J., 1982. Quantification of the contribution of various steps to the control of mitochondrial respiration.. J. Biol. Chem. 257 (6), 2754–2757.
Heinrich, R., Klipp, E., 1996. Control analysis of unbranched enzymatic chains in states of maximal activity. J. Theoret. Biol. 182 (3), 243–252.
Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. Eur. J. Biochem. 42 (1), 89–95.
Heinrich, R., Rapoport, S., Rapoport, T., 1978. Metabolic regulation and mathematical models. Prog. Biophys. Mol. Biol. 1–82.
Heinrich, R., Schuster, S., 2012. The Regulation of Cellular Systems. Springer Science & Business Media.
Hill, T., 1977. Free Energy Transduction in Biology: The Steady-State Kinetic and Thermodynamic Formalism. Elsevier.
Hofmeyr, J.-H.S., Cornish-Bowden, H., 1997. The reversible hill equation: how to incorporate cooperative enzymes into metabolic models. Bioinformatics 13 (4), 377–385.
Hofmeyr, J.-H.S., Cornish-Bowden, A., 2000. Regulating the cellular economy of supply and demand. FEBS Lett. 476 (1–2), 47–51.
Hofmeyr, J.-H.S., Kacser, H., van der Merwe, K.J., 1986. Metabolic control analysis of moiety-conserved cycles. Eur. J. Biochem. 155 (3), 631–640.
Hofmeyr, J.-H.S., Westerhoff, H.V., 2001. Building the cellular puzzle: control in multi-level reaction networks. J. Theoret. Biol. 208 (3), 261–285.
Hornberg, J.J., Binder, B., Bruggeman, F.J., Schoeberl, B., Heinrich, R., Westerhoff, H.V., 2005. Control of MAPK signalling: from complexity to what really matters. Oncogene 24 (36), 5533–5542.
Hui, S., Silverman, J.M., Chen, S.S., Erickson, D.W., Basan, M., Wang, J., Hwa, T., Williamson, J.R., 2015. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. Mol. Syst. Biol. 11 (2), 784.
Ihssen, J., Egli, T., 2004. Specific growth rate and not cell density controls the general stress response in Escherichia coli. Microbiology 150 (6), 1637–1648.
Ihssen, J., Egli, T., 2005. 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. 7 (10), 1568–1581.
Jensen, P.R., Michelsen, O., Westerhoff, H.V., 1993a. Control analysis of the dependence of Escherichia coli physiology on the H (+)-ATPase. Proc. Natl. Acad. Sci. 90 (17), 8068–8072.
Jensen, P.R., Michelsen, O., Westerhoff, H.V., 1995. Experimental determination of control by the H+-ATPase in Escherichia coli. J. Bioenerg. Biomembr. 27 (6), 543–554. http://dx.doi.org/10.1007/BF02111653.
Jensen, P.R., Westerhoff, H.V., Michelsen, O., 1993b. Excess capacity of H (+)-ATPase and inverse respiratory control in *Escherichia coli*. EMBO J. 12 (4), 1277–1282.
Jensen, P.R., Westerhoff, H.V., Michelsen, O., 1993c. The use of lac-type promoters in control analysis. Eur. J. Biochem. 211 (1–2), 181–191.
Kacser, H., 1986. On parts and wholes in metabolism. In: Welch, G.R., Clegg, J.S. (Eds.), The Organization of Cell Metabolism. Springer, pp. 327–337.
Kacser, H., Acerenza, L., 1993. A universal method for achieving increases in metabolite production. Eur. J. Biochem. 216 (2), 361–367.
Kacser, H., Burns, J., 1973. The control of flux. Symp. Soc. Exp. Biol. 27, 65–104.
Kacser, H., Burns, J., 1979. Molecular democracy: who shares the controls? Biochem. Soc. Trans. 7 (5), 1149–1160.
Kahn, D., Westerhoff, H.V., 1991. Control theory of regulatory cascades. J. Theoret. Biol. 153 (2), 255–285.
Kelk, S.M., Olivier, B.G., Stougie, L., Bruggeman, F.J., 2012. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. Sci. Rep. 2 (1), 1–7.
Keren, L., Hausser, J., Lotan-Pompan, M., Slutskin, I.V., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., Segal, E., 2016. Massively parallel interrogation of the effects of gene expression levels on fitness. Cell 166 (5), 1282–1294.
Kholodenko, B.N., Hoek, J.B., Westerhoff, H.V., Brown, G.C., 1997. Quantification of information transfer via cellular signal transduction pathways. FEBS Lett. 414 (2), 430–434.
Kholodenko, B.N., Westerhoff, H.V., 1993. Metabolic channelling and control of the flux. FEBS Lett. 320 (1), 71–74.
King, Z.A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., Palsson, B.Ø., 2015. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLoS Comput. Biol. 11 (8), e1004321.
Kitano, H., 2002. Systems biology: a brief overview. Science 295 (5560), 1662–1664.
Klipp, E., Heinrich, R., 1999. Competition for enzymes in metabolic pathways:: Implications for optimal distributions of enzyme concentrations and for the distribution of flux control. Biosystems 54 (1–2), 1–14.
Kochanowski, K., Okano, H., Patsalo, V., Williamson, J., Sauer, U., Hwa, T., 2021. Global coordination of metabolic pathways in *Escherichia coli* by active and passive regulation. Mol. Syst. Biol. 17 (4), e10064.
Koshland Jr., D.E., Némethy, G., Filmer, D., 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5 (1), 365–385.
Krebs, H., 1957. Control of metabolic processes. Endeavour 16 (63), 125–132.
Kuenen, G., 2019. Continuous cultures (chemostats). In: Encyclopedia of Microbiology. Elsevier, pp. 743–761.
Lu, H., Li, F., Sánchez, B.J., Zhu, Z., Li, G., Domenzain, I., Marcišauskas, S., Anton, P.M., Lappa, D., Lieven, C., et al., 2019. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. Nature Commun. 10 (1), 3586.
Maaløe, O., 2012. An analysis of bacterial growth. Commun. Dev. 3, 33–58.
Maaløe, O., Kjeldgaard, N.O., 1966. Control of Macromolecular Synthesis: A Study of DNA, RNA, and Protein Synthesis in Bacteria. In: Microbial and molecular biology series, W.A. Benjamin.
Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5 (4), 264–276.
Mashego, M.R., Jansen, M.L., Vinke, J.L., van Gulik, W.M., Heijnen, J.J., 2005. Changes in the metabolome of saccharomyces cerevisiae associated with evolution in aerobic glucose-limited chemostats. FEMS Yeast Res. 5 (4–5), 419–430.
Molenaar, D., Van Berlo, R., De Ridder, D., Teusink, B., 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol. Syst. Biol. 5 (1), 323.
Monod, J., 1974. On Chance and Necessity. Springer Verlag.
Monod, J., Changeux, J.-P., Jacob, F., 1963. Allosteric proteins and cellular control systems. J. Mol. Biol. 6 (4), 306–329.
Monod, J., Wyman, J., Changeux, J., 1965. On the nature of allosteric transitions: A plausible model. J. Mol. Biol. 12 (1), 88–118.
Mori, M., Schink, S., Erickson, D.W., Gerland, U., Hwa, T., 2017. Quantifying the benefit of a proteome reserve in fluctuating environments. Nature Commun. 8 (1), 1225.
Mori, M., Zhang, Z., Banaei-Esfahani, A., Lalanne, J.-B., Okano, H., Collins, B.C., Schmidt, A., Schubert, O.T., Lee, D.-S., Li, G.-W., et al., 2021. From coarse to fine: the absolute *Escherichia coli* proteome under diverse growth conditions. Mol. Syst. Biol. 17 (5), e9536.
Müller, S., Regensburger, G., Steuer, R., 2014. Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. J. Theoret. Biol. 347, 182–190.
Newsholme, E.A., 1980. Reflections on the mechanism of action of hormones. FEBS Lett. 117, K121–K134.
Newsholme, P., 2009. Mapping life's reactions: a brief history of metabolic pathways and their regulation. Biochem. Soc. June, 4–7.
Noor, E., Flamholz, A., Bar-Even, A., Davidi, D., Milo, R., Liebermeister, W., 2016. The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. PLoS Comput. Biol. 12 (11), e1005167.
Nordholt, N., van Heerden, J.H., Bruggeman, F.J., 2020. Biphasic cell-size and growth-rate homeostasis by single *Bacillus subtilis* cells. Curr. Biol. 30 (12), 2238–2247.
Nyström, T., 2004. MicroReview: Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? Mol. Microbiol. 54 (4), 855–862.
O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9 (1), 693.
O'Brien, E.J., Utrilla, J., Palsson, B.Ø., 2016. Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. PLoS Comput. Biol. 12 (6), e1004998.
Oldewurtel, E.R., Kitahara, Y., van Teeffelen, S., 2021. Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. Proc. Natl. Acad. Sci. 118 (32), e2021416118.
Orr, H.A., 2009. Fitness and its role in evolutionary genetics. Nature Rev. Genet. 10 (8), 531–539.
Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.Ø., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Mol. Syst. Biol. 7 (1), 535.
Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nature Biotechnol. 28 (3), 245–248.
Papagiannakis, A., Niebel, B., Wit, E.C., Heinemann, M., 2017. Autonomous metabolic oscillations robustly gate the early and late cell cycle. Mol. Cell 65 (2), 285–295.
Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., Palsson, B.Ø., 2004. Comparison of network-based pathway analysis methods. Trends Biotechnol. 22 (8), 400–405.
Pirt, S.J., 1982. Maintenance energy: a general model for energy-limited and energy-sufficient growth. Arch. Microbiol. 133, 300–302.
Planqué, R., Hulshof, J., Teusink, B., Hendriks, J.C., Bruggeman, F.J., 2018. Maintaining maximal metabolic flux by gene expression control. PLoS Comput. Biol. 14 (9), e1006412.

Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2 (11), 886–897.
Pugatch, R., 2015. Greedy scheduling of cellular self-replication leads to optimal doubling times with a log-frechet distribution. Proc. Natl. Acad. Sci. 112 (8), 2611–2616.
Rabbers, I., Bruggeman, F.J., 2022. *Escherichia coli* robustly expresses ATP synthase at growth rate-maximizing concentrations. FEBS J. 289 (16), 4925–4934.
Reder, C., 1988. Metabolic control theory: a structural approach. J. Theoret. Biol. 135 (2), 175–201.
Rohwer, J.M., Meadow, N.D., Roseman, S., Westerhoff, H.V., Postma, P.W., 2000. Understanding glucose transport by the bacterial phosphoenolpyruvate: glycose phosphotransferase system on the basis of kinetic measurements in vitro. J. Biol. Chem. 275 (45), 34909–34921.
Sauro, H.M., Kacser, H., 1990. Enzyme-enzyme interactions and control analysis: 2. The case of non-independence: heterologous associations. Eur. J. Biochem. 187 (3), 493–500.
Savageau, M.A., 1969a. Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions. J. Theoret. Biol. 25 (3), 365–369.
Savageau, M.A., 1969b. Biochemical systems analysis: II. The steady-state solutions for an n-pool system using a power-law approximation. J. Theoret. Biol. 25 (3), 370–379.
Savageau, M.A., 1970. Biochemical systems analysis: III. Dynamic solutions using a power-law approximation. J. Theoret. Biol. 26 (2), 215–226.
Schaechter, M., 2006. From growth physiology to systems biology. Int. Microbiol. 9 (3), 157.
Schaechter, M., Maaløe, O., Kjeldgaard, N.O., 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. Microbiology 19 (3), 592–606.
Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnol. 18 (3), 326–332.
Schuster, S., Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Systems 2 (02), 165–182.
Schuster, S., Kahn, D., Westerhoff, H.V., 1993. Modular analysis of the control of complex metabolic pathways. Biophys. Chem. 48 (1), 1–17.
Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., Hwa, T., 2010. Interdependence of cell growth and gene expression: origins and consequences. Science 330 (6007), 1099–1102.
Scott, M., Hwa, T., 2022. Shaping bacterial gene expression by physiological and proteome allocation constraints. Nat. Rev. Microbiol. 1–16.
Scott, M., Klumpp, S., Mateescu, E.M., Hwa, T., 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Mol. Syst. Biol. 10 (8), 747.
Segel, I.H., 1975. Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems. Wiley New York.
Snoep, J.L., Yomano, L.P., Westerhoff, H.V., Ingram, L.O., 1995. Protein burden in Zymomonas mobilis: negative flux and growth control due to overproduction of glycolytic enzymes. Microbiology 141 (9), 2329–2337.
Solem, C., Koebmann, B.J., Jensen, P.R., 2003. Glyceraldehyde-3-phosphate dehydrogenase has no control over glycolytic flux in Lactococcus lactis MG1363. J. Bacteriol. 185 (5), 1564–1571.
Solem, C., Koebmann, B., Jensen, P.R., 2008. Control analysis of the role of triosephosphate isomerase in glucose metabolism in *Lactococcus lactis*. IET Syst. Biol. 2 (2), 64–72.
Solem, C., Petranovic, D., Koebmann, B., Mijakovic, I., Jensen, P.R., 2010. Phosphoglycerate mutase is a highly efficient enzyme without flux control in *Lactococcus lactis*. Microb. Physiol. 18 (3), 174–180.
ter Kuile, B.H., Westerhoff, H.V., 2001. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. FEBS Lett. 500 (3), 169–171.
Teusink, B., Passarge, J., Reijenga, C.A., Esgalhado, E., Van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., Van Dam, K., Westerhoff, H.V., et al., 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur. J. Biochem. 267 (17), 5313–5329.
Thomas, S., Fell, D.A., 2000. Multisite modulation in the control of glycolysis: Balance of supply and demand? In: Technological and Medical Implications of Metabolic Control Analysis. Springer, pp. 259–266.
Umbarger, H.E., 1956. Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. Science 123 (3202), 848.
Utrilla, J., O'Brien, E.J., Chen, K., McCloskey, D., Cheung, J., Wang, H., Armenta-Medina, D., Feist, A.M., Palsson, B.Ø., 2016. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. Cell Syst. 2 (4), 260–271.
van der Vlag, J., van Dam, K., Postma, P.W., 1994. Quantification of the regulation of glycerol and maltose metabolism by IIAGlc of the phosphoenolpyruvate-dependent glucose phosphotransferase system in *Salmonella typhimurium*. J. Bacteriol. 176 (12), 3518–3526. http://dx.doi.org/10.1128/jb.176.12.3518-3526.1994.
van Heerden, J.H., Berkvens, A.M., de Groot, D.H., Bruggeman, F.J., 2023. Growth consequences of the inhomogeneous organization of the bacterial cytoplasm. bioRxiv http://dx.doi.org/10.1101/2023.04.18.537336.
Van Hoek, P., Van Dijken, J.P., Pronk, J.T., 1998. Effect of specific growth rate on fermentative capacity of baker's yeast. Appl. Environ. Microbiol. 64 (11), 4226–4233.
Varma, A., Boesch, B.W., Palsson, B.Ø., 1993. Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl. Environ. Microbiol. 59 (8), 2465–2473.
Varma, A., Palsson, B.Ø., 1993a. Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. J. Theoret. Biol. 165 (4), 477–502.
Varma, A., Palsson, B.Ø., 1993b. Metabolic capabilities of *Escherichia coli* II. Optimal growth patterns. J. Theoret. Biol. 165 (4), 503–522.
Waley, S., 1964. A note on the kinetics of multi-enzyme systems. Biochem. J. 91 (3), 514.
Walsh, K., Koshland Jr., D.E., 1985. Characterization of rate-controlling steps in vivo by use of an adjustable expression vector. Proc. Natl. Acad. Sci. 82 (11), 3577–3581.
Westerhoff, H.V., Groen, A.K., Wanders, R.J., 1984. Modern theories of metabolic control and their applications. Biosci. Rep. 4 (1), 1–22.
Westerhoff, H.V., Palsson, B.Ø., 2004. The evolution of molecular biology into systems biology. Nature Biotechnol. 22 (10), 1249–1252.
Westerhoff, H.V., Van Dam, K., 1987. Thermodynamics and Control of Biological Free-Energy Transduction. Elsevier Science Limited.
Wortel, M.T., Peters, H., Hulshof, J., Teusink, B., Bruggeman, F.J., 2014. Metabolic states with maximal specific rate carry flux through an elementary flux mode. FEBS J. 281 (6), 1547–1555.
Wu, C., Balakrishnan, R., Braniff, N., Mori, M., Manzanarez, G., Zhang, Z., Hwa, T., 2022. Cellular perception of growth rate and the mechanistic origin of bacterial growth law. Proc. Natl. Acad. Sci. 119 (20), e2201585119.
Young, R., Bremer, H., 1976. Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. Biochem. J. 160 (2), 185–194.