bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Optimal metabolic states in cells

Wolfram Liebermeister[1,2]
[1] INRA, UR1404, MaIAGE, Université Paris-Saclay, France
[2] Institut für Biochemie, Charité – Universitätsmedizin Berlin, Germany

Abstract

Cells need to make an efficient use of metabolites, proteins, energy, membrane space, and time, and resource allocation is also an important aspect of metabolism. How, for example, should cells distribute their protein budget between different cellular functions, e.g. different metabolic pathways, to maximise growth? Cellular resource allocation can be studied by combining biochemical network models with optimality problems that choose metabolic states by their cost and benefit. Various types of resource allocation problems have been proposed. The underlying mechanistic models may describe different cellular systems (e.g. metabolic pathways, networks, or compromises between metabolism and protein production) on different level of detail and using different mathematical formulations (e.g. stoichiometric or kinetic). The optimality problems may use metabolite levels, enzyme levels, or fluxes as variables, assume different cost or benefit functions, and describe different kinds of trade-offs, in which cell variables are either constrained or treated as optimisation objectives. Due to all these differences, optimality problems may be hard to compare or combine. To bring them under one umbrella, I show that they can be derived from a common framework, and that their optimality conditions all show the same mathematical form. This unified view on metabolic optimality problems can be used to justify and combine various modelling approaches and biochemical optimality problems.

Keywords: Metabolic model, resource allocation, state space, optimality condition

1 Introduction

Living cells cannot be understood in their full complexity, and because we cannot do so, we need to describe them by simplified pictures. These pictures – in the simplest case, network graphics including reaction stoichiometries, kinetic rate laws, etc. – may be turned into mathematical models, and these models, nevertheless, can be complex, combining biochemistry, network structures, chemical dynamics, regulation, genetics, and even evolution. Most of them portray cells as chemical reaction systems (as if cells were test tubes) or as information-processing devices (as if cells were computers), thus focusing on mechanisms, dynamics, and regulation. However, cells are believed to also make efficient use of nutrients, proteins, energy, intracellular space, and opportunities provided by their ecological niches, or choosing good compromises between all these needs. This idea of an economical behaviour, or optimal resource allocation, appears in many topical questions in cell biology. For example, what determines the advantages of different pathways, e.g. pathways performing fermentation and respiration for ATP regeneration, or the preferences for specific carbon sources? How does cell growth depend on metabolism and protein production, and how should these two be orchestrated? Such questions are not about simple choices between a few options, but about complicated dependencies between thousands of cell variables. For example, given that an ATP-producing pathway relies on enzymes and that enzyme production itself requires ATP, how should the pathway flux respond to varying ATP demands or to a forced expression of protein somewhere else in the cell?

To address these questions, we may compare a cell to a factory whose production processes are adapted to the market prices of materials, energy, and the goods produced, and metabolic pathways would correspond to

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

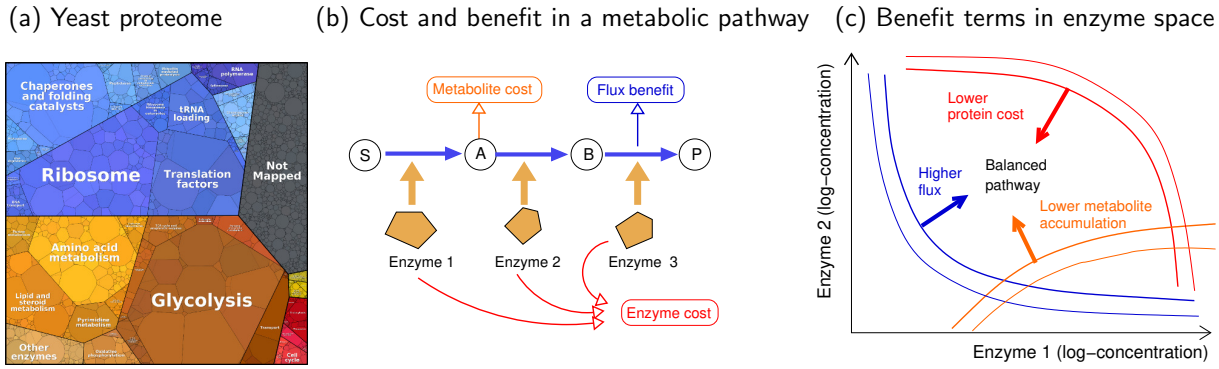


Figure 1: Protein profiles in cells result from a compromise. (a) Measured protein abundances in the budding yeast *S. cerevisiae*, shown as a proteomap [1] (data from [2]). Proteins are represented by polygons (where sizes represent abundance) and arranged by cellular functions (larger areas). (b) Cost and benefit of enzymes in a simple metabolic pathway. The enzyme levels (scored by a cost) determine the metabolite concentrations and fluxes in steady state (scored by cost and benefit) as described by a kinetic model. (c) Space of enzyme profiles, with cost and benefit functions shown as hypothetical benefits (redrawn from [3], where a similar model has been analysed). Enzyme profiles that allow for a high metabolic flux (blue lines) at low metabolite levels (brown lines) and enzyme levels (red lines) provide a selection advantage. If all three terms are described as benefits ("flux", "negative intermediate accumulation", and "negative protein burden"), the sum of benefits is maximised exactly in a point where the three benefit gradients cancel out ("balanced pathway"). While the gradient of the (negative) protein cost is a simple function in enzyme space (assuming that protein cost itself is a simple function), the other two gradients depend on enzyme levels in an indirect way, as described by the kinetic pathway model.

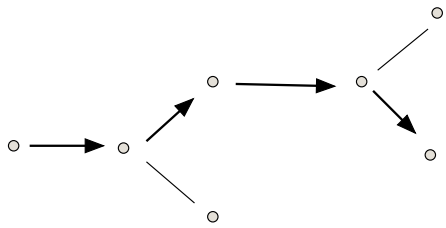
commodity or value chains. The resulting models do not only describe metabolic behaviour, but also the ensuing profits, i.e. growth advantages or the corresponding selection pressures. This is a third level of description beyond metabolic networks and computational models describing the biochemical dynamics, a functional description that states how processes or compounds are used to create benefits. In such a description, we consider a mechanistic cell model, postulate biological objectives, and formulate an optimality problem in which some model variables (e.g. enzyme levels) are not described mechanistically (e.g. by assuming a biochemical model of enzyme production and degradation), but are treated as controllable parameters: we need to *choose* their values in order to optimise some objective function. These types of problems are the topic of this article. On top of the levels of "topics" (describing the "anatomy" of cellular networks) and "dynamics" (describing physiological compound concentrations and reaction rates), these problems introduce a third level of biological function, which may be called "cellular economics". To model resource allocation in cells, mechanistic and economic modelling need to go hand in hand. Here, I focus on such optimality problems, describing how cells *should behave* to achieve some objective, e.g. to maximise their growth rate and how this is reflected in the usage of individual pathways.

Before we model the cell as a whole, let us focus on metabolism. Metabolic performance is greatly enhanced by enzymes, but their production, maintenance, and space requirements put a burden on cells, and the thrust to keep this burden low has a major impact on metabolic states. Therefore, models should describe not only the dynamics, but also the optimal allocation of protein and metabolite levels. For example, Figure 1 (a) shows an overview of protein investments in yeast: almost half of the protein budget is devoted to metabolic enzymes, and a large fraction of them to glycolysis. This raises various questions. How can this pattern be explained? How does it depend on metabolic network structure, enzyme kinetics, enzyme sizes, and metabolic demands? And how can we predict the investment in individual enzymes? For example, some enzymes contain trace elements such as metal atoms; if a metal such as iron is rare, how will this affect the usage of iron-containing and iron-converting proteins, and how will these changes further affect the entire proteome? How should enzyme levels be chosen to maximise cell fitness, and what metabolic states will emerge from these choices? How can optimal enzyme investments be predicted from network structure, enzyme kinetics, and thermodynamics? And how can we use

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

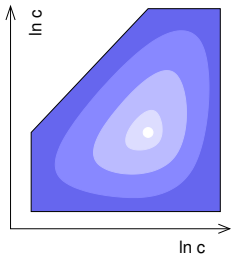
Box 1: Optimality-based metabolic models

Flux balance analysis



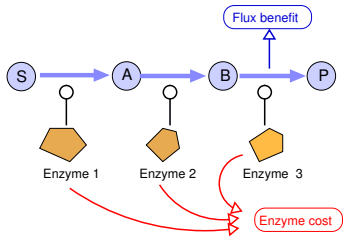
Constraint-based models (CBM) for Flux Balance Analysis (FBA) predict metabolic fluxes from network structure, stationarity, and optimality criteria such as high metabolic production or low enzyme demand. Given cost and benefit functions for fluxes, the aim is to find flux distributions with a maximal benefit under constraints [4], a maximal benefit at a given cost [5], or a minimal cost at a given benefit [6]. Constraint-based models do not describe enzyme kinetics: enzyme levels are either ignored or treated by simplified rules, e.g. assuming that enzyme levels and fluxes are directly proportional. Metabolite concentrations can appear as variables, but usually only to formulate thermodynamic constraints on the fluxes (flux directions depend on the mass-action ratio of reactants).

Cost minimisation in metabolite space



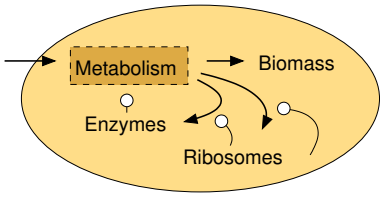
Given a desired metabolic flux distribution, we may search for metabolite and enzyme levels that realise these fluxes at a minimal biological cost. Such enzyme and metabolite levels follow from an optimality problem in metabolite space [11], called enzyme cost minimisation (ECM): we consider a kinetic model, choose a flux distribution, and search for enzyme and metabolite levels that realise these fluxes at a minimal cost. The cost function may include a direct metabolite cost and a cost for the enzyme levels required. Given metabolite levels and fluxes, the enzyme levels are obtained by a simple formula, and the resulting optimisation is a convex optimality problem in log-metabolite concentration space.

Kinetic pathway model



Kinetic models describe the dynamics of compound concentrations and can be used to model metabolism, signalling, gene expression, or the allocation of protein resources to metabolism or ribosomes. Enzyme levels, fluxes, and metabolite concentrations are linked by rate laws. Given all enzyme levels and initial conditions, a metabolic state can be computed by numerical integration. To find optimal metabolic states, the enzyme levels may be optimised[1], e.g. to provide large production fluxes at low total enzyme amounts (see Figure 1) [9, 10]. To model metabolism in growing cells, the dilution of compounds can be taken into account. The growth rate can be treated as a given parameter, as a dynamic variable, or as a variable to be optimised.

Model of growing cells



In microbial whole-cell models, a typical objective is to maximise growth while keeping all cell compounds at physiological levels. To compensate for their constant dilution, the cell needs to reproduce all metabolites, enzymes, ribosomes, and so forth. Whole-cell models can be constraint-based or kinetic. Resource balance analysis [12, 13] is a constraint-based method that covers metabolism, enzyme and ribosome production, and processes such as protein folding by chaperones. Fluxes and enzyme levels are treated as strictly proportional. The aim is to obtain a steady state at a fixed growth rate, and to determine the maximal dilution rate at which this condition can be met. Kinetic whole-cell models are usually much smaller. For example, the aim may be to maximise growth by optimising the production of different sorts of proteins, for protein production (catalysed by ribosomes) or metabolism (catalysed by metabolic enzymes), [14], the kinetic constants that determine the allocation of protein may be optimised.

such a reasoning in practice, to predict fluxes or to engineer metabolic pathways?

Protein allocation has been studied by different types of models, including constraint-based models, kinetic models, and whole-cell models (see Box 1). These models describe metabolic pathways or networks, consisting of compounds, chemical reactions between them, and enzymes that catalyse the reactions. In network models, chemical conversion is described as a "flow of matter": for example, glucose molecules enter the cell and be converted, in various steps, into carbon dioxide; in this process, carbon atoms "flow" through the reaction network. The metabolite concentrations can change in time: if a metabolite is faster produced than consumed, its

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Enzymes	Metabolites	Fluxes	Growth
Maximise benefit (at given cost)	[15]		[5]	[14, 13]
Minimise cost (at given benefit)	[11]	[11]	[6, 16]	
Optimise weighted difference	[9]	[11] (with benefit term)		
Multi-objective optimisation			[17]	

Table 1: Metabolic trade-offs can be modelled in different ways. The control variables (to be varied in optimality problems) can be either enzyme levels, metabolite levels, or fluxes. The metabolic states can be scored by different cost and benefit functions, and trade-offs can be modelled by maximimising a benefit-cost difference, by maximising a benefit at a fixed cost, by minimising a cost at a fixed benefit, or by multi-objective optimisation.

concentration increases; if consumption is faster, its concentration decreases; and if production and consumption are balanced, its concentration remains constant. A steady state is a state in which all internal concentrations remain constant, and to guarantee this, all fluxes have to be balanced. Network models can be coarse-grained or fine-grained and describe smaller or larger parts of the cellular systems. They can cover various of processes, including metabolism, cell signalling, gene expression, protein and mRNA synthesis, and even an entire cell. They can describe *possible* cell states (e.g. constraint-based flux models), *actual* dynamics (e.g. kinetic metabolic models), or *desired* behaviour (e.g. optimality-based models). There are two main types of biochemical network models: kinetic models, which describe the joint dynamics of metabolite concentrations, enzyme concentrations, and fluxes; and flux analysis (or "constraint-based" models (CBM)), which assume stationary flux distributions, but ignore their relation to metabolite or enzyme levels. These different modelling approaches, or "paradigms", make different assumptions, but share some of their equations, e.g. the conditions for steady states.

Usually, economical behaviour in cells implies compromises between opposing needs, e.g. between the costs and benefits of different biochemical processes. The aim is to find a global solution in which all subsystems of the cell interplay in an optimal manner. Mathematically, the search for such states can be formulated as an optimality problem: we model a biochemical system mechanistically, assume various physical and physiological constraints, and optimise objectives such as substrate or enzyme efficiency, use of resources, or maximal cell growth. An example is shown in Figure 1 (b): in a small metabolic pathway, the enzyme levels need to optimise a trade-off between three objectives: a benefit for high flux, a cost for high enzyme levels, and a cost for unfavourable (high or low) metabolite levels. What will the optimal enzyme levels look like?

Metabolic optimality problems can be categorised by several criteria (Table 1). First, by the choice of cell variables, i.e. whether enzyme levels, metabolite levels, and fluxes are treated as parameters, free variables, or dependent variables. Second, by stating how the cell variables are scored by metabolic objectives. And third, by how trade-offs between these objectives are formulated, i.e. which objectives are bounded by constraints, and how they contribute to the overall objective – or whether multi-objective optimisation is applied. Aside from the previous aspects, metabolic optimality problems differ in various other ways. Some models are kinetic, others are constraint-based; some describe metabolism only, others include protein production and other cell processes; some use metabolic objectives (e.g., maximising a pathway flux), others use cell objectives (e.g. maximising cell growth). With all these differences, different models may be hard to compare and combine, and so a unified framework would be desirable. To obtain such a framework, we might ask what is common to the different modelling approaches. First, all these approaches rely on metabolic network models. Second, different models may even describe the same pathways and may use the same biochemical parameter values. Third, despite their different formulations, different models may make exactly the same predictions (e.g. optimisation of enzyme levels, or optimisation of metabolite levels at given optimal fluxes). If different optimality problems, for the same mechanistic cell model, make the same assumptions (e.g. that high enzyme levels are costly), either explicitly (by a penalty term for large enzyme levels) or implicitly (because enzyme production, within the model, consumes resources), then can we assume that they will also yield the same results? And if so, what are the reasons for this equivalence, and where are its limits?

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here I consider different optimality problems for metabolism, and will show how to describe them by a unified framework. Starting from simple optimality problems, as in Figure 1, how to characterise optimal states in general, especially for large metabolic networks. If we reformulate this problem with fluxes or metabolite levels (instead of enzyme levels) as the basic variables, how will these different problems, their optimality conditions, and their solutions be related? And how can different optimality approaches (with their ad hoc assumptions and approximations) be reconciled, and possibly be combined? I describe these different approaches as special cases of a single optimality problem, an optimisation on the set of feasible metabolic states, where each state is characterised by fluxes, metabolite levels, and enzyme levels. These variables can be scored by different metabolic objectives, and there can be trade-offs between them. No matter how we formulate the trade-offs, the resulting optimality problems will share optimality conditions of the same form. The shared optimality conditions allow us to recognise the optimality problems as equivalent.

2 Metabolic optimality problems

Mathematical cell models can rely on different formalisms and describe different sorts of variables. Here, we consider models describung metabolic pathways or networks by their fluxes, metabolite concentrations, and enzyme concentrations (in vectors $\mathbf{v}$, $\mathbf{c}$, and $\mathbf{e}$). The variables are related by kinetic rate laws, and to keep things simple, we only consider steady states and disregard dilution by cell growth (although this could be added if necessary). As a general frame, this covers various kinetic and stoichiometric modelling approaches. To define an optimality problem, we assume that a pathway contributes to cell fitness in three ways: by a flux benefit $b(\mathbf{v})$, a metabolite cost $q(\mathbf{c})$, and an enzyme cost $h(\mathbf{e})$, and that the cells maximises a fitness[2]

$$f(\mathbf{v}, \mathbf{c}, \mathbf{e}) = b(\mathbf{v}) - q(\mathbf{c}) - h(\mathbf{e}). \tag{1}$$

The three terms in this fitness will be called metabolic objectives (in contrast to the "mathematical objectives" described below), and their usage can be justfied by deriving them from whole-cell models with maximisation of growth or other whole-cell objectives. Whether fitness terms are described as benefits (to be maximised) or as costs (to be minimised) is a matter of convention. We may also split an objective into sub-objectives, e.g. separate several cost or benefit terms that score the production or consumption of different compounds, and apply the same mathematical formalism. Maximising the fitness (1) is a way to model trade-offs between our metabolic objectives. The objectives in the formula may also be weighted by prefactors. Other ways of describing trade-offs, including constraints on some of the metabolic objectives, or multi-objective optimisation, will be discussed below. Our fitness function (1) tells us which metabolic states (comprising fluxes, metabolite levels, and enzyme levels) are desirable, but not which states are actually possible (given that state variables depend on each other physically). To describe this, we need a kinetic model that defines a set of feasible states.

A kinetic model describes a flux vector $\mathbf{v}$, a metabolite profile $\mathbf{c}$, and an enzyme profile $\mathbf{e}$, and links them by rate equations and enzymatic rate laws:

$$\frac{\mathrm{d}\mathbf{c}}{\mathrm{d}t} = \mathbf{N}_{\mathrm{int}}\,\mathbf{v}, \qquad \mathbf{v} = \mathbf{v}(\mathbf{c}, \mathbf{e}). \tag{2}$$

Metabolites with fixed concentrations ("*external* metabolites") appear in the rate laws $v_l(\cdot)$ as parameters. Some of the state variables may be restricted by physiological bounds (e.g. minimal and maximal metabolite levels, or an upper bound on their sum). In stationary states (also called steady states), metabolite levels and fluxes are

[2]Assuming separate benefit and cost terms for fluxes and concentrations is a matter of convenience. The approach works for all fitness functions $f(\mathbf{v}, \mathbf{c}, \mathbf{e})$ that are reasonably well behaved and increase monotonically in the enzyme levels.

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

constant in time, and so the vectors $\mathbf{v}$ and $\mathbf{c}$ must satisfy

$$\mathbf{N}_{\text{int}}\,\mathbf{v} = 0, \qquad \mathbf{v} = \mathbf{v}(\mathbf{c}, \mathbf{e}). \tag{3}$$

Two important features of metabolic models – conserved moieties and stationary fluxes – can be inferred from the stoichiometric matrix. The internal stoichiometric matrix $\mathbf{N}_{\text{int}}$ has a left-kernel matrix $\mathbf{G}$ and a right-kernel matrix $\mathbf{K}$, satisfying, respectively, $\mathbf{G}\,\mathbf{N}_{\text{int}} = 0$ and $\mathbf{N}_{\text{int}}\,\mathbf{K} = 0$ (these matrices may also be empty). The rows of $\mathbf{G}$ (and their linear combinations) describe linear conservation relations: for any column $\mathbf{g}$ of $\mathbf{G}^\top$, the product $\mathbf{g} \cdot \mathbf{c}$ will be constant in time. Given the initial values of the conserved moiety concentrations, $\mathbf{c}_{\text{cm}}$, this implies the constraint

$$\mathbf{G}\,\mathbf{c} = \mathbf{c}_{\text{cm}} = \text{const}. \tag{4}$$

Each row describes a conserved moiety (e.g. the total number of phosphate groups in the system, assuming that phosphate groups are only transferred between molecules, and cannot enter or leave the system). In models with moiety conservation (i.e. with an non-empty matrix $\mathbf{G}$), $\mathbf{N}_{\text{int}}$ can be split into a product $\mathbf{N}_{\text{int}} = \mathbf{L}\,\mathbf{N}_{\text{ind}}$ where the rows of $\mathbf{N}_{\text{ind}}$ refer to independent internal metabolites. The link matrix Ł relates these independent metabolites to the set of all metabolites. The columns of $\mathbf{K}$ (and their linear combinations) describe possible stationary fluxes.

If we model only the fluxes, assume that the flux distributions must be stationary (such flux distributions will be called "metabolic flows"), and put bounds on the fluxes (e.g., maximal fluxes or predefined flux directions), we obtain the constraints of Flux Balance Analysis (FBA) [18],

$$\mathbf{N}_{\text{int}}\,\mathbf{v} = 0 \qquad \text{where } \mathbf{v}_{\text{min}} \leq \mathbf{v} \leq \mathbf{v}_{\text{max}}. \tag{5}$$

In FBA, additional optimality criteria are used to select a single flux distribution. In classical FBA, we maximise a linear benefit function $b(\mathbf{v})$, corresponding to the first term in Eq. (1). A typical choice is the biomass production rate, i.e., the flux in the biomass-producing reaction. In FBA with molecular crowding, we additionally consider a weighted sum of the fluxes as a proxy for the presumable enzyme demand and constrain it by an upper bound. In Flux Cost Minimisation (FCM), conversely, we predefine a flux benefit value (e.g. the biomass production rate) and minimise a flux cost $a(\mathbf{v})$ (e.g. the sum of absolute fluxes, in the case of minimal-flux FBA). Flux cost functions may represent a demand for enzymes, which entail a growth deficit. To be meaningful, flux cost functions should increase with the absolute flux $|v_l|$ and show a minimum at $v_l = 0$. This implies that the scaled derivatives $\partial a/\partial v_l)\, v_l$ be positive (for flux cost functions representing optimal enzyme costs, we find in fact that $\partial a/\partial v_l\; v_l = \partial h/\partial u_l\; u_l > 0$). If we put a positive lower bound on $|\partial a/\partial v_l|$, the flux cost must show a kink at $v_l = 0$ (which excludes the sum of quadratic fluxes as a meaningful flux cost function).

The laws of thermodynamics impose that shape the possible metabolic states. The direction of each reaction flux is determined by the ratio of product and substrate concentrations, called mass-action ratio. This relationship holds for any reversible, thermodynamically correct rate law, and shapes the metabolic fluxes in networks. As the mass-action ratio in a reaction increases from small to large values, the flux flips from forward to backward direction. The flip happens at a specific mass-action ratio, called equilibrium constant K_{eq}. We can also describe this by defining the thermodynamic driving force $\theta_l = \ln K_{\text{eq},l} - \sum_l n_{il} \ln c_i$, i.e. the negative Gibbs free energy, measured in units of RT. The force must be positive for reactions with forward flux, and negative for reactions with backward flux (loose sign constraint). If we assume that any non-zero driving forces will lead to a flux, we obtain the strong sign condition. The weak sign condition reflects the fact that the driving force determines the ratio $v+/v- = \mathrm{e}^\theta$ of forward and reverse one-way fluxes [19]: a positive flux (where $v_+ > v_-$) requires a positive force, and a negative flux requries a negative force. The strong sign condition makes an additional assumption: that the one-way fluxes can never vanish, even in the absence of enzyme. Consequences for possible flux patterns

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are described in SI section S1.1.

If flux directions are predefined, they put constraints on the possible metabolite levels. These constraints, together with the physiological ranges, define what metabolite profiles are possible. Some (hypothetical) flux modes cannot be realised by any metabolite profile because they contain loops, and concentrations would have to decrease in a circle [20]. Such flux profiles would represent a *perpetuum mobile* and can be excluded. Other flux distributions are thermodynamically possible, but require unphysiological choices of metabolite levels. Mathematically, the feasible metabolite profiles form a polytope in log-metabolite concentration space (called "M-polytope"). Metabolite profiles outside this polytope would yield the wrong flux directions and can be discarded. If a flux distribution leads to an empty M-polytope (or to a polytope outside the metabolite bounds), it is called *thermo-physiologically infeasible*. If it leads to an empty M-polytope (no matter if metabolite bounds are considered), it is called *thermodynamically infeasible*. Thus, by restricting the metabolite levels to physiological ranges, we can constrain the possible flows to some feasible segments in flux space ("feasible flux patterns"). Flux patterns that imply thermodynamic *loops* can be discarded.

Kinetic and stoichiometric metabolic models combine these assumptions in different ways and add optimality assumptions. However, a pathway model is only meaningful if it reproduces, or at least appoximates, the behaviour that the pathway would show as part of a living cell (or as part of a hypothetical detailed whole-cell model). First, we may check whether our metabolic models contain everything we would need for a whole-cell model (i.e., whether we could obtain a meaningful cell model simply by extending the network to the entire cell). Two main aspects that are still missing are density constraints (e.g. a bound on the weighted sums of all metabolite and enzyme levels within cell compartments) and dilution (i.e. a stationary condition $\mathbf{N}_{\text{int}}\,\mathbf{v} = \lambda\,\mathbf{c}$, where λ is the cell growth rate). Second, if we assume that our pathway objective function (scoring fluxes, metabolite levels, and enzyme levels within the pathway) is a proxy for (or "has been inherited from") a whole-cell objective (e.g. maximal cell growth), we need to understand how these two objectives are related. In the rest of this article, we will focus on metabolic models. That is, we ignore dilution and consider pathway objectives that we pose as a postulate (including effective enzyme costs, which increases with the enzyme levels). However, to make sense of these models, we need to come back to these points, and we will do this in the second part of this article.

3 The metabolic state manifold

Before we get to optimality problems, let us first think about the metabolic states themselves, i.e. all possible combinations of fluxes, metabolite levels, and enzyme levels. These *feasible steady states* – satisfying all physical and physiological constrains described above – form a set in a high.dimensional space, the set on which our optimality problems will be defined. Each point of this set is characterised by fluxes, metabolite concentrations, and enzyme levels, and these state variables are coupled by rate laws, stationarity, thermodynamic laws, and constrained by physiological bounds on concentrations. Mathematically, the set of states is a manifold in $(\mathbf{v}, \mathbf{x}, \mathbf{e})$-space. How can we describe its shape?

Let us consider a simple example, a reversible reaction with fixed product concentration. As shown in Figure 2, the possible states form a 2-dimensional manifold in a 3-dimensional (v, x, e)-space. Projecting this manifold onto the (v, x)-plane yields two patches: one patch (red) for states with positive fluxes and high substate concentrations, and another one (blue) for states with negative fluxes and low substrate concentrations. Each patch corresponds to a possible flux direction and covers the points in metabolite and flux space that agree with this flux direction: mathematically, it is the Cartesian product of two regions, a region in flux space and a region in metabolite space. States outside these patches ("empty" patches shown in grey) would be thermodynamically infeasible, not only for a specific rate law, but for *any* choice of rate laws that respect the thermodynamic sign constraints. The metabolic state manifold itself is a curved surface in (v, x, e)-space. It consists of two sheets that are obtained by lifting the

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

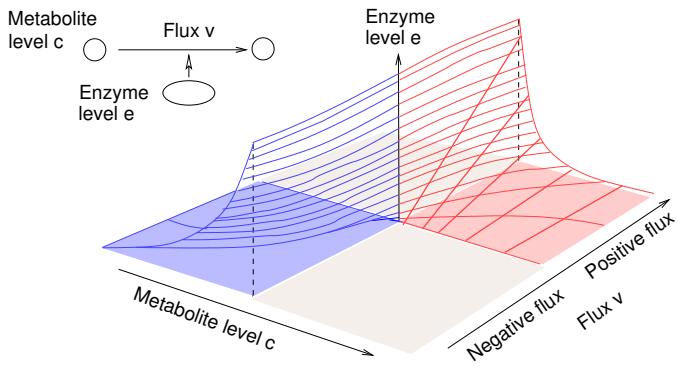


Figure 2: Metabolic state manifold of a single enzymatic reaction with substrate level c, enzyme level e, and reaction rate $v = e\,k(c)$. The product level is assumed to be constant. The two patches in (v, x) space and the corresponding sheets in (v, x, e)-space are shown in blue and red.

patches in e-direction according to the enzyme demand function $e(v, c) = \frac{v}{r(c)}$ (where $r(c)$ is the specific reaction rate). The mapping from a pair (v, x) to its enzyme demand is easy to compute, and the resulting enzyme levels are guaranteed to be positive (because our reversible rate law, within a feasible patch, yields a reaction rate r with the same sign as the flux v). In the state manifold, the enzyme level scales linearly with v, while the contour lines (intersection lines with $c - v$ planes) represent the curves $v = e\,k(c)$ of the rate law (for a fixed value of e). The cell can transit smoothly between all states on the manifold by changing its metabolite and enzyme levels. However, to move from positive to negative fluxes, it needs to pass through a single point in (v, x)-space, the chemical equilibrium state where the two patches touch each other. In (v, x, e)-space, this point corresponds to a semi-infinite line (describing a zero flux with arbitrary enzyme levels) because in the equilibrium state, all enzyme levels are possible. The enzyme demand function $e(v, c)$, used for "lifting" is non-unique is this point[3], and the two sheets are smoothly connected by the line. Altogether, by knowing the flux directions, constructing the flux-metabolite patches, and "lifting" them using the function $e(v, c)$, we can systematically construct all possible states. If we change the model parameters, the shape of the manifold changes as well. A change of the equilibrium constant K_{eq} would change the threshold concentration, which separates the patches in metabolite space. A proportional change of the (forward and backward) k^{cat} values would lead to a scaling in e-direction, and changes of Michaelis-Menten constants would change the shape of the enzyme demand function.

For large metabolic networks, the state manifold is multidimensional and possibly complicated. Nevertheless, we can construct them in the same way as for a single reaction. Of course, we encounter some additional complications: (i) We need to consider a flux pattern (a vector describing all flux directions) instead of a single flux sign, and we obtain polytopes instead of simple line segments in flux and metabolite space. (ii) Stationarity: all internal metabolites must be mass-balanced. (iii) Flows may be constrained to a given flux benefit. Because of the constraints (ii) and (iii), some additional patches may have to be discarded. (iv) The enzyme demand function $\mathbf{e}(\mathbf{v}, \mathbf{c})$ is multidimensional. (v) Furtermore, concentration ranges or fixed concentrations for metabolites may be defined. Knowing all this, we can construct the state manifold for a given network (see Box 2 and Box 3): (i) We enumerate all feasible flux patterns, representing a possible flux-metabolite patch (the Cartesian product of the flux polytope and the metabolite polytope that belong to this flux pattern). (ii) Using the enzyme demand function $\mathbf{e}(\mathbf{v}, \mathbf{c})$, which is easily obtained from the rate laws, we can "lift" the patches into $(\mathbf{v}, \mathbf{x}, \mathbf{e})$ space and obtain the sheets of the manifold. Finally, with additional assumptions about our metabolic system, we can put more constraints on the metabolic state. For example, a given linear flux benefit $b(\mathbf{v}) = b'$ defines a linear hyperplane in flux space, which intersects our metabolite manifold. Notably, the resulting lower-dimensional manifold may not be path-connected, and may thus consist of separate disjoint pieces. Such flux benefit constraints are needed to define flux cost minimisation problems such as "minimising enzyme cost at a given flux benefit".

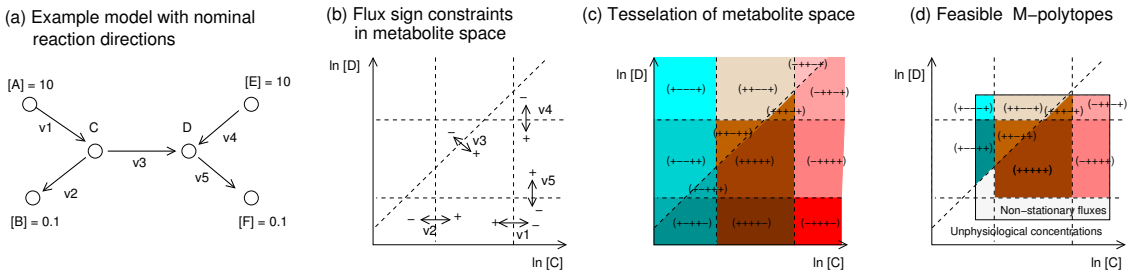
[3] If we additionally assume that cells switch off unnecessary enzymes ("enzyme removal assumption"), we could set the undetermined enzyme levels to $e_l = 0$, and the mapping between states in $(\mathbf{v}, \mathbf{x})$-space and $(\mathbf{v}, \mathbf{x}, \mathbf{e})$-space becomes bijective.

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

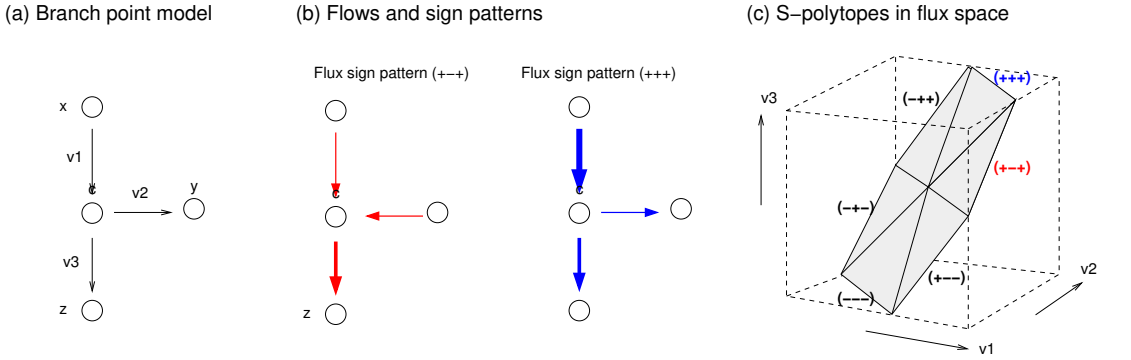
Box 2: Feasible flux and metabolite states

A metabolic network can show a variety of flux/metabolite states. Geometrically, these states form a collection of polytopes, each being the Cartesian product of a flux polytope and a metabolite polytope that comply with the same feasible force pattern sign($\boldsymbol{\theta}$) (see SI) and with box constraints (physiological ranges of fluxes and metabolite levels). A feasible thermodynamic force pattern sign($\boldsymbol{\theta}$) = sign($-\Delta\boldsymbol{\mu}$) must be derivable from feasible chemical potentials $\mu_i = \mu^\circ + RT \ln c_i$, where $\ln c_i$ must, again, be within physiological ranges. In practice, these polytopes can be constructed in the following way:

1. As a first step, we consider the possible metabolite profiles in log-metabolite space. Due to thermodynamic constraints (which go hand in hand with thermodynamically feasible rate laws), each metabolite profile predetermines a set of flux directions (flux pattern), and so the metabolite space is covered by convex polytopes, each related to one of the flux patterns. The physiological concentration ranges define a feasible box in metabolite space. All M-polytopes outside this box can be discarded. Below we will see that some other polytopes must be discarded too. Altogether, we obtain a collection of feasible M-polytopes, defining the possible metabolite profiles and corresponding flux patterns.



2. As a second step, we consider the possible flows in flux space. Each flux pattern corresponds to a segment in flux space (i.e. an orthant or one of its lower-dimensional surfaces). Stationarity, and maybe a predefined linear flux benefit, define a feasible subspace in flux space. Segments that are cut by this subspace are (stationarity)-feasible and contain a feasible flux polytope (S-polytope or B-polytope).



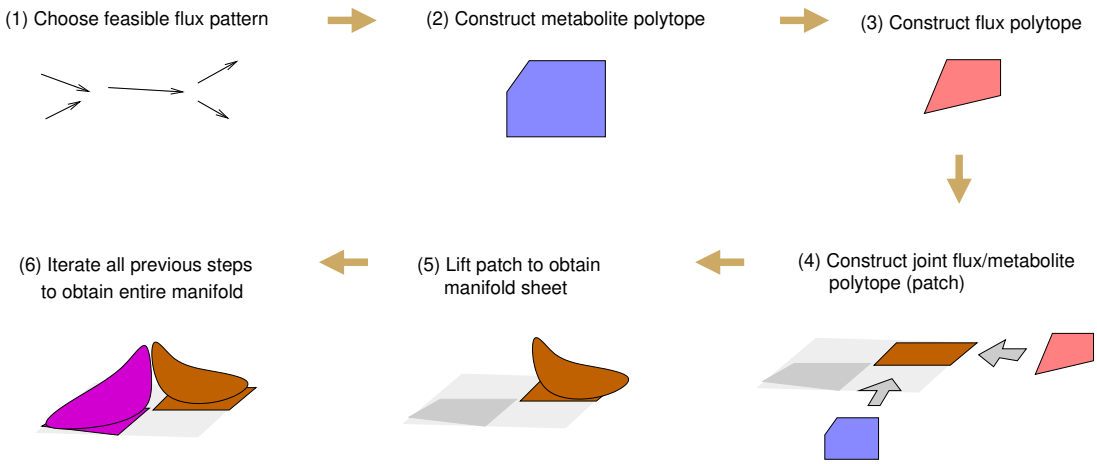
3. Now we combine these conditions. We consider all flux patterns that lead to feasible polytopes in both spaces (i.e. patterns that allow for stationary flows in flux space and for thermo-physiologically feasible metabolite profiles in metabolite space). If a flux pattern does not allow for a feasible flux polytope, the corresponding M-polytope in metabolite space is discarded. If a flux pattern is not realisable by any M-polytope, the corresponding S-polytope in flux space is also discarded. Eventually, we obtain a collection of feasible flux patterns (and corresponding pairs of S-polytopes and M-polytopes) that satisfy all constraints. These flux patterns create a direct correspondence between the polytopes in flux and metabolite space.

Despite its complicated shape, the state manifold has some simple properties: (i) All kinetically possible states (i.e. points of the state manifold) are thermodynamically feasible (i.e. they correspnd to feasible points in flux-metabolite space), and all thermodynamically feasible states can be kinetically realised. There is a unique mapping between the two sets of states, except for states in which a reaction rate vanishes; in this case, any values of the catalysing enzyme are possible. (ii) For a model with n_r enzyme-catalysed reactions and n_{ext} external metabolites, the manifold is an $n_r + n_{ext}$-dimensional in almost all points. (iii) The manifold is differentiable in each point (as long as the rate laws $v_l(\mathbf{c}, \mathbf{e})$ themselves are differentiable); to see this, we simply note that we can parameterise the manifold by $\mathbf{c}$ and $\mathbf{e}$). (iv) The manifold is path-connected, i.e., any two points in the manifold are linked by a path within the manifold. Starting in an initial state, the cell can gradually decrease all enzyme levels and fluxes to zero, change the metabolite profile, and then increase the fluxes and enzyme levels to reach

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Box 3: The state manifold of kinetic models

The state manifold of a kinetic model is a curved manifold in $(\mathbf{v}, \mathbf{x}, \mathbf{e})$-space. To construct it from network structure, rate laws, and constraints, we proceed in the steps shown below:



We first determine all possible flux patterns. Each flux pattern is translated into a patch in flux-metabolite-space. All combinations $(\mathbf{v}, \mathbf{c})$ in this patch can be realised with the chosen laws, and the necessary enzyme levels are easy to determine. By computing the enzyme levels, we "lift" our patch into a sheet in $(\mathbf{v}, \mathbf{m}, \mathbf{e})$-space. We can do this step by step.

the end state. Note that this path leads through a "dead" state (with zero metabolic fluxes); therefore, the proof does not apply to the state manifold restricted to a constant flux benefit. For a detailed description of the state manifold, see SI section S1.

4 Objective functions on the state manifold and effective objectives

We have learned that the set of feasible states can be screened systematically using metabolite levels and fluxes as the free variables. We can now return to our initial question, the search for optimal cell states. Instead of describing the cell as a whole, we will only consider a metabolic network or pathway. To define a metabolic optimality problem, we choose a kinetic model describing our pathway (including physical and physiological constraints), which defines a state manifold. Then, we choose an objective function on the state manifold, e.g. a function of the form Eq. (1), or a simple function such as enzyme cost. Given our objective function, an optimal steady state can be found by screening all combinations of stationary fluxes, metabolite levels, and enzyme levels, and choosing the optimal states $(\mathbf{v}, \mathbf{x}, \mathbf{e})$ that we're looking for, the one that maximises our objective. A simple example, the optimisation of a fitness Eq. (1) for a single chemical reaction, is shown in Figure 3.

An important aspect of these optimality problems are the constraints. First, there are the physical constraints on state variables (e.g., the stationarity condition; the kinetic relationship between enzym levels, metabolite levels, and fluxes; bounds on cell variables; or the constraints of a predefined flux benefit). These constraints determine the shape of the metabolic state manifold or can be used to restrict it (e.g., by imposing a given flux distribution). Second, in our optimality problems, we consider specific state variables called *metabolic objectives*, which can either be optimised (then they are called *mathematical objectives*) or constrained. For example, we may minimise the amount of required enzyme (one metabolic objective) at a fixed rate of biomass production (another metabolic objective). If we constrain metabolic objectives, this means that we consider only a "feasible" subpart of the state manifold. Below we shall consider a generalised version of this as a running example: a minimisation of enzyme

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

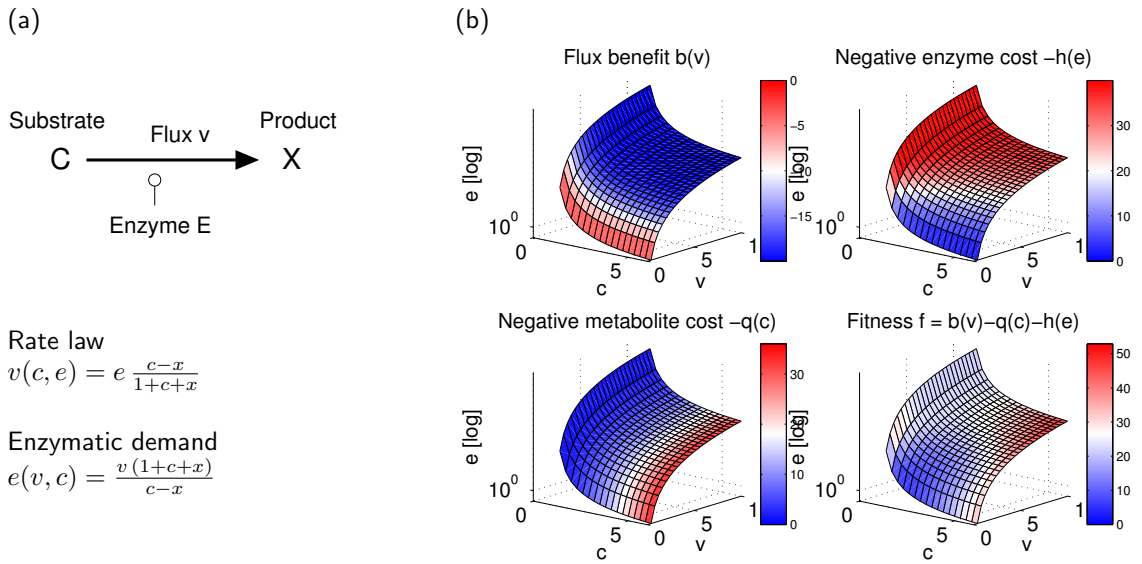


Figure 3: Metabolic states and objective functions for a single-reaction model (compare Figure 2). (a) We consider a reaction C $\leftrightarrow$ X with enzyme level e, substrate level c, constant product level x, and reversible Michaelis-Menten rate law (all rate constants set to 1 for simplicity), which yields the enzyme demand function. (b) By screening the metabolite concentrations and fluxes and solving for the enzyme demand, we obtain the set of feasible metabolic states (c, v, e) (curved surface in panel 1, e on vertical axis, shown on natural logarithmic scale). The flux benefit function (an increasing function of the flux, in arbitrary units) is shown on this manifold in colours (low: red; high: blue). Panels 2 and 3 show the enzyme and metabolite cost (as benefit functions, with a minus sign). The fitness function in panel 4 (flux benefit minus concentration costs) shows a local maximum in the center of the blue region.

cost at a fixed flux benefit[4]. Starting from Eq. (1), this means: we replace the flux benefit term by a constraint, and ignore the metabolite cost. But now we shall try to describe this in subspaces, i.e., to consider enzyme cost as a direct function of fluxes and metabolite levels.

Instead of considering the state manifold as a whole, we may be interested in some of the variables only (either $\mathbf{v}$, $\mathbf{c}$, or $\mathbf{e}$). This means: we may want to "project" our state manifold, including the optimality problem, into flux, metabolite. or enzyme space. In FBA, for example, we only consider fluxes, and ignore the other variables. To "project" our metabolic states (in the manifold) onto flux states (on the FBA flux polytope), we need to eleminate all other variables, but retain all relevant information (e.g., the fact that certain flows, effectively, imply a high enzyme cost). To eliminate a type of variables (fluxes, metabolite levels, or enzyme levels), there are three possibilities: we may (i) use dependencies in the model (e.g. computing enzyme levels directly from fluxes and metabolite levels), (ii) treat some variables as fixed and given (e.g. the metabolic fluxes), or (iii) assume that some (unknown) variables are optimised, given the other (known) variables. For example, to omit the enzyme levels as function arguments, we can either (i) predefine their values, (ii) infer their values from the metabolite levels and fluxes by using the rate law, or (iii) take the optimum value of the utility function across all possible choices of enzyme levels. This means, to project our state manifold onto flux space, we could predefine all metabolite levels (ii) and treat the enzyme levels as functions of fluxes and metabolite levels (i). Alternatively, we could treat enzyme levels as a function of fluxes and metabolite levels (i) and assume that the metabolite are optimised, given the fluxes, for minimal enzyme cost. In both cases, each flow $\mathbf{v}$ would implicitly determine the metabolite and enzyme levels, and therefore an enzyme cost.

[4]There may be good reasons to compare metabolic states at equal flux benefits. First, to compare enzyme cost and flux benefit on the same scale (as in Eq. (1)), they need to show the same measurement units. To make them comparable, relative weights need to be found, which introduces some arbitrariness. Second, if flux benefit $b(\mathbf{v})$ and enzyme cost $q^{\mathrm{app}}(\mathbf{c}|\mathbf{v})$ scale linearly with the flow $\mathbf{v}$, this also holds for the fitness Eq. (1), and this fitness function will have its optima at $\mathbf{v} = 0$ or $\mathbf{v} = \infty$. Instead of maximising this function, we will rather optimise the cost/benefit ratio (which may have a local minimum) or minimise cost at a given benefit.

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

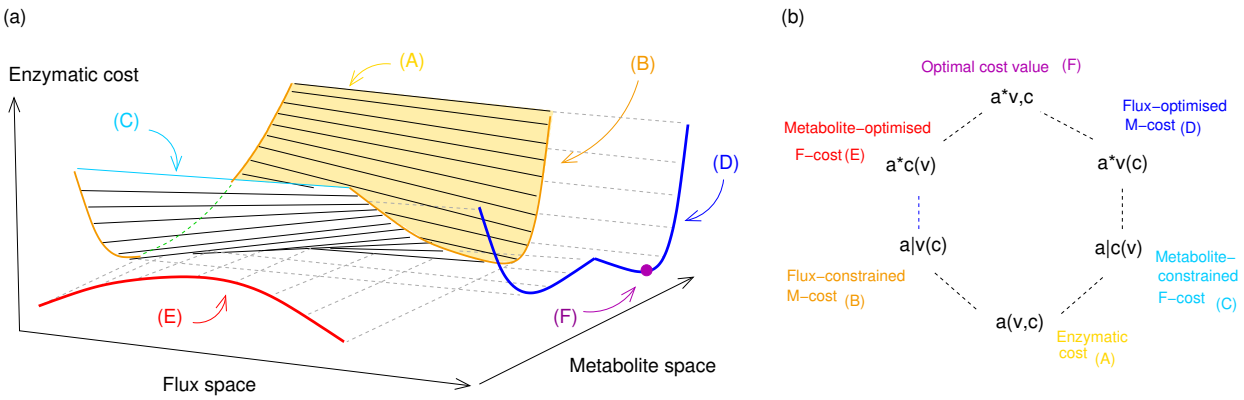


Figure 4: Enzyme cost projected to flux and metabolite space. (a) Enzyme cost functions (schematic example). From the enzymatic cost $h(\mathbf{e}(\mathbf{c},\mathbf{v}))$ as a function of fluxes and metabolite levels (yellow), we obtain projected cost functions ("constrained" enzyme costs and "effective" enzyme costs). (b) Relations between the different enzymatic cost functions shown in (a). Starting from the enzymatic cost $h(\mathbf{e}(\mathbf{c},\mathbf{v}))$ (bottom), constrained and effective costs are derived by predefining or optimising some of the variables.

Let me state this more generally. We assume a utility function $f(\mathbf{v},\mathbf{c},\mathbf{e}) = b(\mathbf{v}) - q(\mathbf{c}) - h(\mathbf{e})$. By convention, let is score fluxes by a benefit, while concentrations are scored by costs. From this objective function, we can now derive apparent objectives that have the same physical meaning, but fewer function arguments. First, using the physical constraints, we can project the objective function from the state manifold to one of the subspaces (e.g. flux-metabolite space). We obtain the *projected objective* and the *projected optimal point*. Second, by predefining some variables, we can consider the objective on a section of the state manifold. We obtain the *conditional objective* and the *conditional optimal point*. And, third, we can replace the objective by a "effective objective", i.e. projecting the objective function into a subspace and choosing, for each projected point, the optimum value along the projection line. This yields the *sihouette objective* and the *sihouette optimal point*. By applying these possibilities in different combinations to fluxes, metabolite levels, or enzyme levels, we can obtain a variety of different functions, all describing the same utility and the same model, but conditioned on different assumptions.

Let us consider enzyme cost, our running example, and how it can be represented in flux or metabolite space. With an enzyme cost function $h(\mathbf{e})$ (linear) and an enzyme demand function $\mathbf{e}(\mathbf{c},\mathbf{v})$, we can systematically define several enzymatic cost functions in metabolite (M) and flux (F) space (see Figure 4). If the enzyme demand $\mathbf{e}(\mathbf{v},\mathbf{c})$ is obtained from thermo-feasible rate laws, it will be non-negative on the thermo-feasible patches. From the enzyme cost $h(\mathbf{e})$, we obtain a number of projected enzymatic cost functions with specific mathematical properties (see Figure 4; the letters in the last column refer to the figure)

Enzymatic cost	$a^{\mathrm{enz}}(\mathbf{c},\mathbf{v}) = h(\mathbf{e}(\mathbf{c},\mathbf{v}))$	biconvex	A
Flux-constrained enzymatic M-cost	$a^{\mathrm{enz}}(\mathbf{c}\vert\mathbf{v}) = a^{\mathrm{enz}}(\mathbf{c},\mathbf{v})$	convex	B
Metabolite-constrained enzymatic F-cost	$a^{\mathrm{enz}}(\mathbf{v}) = a^{\mathrm{enz}}(\mathbf{c},\mathbf{v})$	linear	C
Flux-optimised enzymatic M-cost	$a^{\mathrm{enz}}(\mathbf{c}\vert*\mathbf{v}) = \min_{\mathbf{v}} a^{\mathrm{enz}}(\mathbf{c},\mathbf{v})$	piecewise convex	D
Metabolite-optimised enzymatic F-cost	$a^{\mathrm{enz}}(\mathbf{v}\vert*\mathbf{c}) = \min_{\mathbf{c}} a^{\mathrm{enz}}(\mathbf{c},\mathbf{v})$	concave	E

In contrast to the enzym cost $h(\mathbf{e})$ itself, an enzymatic flux cost is an overhead cost: a function representing the enzyme cost implied by the fluxes.

The projection approach can be applied to any objective function on the state manifold (i.e. any function of fluxes, metabolite levels, and enzyme levels). For example, if we add a convex metabolite cost $a^{\mathrm{met}}(\ln \mathbf{c})$ to the enzyme cost, we obtain a new "kinetic" cost function (representing, e.g., the relative cell volume occupied by both types of compounds). Again, the kinetic cost can be projected into different subspaces:

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kinetic cost	$a^{\mathrm{kin}}(\mathbf{c},\mathbf{v}) = a^{\mathrm{enz}}_{c,v}(\mathbf{c},\mathbf{v}) + a^{\mathrm{met}}(\mathbf{c})$	biconvex	A
Flux-constrained kinetic M-cost	$a^{\mathrm{kin}}(\mathbf{c}\|\mathbf{v}) = a^{\mathrm{kin}}(\mathbf{c},\mathbf{v})$	convex	B
Metabolite-constrained kinetic F-cost	$a^{\mathrm{kin}}(\mathbf{v}\|\mathbf{c}) = a^{\mathrm{kin}}(\mathbf{c},\mathbf{v})$	linear with offset	C
Flux-optimised kinetic M-cost	$a^{\mathrm{kin}}(\mathbf{c}\|*\mathbf{v}) = \min_{\mathbf{v}} a^{\mathrm{kin}}(\mathbf{c},\mathbf{v})$	piecewise convex	D
Metabolite-optimised kinetic F-cost	$a^{\mathrm{kin}}(\mathbf{v}\|*\mathbf{c}) = \min_{\mathbf{c}} a^{\mathrm{kin}}(\mathbf{c},\mathbf{v})$	concave	E

5 Conditions for optimal metabolic states

Starting from optimality problems on the state manifold, and defining conditional or effective objectives, we now reobtain known optimality problems in flux, metabolite, and enzyme space. With further approximations, we obtain methods such as FBA with flux minimisation or FBA with molecular crowding. In metabolic optimality problems, we need to evaluate the terms from Eq. (1) (flux benefit, metabolite cost, enzyme cost), while restricting the variables $\mathbf{v}, \mathbf{c}$, and $\mathbf{e}$ to feasible metabolic states, i.e. points on the state manifold (see again Figure 3). How can we systematically screen this manifold? We know this already. Instead of treating fluxes, metabolite levels, and enzyme levels as independent variables, we choose a subset of independent variables, e.g. fluxes and metabolite levels. Based on the metabolic objective functions on the state manifold, we can formulate metabolic optimality problems (like those shown in Figure 1 and Box 1). We can formulate them in different (yet equivalent) ways, using different kinds of free variables and different choices of the objectives and constraints. With Eq. (1) as a general fitness objective, different ways of screening metabolic states lead to different, equivalent optimality problems.

Let us first consider the enzyme levels (and conserved moiety concentrations) as free variables. With steady-state metabolite levels and fluxes given as functions $\mathbf{c}^{\mathrm{st}}(\mathbf{e})$ and $\mathbf{v}^{\mathrm{st}}(\mathbf{v})$, the fitness becomes a function

$$f(\mathbf{e}) = g(\mathbf{e}) - h(\mathbf{e}), \tag{6}$$

where $g(\mathbf{e}) = b(\mathbf{v}^{\mathrm{st}}(\mathbf{e})) - q(\mathbf{c}^{\mathrm{st}}(\mathbf{e}))$. By maximising this function, we obtain an optimal metabolic state. However, the functions $\mathbf{c}^{\mathrm{st}}(\mathbf{e})$ and $\mathbf{v}^{\mathrm{st}}(\mathbf{e})$ are not explicitly known, so the function $f(\mathbf{e})$ can only be evaluated numerically, and its general mathematical properties are unknown. It may have multiple local optima, which makes the optimisation difficult. Luckily, there is another way to proceed: after writing the enzyme levels as functions of fluxes and metabolite levels, the fitness function can be cast as

$$f(\mathbf{v},\mathbf{c}) = b(\mathbf{v}) - q^{\mathrm{app}}(\mathbf{c}|\mathbf{v}) \tag{7}$$

with $q^{\mathrm{app}}(\mathbf{c}|\mathbf{v}) = q(\mathbf{c}) + h(\mathbf{e}(\mathbf{v},\mathbf{c}))$, and can now be optimised with respect to $\mathbf{v}$ and $\mathbf{c}$. For simplicity, let us assume that the flux benefit $b(\mathbf{v})$ is prescribed, so we consider again our running example, minimising enzyme cost at a given flux benefit. This optimisation can be performed by combining a minimisation in enzyme space with an optimisation in metabolite space [21], and we can do this in two different ways (see Figure 5). First, given a predefined metabolite profile $\mathbf{c}$, each flux profile $\mathbf{v}$ will have an enzyme demand $\mathbf{e}(\mathbf{v})$. The resulting enzyme cost, written as a function $a^{\mathrm{enz}}(\mathbf{v}|\mathbf{c}) = \mathbf{e}(\mathbf{v},\mathbf{c})$ of the fluxes, is called the *metabolite-conditioned enzymatic F-cost*. Next, by minimising this cost in flux space (at a fixed flux benefit) and treating $\mathbf{c}$ as variable again, we obtain a cost $q^{\mathrm{enz/v}}(\mathbf{c})$, a function on the metabolite polytope. This function is called the flux-optimised enzymatic M-cost (see Figure 5). Second, we can also proceed in the opposite order: we fix a flow $\mathbf{v}$ and obtain the (flow-conditioned) enzymatic cost in metabolite space, $q^{\mathrm{enz}}(\mathbf{c}|\mathbf{v})$; then, by minimising with respect to the metabolite profiles $\mathbf{c}$, we obtain the *metabolite-optimised enzymatic F-cost* $a^{\mathrm{enz/c}}(\mathbf{v})$. The same goes for kinetic (enzyme plus metabolite) cost functions.

If a feasible metabolic state performs better than (or equally good as) all feasible states in its neighbourhood, it is called locally optimal. How can we find such locally optimal states? Let us consider again enzyme cost

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

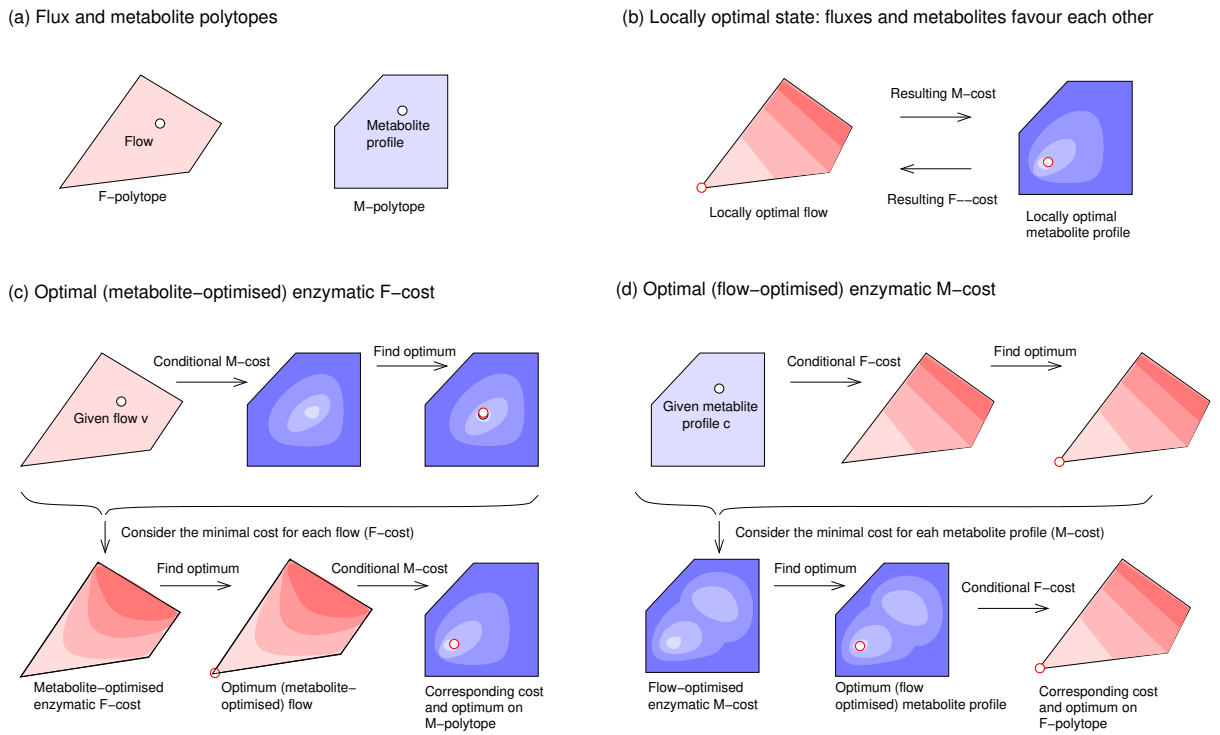


Figure 5: Search for optimal metabolic states in flux and metabolite space. (a) Optimising the M-optimised enzymatic F-cost. For each flow $\mathbf{v}$, the minimal cost is computed by optimising the metabolite levels (top). The resulting cost function (a effective F-cost) is concave on the flux polytope (bottom). (b) Optimising the F-optimised enzymatic M-cost. For each choice of the metabolite profile $\mathbf{m}$, the minimal cost is computed by optimising the fluxes (top). The resulting effective M-cost is piecewise convex on the flux polytope (bottom). (c) Condition for locally optimal states: each of the vectors (flow and metabolite profile) must be a unique optimum given the other one.

minimisation in flux and metabolite space at a given flux benefit. As before, we parametrise the metabolic states by fluxes and logarithmic metabolite levels, i.e. we minimise the enzyme cost $q^{\mathrm{enz}}(\mathbf{x},\mathbf{v}) = h(\mathbf{e}(\mathbf{c},\mathbf{v}))$ in $(\mathbf{v},\mathbf{x})$-space at a given flux benefit. Any locally optimal state will satisfy a simple optimality condition (see Figure 5). Since $q^{\mathrm{eff}}(\mathbf{x},\mathbf{v})$ is biconvex, for a given flow $\mathbf{v}$ there can be only one optimal metabolite profile $\mathbf{x}$, and for each given $\mathbf{x}$ there can be only one optimum in $\mathbf{v}$-direction. Therefore, for $(\mathbf{v}^{\mathrm{opt}},\mathbf{x}^{\mathrm{opt}})$ to be a locally optimal state, $\mathbf{v}^{\mathrm{opt}}$ must be the only favoured flow of $\mathbf{x}^{\mathrm{opt}}$, and $\mathbf{x}^{\mathrm{opt}}$ must be the only favoured metabolite profile of $\mathbf{v}^{\mathrm{opt}}$. Or briefly: in a locally optimal state, metabolite profile and flow must uniquely favoure each other. Computing the favoured flow of a given metabolite profile (by linear FCM), or the favoured metabolite profile of a given flow (by ECM), is relatively easy. In particular, all uniquely favoured flows (obtained by linear FCM) must be corners of the flux polytope, and the same holds for all locally optimal states.

If we formulate the optimality problems in different subspaces, as described above, what are the optimality conditions, and what can we learn from them? Let us consider our two main cases, cost optimisation in flux/metabolite space and a fitness optimisation in enzyme space.

First we consider the problem of optimising metabolite and enzyme levels for a minimal kinetic (metabolite+enzyme) cost at a predefined flow $\mathbf{v}$ [11]. From the optimality condition for the metabolite profile $\mathbf{c}$

$$0 \;=\; -\nabla_{\mathbf{c}}\left[q(\mathbf{c}) - h(\mathbf{e}(\mathbf{c},\mathbf{v}))\right], \tag{8}$$

we obtain a condition for the cost derivatives ("prices") in this optimal state (see Figure 6 (a) and SI section

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

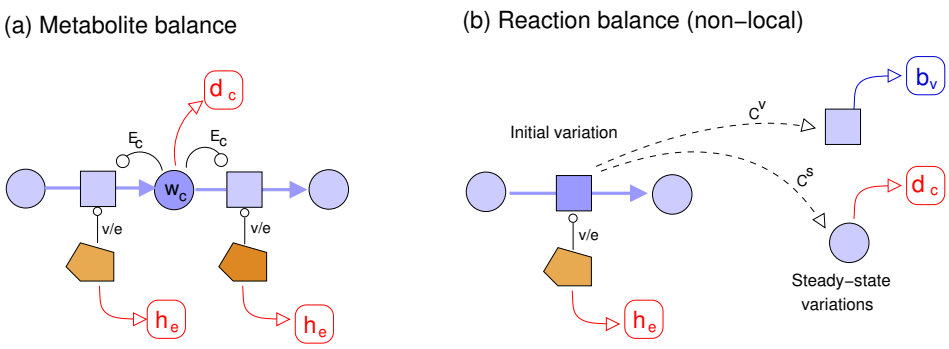


Figure 6: Economic balance conditions. (a) Metabolite balance condition, derived from cost minimisation at predefined fluxes. According to the balance condition, a metabolite variation δc, together with a compensating enzyme variation $\delta e^{\text{comp}} = -\frac{e}{v}\, E, \delta c$ (which, together, would leave all fluxes and all other variables unchanged) would have a zero fitness effect. (b) Reaction balance, derived from fitness optimisation in enzyme space. According to the balance condition, an enzyme variation δe, together with the resulting steady-state variations δv^{st}, δc^{st} (anywhere in the network) would have a zero fitness effect. Both balance conditions are non-local: the metabolite balance in the case of moiety conservation (not shown), and the reaction balance due to the occurrence of metabolic control coefficients (dashed arrows).

S3.3)

$$\mathbf{q}_{\text{c}} \quad = \quad \mathbf{E}_{\text{c}}^{\top}\, \mathbf{E}_{\text{e}}^{-1}\, \mathbf{h}_{e}. \tag{9}$$

This equation, called "metabolite balance condition", links the metabolite prices (in a vector $\mathbf{q}_{\text{c}}$) to the enzyme prices (in a vector $\mathbf{h}_{e}$). Each component of this vector equation refers to a metabolite and to the enzymes that catalyse the surrounding reactions (the reactions in which this metabolite is directly involved). The two sets of prices are linked by the unscaled metabolite elasticities $E^{\text{v}_l}_{\text{c}_i} = \partial v_l / \partial c_i$ and enzyme elasticities $E^{v_l}_{e_l} = \partial v_l / \partial e_l$. The terms in this formula describe local quantities (i.e. the price of a single metabolite or enzyme, or relations $\mathbf{E}_{\text{c}}$ between neighbouring variables). Notably, Eq. (9) also holds in a more general optimality problem: a problem in which the flows are not predefined, but optimised along with the other variables? Above, we saw that the (flux-optimised) enzymatic M-cost, in a region around its optimum point, is given by the (flux-constrained) enzymatic M-cost (with fluxes constrained to the optimal flow). Therefore, Eq. (9) also holds if flows and metabolite profiles are optimised together. What does the balance conditions tell us? If we multiply it by a differential concentration change $\delta\mathbf{c}$, we obtain a change of metabolite cost $\mathbf{q}_{\text{c}}\, \delta\mathbf{c}$ on the left, and a change $\mathbf{h}_{e}^{\top}\, \mathbf{E}_{\text{e}}^{-1}\, \mathbf{E}_{\text{c}}\, \delta\mathbf{c}$ on the right. In fact, we can read $\delta\mathbf{e} = \mathbf{E}_{\text{e}}^{-1}\, \mathbf{E}_{\text{c}}\, \delta\mathbf{c}$ as an *equivalent* enzyme change – a change that would have exactly the same effect on the metabolic state as our change $\delta\mathbf{c}$. This also means: if we apply an arbitrary, small metabolite change $\delta\mathbf{c}$ and *compensate* all its effects on the fluxes by a change $-\delta\mathbf{e}$, then the net fitness change is exactly zero! This makes sense: since we start from an optimal state, the cell cannot further improve this state by varying metabolite and enzyme levels, while keeping the fluxes unchanged.

Second, we consider an optimisation of enzyme levels $\mathbf{e}$ for a maximal fitness Eq. (1), with metabolic objective terms $b(v^{\text{st}}(\mathbf{e}))$ (steady-state flux benefit), $q(c^{\text{st}}(\mathbf{e}))$ (steady-state metabolite cost), and $h(\mathbf{e})$ (enzyme cost). The optimality condition in enzyme space reads

$$0 \quad = \quad \nabla_{\mathbf{e}} \left[b(\mathbf{v}^{\text{st}}(\mathbf{e})) - q(\mathbf{c}^{\text{st}}(\mathbf{e})) - h(\mathbf{e}) \right]. \tag{10}$$

As noted before, and as shown in Figure 1 (c), the three objective gradients must be balanced. With the flux gain $\mathbf{b}_{\text{v}}^{\star} = \nabla_{\mathbf{v}} b$, metabolite price $\mathbf{q}_{\text{c}} = \nabla q$, control coefficient matrices $\mathbf{C}^{\text{S}}$ and $\mathbf{C}^{\text{V}}$, and enzyme elasticity matrix

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\mathbf{E}_\mathrm{e}$, we can rewrite this condition as a reaction balance condition (see Figure 6 (b) and SI section S3.4)

$$\mathbf{C}^{\mathrm{V}\top}\mathbf{b}_\mathrm{v}^\star - \mathbf{C}^{\mathrm{S}\top}\mathbf{q}_\mathrm{c} = \mathbf{E}_\mathrm{e}^{-1}\,\mathbf{h}_e. \tag{11}$$

This optimality condition yields one equation for every enzyme. Again, to see what it means, may consider small state variations: a variation of enzyme levels $\delta\mathbf{e}$, leading to stationary changes $\delta\mathbf{v}^\mathrm{st}$ and $\delta\mathbf{c}^\mathrm{st}$. Again, the balance condition tells us any such variation would be fitness-neutral: the metabolic benefit on the left would be exactly balanced by the enzyme cost on the right. The two balance conditions (11) and (9), in the form given here, and are not fully local. Eq. (9) seems to be local, because it contains only variables linked to the metabolite perturbed and its neighbouring elements. However, if moiety conservation were considered (which we ignored here for simplicity), there would be connections, through conserved moieties, to distant parts of the network. Eq. (11) is clearly non-local because it contains the control matrices $\mathbf{C}^\mathrm{V}$ and $\mathbf{C}^\mathrm{S}$, which relate the local perturbation of a reaction rate to steady-state changes of concentration and fluxes anywhere in the network. To know these control coefficients, we need to consider the entire modes at once, and know its optimal state. This makes it hard to apply to a single reaction or pathway inside a larger, possibly unknown system.

6 Modelling the compromises in cells

If different metabolic objectives are in conflict, optimal compromises need to be found. Mathematically, trade-offs can be described in different ways. First, we can combine different metabolic objectives in a single objective. In Eq. (1), for example, we combine flux benefit, metabolite cost, and enzyme cost by taking their difference. We may also optimise a ratio (e.g., the enzyme cost per flux benefit). Second, we may optimise one metabolic objective while constraining the others: e.g. we maximise flux benefit at a fixed enzyme cost, or minimise enzyme cost at a fixed flux benefit. This can be used for a stepwise or nested "layered" optimisation of different objectives. Third, we may perform multi-objective optimisation. To determine a set of Pareto-optimal solutions, called Pareto front, we score fluxes, metabolite levels, and enzyme levels separately and search for potentially optimal compromises. A solution is optimal if none of the objectives can be further improved without compromising any other objectives. In other words, there is no other solution that scores better in some objectives, and equally well in all others. To generate point on the Pareto front, one may either optimise many linear combinations of the different objectives. In a problem with two objectives, one may also constrain one of the objectives to different values and obtimise the other. Interestingly, no matter how trade-offs are described, the criterion for optimal states is always the same: some convex combination of the three cost gradients must vanish. In a single-objective optimisation, each combination of the objectives (i.e. each choice of the numerical weights) yields a different solution. Together, these solutions cover a region in enzyme space, and this is also exactly the region of Pareto-optimal points. Moreover, by formulating the problem with fluxes and metabolite levels as the free variables, we would get to the same solutions. Thus apparently, no matter how we formulate the optimality problem mathematically, we always obtain the same result!

We saw that compromises between objectives can be modelled in various ways: by combining them into one objective, by optimising one objective while constraining the others, or by multi-objective optimisation. We saw examples of this in the paper. If we associate lower bounds (on metabolic objectives) with benefits and upper bounds with costs, these approaches are closely related, not only in what they assume about cells, but also mathematically: all of them lead to optimality conditions of the same form! If a vector $\mathbf{x}$ contains the cell variables to be optimised (e.g., enzyme levels) and if $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots$ are our metabolic objectives (e.g., enzyme cost, metabolite cost, and flux benefit), the optimality conditions have the form

$$0 = \sum_i \varphi_i\,\nabla_\mathbf{x} f_i(\mathbf{x}). \tag{12}$$

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

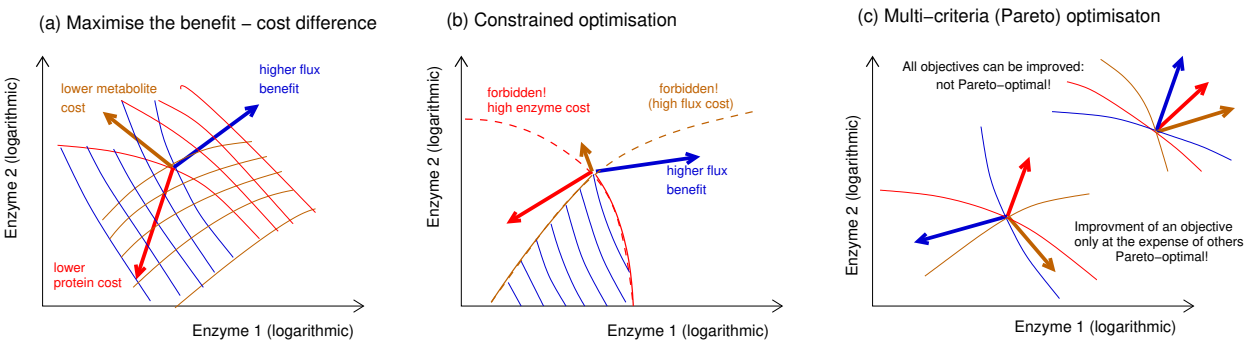


Figure 7: Mathematical ways to describe trade-offs between optimality objectives (in the example, for an optimisation in enzyme space). Like in Figure 2, we consider a metabolic model with flux benefit, metabolite cost, and enzyme cost, but we formulate the costs as metabolite "benefits" (i.e. negative metabolite costs) and enzyme "benefits" (i.e. negative enzyme costs). (a) Optimisation of a single combined objective, the sum of all benefit terms. In the optimal state, the three gradients must cancel each other. If the benefit terms are weighted in the sum, the same weight appear in the sum of gradients. (b) Constrained optimisation. In the example, we maximise flux benefit, while imposing lower bounds on metabolite and enzyme benefit. Again, in the optimal state a weighted sum of the gradients must vanish. Now, some of the prefactors (those that correspond to constraints) are Lagrange multipliers. (c) Multi-objective optimisation. Again, andA optimal compromise (Pareto-optimal state) requires a weighted sum of gradients to be balanced (i.e. to vanish). To illustrate this, a non-optimal state is shown on the right: since the gradients point to similar directions, a small movement in enzyme space could improve some objectives without compromising the others. In the Pareto-optimal state on the left, this is impossible.

This equation assumes that our cell variables $\mathbf{x}$ can be freely varied; if they need to respect constraints (e.g., positivity constraints; or if $\mathbf{x}$ represents interdependent variables such as fluxes, concentrations, and enzyme levels), there will be extra terms with Lagrange multipliers. The weight φ_i can have different meanings, depending on how compromises are described mathematically. In a single objective problem with $f(\mathbf{x}) = \sum_i \varphi_i \; f_i(\mathbf{x})$, the weights φ_i are simply the prefactors in the objective function. In we optimise one objective while constraining the others (e.g., optimising f_1, with bounds on $f_2, .., f_n$), we can set $\varphi_1 = 1$ while all other weights φ_i are Lagrange multipliers (with signs depending on the type of constraints (upper and lower bounds) and optimisation (maximisation or minimisation of f_1), or zero values for inactive constraints). In a multi-objective optimisation, the optimality condition has again the same form, with different values of the φ_i for different Pareto-optimal states (proof see SI section S3.2).

To understand what the optimality conditions mean geometrically, let us return to Figure 1 (c) and consider the balance of gradients in the optimal state. For clarity, we revert the signs of all cost terms and describe them as benefit terms. Again we consider the three ways to describe compromises. If a combined benefit function is maximised, a weighted sum of the benefit cost gradients, with the same weights as in the combined benefit, must vanish (see Figure 7). If we fix two objectives and optimise the third, a weighted sum of the gradients needs to vanish, but now one of the prefactors is 1, while the others are given by Lagrange multipliers. And also if we apply multi-objective optimality, a weighted sum of the gradients has to vanish [22] (proof in SI S3.2). Thus, whether we treat our metabolic objectives as mathematical objectives, as constraints, or in a multi-objective setting, we always obtain the same optimality criterion: a weighted sum of the three benefit gradients must vanish. If metabolic objectives refer to costs or upper bounds (instead of benefits or lower bounds), the weights for their gradient will have minus signs. This condition can be extended to any number of objectives, and to additional constraints between variables, e.g. if we consider separate cost terms for different enzyme fractions (in different cell compartments or membranes). Thus, due to their common optimality conditions, we can say that all the different optimality problems are equivalent!

For convenience, we can remove the prefactors φ_i in Eq. (1) by rescaling our objective functions. Originally,

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the φ_i are either predefined (in single-objective optimisation), or they emerge in the optimal state (as Lagrange multipliers in constrained optimisation, or in multi-objective problems). However, in each specific solution (based on given objectives b, q, and h), the prefactors φ will have specific numerical values, and if we are interested only in this solution, we can replace our problem by a single-objective "proxy" problem with effective objectives $b' = \varphi_b\, b$, $q' = \varphi_q\, q$. This new problem will have the same optimality condition as before, but with prefactors φ_i equal to 1, which can be ignored. Two details should be noted. First, the objectives in our proxy problem will be differently scaled for different solutions of our original problem. For example, if we start from a multi-criterio problem, then the prefactors φ, and thus the objectives of the proxy problem will be differently scaled in every Pareto-optimal point. Second, in our original problem, φ_b will typically be positive (for benefits, described by lower bounds in a maximisation problem) and φ_q and h will typically be negative (for cost, described be upper bounds in a maximisation problem). Accordingly, benefit terms in the resulting proxy problem will have positive signs, and cost terms will have negative signs. This justfies our previous, simplified proxy problems and shows that their optimality conditions hold generally, no matter how compromises are described.

7 Discussion

To unify different metabolic modelling approaches based on optimality principles, I considered general optimality problems on the metabolic state manifold. The resulting framework clarifies how modelling approaches (such FBA, RBA, kinetic models, and simplified whole-cell models) are logically related and how they can be interfaced in modular or layered modelling. Enzyme levels were generally used as a synonym for enzyme activities. While allosteric regulation can be included in the models[5], transcriptional regulation were not considered, because the aim here was to find which enzyme levels would be *profitable* for a cell, and not how these profitable enzyme levels are mechanistically realised. Nevertheless, if optimality calculations reveal quantitative relationships between metabolite levels and enzyme levels across optimal states, one could try to approximate these relationships by functions $e_l^{\mathrm{opt}} = e_l^{\mathrm{opt}}(\mathbf{c}^{\mathrm{opt}})$ and claim that these would be suitable gene regulation functions for the expression of enzymes.

The state manifold is not only practical for optimisation, but is also interesting itself, in order to better understand the sampling of metabolic states. The concept of a state manifold resembles the way state variables (such as temperature, pressure, and volume of a gas) are treated in physics. In classical thermodynamics, we consider a number of variables (e.g., volume, energy, and entropy of an amount of gas) that are related by a state equation, and other variables (e.g., pressure and temperature) that can be computed from them by taking derivatives. Since all these variables depend on each other, we obtain a manifold of possible states (two-dimensional, in this case). There is no "natural" choice of basic variables: in one case, it may be practical to use volume and energy as basic variables (while all others will be dependent on them); in another case, for example, we may prefer pressure and temperature. In metabolic models, we encounter a similar situation: in some cases, we parameterise metabolic states by enzyme levels, in others by metabolite levels and fluxes – but these are just different ways to refer to points on the state manifold.

From our general optimality problems on the state manifold, we can derive a variety of specific optimality problems. For example, the fitness Eq. (1), originally formulated as a function of enzyme levels, metabolite levels, and fluxes, can be optimised with different choices of the free variables. Each formulation has its advantages and disadvantages. If enzyme levels are treated as free variables, they can easily be constrained (e.g. by a bound on the sum of all enzyme levels). A layered flux/metabolite optimisation has its own advantages. First, it can be split into separate, layered optimality problems, a concave problem for fluxes [16] and a convex problem for

[5]Allosteric and post-translational regulation of enzymes can be generally considered costly, and would be excluded by optimality approaches that only optimise state variables, as considered here. The cost of regulation is maybe balanced by the benefits of stabilising metaboic states or making them better controllable (through changes in the Jacobian matrix). This could be included in optimality approaches as another objective in the future.

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

metabolite levels [11], that are easier to study and solve than general non-convex problems. Second, we can easily predefine or constrain the fluxes, the main functional target of metabolism. By running optimisation with predefined fluxes, we can even define plausible flux cost functions $a(\mathbf{v})$ to be used in flux analysis. Importantly, this approach provides a theoretical justification for flux cost functions and for enzyme constraints used in existing methods, e.g. FBA with molecular crowding.

If a cell process (e.g. a reaction flux) causes indirect costs (e.g. costs for enzyme maintenance), and if a model describes them as direct costs, I call these costs "overhead costs". The *enzymatic flux cost* is a good example: it describes the enzyme investment needed to realise certain fluxes, but is written as a function of the fluxes themselves, i.e. as a function in flux space. The fact that it does not arise within a reaction, but elsewhere in the cell (e.g. in enzyme production, or because enzymes occupy space, impinging on other growth-relevant processes) is hidden by the description as flux costs. We can rougly distinguish between two types of overhead costs: costs that arise from the causal *conditions* of a process (e.g. the cost of enzyme and metabolite levels required to realise a flux), and costs that arise from the causal *consequences* of a process (e.g. the cost of a toxic compound that accumulates due to a flux). However, since cellular networks contain cycles and feedback loops, this distinction is not very strict.

We can now classify optimality problems. Many of them are related, and some will define the same optimal states. Related problems may be fully equivalent or may share similarities at different levels. First, two problems may use the same kinetic model and objective function, but treat different cell variables as the free variables (e.g. a direct enzyme optimisation and a flux optimisation with kinetic flux cost). Regarding their biological assumptions, such problems are completely equivalent. Second, two problems may assume the same metabolic objectives, but represent them either as mathematical objectives or as constraints (e.g., minimising a cost at a fixed benefit *versus* maximising this benefit at a limited cost). Third, two problems may describe different, but overlapping biological systems (e.g. a model of a metabolic pathway versus a cell model containing this pathway) and may therefore employ different fitness objectives (maximal enzyme-specific pathway flux versus maximal cell growth) that are equivalent "in disguise" (i.e. leading to the same optimal state in the pathway). Fourth, two problems may rely on different modelling paradigms (e.g. constraint-based versus kinetic models), but employ equivalent objectives (possibly, again in disguise), leading to the same metabolic state. Finally, two optimality problems may describe pathways in the same cell, under the same growth conditions, and our aim is to compare and combine their results.

The different formulations of an optimality model represent different assumptions about the "freedom" of cells to vary some cellular quantities – i.e. whether quantities are assumed to be fixed or whether they could be changed at the expense others. For example, when describing a cellular trade-off between biomass production and enzyme demand, we may assume that the cell can devote a fixed protein budget to metabolism. In any case, we may assume that cells can shift resources between metabolic pathways; so for a single pathway, there will be a way to increase the enzyme amount, at a cost that is not an actual enzyme cost, but a decreased metabolic performance in other pathways due to the reallocation of enzyme, represented by an "overhead" enzyme cost! But we may also argue that cells – if biomass production were really important – could always find ways to direct other protein resources (e.g., resources currently spent on ribosomes) towards metabolism if this increases the biomass production rate. However, increasing enzyme levels will eventually hit some limit: eventually, there will be an optimal compromise between all pathways, including protein translation, and that the metabolic enzyme budget *in this compromise state* defines an enzyme budget to be fixed for our calculation. Multi-objective optimisation, finally, does not constrain a single objective, but all of them together. In this way, it can describe a set of solutions that may be optimal for cells under different external conditions. Formally, each of these situations can also be described as a situation in which some objectives are fixed, giving rise to a single-objective optimality problem with a specific choice of constraints or weights for the metabolic objectives. All this shows that our choice between mathematical objectives and constraints is not a matter of physics or physiology: it is a pure modelling

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assumption, a choice between hypothetical scenarios that we attribute to the cell. Like in our mathematical formulae, constraints can be replaced by cost terms, or cost terms can be replaced by constraints, depending on how we would like to frame our assumptions.

Here we assumed that cells realise optimal metabolic states, with simple criteria for optimality. Notably, the theory does not *claim* optimality, nor does it prove or disprove optimality in real cells. Instead, it poses optimality as an assumption and shows some of the consequences: how optimal metabolic behaviour, according to our optimality criteria, would shape the state of cells. Optimality assumptions, often tacitly made by biologists, can be explicitly studied, and their consequences can be tested by models. We can simulate how well organisms function and under which selection pressures and constraints they evolve. In fact cells' behaviour often appears to be non-optimal. For cells that have not been evolutionarily adaptated to laboratory conditions or to experimental perturbations like gene knockouts, this is not surprising. There are different ways to model such "non-optimal" cell states. On the one hand, we can abandon the optimality assumption, describe enzyme profiles as non-optimal, and quantify the "loss" caused by non-optimality. For example, phenomena such as preemptive expression or variable expression levels in cell populations (as observed in bacterial persistence), may be adaptations to complex environments with varying nutrient supplies or rare, severe challenges by antibiotics. To describe such a behaviour as beneficial, we may modify our optimality problems and include adaptations to uncertain future challenges as side objectives. Then, apparently wasteful enzyme profiles (i.e. wasteful if only the cell's current environment is considered) can be actually economical as bet-hedging strategies (i.e. considering possible future challenges).

The aim proposed in the introduction – a general theory of optimal metabolic states – has been achieved, but only to a certain extent. We have a theory that works for a single pathway (up to the size of the entire metabolic network). But what if we want to model an entire cell? Or, at least, if we'd like to take into account that our pathway is surrounded to other pathways, and that the true fitness objective (e.g., biomass production) is not realised inside our pathway, but elsewhere in the network? We encountered two problems: First, we obtained an optimality condition , but this condition contains the metabolic control coefficients, which relate a reaction in question to all other reactions and metabolites in the system. Since our system of interest is the entire cell, we would need an entire, detailed cell model to use this condition. If we consider a single pathway model, important effects may be missing. What we need to do, for a meaningful theory (that applies to individual pathways, but effectively includes all fitness effects outside the pathway), is to find similar balance equations that are local, i.e. in which all fitness effects are described by local variables. Second, in the state manifold, it was still difficult to consider density constraints and dilution, which are important in whole-cell models (with a constrained protein budget or general denisty constraints). To solve both problems, I propose to take a closer look at the Lagrange multipliers appearing the in the optimality problems, and to interpret them as "economic values", associated with and dual to the physical variables in our model. I will do this in the second part of this article.

Acknowledgements

I thank Mariapaola Gritti, Elad Noor, and Stefan Müller for inspiring discussions. This work was funded by the German Research Foundation (LI 1676/2-2).

References

[1] W. Liebermeister, E. Noor, A. Flamholz, D. Davidi, J. Bernhardt, and R. Milo. Visual account of protein investment in cellular functions. *PNAS*, 111(23):8488–8493, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[2] N. Nagaraj, N.A. Kulak, J. Cox, N. Neuhauser, K. Mayr, O. Hoerning, O. Vorm, and M. Mann. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra hplc runs on a bench top orbitrap. *Mol Cell Proteomics*, 11:M111.013722, 2012.

[3] L. Zelcbuch, N. Antonovsky, A. Bar-Even, A Levin-Karp, U. Barenholz, M. Dayagi, W. Liebermeister, A. Flamholz, E. Noor, S. Amram, A. Brandis, T. Bareia, I. Yofe, H. Jubran, and R. Milo. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Research*, 41(9):e98, 2013.

[4] K. Zhuang, G.N. Vemuri, and R. Mahadevan. Economics of membrane occupancy and respiro-fermentation. *MSB*, 7:500, 2011.

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

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

[7] E. Klipp, R. Heinrich, and H.-G. Holzhütter. Prediction of temporal gene expression. Metabolic optimization by re-distribution of enzyme activities. *Eur. J. Biochem.*, 269:1–8, 2002.

[8] G.M. De Hijas-Liste, E. Klipp, E. Balsa-Canto, and J. R. Banga. Global dynamic optimization approach to predict activation in metabolic pathways. *BMC Systems Biology*, 8(1):1, 2014.

[9] J.G. Reich. Zur Ökonomie im Proteinhaushalt der lebenden Zelle. *Biomed. Biochim. Acta*, 42(7/8):839–848, 1983.

[10] W. Liebermeister. Metabolic economics in kinetic models. *Preprint on arXiv.org: arXiv:1404.5252*, 2014.

[11] E. Noor, A. Flamholz, A. Bar-Even, D. Davidi, R. Milo, and W. Liebermeister. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLoS Comput Biol*, 12(10):e1005167, 2016. doi:10.1371/journal.pcbi.1005167.

[12] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim Biophys Acta*, 1810(10):978–988, 2011.

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

[14] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5:323, 2009.

[15] R. Heinrich and E. Klipp. Control analysis of unbranched enzymatic chains in states of maximal activity. *J. Theor. Biol.*, 182(3):243–252, 1996.

[16] W. Liebermeister. Flux cost functions and the choice of metabolic fluxes. *Preprint on arXiv.org: arXiv:1801.05742*, 2018.

[17] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, 3:119, 2007.

[18] A. Varma and B.Ø. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnology*, 12:994–998, 1994.

bioRxiv preprint doi: https://doi.org/10.1101/483867; this version posted November 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[19] D.A. Beard and H. Qian. Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS ONE*, 2(1):e144, 2007.

[20] N.D. Price, I. Thiele, and B.Ø. Palsson. Candidate states of Helicobacter pylori genome-scale metabolic network upon application of loop law thermodynamic constraints. *Biophysical Journal*, 90:3919–3928, 2006.

[21] M.T. Wortel, E. Noor, M. Ferris, F.J. Bruggeman, and W. Liebermeister. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Computational Biology*, 14(2):e1006010, 2018.

[22] O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, and U. Alon. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science*, 336:1157–1160, 2012.