The enzyme cost of metabolic fluxes

Wolfram Liebermeister[1] and Elad Noor[2]

[1] Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France
[2] Department of Plant and Environmental Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel

Abstract

In this chapter we discuss why certain pathway designs have been selected by evolution, by hypothesizing that some are more beneficial than others – based on several possible criteria and optimization goals: minimizing the number of reactions, maximizing product yield, increasing reaction turnover rates, and avoiding small thermodynamic driving forces. It turns out that all these criteria are related to a single objective: minimizing enzyme demand per product production rate or, equivalently, maximizing "enzyme productivity". We first focus on simple unbranched pathways with predefined flux distributions. We discuss several feasibility and optimality problems where metabolite concentrations are independent variables and solve for the minimal enzyme demand. In this setting, we see how enzyme productivity can be assessed or predicted and how it depends on different system parameters such as kinetics, thermodynamics, and concentrations of enzymes and metabolites. We discuss the difference between growth rate and yield. We then illustrate it by comparing between pathway options for glycolysis.

Keywords: enzyme demand - enzyme cost - rate-yield trade-off - thermodynamic force - enzyme efficiency - metabolic pathway - metabolite cost - factorized rate law - max-min driving force - enzyme cost minimization - choice between pathways

Contributions: This chapter was drafted and written by Wolfram Liebermeister and Elad Noor and has been reviewed by Avi Flamholz, Daan de Groot, Maarten Droste, and Ohad Golan.

To cite this chapter: W. Liebermeister and E. Noor. The enzyme cost of metabolic fluxes (Version January 2025). 10.5281/zenodo.8154381. Chapter from: The Economic Cell Collective (2025). Economic Principles in Cell Biology. No commercial publisher | Online open access book.

The authors are listed in alphabetical order.

Chapter overview

- In this chapter we discuss why certain pathway designs have been selected by evolution, by hypothesizing that some are more beneficial than others – based on several possible criteria and optimization goals: minimizing the number of reactions, maximizing product yield, increasing reaction turnover rates, and avoiding small thermodynamic driving forces.
- It turns out that all these criteria are related to a single objective: minimizing enzyme demand per product production rate or, equivalently, maximizing "enzyme productivity".
- We first focus on simple unbranched pathways with predefined flux distributions. We discuss several feasibility and optimality problems where metabolite concentrations are independent variables and solve for the minimal enzyme demand. In this setting, we see how enzyme productivity can be assessed or predicted and how it depends on different system parameters such as kinetics, thermodynamics, and concentrations of enzymes and metabolites.
- We discuss the difference between growth rate and yield. We then illustrate it by comparing between pathway options for glycolysis.

6.1. What guides evolution to select one pathway over another?

In the previous chapters, we asked what flux distributions are possible in a network, and which are most profitable for a certain task. Now we shall ask, more specifically, what led to the choice of existing pathways, or what makes a pathway variant favorable over another one that exists, or may have existed, in evolution. Of course, the same question plays also an important role in metabolic engineering, when new pathways are added to an organism, typically with the goal of achieving a maximal production, while imposing the smallest possible burden on the cell.

The chemical space is vast and many options exist for the same process, even if we consider only reactions with known enzyme mechanisms and impose thermodynamic constraints. Hence, while evolution had a choice between many pathway variants, only a tiny fraction of these possible variants is actually realized in nature, and a core part of central metabolism almost always follows the exact same design. The few exceptions that exist actually prove the rule, such the two natural variants of glycolysis discussed later in this chapter. How can we understand why a certain variant is used in a certain organism or situation? And why are many variants not used at all? Moreover, some very successful pathways show features that might appear strange at first glance [1]: in glycolysis, an initial investment of ATP is required, and only later it is recovered in higher amounts leading to a net gain. Is this just an evolutionary accident, i.e. a case where the pathway that evolved first is the one that stuck around although it is not necessarily better than all the alternatives? Or, rather, evolution did manage to find the optimal solution and therefore we should try to explain what the advantages of these "engineered" features are?

In this chapter, we assume that it was a selection for functional features, not chance, that determined these pathway "choices", and ask: what guides evolution to select one pathway over another? What are the criteria that make pathways "efficient" or "profitable" for a cell or, alternatively, for a metabolic engineer? To compare pathways, we assume that each pathway comes with a predefined flux distribution, and therefore a predefined product yield, and alternative pathways (yielding the same product) are compared at equal product production rates.

When people talk about natural ecosystems, diversity is usually the first topic discussed. Indeed, evolution through natural selection is almost guaranteed to create diversity where species evolve to occupy biological niches while exploring the vast space of possible phenotypes. Similarly, the world of biochemistry is a vast space of possible reactions. Metabolic enzymes participate in a network of pathways that supply cells with energy, and building blocks for biomass. Scientists have been studying these biochemical reactions for nearly 300 years [2] – so far tens of thousands such reactions have been classified; certainly many more exist in nature. Here are a few online databases where biochemical reaction data are collected or predicted: MetaNetX, KEGG, MetaCyc, BiGG, ModelSEED, ATLAS

Philosophical remark 6.A What do we mean by a "pathway"?

The notion of "pathways" is common in cell biology to describe a set of reactions, proteins, or processes that form a functional unit. However, there is no general definition: in practice, a pathway is often just a subregion of interest within a larger network. In metabolism, "pathways" often lead from some important substrate to some important product, with a simple and predefined flux distribution that consumes substrate(s), generates product(s), and may or may not make use of co-factors. Considering fluxes in specific pathways (instead of flux distributions in the entire network) is often a practical choice and, importantly, a choice that assumes that we can model, understand, manipulate, or engineer such a pathway without strongly affecting the rest of the cell. This has a number of benefits: (i) Instead of studying a huge network, we can look at pathways separately; (ii) there are reasons to believe that the flux distributions in enzyme-efficient metabolic states must be elementary flux modes (see Chapter 4 in [3]). Since EFMs often entail discrete choices between different pathways, it can make sense to study these pathways separately (iii) once we understand the costs and benefits of single pathways (with a single, scalable flux mode), we can apply the same thinking to analysing flux distribution on the entire metabolic network. Thus, in the rest of this chapter, all results about "pathways" will also hold generally for entire networks, as long as a (scalable) flux mode is given. Instead of comparing alternative pathways, we can compare alternative flux modes. In the following chapter, we use this for optimizing over the set of all possible flux modes that a given network can support.

of biochemistry.

To study the choice between pathways variants, we consider alternative pathways leading from A to B (or having a certain net sum formula) and their respective advantages and disadvantages. For simplicity, let us focus on biosynthesis pathways whose main task is more or less clear: producing a precursor molecule. Thus, the theoretical question would be: if a cell needs to make B from A, which pathway should it use? More specifically, how should the metabolic reactions be chosen and in what order? What should their kinetics and how should they be regulated?

If the pathway variant found in nature is due to selection for "good functioning", then what are the features that make existing pathway designs successful? In short, what are criteria for "good" pathways? One possible criterion seems to be simplicity, that is, choosing a short route from pathway substrate and pathway product.

In contrast to the huge diversity that is allowed by the catalytic capabilities of enzymes, a few metabolic pathways are extremely ubiquitous and exist virtually in every living cell. For example, glycolysis is a general term for pathways that convert glucose to pyruvate while producing ATP [1]. One variant of glycolysis, named after Gustav Embden, Otto Fritz Meyerhof, and Karol Parnas (or the EMP pathway for short, see Figure 6.1), was the first metabolic pathway to be discovered by scientists [2]. Often, the pyruvate is reduced to lactate or ethanol, which makes the pathway redox balanced. Therefore, it one of the most common way for producing ATP anaerobically (i.e. without oxygen to serve as an electron acceptor). Another common variant was discovered in 1952 by Nathan Entner and Michael Doudoroff [4] (ED for short). For example, *E. coli* is capable of metabolizing glucose through both the EMP or the ED variants, and often does so simultaneously [5].

More generally, the overall reaction describing glycolysis is:

$$\mathrm{Glucose} + 2\,\mathrm{NAD(P)}^+ + n\,\mathrm{ADP} + n\,\mathrm{Phosphate} \longrightarrow 2\,\mathrm{Pyruvate} + 2\,\mathrm{NAD(P)H} + n\,\mathrm{ATP} + n\,\mathrm{H_2O} \qquad (6.1)$$

where the value of n for the EMP pathway is 2. Ng et al. [6] explored the space of all possible glycolyses (with different values of n), by exhaustively enumerating all glycolytic pathway variants. In order to generate the variants, they adapted a computational method first introduced by Bar-Even et al. [7] for finding alternative carbon fixation cycles – metabolic cycles whose net reaction converts CO_2 into organic compounds. You start by collecting a database of known biochemical reactions (e.g. from a database such as KEGG [8]) and then use a linear-programming algorithm

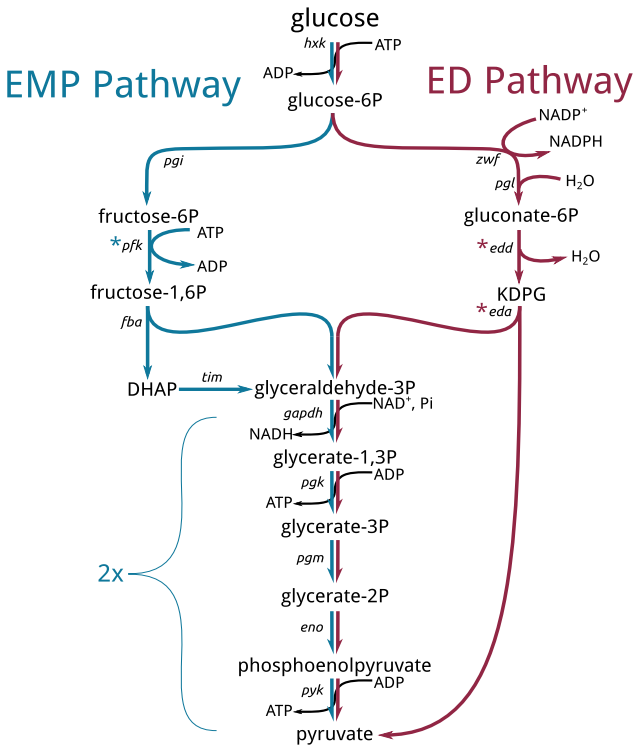

Figure 6.1: Two natural variants of the glycolysis pathway, named after their discoverers: Embden-Meyerhof-Parnas (EMP) and Entner-Doudoroff (ED)

to identify the set of reactions with the minimal sum of fluxes that conform to the predefined net reaction (e.g. 6.1). The objective is somewhat arbitrary, but since solving the LP requires setting an objective, we chose the min-flux as a reasonable proxy for the simplicity of the pathway. In any case, we will soon see how one can iterate through all possible solutions. Ng et al. [6] used this algorithm with the stoichiometry from 6.1 to find all possible glycolysis pathways comprising known enzymatic reactions (see Box 6.B).

The objective set by the linear problem (6.2) is minimizing the sum of fluxes, which corresponds to pathways with fewer reactions and low fluxes in each one. As discussed in 5.2 in [3], this objective is only a crude proxy for the efficiency of a pathway, and its only purpose is to get the pathway solutions in a relatively logical order. Although we have discussed global enzyme constraints in previous chapters (such as molecular crowding and proteome allocation), when comparing pathways we will focus only on the efficiency of the pathway itself. This will allow us to compare pathways without thinking about the rest of the cell or a specific metabolic context. But how can one quantify the efficiency of a pathway? The next section will be dedicated to exactly this question.

6.2. Pathway efficiency - some notions and thoughts

For glycolysis alone, Ng et al. [6] found 11,916 alternatives that produce at least one mole of ATP per mole of glucose. These include, of course, the EMP pathway. Although evolution can explore these options, natural selection typically converges on one or a few efficient variants. This does not mean that every single pathway observed in nature must be optimal, but we generally expect cells hosting highly inefficient pathways to eventually become extinct. Iacometti et al. [10] tested this experimentally by knocking out the EMP pathway from *E. coli* and forcing the cells to use the alternatives that naturally exist in this bacterium. In all cases, growth rates were slower than in the wild-type.

Before we discuss other examples for metabolic pathways, we need to define what we mean by "efficiency". There are several criteria one should consider:

- Low consumption rate of the substrate
- High generation rate of the product
- High regeneration rate or low consumption rate of the co-factor
- Small number of steps [11]
- Higher thermodynamic forces [12, 13]
- High enzyme turnover numbers
- High enzyme saturation levels

Some of these criteria refer to the cost (or investment) of the pathway, while others reflect the benefit (or profit) to the cell. By considering two common scenarios – single nutrient limitation or exponential growth in rich media – we can focus on two simple criteria which provide good measures of efficiency.

When the availability of a single nutrient is limiting growth, maximizing the molar yield (i.e. the number of moles of product generated for each mole of the nutrient) becomes the important feature. Yield is rather straightforward to calculate, as it is a direct outcome of the stoichiometry of the pathway. For example, anaerobic fermentation is often compared to respiration and deemed inefficient since it yields two moles of ATP per glucose, instead of ≈30 [14].

On the other hand, when conditions are good, such as during exponential growth in rich media, minimizing the total number of proteins required is often the objective which determines growth rate. . Here, we will be using the enzyme demand (e.g. in grams of protein) per unit of flux (typically, in mmol per hour per gram of cell dry weight). In fact, the enzyme demand per flux, as an objective, takes into consideration both the cost (protein) and the benefit (flux). Importantly, these two criteria scale linearly with respect to each other: doubling the amount of all enzymes without changing any of the metabolite concentrations would directly double the flux in the pathway. Therefore, this measure of efficiency is independent of the magnitude of the flux in the pathway. But, as we will see shortly, enzyme demand is a non-linear function, making it trickier to compute compared to other constraint-based problems such as ones we've seen in previous chapters.

Notably, these two measures of efficiency are not only useful for evolutionary processes, but for bioengineering as well. Obviously, the molar yield has economical implications when, for example, producing ethanol from sugar. However, the rate of a bioprocess is important as well due to the costs involved, e.g. for maintaining an operational bioreactor. One can imagine a computational model that accurately predicts the enzyme demand per flux of a pathway. Choosing the pathways with the lowest demand would be a good strategy for increasing the overall rate of bioproduction [15].

We define the enzyme demand per unit flux as the total amount of enzyme (in grams of protein) that is required to catalyze all of the pathway reactions at their required rates. We start by deriving a formula for the demand of a single enzymatic reaction. Consider an enzyme-catalyzed reaction:

$$\mathrm{S} \rightleftharpoons \mathrm{P} \tag{6.4}$$

where s and p will be the concentrations of the substrate (S) and product (P) respectively, and E the concentration of the enzyme which catalyzes this reaction (for simplicity, we drop the *tot* subscript from E_{tot}). Here, we will be using the factorized rate law (Eq. 3.10 in [3]), but other kinetic rate laws would produce similar results. The rate of a reaction is given by:

$$v = e \cdot k^{+}_{\mathrm{cat}} \cdot \frac{s/K_{\mathrm{S}}}{1 + p/K_{\mathrm{P}} + s/K_{\mathrm{S}}} \cdot \left(1 - e^{\Delta_{\mathrm{r}}G'/RT}\right) \tag{6.5}$$

where k^{+}_{cat} is the forward turnover rate, K_s and K_p are the Michaelis-Menten constants for the S and product P, and $\Delta_{\mathrm{r}}G'$ is the Gibbs free energy. So, the minimal amount of enzyme that is required for reaching a given rate v is:

$$q \equiv v \cdot h \cdot \frac{1}{k^{+}_{\mathrm{cat}}} \cdot \frac{1 + p/K_{\mathrm{P}} + s/K_{\mathrm{S}}}{s/K_{\mathrm{S}}} \cdot \left(1 - e^{\Delta_{\mathrm{r}}G'/RT}\right)^{-1}, \tag{6.6}$$

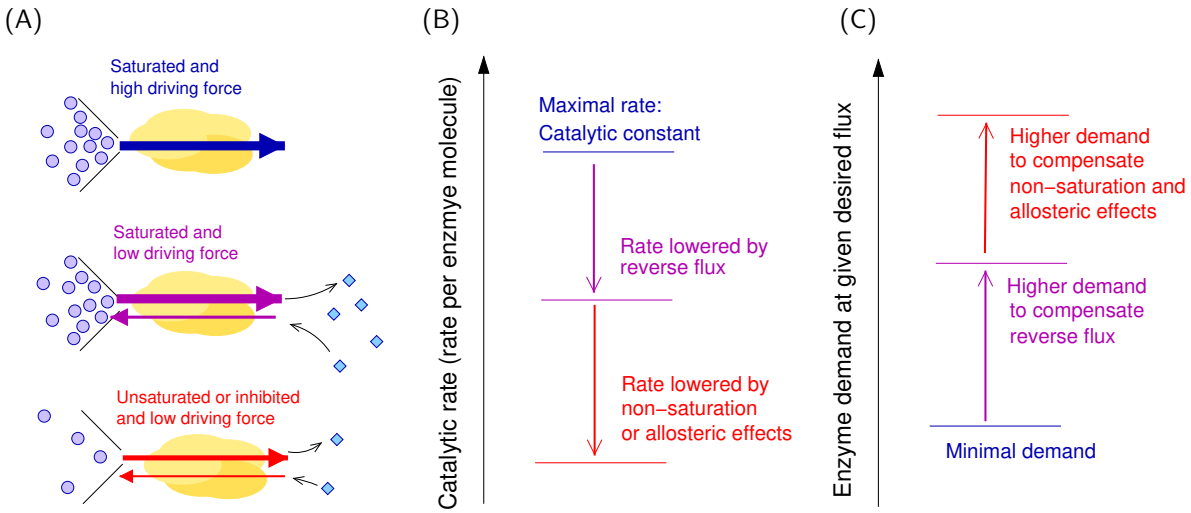

Figure 6.2: Enzyme cost in metabolism – (A) Enzyme-specific flux depends on a number of physical factors. Under ideal conditions, an enzyme molecule catalyses its reaction at a maximal rate given by the enzyme's forward catalytic constant (blue). The rate is reduced by microscopic reverse fluxes (magenta) and by incomplete saturation with substrate, causing waiting times between reaction events, or by enzyme inhibition or incomplete activation (red). (B-C) On a logarithmic scale, catalytic rates and enzyme demand can be split into sums of efficiency terms. With lower catalytic rates, larger amounts of enzyme are required for realizing the same metabolic flux.

where h is a number converting enzyme concentration e into enzyme amount q (for example, the enzyme molecular mass). For an illustration, see Figure 6.2 . Summing up the demand across all the reactions in the pathway (each with its own rate, kinetic parameters, and substrate/product concentrations) will produce the total enzyme demand. Looking at this function, we can already make some interesting observations. First, the kinetic parameters (k^+_{cat}, K_p, and K_s) can be treated as constants since they change only in evolutionary timescales, and we often assume that existing enzymes already have near-optimal kinetics (although that's not always the case). Since we care about the demand *per pathway flux* one can, without loss of generality, set v to 1. However, if the pathway requires a non-trivial ratio between some reactions, the value of v can be different based on the stoichiometry. Finally, the thermodynamic term, i.e. $1 - e^{\Delta_{\mathrm{r}}G'/RT}$ (which we will discuss in more detail in the following section, 6.3), is a function of the metabolite concentrations and the K_{eq}, which is another constant. So, generally speaking, enzyme demand is defined by a set of constants that are unique to each pathway, and variables that represent the metabolite concentrations. Since these concentrations are subject to change depending on the growth conditions, we often treat them as optimization variables and try to find the minimal demand possible within certain constraints. In Section 6.4, we will see a general method for finding the minimal value using convex optimization.

Most of the proposed criteria for good pathways have either to do with material investments (such as substrate, cofactor, or energy demand) or with "machine investments", that is, enzyme demands. Enzyme demands, in turn, depend on pathway length, enzyme masses, and enzyme efficiency, and therefore on rate laws (where k_{cat} values, thermodynamic forces, and metabolite concentrations come into play). In fact, many criteria which we discussed earlier as indicators of efficiency are actually an approximation of the enzyme demand under certain assumptions. For example, the number of steps is proportional to the total demand if all enzymes have exactly the same k^+_{cat}, saturation, and thermodynamics. Therefore, it is quite a useful rule-of-thumb in case not much else is known about the enzymes themselves. A better approximation, denoted *Pathway Specific Activity*, was used by [7] to compare CO_2 fixation cycles. If we assume that all enzymes are fully saturated and irreversible, the demand would be a direct function of the individual enzyme specific activities (specifically, proportional to the sum of all their reciprocal values). But even if we know nothing about the enzyme kinetic parameters, thermodynamics alone can provide us with useful information with which to grade pathways. Specifically, the K_{eq} of a reaction is a universal constant that is not

affected by enzymes, but rather determined solely by the chemical structures of the substrates and products.

In the following sections, we will focus on enzyme use efficiency as a main objective and consider a thermodynamic approximation, relating enzyme demands to thermodynamic forces. For linear metabolic pathways, optimal enzyme profiles (and the associated metabolite profiles and enzyme costs) can be computed with closed formulae. We will also discuss a way to compute optimal enzyme profiles numerically, for networks of any shape and size, as long as the flux mode is known.

6.3. The role of thermodynamics

In general, when considering larger metabolic networks, thermodynamic feasibility can play an important or even crucial role in determining which pathways are used. In this section we will discuss this role more explicitly and see how thermodynamics can still give us useful insights about pathway efficiency even when no other kinetic data is available.

Why are thermodynamic driving forces a meaningful criterion for good pathways? In brief, the driving forces, defined as $\theta \equiv -\Delta_{\mathrm{r}}G'/RT$, play a double role: first, they determine whether or not a pathway flux is feasible at all, given the metabolite concentrations at the pathway boundary (i.e. the metabolites that form connections to the broader metabolic network); and second, in case the pathway *is* feasible, driving forces can affect enzyme efficiency and, consequently, the enzyme demand for a given desired pathway flux. In Chapter 3 in [3], we learned that $\Delta_{\mathrm{r}}G'$, and hence the driving force θ, depends on the equilibrium constant K_{eq} of the reaction and on the substrate and product concentrations. We also learned that for a flux in forward direction, the driving force must be positive. Beyond that, the efficiency of an enzyme is proportional to $\eta^{\mathrm{for}}(\theta) = 1 - \mathrm{e}^{-\theta}$, a function that ranges between 0 (for $\theta = 0$, reactions in thermodynamic equilibrium) and 1 ($\theta \gg 1$, reactions far from equilibrium). Let us now see how this non-equilibrium relation affects pathway efficiency.

6.3.1. Enzyme kinetics and driving forces

We should remind ourselves some of the lessons learned in Chapter 3 in [3]. Specifically, recall the factorized rate law [16] with a reversibility term that is an explicit function of the Gibbs energy (Eq. 3.10 in [3]):

$$v = e \cdot k_{\mathrm{cat}}^{+} \cdot \frac{\prod_i s_i^{\nu_i}/K_s}{1 + \prod_j p_j^{\nu_j}/K_p + \prod_i s_i^{\nu_i}/K_s} \cdot \left(1 - e^{\Delta_{\mathrm{r}}G'/RT}\right). \tag{6.7}$$

The enzyme mechanism behind this formula assumes fast binding and unbinding of substrate and product, and a slow reversible conversion step (of bound substrate into bound product). Note that here we generalize the rate law for cases with more than one substrate and one product, where ν_i and ν_j are the stoichiometric coefficients of substrates and products, respectively[1]. This generalization is one out of many, and corresponds to the assumption that all reactants bind independently to the enzyme (and at random order). We focus on this rate law because it is one of the simplest, but the theoretical results in this chapter apply to most other generalizations as well (e.g. convenience kinetics [17]).

According to the definition of k_{cat}^{+}, and also by noticing that the middle and rightmost terms in Eq. (6.7) are each smaller than 1, the rate of an enzymatic reaction is bounded by $v \leq e \cdot k_{\mathrm{cat}}^{+}$ (see Mathematical Details Box 6.C for a detailed explanation). However, the additional terms are often much lower than 1, which means that the rate does not reach its maximum. If we try to measure the apparent catalytic rate by dividing the rate by the enzyme abundance ($k_{app} = v/E$) we would typically get a value that is lower than k_{cat}^{+}, while only in rare "ideal" cases,

[1] In general, reaction stoichiometries can be arbitrarily scaled. For example, instead of a reaction 2 A $\rightarrow$ B, we may write A $\rightarrow \frac{1}{2}$ B for convenience, which will only lead to a scaling factor in the reaction rate. However, this holds only if reaction stoichiometries are used to describe mass-balance. In cases like Eq. (6.7), where stoichiometries appear in kinetic rate laws or in thermodynamic balances, we do not have this choice. In these cases, the stoichiometries must reflect the molecularities, that is, the actual number of reactant molecules involved in the enzymatic reaction.

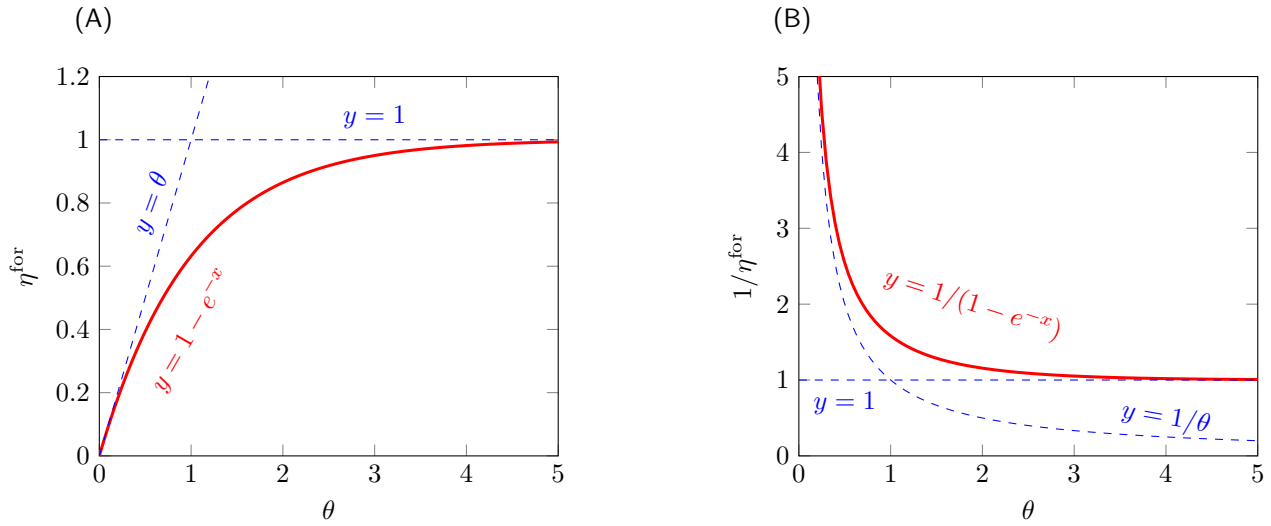

Figure 6.3: The thermodynamic efficiency term η^{for} and some approximations – (A) In a given reaction, the thermodynamic efficiency term $\eta^{\text{for}} = 1 - e^{-\theta}$ (solid line) can vary between 0 and 1 depending on the driving force θ. Small driving forces make the enzyme inefficient, since $\eta^{\text{for}} \to 0$, while for large forces, thermodynamics does not play a role as $\eta^{\text{for}} \to 1$. The dashed lines show two linear approximation that hold always as bounds, but can also be used as good approximations for small or large θ values, respectively: $(1 - e^{-\theta}) < \theta$ and $(1 - e^{-\theta}) < 1$. (B) The reciprocal value $1/\eta^{\text{for}}$ is one of the factors determining enzyme demand. The solid line shows the thermodynamic demand factor $1/\eta^{\text{for}}$, while the dashed lines show the resulting approximations $1/\eta^{\text{for}} > 1/\theta$ and $1/\eta^{\text{for}} > 1$, corresponding respectively to the enzyme demand approximations $E \geq \frac{v}{k_{\text{cat}}\theta}$ and $E \geq \frac{v}{k_{\text{cat}}}$.

k_{app} would approach the k_{cat}^{+}. In fact, this reasoning was used by Davidi et al. [18] to estimate the k_{cat}^{+} values of more than 100 enzymes in *E. coli*, where they sampled many growth conditions and took the maximum k_{app} as the estimate.

As discussed in Section 3.2.3 in [3], the factorized rate law has a thermodynamic perspective based on the flux-force relationship, where we view the reversibility term as a "penalty" for the fact that by lowering the energy barrier, enzymes must catalyze reactions in both directions. When the driving force (θ) is low, the reverse reaction flux can become significant and lower the net flux. On the other hand, if the driving force is large enough, this term can be ignored and the rate law resembles irreversible kinetics .

So far we've seen that increasing the driving force of a single reaction translates to a better enzyme efficiency and lower demand. If we consider whole pathways, ones whose overall driving force is larger have more of it to distribute among the reactions and therefore should also have higher efficiencies overall. However, using "too much" driving force can also have downsides. Using a larger amount of the Gibbs energy to drive the pathway reactions means that less of that energy would go for building biomass or currency metabolites such as ATP. An example for this trade-off between the efficiency of single enzymes (in terms of backward rates) and the overall pathway efficiency (in terms of ATP yield) was demonstrated by Flamholz et al. [19] who analyzed two versions of the famous glycolytic pathway (see Figure 6.1 below).

6.3.2. Driving forces should not be too small

With the factorized rate law 6.7, we can approximate the reaction rates by $v \leq e\, k_{\text{cat}}\, (1 - \mathrm{e}^{-\theta})$ (where we assume positive fluxes by convention). The thermodynamic efficiency $\eta^{\text{for}} = 1 - \mathrm{e}^{-\theta}$ plays a prominent role. As shown in Figure 6.3, this formula yields two important approximations: for small forces θ, that is, close to equilibrium, we obtain $\eta^{\text{for}} \approx \theta$, while for large forces, that is, for strongly forward-driving reactions, we obtain $\eta^{\text{for}} \approx 1$. In fact, both approximations also serve as upper bounds across all θ values. What does this mean? Far from equilibrium, the

thermodynamic term does not play a role and can be ignored. Close to equilibrium, in contrast we obtain a simple approximation for fluxes

$$v < e \cdot k_{\mathrm{cat}}^{+} \cdot (1 - e^{-\theta}) < e \cdot k_{\mathrm{cat}}^{+} \cdot \theta \tag{6.8}$$

and hence for the enzyme demand

$$e > \frac{v}{k_{\mathrm{cat}}^{+} \cdot (1 - e^{-\theta})} > \frac{v}{k_{\mathrm{cat}}^{+} \cdot \theta}\,. \tag{6.9}$$

As θ goes to zero, the enzyme demand (for a given desired flux) goes to infinity. We already know the reason from Chapter 3 in [3]: the driving force determines the ratio of forward and reverse one-way fluxes, $\frac{v^+}{v^-} = \mathrm{e}^{\theta}$. If θ comes close to zero, their relative difference becomes very small, and in order to obtain a given net flux $v = v^+ - v^-$, both v^+ and v^- must grow enormously, which would require an a large amount of enzyme. This effect concerns only very small θ values - for θ much larger than 1 (or $\Delta_{\mathrm{r}}G'$ much smaller than -RT), it can be neglected. Therefore, redistributing driving forces between reactions, to avoid very small forces, can save enzyme costs. The relation between driving forces, enzyme efficiency enzyme demand is shown in more detail in Figure 6.4.

If small driving forces should be avoided to prevent enzyme costs from going infinity, how can this happen in practice? The driving forces themselves depend on metabolite levels, which can vary over several orders of magnitude. While the true metabolite concentrations are usually unknown, we hypothesize that selection favors concentration profiles that prohibit very small driving forces, in order to escape the ensuing large enzyme demands. Of course, completely avoiding small driving forces may be impossible, as there is always a trade-off: if a metabolite concentration decreases, the driving forces of all reactions producing it will increase, but the driving forces of all reactions consuming it will decrease simultaneously. So, all else being equal, the optimal metabolite profile is one that distributes its driving forces as evenly as possible.

6.3.3. Max-Min driving force method

Previously in Chapter 4.3.2 in [3], we discussed adding thermodynamic constraints to constraint-based models in order to comply with the second law of thermodynamics. We can extend that approach in order to implement the idea of avoiding small driving forces. When we talk about the thermodynamic profile of a metabolic pathway, we usually try to visualize it by the cumulative Gibbs energy of reaction: we start at 0 and at each step add the $\Delta_{\mathrm{r}}G'$ of the next reaction, which, assuming the pathway is feasible, is a negative number. The profile therefore has a shape of a downhill slope. The end point represents the total Gibbs energy and depends only on the concentrations of the metabolites that are part of the net reaction. Intermediate metabolites do not affect it, but they do determine the shape of the profile itself (see Figure 6.4). Specifically, each intermediate metabolite typically affects the driving force of two reactions – the one producing it and the one consuming it – with opposite signs. Therefore, changing the concentration of an intermediate can help increase the driving force of one reaction, but always at the expense of another reaction. This strong coupling between $\Delta_{\mathrm{r}}G'$ is why it is not trivial to find the optimal thermodynamic profile of a pathway.

The Max-Min driving force method (MDF) [20] is a method for predicting metabolite concentrations, based on the principle of evenly distributed driving forces. All fluxes are fixed and given, and assumed to be positive. It assumes that each metabolite concentration must remain in a predefined range, converts each choice of metabolite concentrations into the corresponding pattern of driving forces, and determines the *smallest* resulting driving force in the network. If this smallest driving force is negative, the flux distribution cannot be realized thermodynamically. Otherwise, the larger this smallest driving force, the better the overall metabolite profile. Hence, among all possible metabolite profiles, MDF predicts the one that maximizes the value of the minimal driving force across the network. Mathematically, this leads to a linear optimization problem: in the space of logarithmic metabolite concentrations,

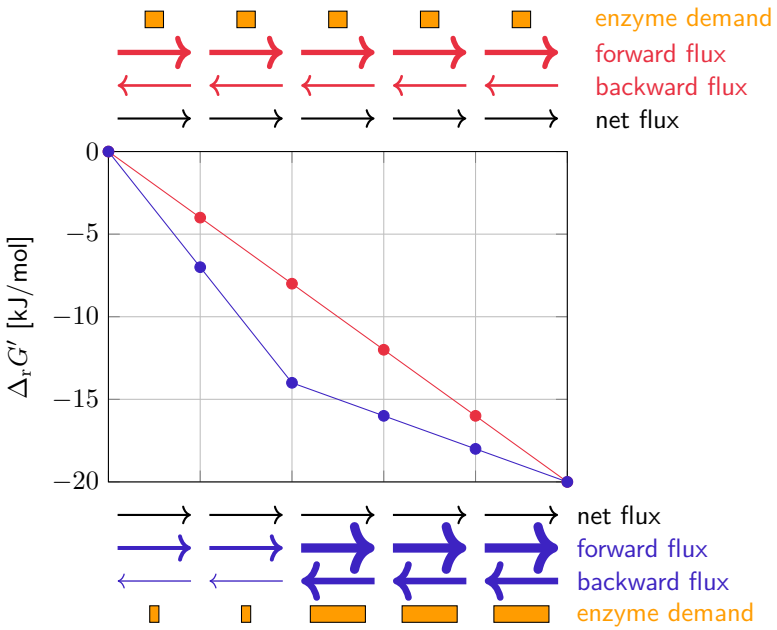

Figure 6.4: Thermodynamic forces, enzyme efficiency, and enzyme demand in a linear chain of reactions – The plot in the center represent two possible profiles of the thermodynamic driving forces (blue and red). The curves describe the cumulative $\Delta_r G'$ values: while the total $\Delta_r G'$ is fixed (and determined by external metabolite concentrations), the shape of the profile can vary. In the optimal profile (in red), small driving forces are avoided. The driving forces determine the ratios of forward and backward one-way fluxes (red arrows), and at a given net flux (black arrows) the enzyme demands. In the suboptimal blue curve, in contrast, the last three reactions show lower forces, and therefore relatively high reverse fluxes (blue arrows); to obtain the same net flux, forward and backward fluxes have to be strongly increased, which increases the enzyme demand.

a lower bound on all driving forces (denoted B) is maximized (Eq. 6.10). An illustrative example is shown in Figure 6.5.

$$\begin{aligned} \text{Maximize}_{\mathbf{x},B} \quad & B \\ \text{Subject to} \quad & -(\mathbf{\Delta_r G'^{\circ}} + RT \cdot \mathbf{N}^{\top}\mathbf{x}) \geq B \\ & \ln(\mathbf{C_{min}}) \leq \mathbf{x} \leq \ln(\mathbf{C_{max}}) \end{aligned} \tag{6.10}$$

MDF is easy to apply: it is based on a simple Linear Programming problem and requires only the following input data: (i) the stoichiometric network; (ii) the flux directions; (iii)) the known equilibrium constants (or equivalently, the standard reaction Gibbs free energies); (iv) physiological ranges for metabolite concentrations. Based on these data alone, metabolite concentrations and driving forces (or $\Delta_r G'$ values) are predicted. An example application can be found in Hädicke et al. [21], where the potential of CO_2 fixation in *E. coli* via endogenous pathways was analyzed using MDF.

A theoretical insight from MDF is the notion of distributed bottlenecks. A simple bottleneck would consist of a single reaction whose driving force cannot be increased because the substrates are at their upper concentration bounds and the products are at their lower concentration bounds. Given the fixed equilibrium constant, nothing can be done to increase the driving force in this reaction. A distributed bottleneck is more complicated: it consists of a series of reactions that all share the same low driving force, which, because of all the concentration constraints in the system, cannot be further increased (e.g. as in Figure 6.4). Even though each single reaction looks "harmless" because its

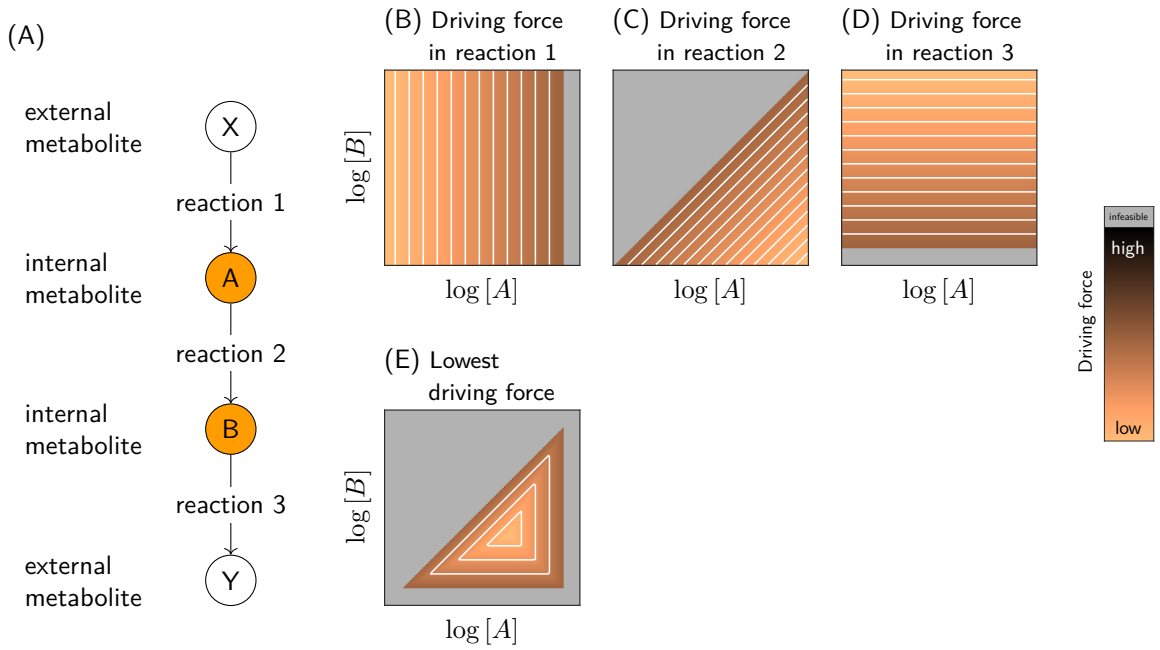

Figure 6.5: Max-Min Driving force method (MDF): an optimality problem in metabolite space – (A) Example pathway with given equilibrium constants and fixed concentrations of the external metabolites X and Y. What are the most favorable concentrations of the internal metabolites A and B? Assuming that small driving forces should be avoided in all reactions, MDF determines the metabolite profile that optimizes a worst case: it maximizes the worst (that is, smallest) driving force among all three reactions. (B) Driving force in reaction 1, as a function of the logarithmic concentrations of A and B, called $\ln a$ and $\ln b$. Higher concentrations of A (the reaction product) lead to smaller driving forces. Above a critical value (where X and A are in equilibrium), the driving force becomes negative, and a forward flux is impossible (gray region). The concentration of B, which does not participate in the reaction, does not play a role. (C) Driving force for reaction 2. Here, it is the ratio b/a that counts. The lower the ratio (lower right), the higher the driving force. If the ratio is higher than the equilibrium constant, the driving force becomes negative (grey region). (C) Driving force for reaction 3. (E) By overlaying the contours in (B), (C), and (D) and taking the minimum value, we obtain the minimal driving force $\theta^{\min}$ among all three reactions. $\theta^{\min}$ is a piecewise linear function of $\ln a$ and $\ln b$ within the feasible range, yielding positive forces in all three reactions. The maximum point of this function is the optimum metabolite profile predicted by MDF. In the example shown, the feasible concentration space is entirely defined by the driving forces themselves, given the external concentrations. In general, physiological concentration ranges for all metabolites could further decrease the solution space and shift the optimum point (not shown).

own driving force could still be increased, this increase would happen at the expense of other driving forces.

6.3.4. The role of thermodynamics for metabolic states

In summary, thermodynamics provides important clues both about the feasibility of pathways fluxes and about their enzyme demand. To use this knowledge, fluxes need to be considered together with metabolite concentrations (to obtain the possible driving forces), but no detailed knowledge of enzyme kinetics is required. Thermodynamics alone yields an upper bound on fluxes (and hence, a lower bound on enzyme demands) that holds for any kinetic rate laws. The only required data (except for the metabolic network itself) are equilibrium constants (or equivalently, standard Gibbs free energies of reactions $\Delta_r G'^{\circ}$), which can be obtained from the eQuilibrator tool (equilibrator.weizmann.ac.il) [22, 23, 24] as well as physiological bounds on metabolite concentrations. Given this information, and given a feasible choice of metabolite concentrations, we can compute the driving forces of all reactions, and from the factorized rate law (and assuming positive fluxes by convention) we can then approximate the reaction rates by $v \leq e\, k_{\text{cat}}\, (1 - \mathrm{e}^{-\theta})$.

We also recall from Chapter 3 in [3] that driving forces are not independent between reactions, but depend on the metabolite concentrations, which creates trade-offs: in a chain A $\stackrel{R_1}{\rightarrow}$ B $\stackrel{R_2}{\rightarrow}$ C, a lower concentration of B will increase

the driving force in R_1, but decrease the driving force in R_2. For high enzyme efficiency (low enzyme demand), all driving forces should in principle be high, but this is most important for low θ values (while for $\theta \gg 1$ it does not even matter). Therefore we may conclude that, to save enzyme, a cell should rearrange its metabolite levels within physiological bounds such that small θ are avoided. Implementing this as an optimality problem, we obtain MDF.

In conclusion, we described (i) a general rule of thumb that poor thermodynamics makes reactions costly; (ii) simple approximations of enzyme cost; and (iii) practical methods (MDF) to obtain metabolite profiles with favorable thermodynamic properties.

6.4. Enzyme cost minimization

6.4.1. Enzyme cost minimization

The problem of minimizing the total enzyme demand (or cost) for a given pathway can be solved numerically, thanks to the fact that they are always convex [25]. Finding the minimum of the convex objective (the total enzyme cost) in a convex set (the set of admissible metabolite profiles, a convex polytope in log-metabolite space) can be done efficiently. In contrast to general optimality problems, such problems have a unique local optimum, which can be found by simple numerical methods. In this section, we demonstrate it with a simple example, the same three-reaction pathway that you already saw in Section 6.3 above.

6.4.2. Enzyme cost landscape of a metabolic pathway

Given the fluxes, kinetics, and concentration bounds in a metabolic pathway model, one can predict the enzyme demand by assuming that cells minimize the enzyme cost in that pathway. In the Enzyme Cost Minimization method A reaction rate $v = e \cdot f(\mathbf{c})$ depends on enzyme level e and metabolite concentrations c_i through the enzymatic rate law, $f(\mathbf{c})$. If the metabolite concentrations were known, we could directly compute enzyme demands $e = v/f(\mathbf{c})$ from fluxes, and similarly calculate the flux-specific enzyme demand $e/v = 1/f(\mathbf{c})$. However, metabolite concentrations are usually unknown and vary between experimental conditions. Therefore, there can be many solutions for e and $\mathbf{c}$ realizing one flux distribution. To select one of them, we employ an optimality principle: we define an enzyme cost function (for instance, total enzyme mass) and choose the enzyme profile with the lowest cost while restricting the metabolite levels to physiological ranges and imposing some thermodynamic constraints. As we shall see below, the solution is in many cases unique.

Let us demonstrate this procedure with a simple example (Figure 6.6 (a)). In the pathway $X \rightleftharpoons A \rightleftharpoons B \rightleftharpoons Y$, the external metabolite levels [X] and [Y] are fixed and given, while the intermediate levels [A] and [B] need to be found. As rate laws for each of the three reactions, we use reversible Michaelis-Menten (MM) kinetics

$$v = E\,\frac{k^+_{\mathrm{cat}}\, s/K_{\mathrm{S}} - k^-_{\mathrm{cat}}\, p/K_{\mathrm{P}}}{1 + s/K_{\mathrm{S}} + p/K_{\mathrm{P}}} \tag{6.11}$$

with enzyme level E, substrate and product levels s and p, turnover rates k^+_{cat} and k^-_{cat}, and Michaelis constants K_{S} and K_{P}. In kinetic modeling, steady-state concentrations would usually be obtained from given enzyme levels and initial conditions through numerical integration. Here, instead, we fix a desired pathway flux v and compute the enzyme demand as a function of metabolite concentrations:

$$E(s,p,v) = v\,\frac{1 + s/K_{\mathrm{S}} + p/K_{\mathrm{P}}}{k^+_{\mathrm{cat}}\, s/K_{\mathrm{S}} - k^-_{\mathrm{cat}}\, p/K_{\mathrm{P}}}. \tag{6.12}$$

Figure 6.6 shows how the enzyme demand in each reaction depends on the logarithmic reactant concentrations. To obtain a positive flux, substrate levels s and product levels p must be restricted: for instance, to allow for a positive flux in reaction 2, the rate law numerator $k^+_{\mathrm{cat}}\,[A]/K_{\mathrm{S}} - k^-_{\mathrm{cat}}\,[B]/K_{\mathrm{P}}$ must be positive. This implies

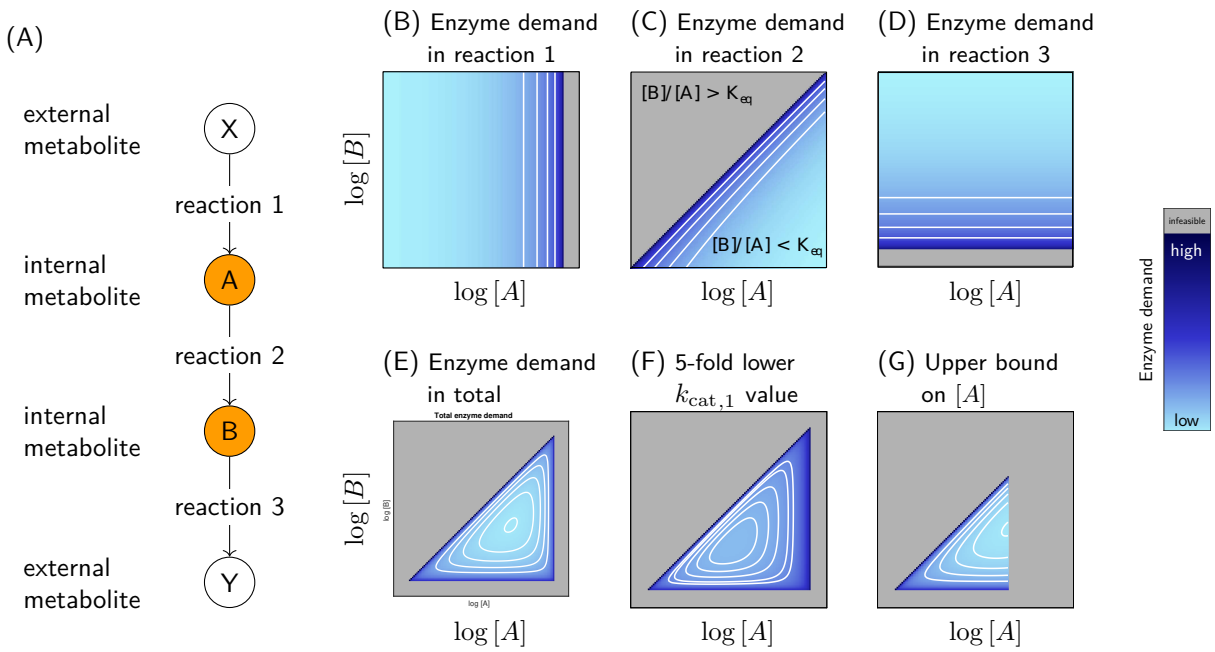

Figure 6.6: Enzyme demand in a metabolic pathway – (A) Pathway with reversible Michaelis-Menten kinetics (equilibrium constants, catalytic constants, and K_{M} values are set to values of 1, $[A]$ and $[B]$ denote the variable concentrations of intermediates A and B in mM). The external metabolite concentrations $[X]$ and $[Y]$ are fixed. Plots (B)-(D) show the enzyme demand of reactions 1, 2, and 3 at given flux $v = 1$ according to Eq. (6.12). Grey regions represent infeasible metabolite profiles. At the edges of the feasible region (where A and B are close to chemical equilibrium), the thermodynamic driving force goes to zero. Since small forces must be compensated by high enzyme levels, edges of the feasible region are always dark blue. For example, in reaction 1 (panel (B)), enzyme demand increases with the level of A (x-axis) and goes to infinity as the mass-action ratio $[A]/[X]$ approaches the equilibrium constant (where the driving force vanishes). (E) Total enzyme demand, obtained by summing all enzyme levels. The metabolite polytope – the intersection of feasible regions for all reactions – is a triangle, and enzyme demand is a cup-shaped function on this triangle. The minimum point defines the optimal metabolite concentrations and optimal enzyme levels. (F) As the k_{cat} value of the first reaction is lowered by a factor of 5, states close to the triangle edge of reaction 1 become more expensive and the optimum point is shifted away from the edge. (G) The same model with a physiological upper bound on the concentration $[A]$. The bound defines a new triangle edge. Since this edge is not caused by thermodynamics, it can contain an optimum point, in which driving forces are far from zero and enzyme costs are kept low. Please note the resemblance to the MDF problem for the same pathway, shown in Figure 6.5.

that $[B]/[A] < K_{\mathrm{eq}}$ where the reaction's equilibrium constant K_{eq} is determined by the Haldane relationship, $K_{\mathrm{eq}} = (k^+_{\mathrm{cat}}/k^-_{\mathrm{cat}}) \cdot (K_{\mathrm{P}}/K_{\mathrm{S}})$. With all model parameters set to 1, we obtain the constraint $[B]/[A] < 1$, i.e. $\ln[B] - \ln[A] < 0$, putting a straight boundary on the feasible region (Figure 6.6 (c)). Close to chemical equilibrium ($[B]/[A] \approx K_{\mathrm{eq}}$), the enzyme demand e_2 approaches infinity. Beyond that ratio ($[B]/[A] > K_{\mathrm{eq}}$) no positive flux can be achieved (grey region). Such a threshold exists for each reaction (see Figure 6.6 (b)-(d)). The remaining feasible metabolite profiles form a triangle in log-concentration space, which we call *metabolite polytope* $\mathscr{P}$ (Figure 6.6 (e)), and Eq. (6.12) yields the total enzyme demand $e_{\mathrm{tot}} = e_1 + e_2 + e_3$, as a function on the metabolite polytope. The demand increases steeply towards the edges and becomes minimal in the center. The minimum point marks the optimal metabolite profile, and via Eq. (6.12) we obtain the resulting optimal enzyme profile.

The metabolite polytope and the large enzyme demand at its boundaries follow directly from thermodynamics. To see this, we consider the unitless *thermodynamic driving force* $\Theta = -\Delta_{\mathrm{r}}G'/RT$ [26] derived from the reaction Gibbs free energy $\Delta_{\mathrm{r}}G'$. The thermodynamic force can be written as $\Theta = \ln \frac{K_{\mathrm{eq}}}{[B]/[A]}$, i.e. the driving force is positive whenever $[B]/[A]$ is smaller than K_{eq}, and it vanishes if $[B]/[A] = K_{\mathrm{eq}}$. How is this force related to enzyme cost? A reaction's net flux is given by the difference $v = v^+ - v^-$ of forward and backward fluxes, and the ratio v^+/v^-

depends on the driving force as $v^+/v^- = \mathrm{e}^{\Theta}$. Thus, only a fraction $v/v^+ = 1 - \mathrm{e}^{-\Theta}$ of the forward flux acts as a net flux, while the remaining forward flux is partially canceled by the backward flux. Close to chemical equilibrium, where the mass-action ratio $[B]/[A]$ approaches the equilibrium constant K_{eq}, the driving force goes to zero, the reaction's backward flux increases, and the flux per enzyme level drops. This is what happens at the triangle edges in Figure 6.6: a reaction approaches chemical equilibrium, the driving force Θ goes to zero, and large enzyme amounts are needed for compensation. Exactly on the edge, the driving force vanishes and no enzyme level, no matter how large, can support a positive flux. The quantitative cost depends on model parameters: for example, by lowering a k_{cat} value, the increase in enzyme cost at the boundary becomes steeper and the optimum point is shifted away from the boundary (see Figure 6.6 (f)).

6.4.3. Enzyme cost as a function of metabolite concentrations

The prediction of optimal metabolite and enzyme levels can be extended to models with general rate laws and complex network structures. In general, enzyme demand depends not only on driving forces and k_{cat} values, but also on the kinetic rate law, which includes K_{M} values and small-molecule regulation. We can conveniently model or approximate these factors by using factorized rate laws. Let us write this rate laws here again in a general form to see the different factors at play. As we learned in Section 6.2, the rate of a reaction depends on enzyme level e, forward catalytic constant k^+_{cat} (i.e. the maximal possible forward rate per unit of enzyme, in s^{-1}), driving force (i.e. the ratio of forward and backward fluxes), and on kinetic effects such as substrate saturation or small-molecule regulation. If all active fluxes are positive, reversible rate laws like the Michaelis-Menten kinetics in Eq. (6.11) can be factorized as [16]:

$$v = e \cdot k^+_{\mathrm{cat}} \cdot \eta^{\mathrm{for}} \cdot \eta^{\mathrm{sat}} \cdot \eta^{\mathrm{reg}}. \tag{6.13}$$

Negative fluxes, which would complicate our formulae, can be avoided by orienting the reactions in the direction of fluxes.

Enzyme demand can be quantified as a concentration (e.g. enzyme molecules per volume) or mass concentration (where enzyme molecules are weighted by their molecular weights). If rate laws, fluxes, and metabolite concentrations are known, the enzyme demand of a single reaction l follows from Eq. (6.13) as

$$e_l(\mathbf{c}, v_l) = v_l \cdot \frac{1}{k^+_{\mathrm{cat},l}} \cdot \frac{1}{\eta^{\mathrm{for}}_l(\Theta(\mathbf{c}))} \cdot \frac{1}{\eta^{\mathrm{sat}}_l(\mathbf{c})} \cdot \frac{1}{\eta^{\mathrm{reg}}_l(\mathbf{c})}. \tag{6.14}$$

To determine the enzyme demand of an entire pathway, we sum over all reactions: $E^{\mathrm{path}}_{\mathrm{tot}} = \sum_l e_l$. Based on its enzyme demands e_l, we can associate each metabolic flux with an enzyme cost $q = \sum_l h_{e_l} e_l$, describing the effort of maintaining the enzymes. The burdens h_{e_l} of different enzymes represent, e.g. differences in molecular mass, post-translational modifications, enzyme maintenance, overhead costs for ribosomes, as well as effects of misfolding and non-specific catalysis. The enzyme burdens h_{e_l} can be chosen heuristically, for instance, depending on enzyme sizes, amino acid composition, and lifetimes. Setting $h_{e_l} = m_l$ (protein mass in grams per mole), q will be in gram protein per gram cell dry weight. Considering the specific amino acid composition of enzymes, we can also assign specific costs to the different amino acids. Alternatively, an empirical cost per protein amount can be established by the level of growth impairment that an artificial induction of protein would cause [27, 28]. Thus, each reaction flux v_l is associated with an enzyme cost q_l, which can be written as a function $q_l(v_l, \mathbf{c}) \equiv h_{e_l} e_l(\mathbf{c}, v_l)$ of flux and metabolite concentrations. From now on, we refer to log-scale metabolite concentrations $s_i = \ln c_i$ to obtain simple optimality problems below. From the factorized rate law Eq. (6.14), we obtain the enzyme cost function

$$q(\mathbf{s}, \mathbf{v}) \equiv \sum_l h_{e_l} e_l(v_l, \mathbf{s}) = \sum_l h_{e_l} \cdot v_l \cdot \frac{1}{k^+_{\mathrm{cat},l}} \cdot \frac{1}{\eta^{\mathrm{for}}_l(\mathbf{s})} \cdot \frac{1}{\eta^{\mathrm{sat}}_l(\mathbf{s})} \cdot \frac{1}{\eta^{\mathrm{reg}}(\mathbf{s})} \tag{6.15}$$

for a given pathway flux $\mathbf{v}$. If the fluxes are fixed and given, our enzyme cost becomes, at least formally, a function

of the metabolite levels. The cost function is defined on the metabolite polytope $\mathscr{P}$, a convex polytope in log-concentration space containing the feasible metabolite profiles. Like the triangle in Figure 6.6, the polytope is defined by physiological and thermodynamic constraints.

Beyond minimizing the total enzyme cost, one can also use Enzyme Cost Minimization to analyze the individual enzyme demands. When the metabolite levels are known, the demand can be directly calculated and each efficiency factor (η) in Eq. (6.15). By omitting some factors or replacing them by constant numbers $0 < \eta \leq 1$, simplified enzyme cost functions with fewer parameters can be obtained. For example, $\eta^{\mathrm{for}} = 1$ would imply an infinite driving force $\Theta \to \infty$ and a vanishing backward flux, $\eta^{\mathrm{sat}} = 1$ implies full substrate saturation, and $\eta^{\mathrm{reg}} = 1$ implies full enzyme activation and no enzyme inhibition (or no small-molecule regulation at all). In these limiting cases, enzyme activity will not be reduced, and enzyme demand will be given by the capacity-based estimate v/k^{+}_{cat}, a lower estimate of the actual demand. Instead of omitting an efficiency factor, it can also be set to a constant value between 0 and 1. Such simplifications and the resulting enzyme cost functions with fewer parameters can be practical if kinetic constants are unknown.

6.4.4. General lessons from Enzyme Cost Minimization

Enzyme cost minimization not only provides numerical solutions, but also some general insights.

1. **Convexity** Enzyme Cost Minimization shows again the importance of the metabolite polytope. The usage of logarithmic metabolite concentrations not only leads to a good search space for feasible metabolite profiles (as in MDF), but also facilitates optimization because enzyme cost is a convex function of the metabolite log-concentrations [29]. Convexity makes this optimization tractable and scalable – unlike a direct optimization in enzyme space. Convexity holds for a wide range of rate laws and for extended versions of the problem, e.g. including bounds on the sum of (non-logarithmic) metabolite concentrations or bounds on weighted sums of enzyme fractions.

2. **Factorized rate laws disentangle individual enzyme cost effects** To see how metabolic states are shaped by different physical factors, we considered factorized rate laws. The different terms in these functions represent specific physical factors and require different kinetic and thermodynamic data for their calculation. By neglecting some terms, one obtains different approximations of the true enzyme cost. By comparing the different scores, we can estimate the enzyme cost that cells "pay" for running reactions at small driving forces (to save Gibbs free energy) or for keeping enzymes beneath substrate-saturation (e.g., to dampen fluctuations in metabolite levels).

3. **Relationship to other optimality approaches** Beyond their practical advantages, factorized enzyme cost functions also allow us to easily compare our method to earlier modeling and optimization approaches. These approaches typically focused on only one or two of the factors that are taken into account in Enzyme Cost Minimization, and many of them can be reformulated as approximations of this method [20, 30, 12].

4. **Enzyme cost is related to thermodynamics** In FBA, thermodynamic constraints and flux costs appear as completely unrelated aspects of metabolism (as is explained in Chapter 5 in [3]). Thermodynamics is used to restrict flux directions, and to relate them to metabolite bounds, while flux costs are used to suppress unnecessary fluxes. In Enzyme Cost Minimization, thermodynamics and flux cost appear as two sides of the same coin. Like in FBA, flux profiles are thermodynamically *feasible* if they lead to a non-empty metabolite polytope, allowing for positive forces in all reactions. However, the values of these forces also play a role in shaping the enzyme cost function on that polytope. Together, metabolite polytope and enzyme cost function (as in Figure 6.6) summarize all relevant information about flux cost.

Many pathways are regulated, for instance by feedback inhibition of enzymes via the end product. While this may stabilise the dynamics and adapt it to current demands, such enzyme regulation comes at a cost, which we can

estimate by following the logic of Enzyme Cost Minimization. Many enzymes are regulated by small molecules that act as competitive or allosteric inhibitors [31], an effective way to implement feedback control, for example to adapt the flux in biosynthesis pathways to current needs. In order for such a regulation to work, the enzyme needs to be partially inhibited on average (because only then, its activity can be increased on demand, by alleviating the inhibition). Therefore, the enzyme efficiency goes down, and the cell needs to provide more enzyme to catalyze the same flux than without the inhibition.

How much will this regulation cost the cell as part of the enzyme budget? From the perspective of Enzyme Cost Minimization, where we start from desired fluxes and compute the enzyme demand, this question is easy to answer: in the inhibited enzyme case, the lower efficiency will be described by a factor $\eta^{\mathrm{reg}} \in [0, 1]$ (Mathematical Details Box 6.C). In the same reaction, the enzyme demand increases by a factor $1/\eta^{\mathrm{reg}}$, so the extra cost is simply $1/\eta^{\mathrm{reg}} - 1$ times the "baseline" cost of this enzyme (without inhibition). Specifically, a non-competitive inhibitor, with efficiency factor $\eta^{\mathrm{reg}} = \frac{1}{1 + c/K_I}$ yields a cost factor $1+c/K_I$. If the metabolite concentrations are fixed, this corresponds to an extra enzyme demand $\Delta e_l = \frac{e_l\, c_i}{K_{I,li}}$. Similarly, an enzyme activation with efficiency factor $\eta^{\mathrm{reg}} = \frac{c/K_A}{1+c/K_A}$ in the rate laws yields a cost factor $\frac{1+c/K_A}{c/K_A} = 1 + K_A/c$ in the formulae for enzyme demands. If the metabolite concentrations are fixed, this corresponds to an extra enzyme demand $\Delta e_l = \frac{e_l\, K_{A,li}}{c_i}$ (where l and i denote the regulated reaction and the regulating metabolite, respectively). As usually in Enzyme Cost Minimization, an optimal rearrangement of enzyme and metabolite concentrations must be taken into account, which will then slightly reduce the overall cost.

The predictions of optimal states by Enzyme Cost Minimization rely on two main inputs: a metabolic model that relates metabolite concentrations, enzyme levels, and fluxes, and an optimality principle based on the assumption that cells realize their production fluxes at a minimal total enzyme cost. To test whether this optimality principle holds at all, Noor et al. [25] compared the predictions from Enzyme Cost Minimization to predictions from the same metabolic model and the same flux distribution, but with randomly sampled metabolite profiles (and the corresponding enzyme profiles). In comparison, metabolite profiles sampled close to the Enzyme Cost Minimization optimum yielded significantly better enzyme level predictions than metabolite profiles sampled more broadly. This strongly supports the idea that *E. coli* metabolism, in the conditions studied, is at least partially optimized for low enzyme cost, and thus supports cost-optimality as a principle in living cells.

6.5. Comparison of alternative pathways

Having clarified our main functional criteria for pathways (substrate productivity and enzyme productivity) and how they depend on pathway details (including outer concentrations), we can now compare alternative pathways by their substrate and enzyme demand per production flux (an example of "cost per benefit") and see which one scores better.

6.5.1. A tale of two glycolyses

One of the canonical examples discussed throughout this book is how cells choose between respiration and fermentation for making their ATP. However, having a precise kinetic model for respiration is difficult, since it involves electron transfer and membrane-bound reactions. Therefore, it is challenging to calculate the enzyme cost of respiration using models like those discussed in this chapter. Flamholz et al. [19] analyzed a similar but simpler case by comparing between the EMP and ED variants of glycolysis, since all the required enzymes are soluble and expressed in the cytoplasm and/or the periplasm and many of their kinetic parameters are measured. The common description of glycolysis ends in pyruvate (e.g., as depicted in Figure 6.1). This means that the pathway is not neutral in terms of redox, since the oxidation state of pyruvate is higher than glucose. In order to simplify the comparison and focus only on ATP yield (rather than NADH), the EMP and ED pathways were extended to end in lactate by including lactate dehydrogenase (*ldh*) as an extra step, making them redox neutral. These could be thought of as the more relevant versions of the pathways in anaerobic conditions.

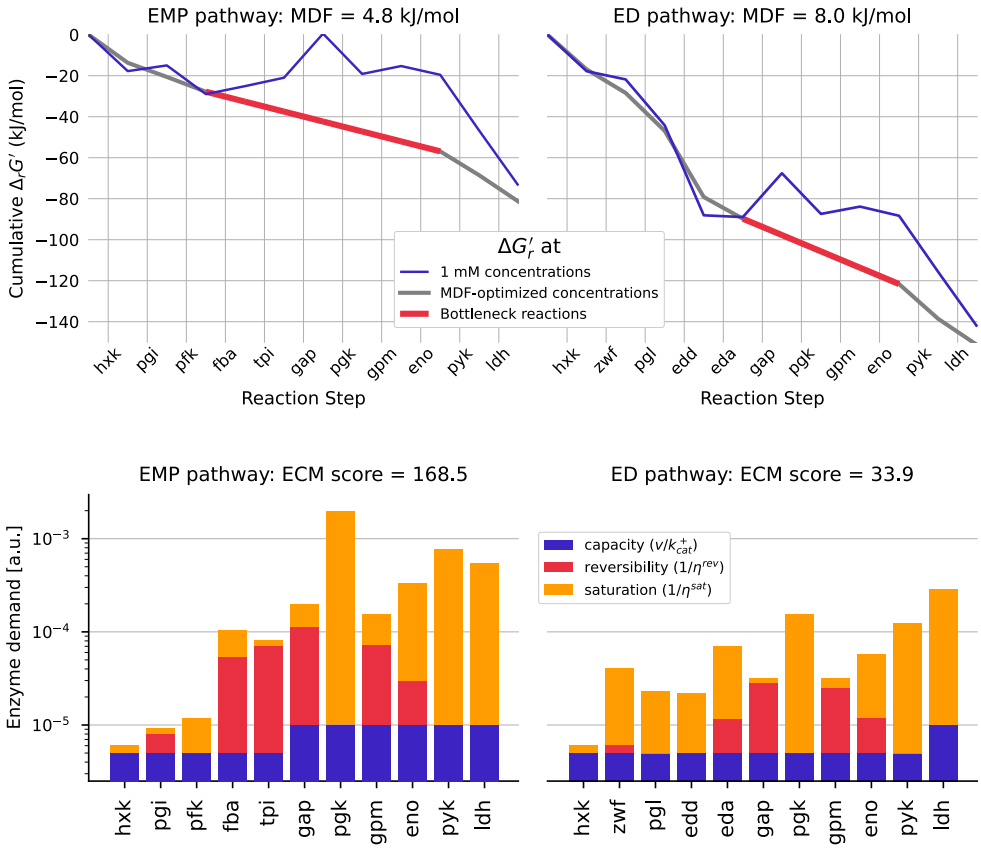

Figure 6.7: Comparing two metabolic pathways using the MDF and the Enzyme Cost Minimization methods – For the MDF analysis (top row), the dark blue line represents the cumulative Gibbs energy along the pathway if all metabolite concentrations were 1 mM. The MDF solution is presented as a gray line, where the bottleneck reactions are highlighted in red. For the Enzyme Cost Minimization analysis (bottom row), we used the same kinetic parameters for all enzymes in both pathways ($k_{\mathrm{cat}} = 200$ s^{-1}, $K_{\mathrm{M}} = 200$ μM, same as in [19]). However, here we used an updated version of Enzyme Cost Minimization with the factorized rate law, therefore the results are not identical. A Jupyter notebook for generating the figure can be found on the book website.

Although EMP-based fermentation is usually described in textbooks as less efficient than respiration, since it produces only 2 moles of ATP per mole glucose instead of ≈ 30, the ED pathway has an even lower yield – 1 mole of ATP. Nevertheless, the ED pathway is quite common among the bacteria. For example, *Zymomonas mobilis* – the bacterium used in fermenting pulque (a.k.a., agave wine [32]) and a promising platform for bio-production [33] – lacks key enzymes from the EMP pathway and uses the ED pathway exclusively to metabolize sugars. These bacteria don't seem to be bothered by the low ATP yield and can achieve high growth rates [34]. This already suggests to us that the ED pathway is probably superior to EMP in other aspects, such as the enzyme demand. Another clue was provided by a study which found that the ED pathway improves *E. coli* growth during glucose up-shifts and that the flux through it increases by 130% [35] (see Economic Analogy Box 6.D)

To see if indeed the models provide predictions that are consistent with the experimental evidence, Flamholz et al. [19] first used the MDF method to compare the two pathways. The ED pathway was found to be substantially more thermodynamically favorable, with a much higher score than the EMP pathway (8.0 versus 4.8 kJ/mol, see Figure 6.7 upper row).

Although the EMP pathway is clearly more favorable, we can still argue that an MDF of 4.8 kJ/mol is good enough, as it means $\theta > 1.9$ for each one of the pathway reactions. In this case, $\eta^{\mathrm{for}} > 0.85$ (see Figure 6.3) and therefore it might be a small price to pay for double the ATP yield. But, as discussed earlier, the efficiency of a pathway is

affected by other factors besides the thermodynamics. Flamholz et al. [19] tried to see whether ED is superior to EMP also in terms of the enzyme cost using the Enzyme Cost Minimization method. Indeed, they found that the ED pathway would require ≈5 times less protein compared to EMP for catalyzing the same flux (see Figure 6.7 bottom row). So, although the ATP yield of the ED pathways is half that of EMP, one can still generate ATP at a higher rate using the same amount of protein, according to the model.

The comparison of EMP and ED provided some insight as to a trade-off that can exists between the yield of a pathway and its cost, or enzyme burden. However, one can expand the question and ask if there are any other theoretically possible glycolysis pathways that might be able to break this trade-off and be more efficient than EMP and ED in both aspects. Ng et al. [6] tried to address this question with an algorithm they called *optStoic* that generates all biochemically feasible routes between glucose and pyruvate, with various ATP/glucose yields. They then ran pathway analysis on all 11,916 options and found that indeed both EMP and ED are both (nearly) Pareto-optimal. This suggests that evolution may indeed select for features such as high yield and low enzyme cost, where one might be more important than the other depending on the context.

6.5.2. Metabolic engineering

Besides the quest for understanding the evolution of existing biochemical pathways, pathway analysis methods like MDF and Enzyme Cost Minimization have also been used by metabolic engineers in order to rank and prioritize different alternative designs. For example, Volpers et al. [36] used the MDF algorithm and the Pathway Specific Activity measure to compare between designs of photo-electro-autotrophic strains. Similarly, Löwe and Kremling [37] used the Enzyme Cost Minimization algorithm to predict the enzyme demand of both natural and synthetic carbon fixation cycles.

6.5.3. Predicting the metabolite concentrations

So far, the examples given in this section focused on analyzing and comparing pathway alternatives in isolation, outside of the context of actual living organisms. However, we should not forget that the motivation for optimization goals such as enzyme demand are derived from physiological and evolutionary principles. Therefore, the optimal solutions coming from MDF and Enzyme Cost Minimization might be good predictions for the actual metabolic state that exists in naturally evolved organisms.

For example, a few years after the *in silico* analysis of the ED pathway [19], Jacobson et al. [38] measured the intracellular concentrations ED intermediates in *Z. mobilis*, and used them to calculate the Gibbs energies of the pathway's reactions. Indeed, they found that they closely fit the predicted values from the MDF solution. Similarly, measured values of enzyme and metabolite concentrations in *E. coli* correlate with predicted values from Enzyme Cost Minimization (when empricial reaction fluxes were obtained from ^{13}C-MFA measurements, Figure 6.8) [25]. In a related paper, Wortel et al. [39] expanded the idea of this method to explore the entire flux polytope.

These results suggest that indeed the optimization process that occurs throughout evolution is somewhat similar to the (much simplified) models presented here. Of course, improving the accuracy of the inputs and accounting for other effects that impact fitness could improve the predictions further. On the other hand, it might be naïve to expect natural systems to be optimal, which would mean that using basic principles to *precisely* predict phenotypes is an impossible task.

6.6. Concluding remarks

Coming back to our initial question, what have we learned from theory about the choice between possible pathways? The "choice between pathways" in a larger network is actually a choice between (network-wide) flux distributions that use different alternative pathways. Here we discussed how to score the usefulness of given flux distributions,

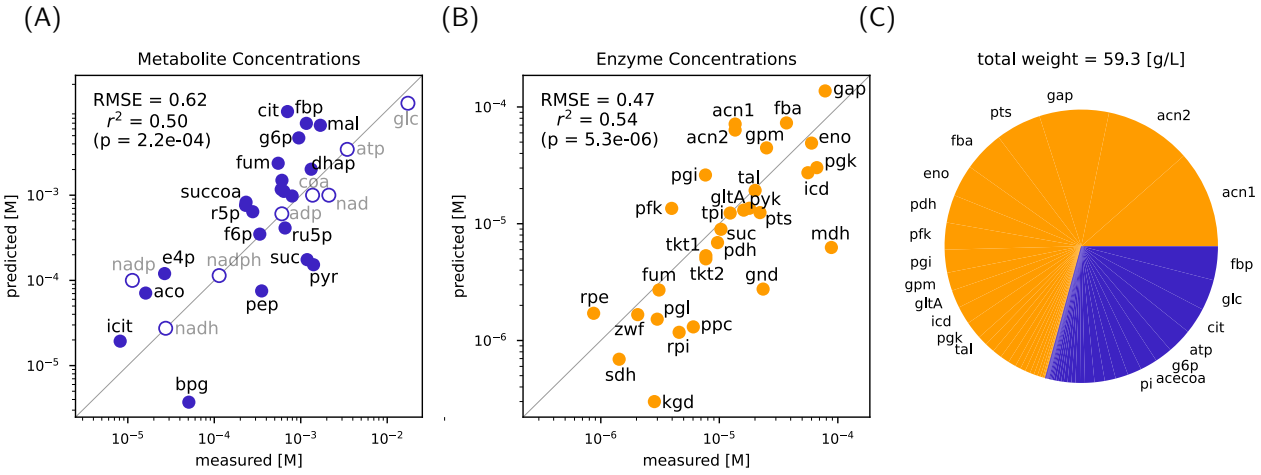

Figure 6.8: Validation of metabolite and enzyme concentrations, predicted by Enzyme Costs Minimization, in the central carbon metabolism of *E. coli* – (A) Comparing predicted and measured metabolic concentrations. The thin diagonal line marks $x = y$, i.e. where the predictions match the measurements. Hollow blue points represent co-factors whose concentration is fixed in the analysis and therefore are not actually predicted. Full blue points are for all other metabolites whose allowed concentration range was set to $1\mu M - 10mM$. The Root Mean Squared Error (RMSE, in $\log_{10}$ scale), r^2 (Pearson correlation), and p-value refer only to the full points. (B) Comparing predicted and measured enzyme concentrations. (C) A pie chart showing the distribution of the predicted absolute mass-concentrations for both enzymes (orange) and metabolites (blue) together. Note that *aconitase* (catalyzing the reactions acn1 and acn2) has a lower specific activity than *glyceraldehyde-3P dehydrogenase* (catalyzing gap), and therefore occupies a higher fraction of the mass-concentration even though the required concentration of the latter enzyme is higher. Labels of enzymes and metabolites that occupy the smallest fractions of the biomass are omitted due to lack of space.

which can also be used to score single pathways.

Importantly, flux distributions are scalable (by scaling all enzyme levels proportionally, and keeping all metabolite levels constant). If we scale the fluxes, this will scale both the flux benefit (for instance, the production of a desired product or biomass) and the required resources (substrates consumed, enzyme budget invested, or toxic byproducts produced). Because of this scaling property, our "quality criteria" mostly have the form of ratios between an output flux (as the benefit) and some (limited) resource (the cost). Such ratios are called "productivities", where in Chapter 4 in [3]-5 in [3] we focused mostly on substrate productivity (or yield on substrate) and in this chapter on enzyme productivity (or enzyme-specific rate) as important criteria. Why these criteria? On the one hand, they are closely related to some big objectives of the entire cell – depending on the type of competition it is facing. On the other hand, they are easy to link to some concrete criteria about metabolic pathways such as product yield, pathway length, k_{cat} values, thermodynamic forces, etc.

Since yield on substrate depends only on the shape of the flux distribution, it can be studied by methods like FBA (see chapters 4 in [3] and 5 in [3]). In this chapter, we focused on the more difficult case, enzyme productivity, where thermodynamics, enzyme kinetics, and the arrangement of metabolite and enzyme concentrations come into play. The factorized law in Eq. (6.7) shows us how the enzyme demand of a flux distribution can be computed if metabolite concentrations are known, and how the demand depends on forward k_{cat}, the thermodynamic force, and enzyme saturation. The only difficulty is that the thermodynamic forces and metabolite concentrations are usually not known. Here we considered some best-case scenarios, assuming that the cell will realize the concentration arrangements that optimize pathway performance. When considering thermodynamics alone (and making some further simplifications), this led to the MDF method. For the full problem, the solution is provided by Enzyme Costs Minimization. This method is directly related to the different pathway criteria we discussed initially (including pathway length, thermodynamic forces, and k_{cat} values) and thus shows how these different factors determine enzyme

demand. As a numerical method, it is relatively easy to use because it is a convex optimization problem. But if little data is available, simpler methods such as MDF, with their lower demand for parameters, may be useful tools to predict pathway usage.

In order to predict optimal metabolic states, we started in the previous chapter with models that optimize the fluxes in an entire network. Howeve, to keep the models linear, kinetics and concentrations were largely ignored. In FBA with molecular crowding, a connection between fluxes and enzyme levels was made via empirical parameters, the apparent catalytic rates or "enzyme efficiencies". We now saw that these parameters are not at all constant parameters, but emerge from kinetics and given concentration profiles, and we also saw how optimal concentration profiles can be computed for a given flux distribution. This means: we now know how to predict optimal fluxes from known enzyme efficiencies, and we know how to predict optimal concentrations (and therefore enzyme efficiencies) from known fluxes. In the next chapter we will put these two things together, in order to predict all variables in the system – fluxes, metabolite concentrations, enzyme efficiencies, and enzyme levels – from a single principle of maximal overall enzyme efficiency.

Recommended readings

A search for efficient pathways, based on different criteria: Arren Bar-Even, Elad Noor, Nathan E. Lewis, and Ron Milo. Design and analysis of synthetic carbon fixation pathways. Proceedings of the National Academy of Sciences, 107(19):88898894, 2010. doi: 10.1073/pnas.0907176107.

The max-min driving force method: Elad Noor, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput. Biol., 10(2):e1003483, 2014. doi: 10.1371/journal.pcbi.1003483.

Enzyme cost minimization: Elad Noor, Avi Flamholz, Arren Bar-Even, Dan Davidi, Ron Milo, and Wolfram Liebermeister. The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. PLoS Comput. Biol., 12(11):e1005167, 2016. doi: 10.1371/journal.pcbi.1005167.

Problems

Problem 6.1 Pathway efficiencies

Estimate pathway efficiencies (i.e. product production rates per total enzyme concentration) from simple back-of the envelope calculations and plausible numbers (refer to the BioNumbers database for realistic values). (a) From pathway length (assuming reasonable apparent k_{cat} values); (b) from given apparent k_{app} values (or given k_{cat} values and $\Delta_r G$). (c) Convert the results into growth rates (assuming realistic estimates of the total protein density; the proteome fraction of metabolic enzymes; the biomass production rate etc). Assume plausible numbers in all cases.

Problem 6.2 Efficiency – dependence on substrate

Compute the reduction of pathway efficiency in a linear chain when decreasing the external substrate concentration (no constraints on metabolite levels)

Problem 6.3 ATP yield in glycolysis

Derive the optimal ATP yield in a glycolysis model with a linear flux-force relationship

Problem 6.4 MDF method

Implement the MDF method in a programming language of your choice.

Problem 6.5 MDF and enzyme cost

The optimality principle of MDF (avoiding small thermodynamic driving forces) can be justified by assuming that low driving forces would entail high enzyme demands. Do you expect that MDF solutions are also Enzyme Costs Minimization solutions (or vice versa)? Otherwise, can you think of an approximation of the Enzyme Costs Minimization problem, such that MDF provides the correct solution? Show how the Enzyme Costs Minimization objective could be approximated step by step, and illustrate this with an example.

Problem 6.6 Cycle of chemical reactions

Assume a cycle of chemical reactions $A \leftrightarrow B \leftrightarrow C \leftrightarrow A$ without co-factors or external inputs/outputs. (a) Show that there is no stationary, thermodynamically feasible flux distribution except for the (trivial) vanishing flux. (b) Explain why, if there were a flux, this would be a perpetuum mobile.

Problem 6.7 Optimal enzyme levels in two-reaction chain

Consider a chain of two reactions $S \leftrightarrow X \leftrightarrow P$ with enzymes e_1 and e_2, $v_1 = e_1(k_{+1}S - k_{-1}X), v_2 = e_2(k_{+2}X - k_{-2}P)$. Compute the steady state flux given e_1, e_2. Let $e_1 + e_2 = E_{\mathrm{tot}}^{\mathrm{path}}$ be fixed. Determine e_1, e_2 such that the flux is maximal. Use Lagrange multipliers. Hint: Assume forward flux where $P/S < (k_{+1}k_{+2})/(k_{-1}k_{-2}) = q_1 q_2$.

Problem 6.8 Flux maximization in a linear pathway

Prove that the function:

$$f(\boldsymbol{e}) = \frac{1}{\sum_i (A_i e_i)^{-1}} \tag{6.16}$$

for a fixed $\mathbf{A}$ and under the constraint $\sum_i e_i = e_{\mathrm{tot}}$, is at its maximum when:

$$e_i = e_{\mathrm{tot}} \cdot \frac{A_i^{-1/2}}{\sum_i A_i^{-1/2}}$$

Problem 6.9 Haldane kinetic rate law

Haldane described an enzyme-catalyzed reaction by three steps, each following a mass-action rate law:

$$\mathrm{S} + \mathrm{E} \underset{\mathrm{k_2}}{\overset{\mathrm{k_1}}{\rightleftharpoons}} \mathrm{ES} \underset{\mathrm{k_4}}{\overset{\mathrm{k_3}}{\rightleftharpoons}} \mathrm{EP} \underset{\mathrm{k_6}}{\overset{\mathrm{k_5}}{\rightleftharpoons}} \mathrm{P} + \mathrm{E}\,. \tag{6.17}$$

The ODE system describing the change in time of each species is:

$$\begin{aligned} \frac{d[ES]}{dt} &= [E]\cdot[S]\cdot k_1 + [EP]\cdot k_4 - [ES]\cdot(k_2 + k_3) \\ \frac{d[EP]}{dt} &= [E]\cdot[P]\cdot k_6 + [ES]\cdot k_3 - [EP]\cdot(k_4 + k_5) \\ \frac{d[E]}{dt} &= -[E]\cdot[S]\cdot k_1 + [ES]\cdot k_2 + [EP]\cdot k_5 - [E]\cdot[P]\cdot k_6 \end{aligned} \tag{6.18}$$

Prove that at quasy-steady-state (where the total enzyme concentration is fixed, and the concentration of each species doesn't change over time), the rate in which $[S]$ is converted to $[P]$ is governed by the following rate law:

$$v = [E_0]\frac{k_{\mathrm{cat}}^{+}[S]/K_{\mathrm{S}} - k_{\mathrm{cat}}^{-}[P]/K_{\mathrm{P}}}{1 + [S]/K_{\mathrm{S}} + [P]/K_{\mathrm{P}}} \tag{6.19}$$

where:

$$K_{\mathrm{S}} = \frac{k_2k_4 + k_2k_5 + k_3k_5}{k_1(k_3 + k_4 + k_5)};\ K_{\mathrm{P}} = \frac{k_2k_4 + k_2k_5 + k_3k_5}{k_6(k_2 + k_3 + k_4)};\ k_{\mathrm{cat}}^{+} = \frac{k_3k_5}{k_3 + k_4 + k_5};\ k_{\mathrm{cat}}^{-} = \frac{k_2k_4}{k_2 + k_3 + k_4}$$

Problem 6.10 The factorized rate law

Use the Haldane relationship:

$$\frac{k^+_{\text{cat}}}{k^-_{\text{cat}}}\frac{K_\text{P}}{K_\text{S}} = \frac{k_1 k_3 k_5}{k_2 k_4 k_6} = K_\text{eq} \tag{6.20}$$

and the definition of Gibbs free energy:

$$\begin{aligned} \Delta_\text{r} G'^\circ &= -R \cdot T \cdot \ln K_\text{eq} \\ \Delta_\text{r} G' &= \Delta_\text{r} G'^\circ + R \cdot T \cdot \ln\left([P]/[S]\right) \end{aligned} \tag{6.21}$$

to prove that Eq. (6.19) is equivalent to the following factorized rate law:

$$v = [E_0] k^+_{\text{cat}} \cdot \left(1 - e^{\Delta_\text{r} G'/RT}\right) \cdot \frac{[S]/K_\text{S}}{1 + [S]/K_\text{S} + [P]/K_\text{P}}. \tag{6.22}$$

Solutions to problems

Problem 6.9 (Haldane kinetic rate law)

First, we add the constraint on the total enzyme concentration ($[E] + [ES] + [EP] = E_\text{tot}$) and rewrite the ODE system in matrix notation:

$$\begin{pmatrix} 1 & 1 & 1 \\ [S]k_1 & -(k_2+k_3) & k_4 \\ [P]k_6 & k_3 & -(k_4+k_5) \\ -[S]k_1 - [P]k_6 & k_2 & k_5 \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = \begin{pmatrix} [E_0] \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{6.23}$$

Note that the last row is linearly dependent on the two previous ones (it is minus their sum). Therefore, we can drop it from the system without loosing information. Then, we will find exlicit expressions for $[E]$, $[ES]$, and $[EP]$ by using Gaussian elimination – a process of eliminating off-diagonal values in the matrix until we reach the identity matrix, while at the same time applying the same operations to the vector on the right-hand side of the equality.

Step 1, elimination the off-diagonal elements on the first column (subtracting the first row times $[S]k_1$ from the 2nd row and the first row times $[P]k_6$ from the 3rd row)

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & -(k_2+k_3) - [S]k_1 & k_4 - [S]k_1 \\ 0 & k_3 - [P]k_6 & -(k_4+k_5) - [P]k_6 \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_\text{tot}] \begin{pmatrix} 1 \\ -[S]k_1 \\ -[P]k_6 \end{pmatrix}.$$

Step 2, dividing the second row by $-(k_2 + k_3 + [S]k_1)$ to have 1 on the diagonal:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{[S]k_1 - k_4}{k_2+k_3+[S]k_1} \\ 0 & k_3 - [P]k_6 & -(k_4+k_5) - [P]k_6 \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_\text{tot}] \begin{pmatrix} 1 \\ \frac{[S]k_1}{k_2+k_3+[S]k_1} \\ -[P]k_6 \end{pmatrix}.$$

Step 3, subtracting the second row from the 1st, and again from the 3rd (after multiplying by $k_3 - [P]k_6$):

$$\begin{pmatrix} 1 & 0 & 1 - \frac{[S]k_1 - k_4}{k_2+k_3+[S]k_1} \\ 0 & 1 & \frac{[S]k_1 - k_4}{k_2+k_3+[S]k_1} \\ 0 & 0 & -(k_4+k_5) - [P]k_6 - \frac{([S]k_1 - k_4)(k_3 - [P]k_6)}{k_2+k_3+[S]k_1} \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_\text{tot}] \begin{pmatrix} 1 - \frac{[S]k_1}{k_2+k_3+[S]k_1} \\ \frac{[S]k_1}{k_2+k_3+[S]k_1} \\ -[P]k_6 - \frac{[S]k_1(k_3 - [P]k_6)}{k_2+k_3+[S]k_1} \end{pmatrix}.$$

which after simplifying becomes:

$$\begin{pmatrix} 1 & 0 & \frac{k_2+k_3+k_4}{k_2+k_3+[S]k_1} \\ 0 & 1 & \frac{[S]k_1-k_4}{k_2+k_3+[S]k_1} \\ 0 & 0 & -\frac{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5}{k_2+k_3+[S]k_1} \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_{\text{tot}}] \begin{pmatrix} \frac{k_2+k_3}{k_2+k_3+[S]k_1} \\ \frac{[S]k_1}{k_2+k_3+[S]k_1} \\ -\frac{[P]k_6k_2+[P]k_6k_3+[S]k_1k_3}{k_2+k_3+[S]k_1} \end{pmatrix}.$$

and we normalize the last row to have 1 on the diagonal:

$$\begin{pmatrix} 1 & 0 & \frac{k_2+k_3+k_4}{k_2+k_3+[S]k_1} \\ 0 & 1 & \frac{[S]k_1-k_4}{k_2+k_3+[S]k_1} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_{\text{tot}}] \begin{pmatrix} \frac{k_2+k_3}{k_2+k_3+[S]k_1} \\ \frac{[S]k_1}{k_2+k_3+[S]k_1} \\ \frac{[P]k_6k_2+[P]k_6k_3+[S]k_1k_3}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \end{pmatrix}.$$

Step 4, we eliminate the off-diagonal values of the third column using the 3rd row:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_{\text{tot}}] \begin{pmatrix} \frac{k_2+k_3}{k_2+k_3+[S]k_1} - \frac{k_2+k_3+k_4}{k_2+k_3+[S]k_1} \cdot \frac{[P]k_6k_2+[P]k_6k_3+[S]k_1k_3}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \\ \frac{[S]k_1}{k_2+k_3+[S]k_1} - \frac{[S]k_1-k_4}{k_2+k_3+[S]k_1} \cdot \frac{[P]k_6k_2+[P]k_6k_3+[S]k_1k_3}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \\ \frac{[P]k_6k_2+[P]k_6k_3+[S]k_1k_3}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \end{pmatrix}$$

Simplifying the expressions on the right-hand side is a lengthy process (which we do not show here) and in the end we get:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} [E] \\ [ES] \\ [EP] \end{pmatrix} = [E_{\text{tot}}] \begin{pmatrix} \frac{k_2k_4+k_2k_5+k_3k_5}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \\ \frac{[P]k_4k_6+[S]k_1k_4+[S]k_1k_5}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \\ \frac{[P]k_2k_6+[P]k_3k_6+[S]k_1k_3}{[S]k_1(k_3+k_4+k_5)+[P]k_6(k_2+k_3+k_4)+k_2k_4+k_2k_5+k_3k_5} \end{pmatrix}$$

Therefore,

$$[E] = [E_{\text{tot}}] \frac{k_2k_4 + k_2k_5 + k_3k_5}{[S]k_1(k_3 + k_4 + k_5) + [P]k_6(k_2 + k_3 + k_4) + k_2k_4 + k_2k_5 + k_3k_5} \tag{6.24}$$

$$[ES] = [E_{\text{tot}}] \frac{[P]k_4k_6 + [S]k_1k_4 + [S]k_1k_5}{[S]k_1(k_3 + k_4 + k_5) + [P]k_6(k_2 + k_3 + k_4) + k_2k_4 + k_2k_5 + k_3k_5} \tag{6.25}$$

$$[EP] = [E_{\text{tot}}] \frac{[P]k_2k_6 + [P]k_3k_6 + [S]k_1k_3}{[S]k_1(k_3 + k_4 + k_5) + [P]k_6(k_2 + k_3 + k_4) + k_2k_4 + k_2k_5 + k_3k_5} \tag{6.26}$$

Σ Math box 6.B Integer cuts for iterating all possible pathway variants

The linear problem can be described by:

$$\begin{aligned} \text{minimize} \quad & \sum_i v_i \\ \text{subject to} \quad & \mathbf{N}\mathbf{v} = 0 \\ & \forall i \quad 0 \leq v_i \leq \beta \\ & v_{\text{glycolysis}} = -1 \end{aligned} \tag{6.2}$$

where $\mathbf{v}$ is the flux variable, and $\mathbf{N}$ is comprised of the universal stoichiometric matrix , and in addition one reaction (whose flux is denoted $v_{\text{glycolysis}}$) which has the stoichiometry of Eq. (6.1). The constraint $v_{\text{glycolysis}} = -1$ ensures that the sum of all active reactions except for $v_{\text{glycolysis}}$ will together form a full glycolysis pathway, since their net reaction has to balance the stoichiometry of $v_{\text{glycolysis}}$ given the mass balance constraint $\mathbf{N}\mathbf{v} = 0$. β given the upper bound on the flux for all reactions. For simplicity, we assume that all fluxes are positive and that reversible reactions are split into their two opposing directionalities . β is a tunable parameter that is an upper bound on all the fluxes in the solution pathways. Setting it too low would exclude solutions with complex stoichiometries. On the other hand, a very high value would increase the complexity of the search and lead to very long run-times. Typically, we choose $\beta = 10$ which is a good balance between the two extremes. Finally, we set the objective function ($\sum_i v_i$) to minimize the sum of fluxes. As we will explain shortly, we can iterate through all possible solutions and therefore the objective will only determine the order at which we find them.

To find all possible glycolysis pathways comprising known enzymatic reactions, Ng et al. [6] iteratively introduced constraints in order to exclude all previous solutions and find the next optimal one [9]: to exclude a solution, they add an *integer cut*, which is an inequality constraint ensuring that the number of active reactions is strictly larger than the sum over their indicator variables (boolean variables that are equal to 1 if the reaction is active, i.e. carries a nonzero flux). Therefore, at least one of those reactions must be inactive in all future solutions. This is quite similar to constrained Minimal Cut Sets (cMCS) which were introduced in Chapter 4.4 in [3] as a way of exploring the flux space.

Formally, if $\{P_0, P_1 \dots P_m\}$ are the set of solutions already discovered by our algorithm (where $\forall j \, P_j \subseteq \{0, \dots, n\}$, i.e. each solution is a set of integers which correspond to indices of active reactions) then the added constraints will be:

$$\begin{aligned} \forall i \quad & z_i \in \{0, 1\} \\ \forall i \quad & v_i - \beta z_i \leq 0 \\ \forall j \quad & \sum_{i \in P_j} z_i < \|P_j\| \end{aligned} \tag{6.3}$$

where $\|P_j\|$ is the length of pathway j (i.e. the number of reactions). The z_i are boolean reaction indicators, i.e. z_i must be equal to 1 if a reaction is active ($v_i > 0$). The final set of constrains eliminate P_j and any pathway which is a superset of P_j from the solution space. Using this extra set of constraints iteratively, each time generating the *next* pathway and adding it to the excluded list, will eventually go through all possible solutions (by increasing order of their sum of fluxes). It is important to note that using integer cuts requires switching to an MILP (Mixed-Integer Linear Program) solver, which is computationally much more demanding and typically requires a commercial license.

Math box 6.C Factorized rate laws and enzyme cost function

According to Eq. (6.13), reversible rate laws can be factorized into five terms that depend on metabolite concentrations in different ways [16]. For a reaction S $\rightleftharpoons$ P with reversible Michaelis-Menten kinetics Eq. (6.11), a driving force $\theta = -\Delta_r G'/RT$, and a prefactor for non-competitive inhibition, the rate law can be written as

$$v = E \cdot k_{\text{cat}}^{+} \cdot \underbrace{[1 - \mathrm{e}^{-\theta}]}_{\eta^{\text{for}}} \cdot \underbrace{\frac{s/K_{\text{S}}}{1 + s/K_{\text{S}} + p/K_{\text{P}}}}_{\eta^{\text{sat}}} \cdot \underbrace{\frac{1}{1 + x/K_{\text{I}}}}_{\eta^{\text{reg}}}$$

Rate	=	enzyme level	·	forward catalytic constant	·	thermodynamic factor	·	saturation factor	·	regulation factor

with inhibitor concentration x. The product of the first two terms, E and k_{cat}^{+}, represents the maximal velocity, i.e. the rate at full substrate-saturation without backward flux and without enzyme inhibition. The following factors decrease this velocity for different reasons: η^{for} describes a decrease due to backward fluxes, η^{sat} – the decrease due to incomplete substrate saturation, and η^{reg} – the decrease due to small-molecule regulation (see Figure b). While k_{cat}^{+} is an enzyme-specific constant (yet, dependent on conditions such as pH, ionic strength, or molecular crowding in cells), the efficiency factors are concentration-dependent, unitless, and can vary between 0 and 1. The thermodynamic factor η^{for} depends on the driving force (and thus, indirectly, on metabolite concentrations), and the equilibrium constant is required for its calculation. The saturation factor η^{sat} depends directly on metabolite levels and contains the K_{M} values as parameters. Enzyme regulation by small molecules yields additive or multiplicative terms in the rate law denominator, which in our example and can be captured by a separate factor η^{reg}. The enzyme cost for a flux v, with an enzyme burden h_e, can be written as

$$q = h_e \cdot E = h_e \cdot v \cdot \frac{1}{k_{\text{cat}}^{+}} \cdot \underbrace{\frac{1}{[1 - \mathrm{e}^{-\theta}]}}_{1/\eta^{\text{for}}} \cdot \underbrace{\frac{1 + s/K_{\text{S}} + p/K_{\text{P}}}{s/K_{\text{S}}}}_{1/\eta^{\text{sat}}} \cdot \underbrace{[1 + x/K_{\text{I}}]}_{1/\eta^{\text{reg}}}$$

and contains the terms from the rate law in inverse form. The first factors, $h_e\, v/k_{\text{cat}}^{+}$, define a minimum enzyme cost, which is then increased by the following efficiency factors. By omitting some of these factors, one can construct simplified enzyme cost functions with higher specific rates, or lower enzyme demands (compare Figure 6.2b). For a closer approximation, the factors may be substituted with constant numbers between 0 and 1.

Economics analogy 6.D The push for fast growth

The ED pathway seems to be useful as a quick response to a sudden increase in abundance of resources (glucose), but less efficient than EMP when the environment is steady. This is somewhat analogous to start-up companies, which burn large amounts of venture capital in order to grow rapidly. However, after reaching a certain scale, the dynamic nature of start-ups often becomes a burden, where overhead costs pile up and signal that it is time to join a larger corporation.

Economics analogy 6.E Two central assumptions: homogeneity and stationarity

In the models described in this chapter, we generally assume that our system (for example, a metabolic pathway in a cell) is spatially and temporally homogeneous, and that it shows stable stationary states. This is clearly a simplification: in reality, cells are inhomogeneous, with compartments, with enzymes unequally distributed across the cell, and with enzymes forming complexes or dedicated compartments like the glycosome (an organelle in some organisms that contains the glycolytic enzymes), which changes (average) enzyme kinetics. Cells are also dynamic on various time scales (chemical noise, metabolic dynamics, protein expression dynamics), which also may change (average) enzyme kinetics. If we ignore this in our models – assuming a timeless steady state – this will not only cause approximation errors in our metabolic model, but much more importantly, we ignore the fact that the cell can exploit spatial inhomogeneity (e.g. compartments or channeling) and non-steady states (e.g. metabolic oscillations, or adaptation to fluctuations in the environment) to further improve its fitness (as compared to a steady-state, constant enzyme model).

Interestingly, classical economic theory makes similar assumptions – e.g. about markets in equilibrium– which ignore the spatio-temporal, dynamic side of real economic systems, which – as in the case of metabolic models – is likely to lead to wrong results.

Bibliography

[1] Arren Bar-Even, Avi Flamholz, Elad Noor, and Ron Milo. Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat. Chem. Biol.*, 8(6):509–517, May 2012. doi: 10.1038/nchembio.971.

[2] Nana-Maria Grüning and Markus Ralser. Glycolysis: How a 300yr long research journey that started with the desire to improve alcoholic beverages kept revolutionizing biochemistry. *Current Opinion in Systems Biology*, 28:100380, December 2021. doi: 10.1016/j.coisb.2021.100380.

[3] The Economic Cell Collective, editor. *Economic Principles in Cell Biology*. Free online book, 2023. doi: 10.5281/zenodo.8156386.

[4] T Conway. The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol. Rev.*, 9(1):1–27, September 1992. doi: 10.1111/j.1574-6968.1992.tb05822.x.

[5] L. Gerosa, B. R. Haverkorn van Rijsewijk, D. Christodoulou, K. Kochanowski, T. S. Schmidt, E. Noor, and U. Sauer. Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Syst*, 1(4):270–82, 2015. ISSN 2405-4712 (Print) 2405-4712 (Linking). doi: 10.1016/j.cels.2015.09.008.

[6] Chiam Yu Ng, Lin Wang, Anupam Chowdhury, and Costas D Maranas. Pareto optimality explanation of the glycolytic alternatives in nature. *Sci. Rep.*, 9(1):2633, February 2019. doi: 10.1038/s41598-019-38836-9.

[7] A Bar-Even, E Noor, N E Lewis, and R Milo. Design and analysis of synthetic carbon fixation pathways. *Proceedings of the National Academy of Sciences*, 107(19):8889–8894, 2010. doi: 10.1073/pnas.0907176107.

[8] Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, 11 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw1092.

[9] Priti Pharkya, Anthony P Burgard, and Costas D Maranas. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.*, 14(11):2367–2376, November 2004. doi: 10.1101/gr.2872004.

[10] Camillo Iacometti, Katharina Marx, Maria Hönick, Viktoria Biletskaia, Helena Schulz-Mirbach, Beau Dronsella, Ari Satanowski, Valérie A Delmas, Anne Berger, Ivan Dubois, Madeleine Bouzon, Volker Döring, Elad Noor, Arren Bar-Even, and Steffen N Lindner. Activating silent glycolysis bypasses in escherichia coli. *BioDesign Research*, 2022, May 2022. doi: 10.34133/2022/9859643.

[11] Elad Noor, Eran Eden, Ron Milo, and Uri Alon. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell*, 39(5):809–820, September 2010. doi: 10.1016/j.molcel.2010.08.031.

[12] S.D. Finley, L.J. Broadbelt, and V.Hatzimanikatis. Computational framework for predictive biodegradation. *Biotechnol. Bioeng.*, 104(6):1086–1097, 2009. doi: 10.1002/bit.22489.

[13] Arren Bar-Even, Avi Flamholz, Elad Noor, and Ron Milo. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta*, 1817:1646–1659, September 2012. doi: 10.1016/j.bbabio.2012.05.002.

[14] P R Rich. The molecular machinery of keilin's respiratory chain. *Biochem. Soc. Trans.*, 31(Pt 6):1095–1105, December 2003. doi: 10.1042/bst0311095.

[15] Steffen Klamt, Stefan Müller, Georg Regensburger, and Jürgen Zanghellini. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metab. Eng.*, 47:153–169, May 2018. doi: 10.1016/j.ymben.2018.02.001.

[16] E. Noor, A. Flamholz, W. Liebermeister, A. Bar-Even, and R. Milo. A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett*, 587(17):2772–7, 2013. ISSN 1873-3468 (Electronic) 0014-5793 (Linking). doi: 10.1016/j.febslet.2013.07.028.

[17] Wolfram Liebermeister and Edda Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.*, 3(1):41, 2006. doi: 10.1186/1742-4682-3-41.

[18] Dan Davidi, Elad Noor, Wolfram Liebermeister, Arren Bar-Even, Avi Flamholz, Katja Tummler, Uri Barenholz, Miki Goldenfeld, Tomer Shlomi, and Ron Milo. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proceedings of the National Academy of Sciences*, 113 (12):3401–3406, 2016. doi: 10.1073/pnas.1514240113.

[19] Avi Flamholz, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.*, 110(24):10039–10044, June 2013. doi: 10.1073/pnas.1215283110.

[20] Elad Noor, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.*, 10(2):e1003483, February 2014. doi: 10.1371/journal.pcbi.1003483.

[21] Oliver Hädicke, Axel von Kamp, Timur Aydogan, and Steffen Klamt. OptMDFpathway: Identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO2 fixation potential of escherichia coli. *PLoS Comput. Biol.*, 14(9):e1006492, September 2018. doi: 10.1371/journal.pcbi.1006492.

[22] Avi Flamholz, Elad Noor, Arren Bar-Even, and Ron Milo. equilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res.*, 40, January 2012. doi: 10.1093/nar/gkr874.

[23] E. Noor, A. Bar-Even, A. Flamholz, Y. Lubling, D. Davidi, and R. Milo. An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics*, 28(15):2037–44, 2012. ISSN 1367-4811 (Electronic) 1367-4803 (Print) 1367-4803 (Linking). doi: 10.1093/bioinformatics/bts317.

[24] M. E. Beber, M. G. Gollub, D. Mozaffari, K. M. Shebek, A. I. Flamholz, R. Milo, and E. Noor. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res*, 50(D1):D603–D609, 2022. ISSN 1362-4962 (Electronic) 0305-1048 (Print) 0305-1048 (Linking). doi: 10.1093/nar/gkab1106.

[25] Elad Noor, Avi Flamholz, Arren Bar-Even, Dan Davidi, Ron Milo, and Wolfram Liebermeister. The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. *PLoS Comput. Biol.*, 12(11): e1005167, November 2016. doi: 10.1371/journal.pcbi.1005167.

[26] D. A. Beard and H. Qian. Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS One*, 2(1):e144, 2007. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi: 10.1371/journal.pone.0000144.

[27] Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436 (7050):588–592, July 2005.

[28] I. Shachrai, A. Zaslaver, U. Alon, and E. Dekel. Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth. *Molecular Cell*, 38:1–10, 2010. doi: 10.1016/j.molcel.2010.04.015.

[29] W Liebermeister and E Noor. The enzyme cost of given metabolic flux distributions, as a function of logarithmic metabolite levels, is convex. *arXiv: 1501. 02454*, 2015. doi: 10.48550/arXiv.1501.02454.

[30] V. Hatzimanikatis, C. Li, J.A. Ionita, C.S. Henry, M.D. Jankowski, and L.J. Broadbelt. Exploring the diversity of complex metabolic networks. *Bioinformatics*, 21(8):1603–1609, 2005. doi: 10.1093/bioinformatics/bti213.

[31] Ed Reznik, Dimitris Christodoulou, Joshua E Goldford, Emma Briars, Uwe Sauer, Daniel Segrè, and Elad Noor. Genome-Scale architecture of small molecule regulatory networks and the fundamental Trade-Off between regulation and enzymatic activity. *Cell Rep.*, 20(11):2666–2677, September 2017. doi: 10.1016/j.celrep.2017.08.066.

[32] P L Rogers, K J Lee, M L Skotnicki, and D E Tribe. Ethanol production by zymomonas mobilis. In *Microbial Reactions*, pages 37–84. Springer Berlin Heidelberg, 1982.

[33] Gerrich Behrendt, Jonas Frohwitter, Maria Vlachonikolou, Steffen Klamt, and Katja Bettenbrock. Zymo-Parts: A golden gate modular cloning toolbox for heterologous gene expression in zymomonas mobilis. *ACS Synth. Biol.*, November 2022.

[34] Tobias Fuhrer, Eliane Fischer, and Uwe Sauer. Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.*, 187(5):1581–1590, March 2005. doi: 10.1128/JB.187.5.1581-1590.2005.

[35] Richard C Law, Glenn Nurwono, and Junyoung O Park. A parallel glycolysis supports rapid adaptation in dynamic environments. *bioRxiv*, page 2022.08.19.504590, August 2022. doi: 10.1101/2022.08.19.504590.

[36] Michael Volpers, J Nico Claassens, Elad Noor, der John van Oost, de Willem M Vos, M Servé W Kengen, and dos Vitor A P Martins Santos. Integrated in silico analysis of pathway designs for synthetic Photo-Electro-Autotrophy. *PLoS One*, 11, June 2016. doi: 10.1371/journal.pone.0157851.

[37] Hannes Löwe and Andreas Kremling. In-Depth computational analysis of natural and artificial carbon fixation pathways. *BioDesign Research*, 2021, September 2021. doi: 10.34133/2021/9898316.

[38] Tyler B Jacobson, Paul A Adamczyk, David M Stevenson, Matthew Regner, John Ralph, Jennifer L Reed, and Daniel Amador-Noguez. 2H and 13C metabolic flux analysis elucidates in vivo thermodynamics of the ED pathway in zymomonas mobilis. *Metab. Eng.*, 54:301–316, July 2019. doi: 10.1016/j.ymben.2019.05.006.

[39] Meike T Wortel, Elad Noor, Michael Ferris, Frank J Bruggeman, and Wolfram Liebermeister. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Comput. Biol.*, 14(2):e1006010, February 2018. doi: 10.1371/journal.pcbi.1006010.