How Biochemical Constraints of Cellular Growth Shape Evolutionary Adaptations in Metabolism

Jan Berkhout,*,† Evert Bosdriesz,*,‡ Emrah Nikerel,†,§ Douwe Molenaar,*,† Dick de Ridder,†,§ Bas Teusink,*,†,†† and Frank J. Bruggeman*,**,††,‡‡,1

*Systems Bioinformatics, The Centre for Integrative Bioinformatics VU, VU University, 1081 HV, Amsterdam, The Netherlands, †Kluyver Centre for Genomics of Industrial Fermentation, 2600 GA, Delft, The Netherlands, ‡The Netherlands Bioinformatics Centre, 6500 HB, Nijmegen, The Netherlands, §The Delft Bioinformatics Lab, Delft University of Technology, 2600 GA, Delft, The Netherlands, **Netherlands Institute for Systems Biology, 1098 XH, Amsterdam, The Netherlands, ††Amsterdam Institute for Molecules, Medicines and Systems, VU University, 1081 HV, Amsterdam, The Netherlands, and ‡‡Life Sciences, Centre for Mathematics and Computer Science, 1098 XH, Amsterdam, The Netherlands

ABSTRACT Evolutionary adaptations in metabolic networks are fundamental to evolution of microbial growth. Studies on unneeded-protein synthesis indicate reductions in fitness upon nonfunctional protein synthesis, showing that cell growth is limited by constraints acting on cellular protein content. Here, we present a theory for optimal metabolic enzyme activity when cells are selected for maximal growth rate given such growth-limiting biochemical constraints. We show how optimal enzyme levels can be understood to result from an enzyme benefit minus cost optimization. The constraints we consider originate from different biochemical aspects of microbial growth, such as competition for limiting amounts of ribosomes or RNA polymerases, or limitations in available energy. Enzyme benefit is related to its kinetics and its importance for fitness, while enzyme cost expresses to what extent resource consumption reduces fitness through constraint-induced reductions of other enzyme levels. A metabolic fitness landscape is introduced to define the fitness potential of an enzyme. This concept is related to the selection coefficient of the enzyme and can be expressed in terms of its fitness benefit and cost.

ENVIRONMENTAL conditions set the selective pressures acting on unicellular organisms. Microbial fitness is often related to growth properties, such as biomass yield, growth rate, or antibiotic resistance. As a large part of the available resources is spent on the synthesis of metabolic machinery, regulation of the levels of metabolic enzymes can have large influences on fitness (Dean 1989; Dong *et al.* 1995; Dekel and Alon 2005; Stoebel *et al.* 2008). Selection on growth rate may then direct the evolution of microorganisms to optimal allocation of resources for fitness enhancement (Dekel and Alon 2005; Molenaar *et al.* 2009). Alternatively, evolution may be directed by metabolic trade-offs (Beardmore *et al.* 2011; Wenger *et al.* 2011), which may cause sympatric speciation (Friesen *et al.* 2004). To improve our understanding of the driving processes of metabolic evolution, the interplay between selective pressures and the biochemistry and organization of metabolic networks must be taken into account.

Studies on the growth effects of unneeded-protein expression, sometimes called gratuitous or nonfunctional protein expression, indicate significant reductions in growth rate in batch cultivations of *Escherichia coli* (Novick and Weiner 1957; Dong *et al.* 1995; Dekel and Alon 2005; Shachrai *et al.* 2010) and *Zymomonas mobilis* (Snoep *et al.* 1995) and strong selective disadvantages in chemostat cultivations using *E. coli* (Dean *et al.* 1986; Dean 1989; Lunzer *et al.* 2002; Stoebel *et al.* 2008). In *Saccharomyces cerevisiae*, a trade-off related to unneeded-protein expression was found (Lang *et al.* 2009). Dong, Nilsson, and Kurland found that unneeded protein can be expressed up to 30% of the total protein content before *E. coli* growth halts (Dong *et al.* 1995). They concluded that growth reduction was caused by competition for protein synthesis machinery between nonfunctional and growth-promoting proteins (*cf.* Vind *et al.* 1993). They also found significant reductions of ribosomal activity at high unneeded-protein expression, as if the cells experience a nutrient downshift (Dong *et al.* 1996). Stoebel *et al.* (2008) discovered that the costs of

doi: 10.1534/genetics.113.150631
Manuscript received February 20, 2013; accepted for publication March 21, 2013
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150631/-/DC1.
[1]Corresponding author: Systems Bioinformatics, IBIVU, VU University, 1081 HV Amsterdam, The Netherlands. E-mail: f.j.bruggeman@vu.nl

unneeded-protein synthesis of *E. coli*'s *lac* operon in chemostat cultures is due to the transcription and translation process, *e.g.*, the competition for RNA polymerases and ribosomes, rather than due to toxic effects or excessive usage of nucleotide or amino acid precursors. Other studies (Dong *et al.* 1995; Vind *et al.* 1993) also indicate that unneeded-protein synthesis is at the expense of the synthesis of other proteins that have growth-related activities; hence, these are all experimental indications of the existence of a cellular constraint that limits the cellular protein content. Several groups (Dong *et al.* 1995; Snoep *et al.* 1995; Stoebel *et al.* 2008; Shachrai *et al.* 2010) measured a linear dependency of the reduction of the growth rate on the unneeded-protein fraction whereas Dekel and Alon (2005) found a quadratic dependency of the growth-rate reduction on unneeded *lac* operon expression by *E. coli*. In all cases, strong dependencies of fitness on unneeded (or excess) protein synthesis was found.

From an evolutionary perspective, the high cost of unneeded-protein synthesis suggests that adjustments of protein partitioning over growth processes is an important mechanism for fitness enhancement of bacteria (Dong *et al.* 1996; Dekel and Alon 2005). Studies on translation control and the correlation between ribosome content and growth rate of *E. coli* indeed suggest that this organism aims to maximize its growth rate by optimal partitioning of protein over growth processes (Ehrenberg and Kurland 1984; Klumpp and Hwa 2008; Klumpp *et al.* 2009; Zaslaver *et al.* 2009). This could explain why gratuitous protein expression causes such drastic reductions in growth rate. That nonfunctional protein expression leads to growth-rate reduction is also supported by metabolic control theory, which proves that fluxes through metabolic networks scale linearly with the entire (functional) metabolic protein content (Kacser *et al.* 1973; Westerhoff and van Dam 1987; Snoep *et al.* 1995).

The evolutionary importance of optimal protein partitioning over growth processes indicates that a better understanding of the molecular basis and consequences of the cost of unneeded-protein synthesis is needed. We offer a theoretical framework for studying metabolic pathway evolution while the entire organism is under selection of the maximal specific growth rate in a batch cultivation. We start with an evolutionary analysis of metabolic enzyme levels. We derive how the optimal level of a metabolic enzyme can be understood in terms of its direct contribution to fitness (benefit) and its cost (usage of resources at the expense of other needed enzymes). We show that misbalancing of these quantities leads to reductions in fitness and sets the fitness potential (selection coefficients) of individual enzymes. Central to our framework is the concept of limiting growth process, *e.g.*, transcription or translation machinery, which bounds the maximal protein level that can be attained by a cell and is known to be major factor in metabolic evolution (Dekel and Alon 2005; Stoebel *et al.* 2008). We show with mechanistic biochemical models that alternative processes for setting the limits of the cellular protein content, *i.e.*, transcription, translation, or energy usage, all lead to a protein-constraint relation that is linearly dependent on protein concentrations. Together with basic mathematical properties of enzyme kinetics and metabolic pathways, this leads to a cost of unneeded-protein synthesis that is a linear function of the protein concentration, which is in agreement with most experimental data (Dong *et al.* 1995; Snoep *et al.* 1995; Stoebel *et al.* 2008; Shachrai *et al.* 2010). We define a fitness potential for every enzyme in the fitness landscape. This concept indicates the importance of every enzyme for enhancing fitness and can be expressed in terms of enzyme benefit and cost. We show that this measure is a specific formulation of the selection coefficient used in experimental studies (Stoebel *et al.* 2008) and is related to a control coefficient of metabolic control analysis (Kacser *et al.* 1973; Heinrich and Rapoport 1974).

Results

Growth processes set a limit to the cellular protein content

Generally, evolutionary optimization of fitness occurs under constraints. Metabolic networks have functional limits set by biochemical kinetics, thermodynamics, and physics. For instance, total available ATP sets a limit to biomass synthesis (Teusink *et al.* 2006), diffusion time scales limit reaction rates (Berg and Von Hippel 1985), and available membrane space sets the maximal nutrient uptake rate (Molenaar *et al.* 2009). Typically, several constraints act simultaneously.

We are interested in the constraints that set a bound to cellular protein content. Such a constraint, denoted by Φ_j, depends in principle on all the enzyme levels, e_i, and some weight factor for every enzyme, ω_i; thus, we obtain $\Phi_j(\omega, \mathbf{e})$ (boldface letters denote vectors). We give each constraint a bound $\Phi_j(\omega, \mathbf{e}) \leq R_j$ and generally refer to them as resource bounds. Every weight factor, ω_i, can be interpreted as the specific resource requirement of the associated metabolic enzyme. In the simplest case, only one constraint function occurs; defined as a weighted sum of enzyme levels,

$$\Phi(\omega, \mathbf{e}) = \sum_{i=1}^{n} \omega_i e_i \leq R. \quad (1)$$

The total number of metabolic enzymes equals n. This equation implicitly sets the total cellular protein content: $e_T = \sum_{i=1}^{n} e_i$. This constraint immediately suggests that unneeded-protein synthesis lowers the level of growth-related proteins. This is in agreement with experimental findings (Vind *et al.* 1993; Dong *et al.* 1995). These authors found reduced protein expression as a response to unneeded-protein synthesis. Below we quantify this effect and discuss several of its origins.

Regardless of whether a limitation on the number of RNA polymerases (or ribosomes) or in available energy (*e.g.*, in terms of ATP equivalents) is assumed or whether toxic effects resulting from protein synthesis are considered, Equation 1 emerges in each of these cases. The mathematical derivation of these scenarios can be found in the Supporting Information,

File S1. The only difference between these protein-limitation scenarios is that the ω's in Equation 1 have a different biochemical interpretation. Each of these alternative limitation scenarios have been suggested in the literature for setting the protein content per cell (Vind *et al.* 1993; Snoep *et al.* 1995; Klumpp and Hwa 2008; Stoebel *et al.* 2008; Klumpp *et al.* 2009; Molenaar *et al.* 2009; Zaslaver *et al.* 2009). Here we show that they can in principle all be described with the same constraint equation.

Equation 1 has an important consequence for the evolution of metabolic networks. It is an established fact that every flux in a metabolic network, J, increases with a factor α if the entire protein content of the network is increased with this factor α; this is one of the findings of metabolic control theory (Kacser *et al.* 1973; Westerhoff and Van Dam 1987; Giersch 1988). Mathematically, this means that the flux is a first-order homogeneous function of the cellular protein content. As a consequence, having more protein expressed in a metabolic network leads to higher fluxes. However, the total protein content is limited by constraints such as Equation 1. Given the existence of these constraints, if metabolic fluxes are to be optimized in evolution, it is the partitioning of proteins over the entire metabolic network that is being optimized.

Equation 1 implies a trade-off for metabolic systems and indicates the existence of an optimal combination of enzyme levels that maximizes fitness. The constraint $\Phi(\omega, \mathbf{e}) \leq R$ forces the use of cheap enzymes with low ω's as this allows for an increased cellular protein content and, as a consequence, higher fluxes. In a metabolic network, this will inevitably lead to expensive enzymes becoming progressively more limiting and eventually a requirement for an increase in their concentration. This causes a reduction of the total enzyme level in the network and a reduction of metabolic fluxes. Hence, there must exist some optimal combination of enzyme concentrations that balances these opposing forces and maximizes a specific flux given the constraint R. Thus, the optimization problem of a metabolic flux under a resource constraint involves maximizing J/R, which can be interpreted as maximizing the return (J) of investment (R).

Specific growth-rate optimization requires optimal protein allocation over growth processes: Maximization of J/R

The selection pressure in serial dilution experiments of batch cultivations at midexponential growth rate (balanced growth) is the maximal specific growth rate (Vasi *et al.* 1994). Under these conditions, all the nutrients are in excess. The specific growth rate, denoted by μ (per hour, hr^{-1}), is a J/R measure and, therefore, it directly applies to the constraint optimization problem we have just introduced (this is illustrated in-depth in File S2). To understand this, it is instructive to analyze the units of μ. The specific growth rate of a microorganism equals the production rate of biomass ("itself") expressed as gram biomass per hour *per* gram biomass or, equivalently the synthesis rate of protein by the cell divided by total cellular protein content. In other words, the specific growth rate just gives the rate at which one unit organism is produced by one unit organism; *i.e.*, the reciprocal growth rate is directly related to the generation time (t_g; $t_g = \ln(2)/\mu$). Thus, the specific growth rate is a self-replication rate. Therefore, selection for the maximal specific growth rate is directly related to the total (functional or needed) protein content of a cell and synthesis of unneeded protein will only reduce it.

Strictly speaking, the maximal specific growth rate is the selection pressure in (serial) batch cultivation; the theory that we present below does therefore not directly apply to chemostat cultivation. For evolution in chemostats, a different selection pressure that is not obviously related to cellular protein content applies; we return to this point in the discussion.

Operational definition of enzyme benefits and costs

Maximization of the specific growth rate is achieved by expressing every metabolic enzyme to the right level, such that none of the resources are wasted on the wrong enzymes and none of the enzymes are expressed at a too-low level. How can we figure out what the right expression level is for a specific metabolic enzyme? Intuitively, the right level of an enzyme is the enzyme amount at which the benefit minus the cost of the enzyme is largest. But what would be the biochemical definitions of enzyme benefit and cost such that if their difference is maximized the enzyme has attained its optimal level?

In our theory, pure cost originates from protein burden without function (unneeded-protein expression). Pure benefit originates from function without burden. Accordingly, the cost of an enzyme would be equal to the fractional reduction in flux (fitness) when a certain amount of the inactive form of the enzyme would be added, and the remaining enzyme concentrations (including the active form of the enzyme under consideration) would redistribute according to the corresponding reduced resource constraint to a new (and necessarily lower) flux optimum. This concept of cost exactly matches the definition of cost used in the analysis of the influences of unneeded-protein synthesis on growth (Vind *et al.* 1993; Dong *et al.* 1995; Snoep *et al.* 1995; Dekel and Alon 2005; Stoebel *et al.* 2008). When the flux and the constraint are homogeneous functions of the enzyme concentrations to the first order, the new optimum simply corresponds to the same fractional distribution of all active enzymes, but now with a reduction in the available resource corresponding to the amount of inactive enzyme. This predicts a linear relationship between cost and enzyme concentrations:

$$C_i(e_i, R) = -\frac{J(R - \omega_i \bar{e}_i) - J(R)}{J(R)} = \frac{\omega_i \bar{e}_i}{R}. \qquad (2)$$

The derivation of this equation can be found in File S3. It is based on the assumptions that the constraint function depends linearly on the enzyme concentration (Equation 1) and the flux is a first-order homogeneous function of the cellular protein content, which is generally valid in metabolic networks (Kacser *et al.* 1973; Westerhoff and van Dam 1987; Giersch 1988). The notation $\bar{e}_i$ signifies that the enzyme

is expressed in a nonfunctional form; it cannot contribute to fitness. The resource amount $R - \omega_i \bar{e}_i$ corresponds to the residual resource amount after having spent $\omega_i \bar{e}_i$ resource units on unneeded protein.

Experimentally, the cost can be determined by a measurement of the reduction in fitness upon expressing the enzyme of interest under a condition where it is not used (Dong *et al.* 1995; Dekel and Alon 2005; Stoebel *et al.* 2008). Equation 2 indicates that the cost of an enzyme equals its fractional resource usage. The exact reduction in resources is dependent on the enzyme properties as captured in the enzyme's ω coefficient. Interestingly, we find that enzyme cost is entirely independent of metabolic enzyme kinetics.

In a similar fashion, we define the benefit of an enzyme as the fractional increase in flux when the enzyme's specific activity would be increased by a certain fraction without a reduction in the available resource (R remains fixed) while all other enzymes remain at their (optimal) levels. Thus, the benefit of an enzyme is defined as the fractional increase in fitness upon an increase of active enzyme. This increase is not at the expense of any of the available total resources; this is done cost free. In File S4, we derive that the benefit equals the following relationship,

$$\mathcal{B}_i(e_i, R) = \frac{J\left(\hat{e}_{\mathbf{opt}}(e_i, R)\right)}{J\left(\mathbf{e}_{\mathbf{opt}}(R)\right)}. \tag{3}$$

The notation $J(\mathbf{e}_{\mathbf{opt}}(R))$ indicates the metabolic flux when all enzymes are expressed at their optimal level; $J(\hat{e}_{\mathbf{opt}}(e_i, R))$ indicates the flux when all enzymes, except for enzyme i, are kept at their optimal level and enzyme i is expressed to level, e_i. This means that the benefit equals 1 only when enzyme i is at its optimal level.

We emphasize that the benefit definition does not necessarily have to be defined with reference to the optimal levels of all the other enzymes in the system. We take this perspective here to simplify the discussion of the relationship of the enzyme benefit and its fitness potential that follows later. For most practical purposes the enzyme benefit can be better defined as: $\mathcal{B}_i(e_i, R) = J(\hat{e}(e_i, R))/J(\mathbf{e}(R))$. In this formulation, the benefit can become >1 if the expression level e_i was not optimal in the reference condition. When the benefit is measured in an experiment, as done by Dekel and Alon (2005), it this definition of benefit that is most relevant.

The benefit, in contrast to the cost, does depend on enzyme kinetics and requires consideration of the entire metabolic system. Dekel and Alon (2005) measured the benefit for the *lac* operon in *E. coli*. The benefit can be straightforwardly calculated for a mathematical model of a metabolic pathway. First, the reference flux is calculated given enzyme kinetic parameters, a characterization of the environment, and the resource constraint. The benefit curve for each enzyme is then calculated by determining the steady-state flux as function of enzyme level while all other enzymes remain fixed at their optimal values. Typically, the benefit of an enzyme will display saturation behavior with increasing concentrations. Here we have generalized Dekel and Alon's definitions of enzyme benefit to make them applicable to general metabolic pathways.

The enzyme benefit minus cost is maximized at the optimal enzyme level

What remains to be shown at this stage is that a maximization of the return on investment, *i.e.*, of J/R, indeed implies a maximization of benefit minus cost. This we derive in File S6 by showing that the optimization of the flux, J, under the constraint given by Equation 1, indeed gives rise to a maximization of benefit minus cost when the enzyme level is at its optimal level.

The fact that the benefit minus cost is maximal at the optimal level of the enzyme can also be derived from a different perspective. The derivative of benefit minus cost with respect to the enzyme level in the optimum should equal 0:

$$\frac{\partial \mathcal{B}_i\left(e_i^{\mathrm{opt}}, R\right)}{\partial e_i} - \frac{\partial C_i\left(e_i^{\mathrm{opt}}, R\right)}{\partial \bar{e}_i} = \frac{\partial \ln J\left(\mathbf{e}_{\mathbf{opt}}(R)\right)}{\partial e_i} - \frac{\omega_i}{R} = 0. \tag{4}$$

Multiplication of this equation with e_i^{opt} gives rise to the following expression at the optimal state,

$$\frac{\partial \ln J\left(\mathbf{e}_{\mathbf{opt}}(R)\right)}{\partial \ln e_i} = \frac{\omega_i e_i^{\mathrm{opt}}}{R}. \tag{5}$$

On the left-hand side we identify the scaled flux control coefficient of enzyme i, C_i^J, as defined in metabolic control analysis (MCA) (Kacser *et al.* 1973; Heinrich and Rapoport 1974). Interestingly, this result is in agreement with findings from Heinrich and co-workers, who arrived at the same relationship by maximizing the flux through a metabolic pathway under the constraint of fixed total enzyme concentration; *i.e.*, maximization of J/R (Klipp and Heinrich 1994, 1999; Heinrich and Klipp 1996; Heinrich and Schuster 1998). Examples of this relation for other constraint functions are shown in File S7. We retrieve this equation via a different route: through maximization of the difference between benefit and cost. Thus, C_i^J is related to its fractional resource usage at an optimal metabolic state. In a subsequent section, we use from the concept of a fitness landscape to show that this coefficient is also related to the fitness contribution of enzyme i.

A cost–benefit analysis differentiates the importance of enzyme kinetics and process costs

The definitions of enzyme benefits and costs address different aspects of protein expression. The benefit exclusively addresses the contribution of the enzyme activity to fitness (and is therefore zero for an unneeded or nonfunctional protein) without consideration of the cost, *i.e.*, the consequent reduction in the levels of other proteins upon protein expression due to a constraint (*cf.* Vind *et al.* 1993 for experimental evidence). The cost considers the reduction in fitness upon expression of the enzyme when it does not contribute to fitness. In Figure 1, the influence of

pathway kinetics and specific enzyme costs (process costs for transcription and translation, for instance) on the optimal enzyme level is illustrated. Changes in the benefit curve, due to changes in kinetic parameters of any of the pathway enzymes, changes not only the enzyme of interest but also others. This can cause changes in the optimal enzyme level or they derive from changes in enzyme costs. Changes in specific enzyme cost can, for instance, be introduced by decreasing the lifetime of the enzyme such that at steady state more ribosomes are required to sustain the enzyme level. The optimal level of the enzyme occurs in this plot when the slope of the benefit and cost curve are equal (Equation 4). The cost slope depends linearly on the specific cost of the enzyme. The benefit slope depends in a nonlinear manner on kinetic properties in the metabolic network. A sensitivity analysis of the kinetic parameters and the specific enzyme costs on the optimal enzyme level will give additional insight into kinetic and cost contributions in the optimum, *i.e.*, whether the catalytic enzyme constants (K_M or k_{cat}'s) force particular enzyme levels or whether transcription or translation costs dominate.

We emphasize that in Figure 1 the enzyme costs are expressed with respect to the total resource allocated to the pathway. Therefore, the costs in Figure 1 vary from 0 to 1. Alternatively, the total cellular resource amount could have been considered and then the costs would have been much smaller than 1; hence, the cost curves would have been much less steep. Typically, the total cellular resource requirement is unknown but the resource expended on a particular pathway can in principle be deduced from experiments at the pathway level. With mathematical models of metabolic pathways the resource allocated to the pathway can also be predicted, by fixing the metabolic pathway flux and the subsequent minimization of the resource requirement to achieve this flux. The predicted enzyme levels can then be compared to measurements. If the boundary metabolites of the pathway are fixed to measured concentrations and the metabolic pathway flux is also chosen according to the same experiment, then the optimal enzyme levels obtained by minimizing the resource allocated to the pathway will be in agreement with the situation when the entire cellular resource is minimized. Thus when costs are expressed in terms of pathway level resource allocation, the optimal enzyme levels are the same as when the costs are defined with respect to the total cellular resource. The enzyme cost at the cellular level is then obtained from $(R_{\text{path}}/R_{\text{cell}})(\omega_i e_i/R_{\text{path}})$. The flux control coefficient of the enzyme in the optimal state at the level of the entire cell then reduces by the same factor relative to its pathway level value; hence, the flux control coefficient becomes $(R_{\text{path}}/R_{\text{cell}})C_i^J$.

Exploring the relationships between selection coefficients, enzyme costs, and benefits, using a fitness landscape

A cost–benefit analysis sheds light on the optimal distribution of enzyme levels, under a given resource constraint, in terms of enzyme costs and metabolic system kinetics (benefit). What we do not understand at this stage is why some enzymes are more important than others for achieving the optimal state, *i.e.*, for changing (or adapting) fitness. Some enzymes therefore have a higher selection coefficient: for the same change in enzyme level some enzymes will have a higher influence on fitness than others. To address this issue we study the fitness landscape.

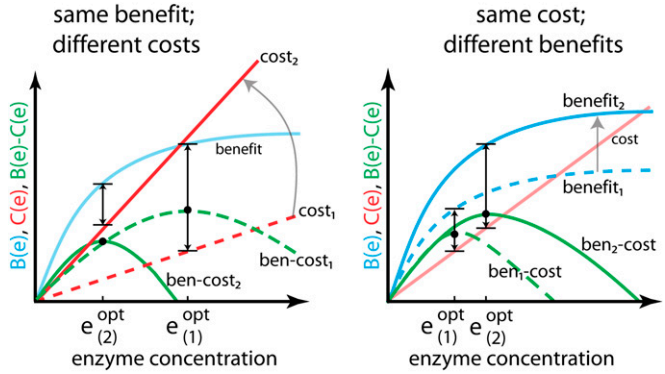

Figure 1 Influences of enzyme kinetics and specific enzyme cost on optimal enzyme levels. Schematic of the effect of different specific enzyme cost (left) and different benefit (right) on optimal enzyme expression. Enzyme kinetic parameters exclusively influence the benefit curve (blue) and the enzyme costs (red) depend only on the process costs for enzyme synthesis. These two factors can independently influence the optimal enzyme level. The difference between benefit and cost is shown by the green curves. The largest difference between benefit and cost is indicated for each scenario and corresponds to the optimum of the green curves. The optimal enzyme level occurs when the slope of the benefit and cost curve in these plots are equal (Equation 4).

We take the steady-state flux through a metabolic pathway as our fitness function. The fitness landscape is defined as the dependency of the pathway flux on all the enzyme levels given a resource constraint that limits total cellular protein content. The constraint bounds the fitness landscape $(\forall i : 0 \le e_i \le R/\omega_i)$. At the optimal combination of enzyme levels, the metabolic pathway flux is maximal and the fitness landscape displays a maximum. One sensible way to obtain an impression of such a multidimensional fitness landscape is to look along an enzyme concentration axis, say of enzyme j, and see how the maximal flux depends on the concentration of this enzyme, taking into account the resource constraint. This can be achieved by fixing enzyme j at some value e_j and then optimizing the flux under the "residual" constraint $R - w_j e_j$ (and $0 < e_j < R/w_j$). Only when e_j equals its optimal value, $e_j = e_j^{\text{opt}}$, the maximal flux J_{opt} is recovered and all enzyme levels will be at their optimal value. The dependency of the optimal flux on e_j, resulting from this procedure, is defined as the fitness landscape of e_j. We define the scaled slope of this fitness landscape for enzyme j at some level of e_j, *i.e.*, $\partial \ln J_{\text{opt}}/\partial \ln e_j$, as the fitness contribution of enzyme j at level e_j. We next describe an analytical expression for this fitness contribution that we derive in File S8.

We are interested in determining the slope of the dependency of the optimal flux (J_{opt}) around the optimal enzyme distribution $\mathbf{e_{opt}}$ for enzyme j. If in this region $J(\hat{e}_{\mathbf{opt}}(e_j, R)) \approx J_{\text{opt}}$ then changes in e_j hardly affect the optimal

flux as the dependency of $J(\hat{e}_{\mathbf{opt}}(e_j,R))$ on e_j is flat and, hence, $\partial J_{\mathrm{opt}}/\partial e_j$ will be small in this region. Alternatively, if $\partial J_{\mathrm{opt}}/\partial e_j$ is large, the dependency is steep and changes in the level of the enzyme have a large effect near the optimal flux. This suggests that this enzyme should evolve if it is not at its optimal expression level. The fitness contribution of enzyme j, $\mathcal{F}_j(e_j)$, is given by (see File S8 for derivation and application to a toy model)

$$\mathcal{F}_j(e_j) = \frac{\partial \ln J_{\mathrm{opt}}}{\partial \ln e_j} = \frac{C_j^J - C_j(e_j,R)}{1 - C_j(e_j,R)} = \frac{\partial \mathcal{B}_j/\partial \ln e_j - \partial C_j/\partial \ln e_j}{1 - C_j(e_j,R)}. \tag{6}$$

This equation has an intuitive interpretation. If the term $C_j^J - C_j(e_j,R)$ is large, a large change in the flux can be obtained at the expense of little resource investment in a change in the enzyme concentration. This signifies an enzyme with large evolutionary potential. Unneeded-protein synthesis can also be studied with this equation when the flux control coefficient is set equal to 0. We note that a conservation relationship exists for the fitness contributions at every point in the fitness landscape (File S8).

The expression for the fitness contribution of enzyme j has some insightful properties. It equals zero when $e_j = e_j^{\mathrm{opt}}$ because then $C_j^J = e_j^{\mathrm{opt}}\omega_j/R$ (Equation 5). The fitness contribution should be positive when $e_j < e_j^{\mathrm{opt}}$ and negative when $e_j > e_j^{\mathrm{opt}}$ (because we are considering a maximum). The denominator is always positive. Therefore, we need to have $C_j^J > e_j\omega_j/R$ when $e_j < e_j^{\mathrm{opt}}$ and $C_j^J < e_j\omega_j/R$ when $e_j > e_j^{\mathrm{opt}}$, exactly as intuition would suggest. When $C_j^J = 1$ (note that typically, $0 < C_j^J < 1$) we retrieve the largest fitness contribution. This means that a high C_j^J suggests a large fitness potential.

Discussion

In this work, we studied the biochemical basis of the constraints that limit the evolutionary adjustments in protein levels required to enhance fitness in batch growth conditions, *i.e.*, when selection acts on the specific maximal growth rate. This fitness objective, maximization of the specific maximal growth rate, can be interpreted as a maximization of the cellular self-replication rate. This becomes clear from its definition, as the biomass synthesis flux per unit biomass or synthesis of the growth machinery of a cell per growth machinery per cell. Hence, production of protein that does not contribute to growth would enhance the amount of protein machinery but not the synthesis flux of new synthesis machinery and therefore cause enhanced protein costs. Thus, for this selective pressure the minimization of resource usage to attain a particular growth rate—*i.e.*, effectively minimizing unneeded-protein synthesis—is a relevant hypothesis with ample experimental support (see Introduction).

In this work, we found that several models of biochemical constraints acting on protein synthesis lead to linear relations between the total available (limited) resources and the protein concentration (Equation 1 and File S1). These models differ in the mechanistic and kinetic interpretation of the specific resource requirement of an enzyme, its ω coefficient. What we did not consider in these models is that at high degrees of enzyme overexpression, additional protein toxicity influences can play a role, such as limitations of specific amino acids, cell morphological influences, or the formation of growth-inhibiting protein aggregates. We omitted these phenomena from our theoretical models because we believe their negative influences on growth will typically be negligible when enzyme levels are changed close to their natural expression level (typically <1% of total cellular protein content). Such experiments are required to determine enzyme costs of wild type or evolved strains in evolutionary studies. However, if enzyme costs are determined from experiments with significant overexpression, protein toxicity cannot be excluded. Moreover, under those conditions it cannot be ruled out that the enzyme cost is no longer a linear function of the enzyme concentration.

In chemostats, the synthesis of unneeded protein also caused a fitness reduction (Dean *et al.* 1986; Dean 1989; Lunzer *et al.* 2002; Stoebel *et al.* 2008) as measured by the selection coefficient. This is surprising, because the selective pressure in a chemostat is not directly linked to resource usage in contrast to the selective pressure in batch, which is essentially expressed in terms of total *functional/needed* protein. The selective pressure in chemostat is the ability to grow at the specific growth rate set by the dilution rate at the lowest possible concentration of the limiting nutrient in the bioreactor. Essentially, the selection pressure acts on the affinity (or more precisely $\mu_{\mathrm{max}}/K_{\mathrm{S}}$), where the selection pressure for the affinity for the substrate is most pronounced at low growth rate (far below the maximal specific growth rate of the organism, *e.g.*, Lunzer *et al.* 2002). It is not immediately evident that under these conditions, fitness can be enhanced by adjusting protein partitioning and whether the functional protein content should be maximized. This is partially because selection acts on substrate affinity and not on reproduction rate (and also not on the number of offspring, not per unit time and in terms of yield). Perhaps, unneeded-protein synthesis in chemostats leads to fitness reduction because the nonfunctional protein produced also goes at the expense of transporter protein, which can be expected to be important under chemostat selection conditions at low growth rates. Alternatively, the fitness in chemostats is enhanced by increased maximal growth rate, which is unlikely at low dilution rates, but cannot be ruled out. However, the basic biological explanations of the importance of protein constraints in chemostat selection are not straightforward, which is why we focused in this article on selection in batch cultures. The role of protein constraints in chemostat evolution experiments deserves more attention in future studies.

The enzyme fitness potential that we have proposed is intimately linked to the selection coefficient used in growth studies. Suppose two mutants, x and y, occur simultaneously in a batch reactor at the same time and they differ in their fitness. Typically a selection coefficient is defined, which

addresses how quickly the fitter mutant outgrows the other mutant. This is done by plotting the time evolution of the quantity $\ln(x/y)$. If mutant x derives from genotype y and differs only in the expression level of one enzyme, e_j, the rate of change of the selection coefficient, $(d/dt)\ln(x/y)$, equals $\mu(e_j + \Delta e_j) - \mu(e_j) \approx (\partial\mu/\partial e_j)\delta e_j$. In the absence of a constraint that limits the cellular protein content, $\partial\mu/\partial e_j$ would be the unscaled control coefficient of the enzyme j on fitness. In the presence of the protein constraint, $\partial\mu/\partial e_j$ equals the unscaled fitness potential of the enzyme (*i.e.*, $(J_{\text{opt}}/e_j)\mathcal{F}_j$). This indicates that the selection coefficient in a serial batch experiment is related to the fitness potential of the mutated enzymes and their benefits and costs. In fact, this correspondence between the control coefficient and the selection coefficient was exploited by Steve Oliver's group (Castrillo *et al.* 2007; Pir *et al.* 2012) when they measured the control coefficient of hundreds of enzymes on growth rate in chemostat using a single-allele knockout library. Note that for many applications, the fitness potential will be close to the control coefficient ($\mathcal{F}_j \approx C_j^J$), because the resource usage of enzyme will often be very small $R \gg \omega_j e_j$ and negligible compared to the value of C_j^J.

The theory described in this article can be useful for rationalizing experimental data of metabolic evolution of microorganisms. This theory extends earlier work on the application of metabolic control analysis to study metabolic fitness (Dykhuizen *et al.* 1987).

Acknowledgments

We thank Johan van Heerden for critical reading of the manuscript. We also thank anonymous reviewers for helping us improving the manuscript. This work was supported by Netherlands Organization for Scientific Research (NWO) Computational Life Sciences Project (CLS-635.100.18) and supported by Netherlands Bioinformatics Centre (NBIC, BioRange 4.10). This work is a contribution by the Kluyver Centre for Genomics of Industrial Fermentation, which is financially supported by the Netherlands Genomics Initiative. F.J.B. thanks the Netherlands Institute for Systems Biology (NISB), the Centre for Mathematics and Computer Science (CWI), and the Dutch Organization for Scientific Research (NWO, Grant number 837.000.001) for funding.

Literature cited

Beardmore, R. E., I. Gudelj, D. A. Lipson, and L. D. Hurst, 2011 Metabolic trade-offs and the maintenance of the fittest and the flattest. Nature 472(7342): 342–346.

Berg, O. G., and P. H. von Hippel, 1985 Diffusion-controlled macromolecular interactions. Annu. Rev. Biophys. Biophys. Chem. 14: 131–160.

Castrillo, J. I., L. A. Zeef, D. C. Hoyle, N. Zhang, A. Hayes *et al.*, 2007 Growth control of the eukaryote cell: a systems biology study in yeast. J. Biol. 6: 4.

Dean, A. M., 1989 Selection and neutrality in lactose operons of *Escherichia Coli*. Genetics 123: 441–454.

Dean, A. M., D. E. Dykhuizen, and D. L. Hartl, 1986 Fitness as a function of beta-galactosidase activity in *Escherichia coli*. Genet. Res. 48(1): 1–8.

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

Dong, H., L. Nilsson, and C. G. Kurland, 1995 Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J. Bacteriol. 177(6): 1497–1504.

Dong, H., L. Nilsson, and C. G. Kurland, 1996 Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J. Mol. Biol. 260(5): 649–663.

Dykhuizen, D. E., A. M. Dean, and D. L. Hartl, 1987 Metabolic flux and fitness. Genetics 115(1): 25–31.

Ehrenberg, M., and C. G. Kurland, 1984 Costs of accuracy determined by a maximal growth rate constraint. Q. Rev. Biophys. 17 (1): 45–82.

Friesen, M. L., G. Saxer, M. Travisano, and M. Doebeli, 2004 Experimental evidence for sympatric ecological diversification 2 to frequency-dependent competition in *Escherichia coli*. Evolution 58(2): 245–260.

Giersch, C., 1988 Control analysis of metabolic networks. 1. Homogeneous functions and the summation theorems for control coefficients. Eur. J. Biochem. 174(3): 509–513.

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

Heinrich, R., and T. A. Rapoport, 1974 A linear steady-state treatment of enzymatic chains. Control 95: 89–95.

Heinrich, R., and S. Schuster, 1998 The modelling of metabolic systems. Structure Control Optimality. Biosystems 47(1–2): 61–77.

Kacser, H., and J. A. Burns, 1973 The control of flux. Symp. Soc. Exp. Biol. 27: 65–104.

Klipp, E., and R. Heinrich, 1994 Evolutionary optimization of enzyme kinetic parameters: effect of constraints. J. Theor. Biol. 171(3): 309–323.

Klipp, E., and R. Heinrich, 1999 Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. Biosystems 54(1–2): 1–14.

Klumpp, S., and T. Hwa, 2008 Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proc. Natl. Acad. Sci. USA 105(51): 20245–20250.

Klumpp, S., Z. Zhang, and T. Hwa, 2009 Growth rate-dependent global effects on gene expression in bacteria. Cell 139(7): 1366–1375.

Lang, G. I., A. W. Murray, and D. Botstein, 2009 The cost of gene expression underlies a fitness trade-off in yeast. Proc. Natl. Acad. Sci. USA 106(14): 5755–5760.

Lunzer, M., A. Natarajan, D. E. Dykhuizen, and A. M. Dean, 2002 Enzyme kinetics, substitutable resources and competition: from biochemistry to frequency-dependent selection in lac. Genetics 162: 485–499.

Molenaar, D., R. van Berlo, D. de Ridder, and B. Teusink, 2009 Shifts in growth strategies reflect tradeoffs in cellular economics. Mol. Syst. Biol. 5(323): 323.

Novick, A., and M. Weiner, 1957 Enzyme induction as an all-or-none phenomenon. Proc. Natl. Acad. Sci. USA 43(7): 553–566.

Pir, P., A. Gutteridge, J. Wu, B. Rash, D. B. Kell *et al.*, 2012 The genetic control of growth rate: a systems biology study in yeast. BMC Syst. Biol. 6: 4.

Shachrai, I., A. Zaslaver, U. Alon, and E. Dekel, 2010 Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. Mol. Cell 38(5): 758–767.

Snoep, J., L. P. Yomano, H. V. Westerhoff, and L. Ingram, 1995 Protein burden in *Zymomonas mobilis*: negative flux

and growth control due to overproduction of glycolytic enzymes. Microbiology 141(9): 2329–2337.

Stoebel, D. M., A. M. Dean, and D. E. Dykhuizen, 2008 The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. Genetics 178: 1653–1660.

Teusink, B., A. Wiersma, D. Molenaar, C. Francke, W. M. de Vos *et al.*, 2006 Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. J. Biol. Chem. 281(52): 40041–40048.

Vasi, F., M. Travisano, and R. E. Lenski, 1994 Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to seasonal environment. Am. Nat. 144 (3): 432–456.

Vind, J., M. A. Sørensen, M. D. Rasmussen, and S. Pedersen, 1993 Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: expression from reporter genes does not always reflect functional mRNA levels. J. Mol. Biol. 231(5): 678–688.

Wenger, J. W., J. Piotrowski, S. Nagarajan, K. Chiotti, G. Sherlock *et al.*, 2011 Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. PLoS Genet. 7(8): e1002202.

Westerhoff, H. V., and K. van Dam, 1987 *Thermodynamics and Control Of Biological Free-Energy Transduction*. Elsevier, Amsterdam.

Zaslaver, A., S. Kaplan, A. Bren, A. Jinich, A. Mayo *et al.*, 2009 Invariant distribution of promoter activities in *Escherichia coli*. PLOS Comput. Biol. 5(10): e1000545.

Communicating editor: J. Lawrence

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150631/-/DC1

How Biochemical Constraints of Cellular Growth Shape Evolutionary Adaptations in Metabolism

Jan Berkhout, Evert Bosdriesz, Emrah Nikerel, Douwe Molenaar, Dick de Ridder, Bas Teusink, and Frank J. Bruggeman

DOI: 10.1534/genetics.113.150631

Contents

S1 Derivation of enzyme production cost based on underlying transcription and translation network

Evolutionary optimization of fitness will in general occur under constraints. Those constraints can have various origins, such as physical, biochemical or thermodynamic. For instance, total available ATP for biomass synthesis, diffusion timescales, available membrane space, or available cell volume can act as constraints. We here aimed to derive such constraints from underlying biochemical networks. We considered three constraints: (1) a resource (or energy) constraint; (2) a ribosome capacity constraint and (3) a RNA polymerase abundance constraint. In the next sections we will derive the origins of these constraints.

S1.1 Derivation of enzyme production cost based on energy or resources constraint

Let us consider the gene network as depicted in Figure S1. In this network, the mRNA product of gene i, indicated as m_i, is synthesized and degraded with rate constants: k_s^m and k_d^m, respectively. The mRNA m_i stimulates protein synthesis of p_i and p_i is degraded with rate constant k_d^p. The conversion of substrate S into product P is catalyzed by p_i, with steady state flux J.

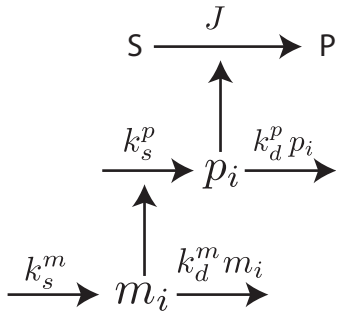

Figure S1. Overview of biochemical network. The product of gene i, m_i stimulate the synthesis of protein p_i, which in turn catalyses the conversion from S to P at the metabolic network. Both, m_i and p_i have specific synthesis (s) and degradation (d) rates, with first-order rate constant k.

The enzyme production cost for enzyme i is defined as the amount of resources required per unit enzyme per unit time. Therefore, the enzyme production cost consists of the cost for transcription of the corresponding gene(s) and the cost of translation of the mRNA(s). For the network depicted in Figure S1 this leads to:

$$\omega_i = \frac{J^p \omega_i^p}{p_i} + \frac{J^m \omega_i^m}{p_i} \tag{S1}$$

where J^m and J^p are respectively the rate for transcription and translation, and ω_i^m and ω_i^p, the corresponding cost of transcription and translation per unit rate. Assuming mass-action kinetics for transcription and translation we can write:

$$\omega_i = \frac{k_d^p\ p_i\ \omega_i^p}{p_i} + \frac{k_d^m\ m_i\ \omega_i^m}{p_i} \tag{S2}$$

The characteristic life times (τ) will generally equal one over a degradation rate constant: $1/k_d$. As a result, we obtain the equation that characterizes the enzyme production cost:

$$\omega_i = \left(\frac{\omega_i^m m_i}{\tau_i^m p_i} + \frac{\omega_i^p}{\tau_i^p} \right) \tag{S3}$$

This equation can easily be extended with relevant properties for other metabolic networks, such as, for instance, an additional cost term for the misfolded proteins. Note that the ratio m_i/p_i can be expressed in terms of rate constants, k_d^p/k_s^p.

S1.2 Derivation of enzyme production cost based on ribosomal occupancy constraint

Here we assume that the amount of cellular protein is limited by the availability of ribosomes and consequently by the competition of mRNA's for ribosomes. The total amount of ribosome, r, equals the free amount, r_F, and the total bound pool $mr = \sum_j m_j r$ then,

$$r = r_F + mr = r_F + \sum_j m_j r \tag{S4}$$

Note that r depends on the specific growth rate μ, we don't need this dependency now so it is omitted. The total fraction of occupied ribosome, ϕ, is approximately fixed in *E. coli* and several other microorganisms across growth conditions and equals,

$$\phi = \frac{mr}{r} = \sum_j \frac{m_j r}{r} \tag{S5}$$

The resource requirement of a mRNA, later we determine the requirement for its protein product, is given by the fraction of all ribosomes it occupies,

$$C_j = \frac{m_j r}{r} \tag{S6}$$

So, we have

$$\phi = \sum_j C_j \tag{S7}$$

The concentration of $m_j r$ can be approximated by with K_j as the dissociation constant of the j-th mRNA for the ribosome,

$$m_j r = r \frac{\frac{m_j}{K_j}}{1 + \sum_k \frac{m_k}{K_k}} \tag{S8}$$

So, we have

$$C_j = \frac{m_j r}{r} = \frac{\frac{m_j}{K_j}}{1 + \sum_k \frac{m_k}{K_k}} \tag{S9}$$

The concentration of the correspond protein p_j with translation rate constant k_j and degradation rate constant δ_j equals,

$$p_j = \frac{k_j}{\delta_j} m_j r \tag{S10}$$

The ribosome resource consumption expressed for protein now becomes,

$$C_j = \frac{m_j r}{r} = \frac{p_j \delta_j}{k_j r} \tag{S11}$$

So, the constraint equation becomes

$$r\phi = \sum_j r C_j = \sum_j \omega_j p_j \tag{S12}$$

This last equation sets a bound to the protein concentration in a cell and derives from the limited availability of ribosome for translation of mRNA. The ω_j, the specific resource consumption of protein j now equals, $\frac{\delta_j}{k_j}$.

S1.3 Derivation of enzyme production cost based on RNA polymerase availability constraint

The derivation of the biochemical constraint resulting from the competition for a limiting pool of RNA polymerase is equivalent to the derivation in the previous section. We start from the total pool, r, which equals the sum of the unbound polymerases, r_F, and the bound pool, $\sum_i r_i$,

$$r = r_F + \sum_i r_i \tag{S13}$$

At steady state, the level of mRNA j equals,

$$m_j = \frac{k_j^m r_j}{\delta_j^m}, \tag{S14}$$

where k_j^m equals the transcription activity constant and δ_j^m the mRNA degradation rate constant. The protein level of enzyme j equals,

$$p_j = \frac{k_j^p m_j}{\delta_j^p} = \frac{k_j^m k_j^p r_j}{\delta_j^m \delta_j^p} \tag{S15}$$

where k_j^p equals the translation rate constant and δ_j^p the proteins degradation rate constant. From the last equation we obtain an expression for r_j in term of the protein level,

$$r_j = \frac{\delta_j^m \delta_j^p}{k_j^m k_j^p} p_j \tag{S16}$$

If we assume that the bound RNA polymerase pool is approximately fixed then the bound fraction ϕ is a constant,

$$r\phi = r\frac{\sum_i r_i}{r} = \sum_i \omega_i p_i, \tag{S17}$$

where ω_i is defined as $\frac{\delta_j^m \delta_j^p}{k_j^m k_j^p}$. Again the linear protein constraint emerges from a biochemical constraint.

S2 The specific growth rate is a J/R measure

A central argument in our work is that the selection pressure in batch cultures of bacteria is the specific growth rate and that this quantity is maximized in evolution by optimal partitioning of proteins over growth processes, given a bound on the cellular protein content that derives from a biochemical constraint in the growth process, such as limited availability of RNA polymerases, ribosomes or energy. We will now illustrate why the specific growth rate is so tightly linked to the biochemical constraint, which sets a limit to a resource (R) for growth.

The specific growth rate, μ (hr^{-1}), is related to the yield on the growth substrate, (S), denoted by $Y_{X/S}$ (in gram dry weight per mol substrate) and the uptake rate of the growth substrate, J_S, in mol substrate per gram dry weight per hr,

$$\mu = Y_{X/S} J_S = \frac{J_S}{1/Y_{X/S}} = \frac{J_S}{R^*} \tag{S18}$$

R^* is also a resource requirement but then for growth substrate, i.e. it's units are mol S per gram dry weight. But we are interested in an intracellular growth resource such as ATP equivalents. To make this explicit we decompose R^* into separate factors,

$$R^* = \frac{1}{Y_{ATP/S}} \frac{1}{Y_{X/ATP}} = \frac{1}{Y_{ATP/S}} \frac{R}{\alpha e_T} \tag{S19}$$

$\frac{1}{Y_{ATP/S}}$ equals the amount of substrate required to make one mole of ATP; this quantity depends on the pathways that the organism uses to make ATP. $Y_{X/ATP}$ is the yield of biomass expressed as gram dry weight per 1 mol of ATP. $\frac{1}{Y_{X/ATP}}$ is the ATP resource requirement $R = \sum_i \omega_i e_i$ divided by 1 gram dry weight; 1 gram dry weight can be expressed in terms of the total protein content $e_T = \sum_i e_i$ (e_i are now enzyme concentrations) as αE where α contains the cellular protein fraction and protein weight.

Consideration of equation S18 and S19 gives,

$$\mu = Y_{X/S} J_S = \frac{J_S}{\frac{1}{Y_{ATP/S}} \frac{R}{\alpha e_T}}, \tag{S20}$$

which shows that μ and J_S/R are linked as described in the main text.

Since, the resource equation $R = \sum_i \omega_i e_i$ does not set the protein content e_T for the same amount of resource, different protein amounts can obtained leading to changes in flux, J. Because J is first-order homogeneous function of the protein content, i.e. $\lambda J = J(\lambda e_T)$, we have, if the $Y_{X/S}$ is fixed, $\lambda\mu = \mu(\lambda e_T)$, which shows that analysis of flux gives rise to same results for the cost and benefit as the analysis of μ. Therefore, the specific growth rate can change because the flux changed at a constant yield, which is a stoichiometric property (calculable with FBA),

$$\delta\mu = Y_{X/S} \delta J_S = Y_{X/S} \sum_i \frac{\partial J}{\partial e_i} \delta e_i \tag{S21}$$

If the $Y_{X/S}$ is fixed then R^* is fixed,

$$\delta R^* = \frac{\partial R^*}{\partial Y_{ATP/S}} \delta Y_{ATP/S} + \frac{\partial R^*}{\partial R} \delta R + \frac{\partial R^*}{\partial \alpha} \delta\alpha + \frac{\partial R^*}{\partial e_T} \delta e_T = 0 \tag{S22}$$

We also assumed R to be fixed, i.e. same amount of ATP is available, and if we assume in addition that the pathways are used for ATP from S then also $Y_{ATP/S}$ is fixed and, therefore,

$$\frac{\partial R^*}{\partial \alpha} \delta\alpha = -\frac{\partial R^*}{\partial e_T} \delta e_T \tag{S23}$$

The protein content of the cell (part of α) must have been changed as a result of a change in the total amount of protein per cell. Then, we have $\delta e_T = \sum_i e_i$ which leads to a flux change of $\delta J = \sum_i \frac{\partial J}{\partial e_i} \delta e_i$.

This section therefore showed the linkage between μ and R, and how changes in the protein content δe_T at fixed R can lead to an uptake flux change δJ, which causes a change in the specific growth rate μ at fixed biomass yield on substrate, $Y_{X/S}$. This indicates that the μ in equation S20 changes because J_s changes due to a change in the e_T and a compensating change in α. This is the mechanism for evolutionary change of metabolic pathway activity resulting from specific growth rate selection given a biochemical constraint that we discuss in this paper.

S2.1 A subnetwork perspective

One of the questions asked in the main text is whether we can study evolutionary adaptation in μ at the level of a single pathway by only considering the resource amount of the pathway and repartitioning this over pathway enzymes to enhance fitness. Then $\delta R = 0 = \delta R_{pathway} + \delta R_{rest}$. If we then demand only enzyme changes pathway then $\forall i: \ \delta R_{rest,i} = 0$ and all enzymes in the remained of the network stay fixed. Then, the enzyme changes considered in the previous section only concern pathway enzymes.

S3 Derivation of the cost function

We define a vector of optimal enzyme concentrations, $\mathbf{e}_{opt}$, given an available amount of resource R that maximizes the pathway flux J of interest to the value J_{opt},

$$J_{opt}(R) = J(\mathbf{e}_{opt}(R)).$$

For metabolic pathways, this flux is a first-order homogeneous function of the enzyme concentration such that $\alpha J = J(\alpha \mathbf{e})$ (1). This property holds for general enzyme kinetics as long as there are no complexes of different enzymes catalyzing metabolic reactions; i.e. in the absence of metabolic channeling (8).

Then the cost of an enzyme i is defined as,

$$\mathcal{C}_i(e_i, R) = \frac{J(\mathbf{e_{opt}}(R)) - J(\mathbf{e_{opt}}(R - \omega_i \bar{e_i}))}{J(\mathbf{e_{opt}}(R))} \tag{S24}$$

where $J(\mathbf{e_{opt}}(R - \omega_i \bar{e_i}))$ denotes the maximal flux at the optimal enzyme distribution but at a lower value of R due to the expression of dummy protein. The reduction in "useful" resource $R - \omega_i \bar{e_i}$ can be written as a multiplication of the available resource R by the factor $1 - \frac{\omega_i \bar{e_i}}{R}$. Because J and R are each first order homogeneous functions with respect to the enzyme concentration ($\alpha R = R(\alpha \mathbf{e})$), dummy enzyme expression leads to a reduction of J_{opt} by the factor $1 - \frac{\omega_i \bar{e_i}}{R}$. Substitution of this relationship into equation S24 reveals that the functional cost of enzyme i equals its fractional resource usage,

$$\mathcal{C}_i(e_i, R) = \frac{\omega_i \bar{e_i}}{R} \tag{S25}$$

S4 Derivation of the benefit function

To define the benefit of enzyme i at concentration e_i we define the fractional change in the flux relative to the optimal state,

$$f(e_i) = \frac{J(\hat{\mathbf{e}}_{\mathbf{opt}}(e_i, R)) - J(\mathbf{e}_{\mathbf{opt}}(R))}{J(\mathbf{e}_{\mathbf{opt}}(R))}$$

where $J(\hat{\mathbf{e}}_{\mathbf{opt}}(e_i, R))$ denotes the flux vector where all enzymes concentrations are at their optimum corresponding to resource constraint R except for enzyme i, which is at concentration e_i. Accordingly, we define the enzyme vector,

$$\hat{\mathbf{e}}_{\mathbf{opt}}(e_i, R) = \left\{e_1^{opt}, \ldots, e_i, \ldots, e_n^{opt}\right\}$$

Hence, the following relationship holds: $\hat{\mathbf{e}}_{\mathbf{opt}}(e_i^{opt}, R) = \mathbf{e}_{\mathbf{opt}}(R)$. Using this definition for f, it follows that the relative flux difference is zero, when $e_i = e_i^{opt}$ (see Figure S3). It can be seen that f becomes negative when $e_i < e_i^{opt}$. To avoid this, we define the benefit of enzyme i, $\mathcal{B}_i(e_i)$ as,

$$\begin{aligned} \mathcal{B}_i(e_i, R) &= f(e_i) + f(0) \\ &= \frac{J(\hat{\mathbf{e}}_{\mathbf{opt}}(e_i, R))}{J(\mathbf{e}_{\mathbf{opt}}(R))} \end{aligned} \tag{S26}$$

Here we used a property of linear metabolic pathways: $J(\hat{\mathbf{e}}_{\mathbf{opt}}(e_i = 0, R)) = 0$. In this formulation, benefit corresponds to intuition: it is positive and will typically be an increasing function with enzyme concentrations.

Since we used the optimal state as reference state, the relative flux difference is zero at e_i^{opt}, as indicated by the arrow (Figure S3). Consequently, at concentration of $e_i < e_i^{opt}$ the relative flux difference is negative (dashed line). To prevent this we add the term, without e_i being present, which is denoted by $f(0)$. The benefit function consists thus of two terms: $f(e_i) + f(0)$. The addition of $f(0)$ does not influence the optimal concentration as indicated by the vertical dashed line.

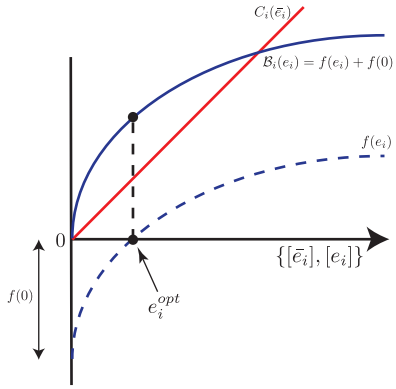

Figure S3. Illustration of the cost and benefit function as function of dummy ($\bar{e}_i$) and useful enzyme (e_i), respectively.

S5 Disentangling of the interplay between metabolic system kinetics and enzyme production cost using cost and benefit functions

With the general definitions of cost and benefit as described in the main text, the interplay between pathway kinetics and enzyme costs will be addressed. We will consider a simple toy model, which consists of two enzymatic steps. Both catalyzed by irreversible product-sensitive Michaelis-Menten kinetics. We address three different issues, all relevant for biochemical pathways: i) changing the affinity of the first enzyme for it's external substrate, ii) changing the maximal capacity of the first enzyme and iii) different enzyme production costs for the enzymes in the pathway.

All simulations are carried out with the constraint function $\Phi(\mathbf{e}) = \omega_1 e_1 + \omega_2 e_2$, and $\Phi(\mathbf{e}) \leq R$. The resource availability R and all kinetic parameter values are fixed, except for the perturbed parameter as just described. We then optimize the steady state flux using the enzyme concentrations and corresponding metabolite levels as variables. We explain these "direct" optimizations in terms of a cost and benefit analysis. For all three perturbations we present our results as a panel of three plots: in the first plot the cost (red) and benefit (blue) functions are plotted, the second plot shows the benefit minus cost (green) function and the third plot corresponds to the direct optimization where we plot the flux relative to the Vmax. In all three plots the curves are plotted as function of the first enzyme. The solid thick line corresponds to the original parameter values, with a black dot indicating the optimum, and the dashed line to the perturbed case, with a black triangle indicating the optimum. The maximal difference between the benefit and cost curve(s) is indicated with a dashed gray line (Figure S4).

To simulate a change in the affinity of the first enzyme, we decreased the affinity constant with a factor 10 (from 1 to 0.1). Note that a decrease in Km reflects a higher affinity. We find that an increase in affinity leads to a lower optimal expression level. This is reflected by a steeper benefit function, which also saturates at a lower level, indicating that there exists a trade-off between affinity and maximal activity.

The perturbation in kcat was performed by an increase from 2 to 20 (note, that the kcat of e_2 is equal to 10). Increasing the kcat of e_1, results in a lower concentration in the optimum. However, due to the resource constraint, investing a lot resources in an enzyme with a low catalytic capacity is at the expense of other enzyme(s). Therefore, we observe a higher optimum at the benefit minus curve but at the same time a lower steady-state flux.

We also simulated the effect of different enzyme cost strategies: reference values used are $\omega_1 = 0.25, \omega_2 = 0.75$, and for the perturbed case we set these parameters to: $\omega_1 = 0.9, \omega_2 = 0.1$. We found an inverse relation between enzyme cost and its optimal level: upon an increase of the ω optimal level decreased. However, the story is a bit more complicated because due to the resource constraint; a change in the $\omega's$ will lead to different total enzyme levels and hence different optimal flux values. This is also reflected by the plot showing the direct optimization: if $e_1 = e_{total}$ then e_2=0 and hence $J = 0$. For the reference conditions this is achieved at $e1 = (R - e2 \cdot w2)/w1 = (1 - 0 \cdot 0.75)/0.25 = 4$. Following a similar calculation we obtain $e_1 = 10/9$ for the perturbed scenario. The explanation for this behavior becomes apparent from inspecting the cost and benefit curves: a higher specific enzyme costs makes a steeper cost function and consequently, the optimal enzyme concentration decreases.

Note that for the first two perturbations, the cost line (red) is not affected, e.g. it is entirely independent of enzyme kinetics. This is also what we found in the main text. The benefit functions on the other hand does change for all three perturbations.

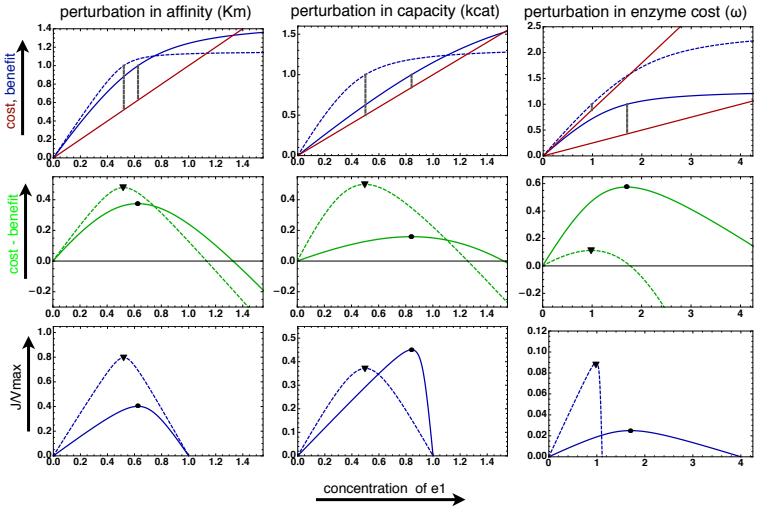

Figure S4. Interplay between enzyme kinetics and enzyme production cost on the optimal enzyme distribution. For a two-enzyme metabolic network three perturbations were applied: a change in the affinity of the first enzyme for it's substrate (left panel), a change in the catalytic capacity of the first enzyme (middle panel) and a change in the enzyme production costs (right panel). In all plots the solid line corresponds to the original parameter set and the solid line the to perturbed scenario. The upper row of plots shows the cost (red) and benefit (blue) as function of the concentration of e_i. The gray dashed lines indicated the maximum between benefit and cost. The middle row of plots corresponds to the benefit minus cost functions. The optimum of these curves is indicated by a solid black circle (reference parameters) and triangle (perturbed parameters, and corresponds to the maximal difference between benefit and cost curves. The lower row of plots shows the flux relative to the Vmax of the first enzyme. Simulations were performed using the following reaction rate: $v_i = kcat_i \cdot e_i \cdot s/(1 + s/K_m^s + p/K_m^p)$ and unless otherwise noted, parameter values were chosen as follows: $kcat_{1,2} = 10; K_{m1,2}^S = 1$ and $0.1; K_{m1,2}^P = 5; S = 1, P = 1; R = 1, \omega_{1,2} = 1$

S6 Maximization of the return of investment

Consider the return on investment of a metabolic pathway, defined $J(\mathbf{e})/R(\mathbf{e})$, where $\mathbf{e} = \{e_i, \ldots, e_r\}$ is the vector of enzyme concentrations, J is the steady state flux though the pathway and R is the amount of resources needed to maintain $\mathbf{e}$.

We are interested the enzyme levels that maximize the return on investment:

$$\text{Maximize}: \quad \frac{J(\mathbf{e})}{R(\mathbf{e})} \tag{S27}$$

Note that there is a subtle difference between this objective and maximizing J for a giving amount of R. However, as we will show below, for linear cost functions the optima of these two objectives are equivalent. Since $\ln x$ is a monotonically increasing function, $f(x)$ and $\ln(f(x))$ generally have the same maximum, i.e. $\ln \arg(\max(\ln f(x))) = \arg(\max(f(x)))$. Thus, equation S27 is equivalent to

$$\text{Maximize}: \quad \ln(\frac{J(\mathbf{e})}{R(\mathbf{e})}) = \ln(J(\mathbf{e})) - \ln(R(\mathbf{e})) \tag{S28}$$

Clearly, maximizing equation S28 for enzyme i requires

$$0 = e_i \left(\frac{\partial \ln J(\mathbf{e})}{\partial e_i} - \frac{\partial \ln R(\mathbf{e})}{\partial e_i} \right). \tag{S29}$$

The first term we identify as the flux control coefficient of e_i, C_i^J. Now, consider the linear cost function

$$R = \sum_{j=1}^{r} w_j e_j. \tag{S30}$$

for which we see that $e_i \frac{\partial \ln R(\mathbf{e})}{\partial e_i} = w_i e_i / R$. The condition for optimality of e_i is thus:

$$C_i^J = \frac{w_i e_i}{R}. \tag{S31}$$

Note that this condition does not require the other enzyme, $e_{j \neq i}$, to be optimal. Condition S31 gives the optimal level of e_i for any distribution of the other enzymes.

S6.1 Maximal difference between benefit and cost corresponds to maximization of return of investment

What remains to be shown at this stage is that a maximization of the return on investment, i.e. of J/R, indeed implies a maximization of $\mathcal{B}_i(e_i, R) - \mathcal{C}_i(e_i, R)$. At the optimal state (at $\mathbf{e_{opt}}(R)$) we obtain,

$$\frac{\partial \mathcal{B}_i(e_i^{opt}, R)}{\partial e_i} - \frac{\partial \mathcal{C}_i(e_i^{opt}, R)}{\partial \bar{e_i}} = \frac{\partial \ln J(\mathbf{e_{opt}}(R))}{\partial e_i} - \frac{\omega_i}{R} = 0$$

Multiplication of this equation with e_i^{opt} gives rise to the following expression at the optimal state:

$$\frac{\partial \ln J(\mathbf{e_{opt}}(R))}{\partial \ln e_i} = \frac{\omega_i e_i^{opt}}{R}. \tag{S32}$$

On the left hand side we identify the scaled flux control coefficient of enzyme i, C_i^J, as defined in metabolic control analysis (MCA) (4). Interestingly, this result is in agreement with findings from Heinrich and co-workers, who arrived at the same relationship by maximizing the flux through a metabolic pathway under the constraint of fixed total enzyme concentration; i.e. maximization of J/R (7; 6; 2; 3) (examples of this relation for other constraint functions are shown below in Section S7).

S7 Derivation of relationship between flux control coefficient and constraint function

Let us consider a linear metabolic network of r reactions under the following constraint,

$$g(\mathbf{e}) = \sum_{i=1}^{r} \omega_i e_i = \sum_{i=1}^{r} c_i = c \tag{S33}$$

We assume that the flux is maximal and that the above constraint is true. Using the Lagrange multiplier, λ, we can find a relationship between the flux derivatives to the enzyme concentrations and the enzyme constraint,

$$\frac{\partial}{\partial e_i}\left(J - \lambda\left(\sum_{i=1}^{r} \omega_i e_i - c\right)\right) = 0 \tag{S34}$$

This gives then for every e_i,

$$\frac{\partial J}{\partial e_i} - \omega_i \lambda = 0 \tag{S35}$$

From the definition of the flux control coefficient we obtain,

$$C_i^J = \frac{\partial J}{\partial e_i}\frac{e_i}{J} = \omega_i \lambda \frac{e_i}{J} \tag{S36}$$

The summation theorem of flux control coefficients leads to,

$$\sum_{i=1}^{r} C_i^J = \frac{\lambda}{J}\sum_{i=1}^{r} \omega_i e_i = 1 \tag{S37}$$

Equation S33 then shows that,

$$\frac{J}{\lambda} = c \tag{S38}$$

So the control coefficients at optimal states becomes,

$$C_i^J = \frac{w_i e_i}{c} = \frac{c_i}{c} \tag{S39}$$

If all w_i's are 1 and $c = e_{total}$ (total enzyme concentration), we obtain the familiar relationship $C_i^J = e_i/e_{total}$ (2).

S7.1 Derivation of the relationship between flux control coefficients and the constraint function

Now we consider the same pathway with the more general constraint,

$$g(\mathbf{e}) = c \tag{S40}$$

The Langrange multiplier now relates to the flux derivative in the optimum as,

$$\frac{\partial J}{\partial e_i} - \frac{\partial g}{\partial e_i}\lambda = \frac{\partial J}{\partial e_i} - \lambda = 0 \tag{S41}$$

From the definition of the flux control coefficient we obtain,

$$C_i^J = \frac{\partial J}{\partial e_i}\frac{e_i}{J} = \frac{\partial g}{\partial e_i}\lambda\frac{e_i}{J} \tag{S42}$$

The summation theorem of flux control coefficients leads to,

$$\sum_{i=1}^{r} C_i^J = \frac{\lambda}{J} \sum_{i=1}^{r} \frac{\partial g}{\partial e_i} e_i = 1 \tag{S43}$$

Since,

$$\frac{J}{\lambda} = \sum_{i=1}^{r} \frac{\partial g}{\partial e_i} e_i \tag{S44}$$

and the flux control coefficient then becomes,

$$C_i^J = \frac{\frac{\partial g}{\partial e_i} e_i}{\sum_{i=1}^{r} \frac{\partial g}{\partial e_i} e_i} \tag{S45}$$

S8 Derivation of an enzyme fitness landscape

S8.1 Introduction to the main concept and equations

We study the slope of the dependency of the optimal flux (J_{opt}) around the optimal enzyme distribution $\mathbf{e_{opt}}$ for enzyme j. If in this region $J(\hat{\mathbf{e}}_{\mathbf{opt}}(e_j, R)) \approx J_{opt}$ then changes in e_j hardly affect the optimal flux as the dependency of $J(\hat{\mathbf{e}}_{\mathbf{opt}}(e_j, R))$ on e_j is flat and, hence, $\partial J_{opt}/\partial e_j$ will be small in this region. Alternatively, if $\partial J_{opt}/\partial e_j$ is large, the dependency is steep and changes in the level of the enzyme have a large effect near the optimal flux. This suggests that this enzyme should evolve if it is not at its optimal expression level. The fitness contribution of enzyme j, $\mathcal{F}_j(e_j)$, is given by (see Section S8.4 for application to a toy model),

$$\mathcal{F}_j(e_j) = \frac{\partial \ln J_{opt}}{\partial \ln\, e_j} = \frac{C_j^J - \frac{e_j\omega_j}{R}}{1 - \frac{e_j\omega_j}{R}} = \frac{C_j^J - \mathcal{C}_i(e_i, R)}{1 - \mathcal{C}_i(e_i, R)} \tag{S46}$$

This equation has an intuitive interpretation. If the term $C_j^J - \mathcal{C}_i(e_i, R)$ is large, a large change in the flux can be obtained at the expense of little resource investment in a change in the enzyme concentration. This signifies an enzyme with large evolutionary potential.

A conservation relationship exists for the fitness contributions at every point in the fitness landscape,

$$\sum_{k=1}^{n} \mathcal{F}_k(e_k) = 0 \tag{S47}$$

This relationship holds because $\sum_{k=1}^{n} C_k^J = 1$ (5; 4) and $\sum_{k=1}^{n} \frac{\omega_i e_i}{R} = 1$ (equation **??**). This equation agrees with intuition if the total cellular protein content would be increased the specific growth rate would remain the same.

S8.2 Derivation of the enzyme fitness potential

We define the steady state flux of a metabolic pathway as function of the enzyme concentrations in this pathway as our fitness function, $J(\mathbf{e})$, and demand a maximal flux given a constraint on the enzyme concentrations, $R = \sum_{k=1}^{r} w_k e_k$. We assume that there exists a combination of enzyme levels that satisfies the constraint and maximizes the flux, this vector of enzyme concentrations $\mathbf{e^o}$ is defined as,

$$\mathbf{e^o} = \arg\max \left(J(\mathbf{e}|R = \sum_{k=1}^{r} w_k e_k) \right) \tag{S48}$$

Here "arg" denotes argument, so "arg $f(x) = x$" in mathematical notation, and "$|R = ...$" means "subject to $R = ...$" or "under the constraint $R = ...$". We define the maximal flux J^o as $J(\mathbf{e^o}|R = \sum_{k=1}^{r} w_k e_k)$.

Now that we have defined the optimum we will define a fitness landscape and study some of its properties. One sensible way to define a fitness landscape is to look along an enzyme concentration axes, say of enzyme j, and see how the maximal flux depends on the concentration of this enzyme, e_j. A convenient way of doing this is is by optimizing the flux while e_j is fixed at some value and the flux is optimized by adjusting the other enzyme levels under the constraint $R - w_j e_j$ (and $0 < e_j < R/w_j$). Only when e_j equals it optimal value, $e_j = e_j^o$, do we recover the maximal flux, J^o, and for all concentrations of $e_j \neq e_j^o$ we have a flux smaller J_o. The dependency of the optimal flux on e_j we define as the fitness landscape of e_j; more strictly

$$J_o(e_j^*) = J(\mathbf{e}|R - w_j e_j^* = \sum_{\substack{k=1 \\ k\neq j}}^{r} w_k e_k \cup e_j = e_j^*) \tag{S49}$$

And therefore $J_o = J_o(e_j^o) = J(\mathbf{e^o}|R = \sum_{k=1}^{r} w_k e_k)$.

S8.3 Ordering enzymes according to evolutionary urgency

Clearly if $J_o(e_j^*)$ is studied around $\mathbf{e_o}$ and in this region $J_o(e_j^*) \approx J_o$ we are looking at an enzyme that does not set the optimal flux to a great extent. So, then $\frac{\partial J_o}{\partial e_j}$ is small in this region. On the other hand, if $\frac{\partial J_o}{\partial e_j}$ is large around the optimum the enzyme is important for the optimal flux value, J_o, and is expected to evolve if it is not yet at its optimal expression level.

The slope of the fitness landscape is defined as,

$$\delta \ln J_o = \left(C_j^J + \sum_{\substack{k=1 \\ k \neq j}} C_k^J \frac{\partial \ln e_k}{\partial \ln e_j} \right) \delta \ln e_j \tag{S50}$$

Note that $\frac{\partial \ln J_o}{\partial \ln e_j} = \frac{\delta \ln J}{\delta \ln e_j} = C_j^J + \sum_{\substack{k=1 \\ k \neq j}} C_k^J \frac{\partial \ln e_k}{\partial \ln e_j}$, which we will abbreviate with $\mathcal{F}_j^J$. When e_j is fixed to some value e_j^* the other enzymes are adjusted to reach the maximum flux level under the constraint $R - w_j e_j^*$. At those states, we can use Lagrange multipliers to determine the control coefficients, C_k^J, of the non-fixed enzymes. We then have the following Langrange function,

$$L = J(e_k) + \lambda \left(\sum_{\substack{k=1 \\ k \neq j}}^{r} w_k e_k - (R - w_j e_j) \right) \tag{S51}$$

For every k we have

$$\frac{\partial L}{\partial e_k} = 0 \tag{S52}$$

in the optimum. This leads to,

$$\frac{\partial J}{\partial e_k} = w_k \lambda \tag{S53}$$

and

$$C_k^J = \frac{w_k \lambda e_k}{J} \tag{S54}$$

In addition, the summation theorem of flux control coefficients dictates,

$$\sum_{\substack{k=1 \\ k \neq j}}^{r} C_k^J = \frac{\lambda (R - w_j e_j)}{J} = 1 - C_j^J \tag{S55}$$

Therefore

$$\lambda = \frac{J(1 - C_j^J)}{R - w_j e_j} \tag{S56}$$

And the control coefficients for the adjustable enzymes (excluding enzyme j) at their optimal levels become,

$$C_k^J = \frac{w_k e_k (1 - C_j^J)}{R - w_j e_j} = \frac{w_k e_k}{R} \frac{1 - C_j^J}{1 - \frac{w_j e_j}{R}} \tag{S57}$$

This we can substitute in equation S50,

$$\delta \ln J_o = \left(C_j^J + \frac{(1 - C_j^J)}{R - w_j e_j} \sum_{\substack{k=1 \\ k \neq j}} w_k e_k \frac{\partial \ln e_k}{\partial \ln e_j} \right) \delta \ln e_j \tag{S58}$$

This equation we can simplify further

$$\begin{aligned}
\delta \ln J_o &= \left(C_j^J + \frac{(1-C_j^J)}{R - w_j e_j} \sum_{\substack{k=1 \\ k\neq j}} w_k e_k \frac{\partial \ln e_k}{\partial \ln e_j} \right) \delta \ln e_j \\
&= C_j^J \delta \ln e_j + \frac{(1-C_j^J)}{R - w_j e_j} \sum_{\substack{k=1 \\ k\neq j}} w_k \frac{\partial e_k}{\partial e_j} \delta e_j \\
&= C_j^J \delta \ln e_j + \frac{(1-C_j^J)}{R - w_j e_j} \sum_{\substack{k=1 \\ k\neq j}} w_k \delta e_k \\
&= C_j^J \delta \ln e_j - \frac{(1-C_j^J)}{R - w_j e_j} w_j \delta e_j \\
&= C_j^J \delta \ln e_j - \frac{(1-C_j^J)}{R - w_j e_j} w_j e_j \delta \ln e_j \\
&= \left(C_j^J - (1-C_j^J) \frac{w_j e_j}{R - w_j e_j} \right) \delta \ln e_j \\
&= \left(\frac{C_j^J - \frac{e_j w_j}{R}}{1 - \frac{e_j w_j}{R}} \right) \delta \ln e_j
\end{aligned}$$

Here we have used $\sum_{\substack{k=1 \\ k\neq j}} w_k \delta e_k = -w_j \delta e_j$ because of the conservation of resource R. Thus, we find for the slope of the fitness landscape for enzyme j,

$$\mathcal{F}_j^J(e_j) = \frac{\partial \ln J_o}{\partial \ln e_j} = \frac{C_j^J - \frac{e_j w_j}{R}}{1 - \frac{e_j w_j}{R}} \tag{S59}$$

Let's study this equation a bit. It is not a classical control coefficient, because all other enzymes are allowed to change upon the perturbation in the enzyme level, e_j. $\mathcal{F}_j^J$ is zero when $e_j = e_j^o$ because then $C_j^J = \frac{e_j^o w_j}{R}$; this makes sense. If should be positive when $e_j < e_j^o$ and negative when $e_j > e_j^o$ (because we are considering a maximum). The denominator is always positive. Therefore we need to have $C_j^J > \frac{e_j w_j}{R}$ when $e_j < e_j^o$ and $C_j^J < \frac{e_j w_j}{R}$ when $e_j > e_j^o$. Again this makes sense; at low values of e_j we have to little enzyme and the flux will increase upon an increase in the enzyme level and at high concentration $e_j > e_j^o$ we have to much and the flux will decrease upon an increase in e_j as this will be at the expense of another enzyme that has become rate limiting. When $C_j^J = 1$ (and it cannot get larger, $0 < C_j^J < 1$) the slope is 1; this is the maximal slope. This means that the condition for a high slope of the fitness landscape is: *a high control coefficient*! This sounds like a trivial result but this is not the case, all the enzyme levels (except for enzyme j) are allowed to change when a change in the level of enzyme j is made of size $\delta \ln e_j$ and the chosen change in those enzyme levels is the one that maximizes the flux at $e_j + \delta e_j$. The fractional change in the optimal flux is then given by equation S59.

Comparison of equation S57 and S59 shows the relation between the control coefficient of the fixed enzyme, e_j, and all the remaining enzymes, which are allowed to attain optimal level given set a specific fixed level of enzyme j under the "residual" resource constraint, $R - w_j e_j$. Solving for C_k^J and C_j^J from these two equations gives some more insight into the relation between the control coefficient and the slope of the fitness landscape, $\mathcal{F}_j^J$,

$$C_k^J = \frac{e_k w_k}{R}(1 - \mathcal{F}_j^J) \tag{S60}$$

$$C_j^J = \mathcal{F}_j^J + \frac{e_j w_j}{R}(1 - \mathcal{F}_j^J) \tag{S61}$$

Note that these equations agree with the summation theorem of flux control as they should, i.e. $\sum_{\substack{k=1\\k\neq j}}^{r} C_k^J + C_j^J = 1$ (note that: $\sum_{\substack{k=1\\k\neq j}}^{r} \frac{e_k w_k}{R} = 1 - \frac{e_j w_j}{R}$). If $\mathcal{F}_j^J = 1$ then $C_j^J = 1$ and $C_k^J = 0$. The relative control amongst the variable enzymes is given by their relative costs (their $e_k w_k$ ratio's). Note also that in the optimum, $\mathcal{F}_j^J = 0$, and therefore the control coefficients are defined as derived in the main text in the cost and benefit analysis, i.e. as $C_l^J = \frac{w_l e_l}{R}$.

S8.4 Illustration of the slope of a fitness landscape

To illustrate the equations for the slope of a fitness landscape, we will apply them in this section to a toy-model. The model consists of 4 enzyme-catalyzed reactions (see Figure S5, for model details see caption). We will create the fitness landscape for the second enzyme, e_2. That means that we will make e_2 a parameter and optimize the other three enzymes of the model with the remaining resources. All $\omega' s$ are set to 1, resulting in the e_{total} of 1. The fitness landscape we obtain is shown in Figure S5, and reveals that the $e_2^{opt} = 0.19$. Next we calculated the slope for two different values of e_2: 0.05 and 0.6 (indicated by the black dots) using equation. S59. The lines corresponding to the slope at $e_2 = 0.05$ is shown by the orange line and the red line corresponds to the slope of the fitness landscape where $e_2 = 0.6$.

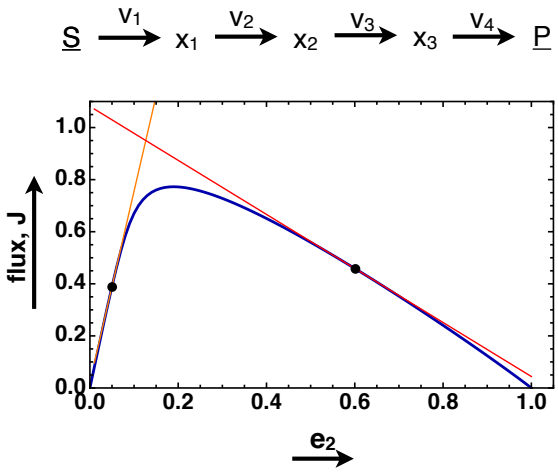

Figure S5. Slope of the fitness landscape. For a four enzyme metabolic system, we calculate the fitness landscape for the second enzyme. The model consists of 4 irreversible reactions, with a general rate equation as given by: $v_i = kcat_i \cdot e_i \cdot s/(1 + s/K_m^s + p/K_m^p)$, with parameter values: $kcat_1 = 3; K_{m,1}^S = 0.1; K_{m,1}^{x1} = 1; kcat_2 = 8; K_{m,2}^{x1} = 0.5; K_{m,2}^{x3} = 3; kcat_3 = 5; K_{m,3}^{x2} = 1; K_{m,3}^{x3} = 0.5; kcat_4 = 6; K_{m,4}^{x3} = 2; K_{m,4}^{p} = 0.75; S = 1; P = 0.1; \omega_{1,2,3,4} = 1; R = 1$. Shown is the fitness landscape for e_2 by the blue line. In orange and red, are two tangents shown, which have a slope that corresponds to the fitness landscape at $e_2 = 0.05$ and 0.6, respectively.

References

[1] C Giersch. Control analysis of metabolic networks. 1. Homogeneous functions and the summation theorems for control coefficients. *Eur. J. Biochem.*, 174(3):509–513, 1988.

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

[3] R Heinrich and S Schuster. The modelling of metabolic systems. Structure, control and optimality. *Bio Systems*, 47(1-2):61–77, 1998.

[4] Reinhart Heinrich and Tom A Rapoport. A Linear Steady-State Treatment of Enzymatic Chains. *Control*, 95:89–95, 1974.

[5] H Kacser, J A Burns, and Burns JA. Kacser H. The control of flux. *Symp Soc Exp Biol*, 27:65–104, 1973.

[6] E Klipp and R Heinrich. Evolutionary optimization of enzyme kinetic parameters; effect of constraints. *J Theor Biol*, 171(3):309–323, 1994.

[7] E Klipp and R Heinrich. Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *Bio Systems*, 54(1-2):1–14, 1999.

[8] H V Westerhoff and K van Dam. *Thermodynamics and control of biological free-energy transduction*. Elsevier, Amsterdam., 1987.