Evolution of enzyme levels in metabolic pathways: A theoretical approach

Charlotte Coton[†], Grégoire Talbot[†], Maud Le Louarn^{*},
Christine Dillmann[†], Dominique de Vienne[†]

May 4, 2021

[†] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France.

Abstract

The central role of metabolism in cell functioning and adaptation has given rise to countless studies on the evolution of enzyme-coding genes and network topology. However, very few studies have addressed the question of how enzyme concentrations change in response to positive selective pressure on the flux, considered a proxy of fitness. In particular, the way cellular constraints, such as resource limitations and co-regulation, affect the adaptive landscape of a pathway under selection has never been analyzed theoretically. To fill this gap, we developed a model of the evolution of enzyme concentrations that combines metabolic control theory and an adaptive dynamics approach, and integrates possible dependencies between enzyme concentrations. We determined the evolutionary equilibria of enzyme concentrations and their range of neutral variation, and showed that they differ with the properties of the enzymes, the constraints applied to the system and the initial enzyme concentrations. Simulations of long-term evolution confirmed all analytical and numerical predictions, even though we relaxed the simplifying assumptions used in the analytical treatment.

Keywords

Adaptive landscape. Metabolism. Enzyme concentration. Co-regulation. Evolutionary equilibrium. Selective neutrality.

Running title

Evolution of enzyme concentrations

^{*}This article is dedicated to the memory of Maud Le Louarn, who contributed so inventively and enthusiastically to this work

1 Introduction

Metabolism is a property of living systems that is central to cell functioning and physiological and genetic adaptation (Cornish-Bowden and Cárdenas 2020). The metabolic networks we see today, comprising an extraordinary diversity of enzymes, cofactors and metabolites, are the product of billion years of evolution that have shaped network topology and enzyme properties.

The importance of metabolism has inspired a wealth of studies on its evolutionary dynamics over different timescales, which follow three main lines of research.

The first line of research relates to the evolution of network structure. Over evolutionary time, continuous selective pressures in diverse environments have resulted in the successive recruitment of new metabolites and enzymes that then coevolve (Noda-Garcia et al. 2018), have modified substrate specificity (Newton et al. 2018) and have determined the general features of the networks, namely functional redundancy, modularity, flux coupling and exaptation (Sambamoorthy et al. 2019). Adaptation to a new environment may involve gain/loss-of-functions and horizontal gene transfers, which rewire the metabolic networks (e.g. Pál et al. 2005; Lee and Palsson 2010; D’Souza and Kost 2016; Morrison and Badyaev 2017).

The second line of research investigates the factors that determine the degree of polymorphism and the rate of evolution of enzyme-coding genes. In addition to classical factors, such as gene dispensability, level of gene expression, protein length, protein folding stability, intron number and protein-protein interaction (see e.g. review by Larracuent et al. 2008), various systemic constraints also influence the degree of protein polymorphism and evolution (Pál et al. 2006). Enzymes at branching points of the network or that are highly connected, and thus have pleiotropic effects, are assumed to be more subject to purifying selection; as a consequence, their coding genes have low levels of polymorphism or slow evolutionary rates (Rausher et al. 1999; Vitkup et al. 2006; Greenberg et al. 2008; Rausher 2013). Flux levels and flux control coefficients are additional factors that influence protein evolution: the genes encoding high-flux enzymes have lower evolutionary rates (Colombo et al. 2014), possibly related to enzyme abundance (Aguilar-Rodriguez and Wagner 2018); higher flux control coefficients of enzymes that are upstream in the pathway result in stronger purifying or positive selection (Flowers et al. 2007; Wright and Rausher 2010; Sellis and Longo 2015).

The third line of research is the study of the genetic variability of enzyme properties. The response of organismal fitness to selection pressures depends on genetically variable enzyme properties, namely kinetic parameters, thermodynamic stability and concentration. In the context of the historical debate between neutralism and selectionism, a number of emblematic examples of variation in enzyme properties have been studied in detail. For instance, in *Drosophila*, the fast and slow allozymes of alcohol dehydrogenase differ both in the value of their catalytic constant (the turnover number, k_{cat}) and their abundance; also in *Drosophila* the A and B variants of G6PD have different quaternary structures, thermal stability and K_m values; in *Fundulus heteroclitus*, the allozymes of LDH-B display kinetic, abundance and stability differences, etc. (reviewed in Eanes 1999). Furthermore, quantitative proteomic studies in various species have revealed the breadth of the genetic variability, the polygenic nature and the relationship with integrated phenotypic traits of enzyme concentrations (Damerval et al. 1994; Blein-Nicolas et al. 2013; Albert and Kruglyak 2015; Chick et al. 2016; Jiang et al. 2019).

In this context, the evolution of enzyme properties in response to selective pressures on the flux, considered to be a proxy for fitness, is an important issue. This question

has been addressed in relation to protein folding stability, catalytic constant values and enzyme concentration, but to our knowledge the number of studies is quite low. Mutations that increase protein folding stability are favored by selection, with evolution towards near neutrality up to the mutation–selection balance steady state (Serohijos and Shakhnovich 2014; Bershtein et al. 2017). Regarding enzyme activity, Kacser and Beeby (1984), using a systems biology approach, proposed a model of early enzyme evolution in which selection for faster growth causes a transition from few, multifunctional, poorly active enzymes to many, differentiated, monofunctional enzymes with high k_{cat} values. However, Heckmann et al. (2018) recently showed that the systemic properties of a network result in epistatic interactions that maintain k_{cat} values far from their theoretical maximum, preventing fitness from reaching its optimum. This result can be explained by the relationship of diminishing returns between turnover number and flux: the effect of a given increase in k_{cat} on the flux vanishes as k_{cat} increases, so that the flux eventually reaches a plateau.

Interestingly, the same type of saturation curve was invoked by Hartl et al. (1985) to account for the widespread selective neutrality of enzyme polymorphisms. These authors analyzed the evolution of enzyme quantity, equated to activity, under selection for higher flux, and showed that successive mutations affecting a particular enzyme drive the enzyme concentration towards the plateau of the concave flux-enzyme curve, where the mutations do not have any detectable effect, *i.e.* they are selectively nearly neutral. This geometrical model of the "natural selection of selective neutrality" is attractive, but surprisingly it has not been assessed with an adaptive dynamics approach. In addition, this model considered only a single enzyme in a given pathway and imposed no upper limits on enzyme quantities, even though there are well-known constraints on space and energy in the cell (Koehn 1991; Kurland and Dong 1996; de Vienne et al. 2001b; Ellis 2001; Eguchi et al. 2018; Klumpp et al. 2019). Although these points were raised, the evolutionary implications were not analyzed by Hartl et al. (1985). The question of the optimization of enzyme kinetic parameters and concentrations in metabolic pathways was analyzed by Klipp and Heinrich (1994, 1999). However, their approach did not include any evolutionary genetics concept and did not allow them to grasp the dynamics of the evolutionary process.

Here, we develop the first adaptive dynamics model of the evolution of enzyme concentrations when there is directional selection for increased flux in a pathway. Our model includes two biologically realistic features: (i) the concentration of each enzyme in the pathway can be altered by mutation, and (ii) constraints, specifically competition for cellular resources and/or co-regulation, can prevent enzyme concentrations from varying freely. To formalize the relationship between enzyme concentrations and flux, we used the metabolic control theory (MCT) framework (Kacser and Burns 1973; Heinrich and Rapoport 1974; Kacser et al. 1995), as generalized by Heinrich et al. (1991) and Lion et al. (2004) to take into account the possible interdependence of enzyme concentrations.

First, we present a system of ordinary differential equations that describes the evolution of enzyme concentrations in a pathway. We then show how the flux evolves and how it determines the evolutionary equilibria of the relative enzyme concentrations. We demonstrate analytically, and validate using computer simulations, that a theoretical equilibrium of relative enzyme concentrations exists whatever the conditions, but that this equilibrium is not necessarily reached when the flux is limited by constraints on enzyme concentration variation. Thus, we differentiate the *effective* equilibrium, *i.e.* the equilibrium that is actually reached, from the *theoretical* equilibrium. When there is co-regulation in addition to competition, the initial enzyme concentrations affect the evolutionary outcome. Finally, we show that multi-enzyme selective neutrality, *i.e.* the selective neutrality of all

enzymes, is achieved whatever the constraints, but that the size of the range of neutral variation largely depends on constraint-specific parameters and possibly on the initial state of the system.

2 Material and Methods

2.1 Theoretical developments

2.1.1 The metabolic model

2.1.1.1 Flux expression We considered a metabolic pathway of n unsaturated Michaelian enzymes transforming substrate X_0 into product X_n via a series of unimolecular reversible reactions:

At steady state, the metabolic flux J through this pathway is (Kacser and Burns 1981):

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}} \quad (1)$$

where:

- $X = X_0 - X_n/K_{1,n}$. $K_{1,n}$ is the product of the successive equilibrium constants of all reactions. X is assumed to be constant and symbolizes that the environment is constant.
- $A_j = \frac{k_{catj}}{K_{Mj}} K_{1,j-1}$, where k_{catj} is the turnover number of enzyme j , K_{Mj} is the Michaelis-Menten constant of the enzyme j for its substrate and $K_{1,j-1}$ is the product of the successive equilibrium constants up to the $(j-1)^{th}$ reaction: $\prod_{k=1}^{j-1} K_k$. Thus A_j is a composite parameter that includes both catalytic efficiency $\frac{k_{catj}}{K_{Mj}}$ and a product of thermodynamic constants that depends on the enzyme's position in the pathway. Hereafter, A_j is referred to as the "pseudo-activity" of enzyme j .
- E_j is the concentration of enzyme j . Note that E_j in italics denotes the concentration of enzyme j while E_j in Roman letter denotes enzyme j (see Glossary of mathematical symbols in Appendix A).

Considering the total enzyme concentration, $E_{tot} = \sum_{j=1}^n E_j$, and the relative concentration of enzyme j , $e_j = \frac{E_j}{E_{tot}}$, equation 1 becomes:

$$J = E_{tot} \frac{X}{\sum_{j=1}^n \frac{1}{A_j e_j}} \quad (2)$$

The flux can be modified by mutations affecting $\frac{k_{catj}}{K_{Mj}}$ or E_j . We considered only the latter, because the variability of enzyme concentrations is by far the main source of variation in enzyme activity in populations and species (see Discussion).

2.1.1.2 The effects of mutation when enzyme concentrations are interdependent

Let a mutation of small effect ν targeting concentration E_i of enzyme i (ν is positive or negative). We assume that the probability law of mutation effect ν is identical for all enzymes, so ν is not indexed. The concentration of enzyme i after a mutation is:

$$E_i^{(1)} = E_i + \nu \quad (3)$$

Because of constraints such as resource limitations or complex interactions between cell components, this change may not correspond to the actual effect of the mutation on the concentration of enzyme i and/or it may modify the concentration of other enzymes. Considering two causes of interdependence of enzyme concentrations, separately or jointly, we formalized the consequences of a mutation in the following way:

1. *Co-regulations.* Various direct and indirect molecular mechanisms result in co-regulation (either positive or negative) of enzyme concentrations. Thus, mutations can jointly modify the concentrations of co-regulated enzymes (*see Discussion*). To express formally the effect of a mutation targeting the concentration of enzyme i (equation 3) on the concentration of enzyme j , we write:

$$E_j^{(1)} = E_j + \beta_{ij}\nu \quad (4)$$

where β_{ij} , the slope of the relation between E_i and E_j , is the *co-regulation coefficient*. We assume linearity between co-regulated enzymes, so we use constant co-regulation coefficients and we pose $\beta_{ii} = 1$, $\beta_{ij} = 1/\beta_{ji}$ and, for any triplet of co-regulated enzymes (i, j, k) , $\beta_{ij} = \beta_{ik}\beta_{kj}$. If enzymes i and j are not co-regulated, we have $\beta_{ij} = \beta_{ji} = 0$. All pairwise co-regulation coefficients can be grouped in a co-regulation matrix, noted $\mathbf{M}_\beta$ (Supporting Information [SI.B.1.1](#)).

For n co-regulated enzymes, the total enzyme concentration after a mutation affecting enzyme i is:

$$\begin{aligned} E_{\text{tot}}^{(1)} &= \sum_{j=1}^n E_j^{(1)} = \sum_{j=1}^n (E_j + \beta_{ij}\nu) \\ &= E_{\text{tot}} + B_i\nu \end{aligned} \quad (5)$$

where $B_i = \sum_{j=1}^n \beta_{ij}$ quantifies the impact on E_{tot} of a mutation targeting E_i . Indeed, from equations 3 and 5, we have:

$$\frac{\Delta E_{\text{tot}}}{\Delta E_i} = \frac{E_{\text{tot}}^{(1)} - E_{\text{tot}}}{E_i^{(1)} - E_i} = B_i \quad (6)$$

B_i is called the *global co-regulation coefficient* (Supporting information [SI.B.1.2](#)).

2. *Competition for resources.* Cellular resources are limited in terms of space, energy, the availability of ribosomes, etc. The total enzyme concentration a cell allocates to a given pathway is necessarily bounded, so that an increase (resp. decrease) in the concentration of an enzyme due to a mutation can result in a decrease (resp. increase) in the concentration of other enzymes ([Snoep et al. 1995](#); [Albertin et al. 2013](#)). We modeled this trade-off by fixing E_{tot} . As a consequence, if a mutation of effect ν targets enzyme i , the concentrations of all the other enzymes will be

modified, as will be the effect of the mutation on enzyme i . This can be viewed as a two-step process:

$$E_i \xrightarrow[\text{mutation of } E_i]{} E_i^{(1)} = E_i + \nu \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_i^{(2)} = E_i + \delta_i$$

where δ_i is the mutation effect after applying the constraint on total enzyme concentration, which changes from $E_{\text{tot}}^{(1)} = E_{\text{tot}} + \nu$ to E_{tot} . This constraint is assumed to respect the proportionality between concentrations, *i.e.*

$$\frac{E_i^{(1)}}{E_{\text{tot}}^{(1)}} = \frac{E_i^{(2)}}{E_{\text{tot}}} \quad (7)$$

From this relation, the expression of δ_i can be deduced (Supporting Information SI.B.1.3):

$$\delta_i = \frac{1 - e_i}{\frac{1}{\nu} + \frac{1}{E_{\text{tot}}}} \quad (8)$$

where $e_i = \frac{E_i}{E_{\text{tot}}}$. We call ν the *canonical effect* of the mutation and δ_i its *actual effect*.

For enzymes $j \neq i$, we have

$$E_j \xrightarrow[\text{mutation of } E_i]{} E_j^{(1)} = E_j \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_j^{(2)}$$

where $E_j^{(2)}$ is the concentration of enzyme j after applying the constraint on total enzyme concentration. To find the expression of $E_j^{(2)}$, we applied the proportionality rule:

$$\frac{E_j^{(1)}}{E_{\text{tot}}^{(1)}} = \frac{E_j^{(2)}}{E_{\text{tot}}} \quad (9)$$

and using equation 8 we get (Supporting Information SI.B.1.3):

$$E_j^{(2)} = E_j - \frac{e_j}{1 - e_i} \delta_i \quad (10)$$

3. *Co-regulation and competition.* If there is both competition and co-regulation between all enzymes, the two-step process is the same as above for enzyme i targeted by a mutation:

$$E_i \xrightarrow[\text{mutation of } E_i]{} E_i^{(1)} = E_i + \nu \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_i^{(2)} = E_i + \delta_i$$

but here $E_{\text{tot}}^{(1)} = E_{\text{tot}} + B_i \nu$ (equation 5). So applying the proportionality argument (equation 7 and Supporting Information SI.B.1.4), we get:

$$\delta_i = \frac{1 - e_i B_i}{\frac{1}{\nu} + \frac{B_i}{E_{\text{tot}}}} \quad (11)$$

For enzymes $j \neq i$, we have

$$E_j \xrightarrow[\text{mutation of } E_i]{} E_j^{(1)} = E_j + \beta_{ij}\nu \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_j^{(2)}$$

where $E_j^{(2)}$ is the concentration of enzyme j after co-regulation *and* reduction of the total enzyme concentration from $E_{\text{tot}}^{(1)}$ to E_{tot} . Using equations 9 and 11, we get (Supporting Information SI.B.1.4):

$$E_j^{(2)} = E_j + \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i} \delta_i \quad (12)$$

In summary, non-independence of enzyme concentrations results in changes in the concentration of enzymes that are not the primary targets of the mutation. The amplitude of concentration change of non-target enzymes depends on the cause of interdependence:

Co-regulation (equation 4)	$E_j^{(1)} = E_j + \beta_{ij}\delta_i$ where $\delta_i = \nu$
Competition (equation 10)	$E_j^{(2)} = E_j - \frac{e_j}{1 - e_i} \delta_i$
Co-regulation and competition (equation 12)	$E_j^{(2)} = E_j + \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i} \delta_i$

Thus the concentration of the mutant enzyme E_j^m can be expressed as the sum of E_j and a *pseudo-mutation* effect:

$$\forall j, \quad E_j^m = E_j + \alpha_{ij}\delta_i$$

This *pseudo-mutation* effect is the product of the actual mutation effect δ_i by a coefficient called the *redistribution coefficient*, noted α_{ij} (note that $\forall i, \alpha_{ii} = 1$).

The general expression of the redistribution coefficient is $\alpha_{ij} = \frac{\Delta E_j}{\Delta E_i}$, where ΔE_j is the change of non-target enzyme j and $\Delta E_i = \delta_i$ is the actual effect of the mutation on target enzyme i . Going to the limit when δ_i is very small, we obtain:

$$\alpha_{ij} = \frac{\partial E_j}{\partial E_i} \quad (13)$$

which is the redistribution coefficient α_{ij} defined by Lion et al. (2004).

Depending on the situations, we have:

1. Independence: $\forall i, j \neq i, \alpha_{ij} = 0$
2. Co-regulation: $\forall i, j \neq i, \alpha_{ij} = \beta_{ij} \neq 0$
3. Competition: $\forall i, j \neq i, \alpha_{ij} = \frac{-e_j}{1 - e_i}$
4. Co-regulation and competition: $\forall i, j \neq i, \alpha_{ij} = \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i}$

2.1.1.3 Quantifying the effect of enzyme concentration variation on the flux

To quantify the effect on the flux of variation in enzyme concentration E_i , we used the *combined response coefficient* (hereafter, the flux response coefficient) (de Vienne et al. 2001a)

$$R_{E_i}^J = \frac{\partial J}{\partial E_i} \frac{E_i}{J} \quad (14)$$

The flux response coefficient is more general than the classical flux control coefficient (Kacser et al. 1995), the latter being valid only if enzymes are independent. The flux response coefficient of enzyme i includes both the flux control coefficient $C_{E_i}^J$ of enzyme i and a term that accounts for enzyme interdependence, which contains the flux control coefficients of the other enzymes and the redistribution coefficients α_{ij} (Lion et al. 2004):

$$R_{E_i}^J = C_{E_i}^J + E_i \sum_{j=1}^n C_{E_j}^J \frac{\alpha_{ij}}{E_j}$$

In Supporting Information SI.B.2, we present another expression of $R_{E_i}^J$ that is less intuitive but more convenient for mathematical treatment.

2.1.2 The evolution model

2.1.2.1 Adaptive dynamics To analyze the evolution of enzyme concentrations, we used an adaptive dynamics approach (Brännström et al. 2013), assuming a constant environment.

We assumed that evolution proceeds by the successive fixation of mutations modifying enzyme concentration. We applied the mutual exclusion principle, *i.e.* two phenotypes cannot coexist because the mutations are assumed to be instantaneously fixed or eliminated (Dieckmann and Law 1996). On an evolutionary timescale, if a mutation produces a novel phenotype, either it is fixed, becoming the new resident, or it is eliminated. The vector of enzyme concentrations of the resident is noted $\mathbf{E}^r = (E_1^r, \dots, E_n^r)$ and that of the mutant $\mathbf{E}^m = (E_1^m, \dots, E_n^m)$.

Mutations appear at rate μ per time unit. They are independent and each mutation targets only one enzyme. The actual effect of a mutation on the target enzyme i is δ_i . Target enzyme concentration after a mutation is:

$$E_i^m = E_i^r + \delta_i$$

The concentration of each of the other enzymes is:

$$E_j^m = E_j^r + \alpha_{ij}\delta_i$$

The selection coefficient of a mutation targeting enzyme i is:

$$s_i = \frac{w_i^m - w^r}{w^r}$$

where w is fitness. We assumed that fitness w is proportional to flux J , so we have:

$$s_i = \frac{J_i^m - J^r}{J^r} \quad (15)$$

The phenotype under selection, namely the flux, is calculated following equation 1.

Since we considered a haploid population of constant effective size N with non-overlapping generations, the fixation probability of a mutation is (Crow and Kimura 1970):

$$P_f \approx \frac{2s_i}{1 - e^{-2Ns_i}} \quad (16)$$

when s_i is small. This expression was used in our computer simulations.

In our analytical developments, we neglected deleterious mutations because their negative selection coefficients result in negligible fixation probability. Since we assumed large population sizes (for all i , $Ns_i \gg 1$), we could use the approximation:

$$P_f \approx 2s_i$$

The higher the mutant flux J^m is relative to the resident flux J^r , the higher the fixation probability of the mutation. After fixation or loss, a new mutation appears, and the process is repeated.

2.1.2.2 The differential equation system of enzyme evolution On a continuous timescale, the variation in concentration of enzyme i can be represented as a stochastic process. At each time t , this variation is a random variable that can be described with a complete event system:

$$\left(\frac{\partial E_i}{\partial t}\right)^{(t)} = \begin{cases} \mu 2s_i^{(t)} & \delta_i^{(t)} & \text{Fixation of mutation targeting } i \\ \mu 2s_j^{(t)} & \alpha_{ji}^{(t)} \delta_j^{(t)} \quad \forall j \neq i & \text{Fixation of mutation targeting } j \\ 1 - \sum_{k=1}^n \mu 2s_k^{(t)} & 0 & \text{No mutation or no fixation} \end{cases} \quad (17)$$

where μ is the probability of occurrence of a mutation, assumed to be independent of the target enzyme. Because this relation is valid at each time t , we will not use the superscript t thereafter. At each time, variation in E_i can be caused by the fixation of either a mutation affecting enzyme i (probability $2s_i$) or a mutation affecting any other enzyme (probability $2s_j$), with mutation effect δ_i and $\alpha_{ji}\delta_j$, respectively. If none of these events happen (probability $1 - \sum_{k=1}^n \mu 2s_k$), E_i does not change. Therefore, the average variation in concentration of enzyme i at each time t is:

$$\left(\frac{\partial E_i}{\partial t}\right) = \mu 2s_i \delta_i + \sum_{\substack{j=1 \\ j \neq i}}^n \mu 2s_j \alpha_{ji} \delta_j \quad (18)$$

This relation is valid at all times t , and because $\alpha_{ii} = 1$, we can write:

$$\frac{\partial E_i}{\partial t} = 2\mu \sum_{j=1}^n s_j \alpha_{ji} \delta_j \quad (19)$$

This differential system allowed us to search for steady states – called evolutionary equilibria – of enzyme concentrations when there is selection for increased flux. When enzymes are independent, $\partial E_i / \partial t$ is strictly positive, and thus enzyme concentrations increase indefinitely and do not reach equilibrium. Consequently, we searched for evolutionary equilibria of relative enzyme concentrations, which is given by:

$$\forall i, \quad \frac{\partial e_i}{\partial t} = 0 \quad (20)$$

2.1.2.3 Selective neutrality of enzyme concentration variation We addressed the question of the neutrality of the variations in enzyme concentrations near evolutionary equilibrium. The "neutral zone" is the area where the absolute value of the selection coefficient is close to zero – classically under a threshold of $1/N$ for haploid populations (Kimura 1985).

To analytically study the way constraint-dependent factors influence the neutral zone at evolutionary equilibrium, we used an expression of s_i that explicitly contains the three relevant variables, *i.e.* $R_{E_i}^J$, δ_i and E_i . Assuming that the mutation effect δ_i is small compared to the concentration E_i ($\delta_i \ll E_i$), we can write $\frac{\Delta J}{\Delta E_i} \approx \frac{\partial J}{\partial E_i}$. Therefore, from equation 14, the selection coefficient is given by (Supporting Information SI.B.3.1):

$$s_i \approx R_{E_i}^J \frac{\delta_i}{E_i} \quad (21)$$

The size of the neutral zone determines what we call the *range of neutral variation* (RNV) of the concentration, the amplitude of which is:

$$RNV = |\delta_i^{\sup} - \delta_i^{\inf}|,$$

where $\delta_i^{\sup}$ (resp. $\delta_i^{\inf}$) is the actual effect of a mutation such that $s_i = +\frac{1}{N}$ (resp. $s_i = -\frac{1}{N}$), s_i being the superior (resp. inferior) limit of the neutral zone. To determine the δ_i 's, we used the initial equation $s_i = \frac{J_i^m - J^r}{J^r}$ (equation 15), where J^r is the resident phenotype and J_i^m the mutant phenotype following a mutation targeting enzyme i . From equation 1, and after transformations, we obtain (Supporting Information SI.B.3.2):

$$\sum_{j=1}^n \frac{1}{A_j(E_j + \alpha_{ij}\delta_i)} - \frac{1}{s_i + 1} \sum_{j=1}^n \frac{1}{A_j E_j} = 0 \quad (22)$$

The solutions of this equation for $s_i = +1/N$ and $s_i = -1/N$ give $\delta_i^{\sup}$ and $\delta_i^{\inf}$, respectively. Due to the complexity of the resulting equations, we wrote a script that numerically computes the limits $\delta_i^{\sup}$ and $\delta_i^{\inf}$ of the RNVs for different scenarios of enzyme interdependence (function `uniroot` of the R Stats package [(R Development Core Team 2008)]).

2.2 Computer simulations

To assess the validity of the mathematical predictions, we performed Markov Chain Monte-Carlo simulations using the R software (R Development Core Team 2008). We considered haploid populations of size $N = 1\,000$ evolving in a constant environment ($X = 1$). At a given time step, all individuals in the population are identical, *i.e.* have the same phenotype J and the same enzyme concentrations (the "genotype") $\mathbf{E} = (E_1, E_2, \dots, E_n)$ and the same pseudo-activities $\mathbf{A} = (A_1, A_2, \dots, A_n)$. Simulations were performed for a three-enzyme metabolic pathway ($n = 3$), with invariable pseudo-activities $A_1 = 1$, $A_2 = 10$ and $A_3 = 30$ and an initial total enzyme concentration $E_{\text{tot}}^0 = 100$ in arbitrary units (see Figure 1A).

The vector of initial concentrations $\mathbf{E}^0$ was changed for each simulation. Initial concentrations E_i^0 were drawn in $\mathcal{U}(0, 100)$, and redistributed proportionally to have $E_{\text{tot}}^0 = 100$. E_{tot} varied over time in the case of independence and co-regulation, but was fixed at $E_{\text{tot}} = 100$ in the case of competition.

Co-regulation coefficient values were $\beta_{12} = 0.1$, $\beta_{23} = 2$ and hence $\beta_{13} = 0.2$ for the positive co-regulations and $\beta_{12} = 0.32$, $\beta_{23} = -0.43$ and hence $\beta_{13} \approx -0.138$ for the negative co-regulations. The resulting co-regulation matrices $\mathbf{M}_\beta$ were computed (Supporting Information SI.B.1.1).

At each time step, a mutation randomly affected one of the enzymes. The mutation effect ν was drawn in $\mathcal{U}(0, 1)$; the sign of ν was randomly drawn (1:1 ratio). Thus the concentration of target enzyme i changed from E_i^r to $E_i^m = E_i^r \pm \nu$; the concentration of all enzymes varied according to the type of dependence considered (see the two-steps process described in paragraph 2.1.1.2). If a concentration became negative, it was set to zero. Finally, we computed the values of the flux, the selection coefficient and the fixation probability of the mutant (equations 1, 15 and 16, respectively).

To determine whether a mutation was fixed or lost, we drew a number in $\mathcal{U}(0, 1)$, and if this number was smaller than P_f the mutation was fixed, otherwise it was lost. If the mutation was fixed, the mutant became the new resident, *i.e.* all individuals displayed the same genotype and phenotype. If the mutation was lost, genotype and phenotype were unchanged. We then proceeded to the next time step. Thus, a time step is the interval between two successive mutations. Populations evolved over 125,000 time steps, and the system values were recorded every 250 steps.

These simulations allowed us to relax some of the assumptions of the mathematical analysis, namely small mutation effects, large population size and the absence of fixation of deleterious mutations.

Data and code availability Scripts of the simulations and custom functions are written in the R language (R Development Core Team 2008). Along with the simulation data, they are included in a home-made package called `SimEvolEnzCons`.

3 Results

We have developed a theoretical framework that describes the evolution of enzyme concentrations in a pathway under selection for increased metabolic flux. We considered that: (i) the flux can be modified by mutations targeting any enzyme in the pathway; (ii) variation in enzyme concentrations can be not independent as a result of competition for cellular resources and/or co-regulation. We based our developments on the principles of adaptive dynamics (Brännström et al. 2013), using the enzyme-flux relationship of the Metabolic Control Theory (MCT) (Kacser and Burns 1973; Heinrich and Rapoport 1974; Kacser et al. 1995) expanded by Heinrich et al. (1991) and Lion et al. (2004) to take into account possible dependencies between enzyme concentrations.

Flux J through a linear metabolic pathway of n Michaelian enzymes was expressed as a function of the parameters of all enzymes (equation 1 in Material & Methods):

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}}$$

where A_j is the *pseudo-activity* of enzyme j , assumed to be constant, and E_j is the concentration of enzyme j that can be affected by mutation (see Glossary of mathematical symbols in appendix A).

If there is selection for increased flux, the evolutionary rate of enzyme concentrations can be expressed by a differential equation system (equation 19), allowing us to: (i) search

for steady states of relative enzyme concentrations e_i ($\forall i, \partial e_i / \partial t = 0$), which we call *evolutionary equilibria*; (ii) analyze flux variation during the evolutionary process.

To assess the validity of our analytical results, we performed simulations of long-term evolution using less stringent assumptions than in the analytical approach.

Finally, we analyzed the extent to which selection for increased metabolic flux results in the selective neutrality of mutations targeting enzyme concentrations.

3.1 Equilibria of relative enzyme concentrations under directional selection for increased flux

We searched for evolutionary equilibria of relative enzyme concentrations in four situations: independence between enzyme concentrations, fixed total enzyme concentration (due to competition), co-regulation between all enzyme concentrations and both competition and co-regulation.

3.1.1 Case 1: Independence between enzymes

When the enzyme concentrations are independent, the evolutionary equilibrium of relative enzyme concentrations is (Supporting Information SI.B.4):

$$\forall i, \quad e_i^* = \frac{A_i^{-1/3}}{\sum_{j=1}^n A_j^{-1/3}} \quad (23)$$

Simulation results of a three-enzyme pathway (Figure 1A) are consistent with this theoretical equilibrium: the evolutionary trajectories converge towards the point of the space of relative concentrations defined by equation 23 (Figure 1B). At this point, the flux response coefficients, which quantify the sensitivity of the flux to variation in enzyme concentrations, are (Supporting Information SI.B.4):

$$\forall i, \quad R_{E_i}^{J*} = \frac{A_i^{-2/3}}{\sum_{j=1}^n A_j^{-2/3}} \quad (24)$$

Because the flux response coefficients are strictly positive and there is no upper limit on enzyme concentration, the flux is continuously increasing (Figure 1C) while the proportions of the different enzymes remain constant at equilibrium (equation 23).

From equations 23 and 24, we see that the relative concentrations and the flux response coefficients at equilibrium are inversely related to the pseudo-activities: enzymes with low pseudo-activity have a high concentration relative to others enzymes and exert a high control on the flux. For these enzymes, variation in abundance is less likely to be selectively neutral (see Section 3.2.2).

3.1.2 Case 2: Competition for cellular resources

In order to model the effects of limited cellular resources, the total enzyme abundance E_{tot} in the pathway was considered to be constant. When a mutation targets the concentration of an enzyme, the concentrations of all enzymes are assumed to be redistributed proportionally to maintain constant the total concentration. As a consequence, the relationship between relative enzyme concentrations and flux is a dome in the multidimensional space

of relative enzyme concentrations, the shape of which is determined by the vector of pseudo-activities $\mathbf{A}$ (Figure 2A). The flux cannot exceed a maximum value $J_{\max}$, which is the flux value at evolutionary equilibrium.

At $J_{\max}$ the flux response coefficients are:

$$\forall i, \quad R_{E_i}^J = 0$$

375 because at the top of the dome, the derivative of the flux with respect to E_i is null in all dimensions. From this equality, the evolutionary equilibrium of relative enzyme concentrations at $J_{\max}$ can be found (Supporting Information SI.B.5):

$$\forall i, \quad e_i^* = \frac{A_i^{-1/2}}{\sum_{j=1}^n A_j^{-1/2}} \quad (25)$$

As in the case of independence, simulation results were consistent with this theoretical equilibrium (Figure 2B). As previously, the higher the pseudo-activity of the enzyme, the 380 lower its relative concentration, but due to the difference in the value of the power of A_i ($-1/2$ instead of $-1/3$), the relative enzyme concentrations at evolutionary equilibrium were more dispersed when there was competition than when there was independence.

From equation 25, we can calculate $J_{\max}$, which is linearly related to E_{tot} (Supporting Information SI.B.5):

$$J_{\max} = E_{\text{tot}} \frac{X}{\left(\sum_{j=1}^n A_j^{-1/2} \right)^2} \quad (26)$$

385 Figure 2C shows the results of the simulations of flux evolution when E_{tot} is fixed. The expected $J_{\max}$ was reached whatever the value of the initial enzyme concentrations.

3.1.3 Case 3: Co-regulation between enzyme concentrations

When there is co-regulation between enzyme concentrations, in the absence of competition, a mutation targeting a given enzyme results in variation in the concentration of 390 the other enzymes. To model this phenomenon, we used the *co-regulation coefficient* β_{ij} – the slope of the linear relation between E_i and E_j –, which quantifies the variation in E_j when a mutation affects E_i .

When all enzyme concentrations are related to each other ($\forall(i, j), \beta_{ij} \neq 0$), the evolutionary equilibrium of the relative enzyme concentrations is (Supporting Information 395 SI.B.6.1.1):

$$\forall i, \quad e_i^* = \frac{1}{B_i} \quad (27)$$

where $B_i = \sum_{j=1}^n \beta_{ij}$, the *global regulation coefficient* of enzyme i , quantifies the effect of the variation of E_i on the total enzyme concentration (equation 6): $B_i = \Delta E_{\text{tot}} / \Delta E_i$. This equilibrium has a non-intuitive property: at this point, mutations, whatever their effect, no longer modify the *relative* enzyme concentrations (Supporting Information SI.B.6.1.2). 400 Metaphorically, we will call this point an *evolutionary black hole*.

The enzyme concentrations are assumed to be linearly related. Consequently, in the multidimensional space of relative enzyme concentrations, the system can be described as a point moving on a straight line $\mathcal{E}$. The parametric equation of $\mathcal{E}$ can be determined

from two known points, the initial relative enzyme concentrations $\mathbf{e}^0$ and the theoretical equilibrium $\mathbf{e}^*$ (Supporting Information SI.B.6.2):

$$\mathcal{E}: \quad \forall i, \quad e_i = \tau(e_i^* - e_i^0) + e_i^0 = \tau \left(\frac{1}{B_i} - e_i^0 \right) + e_i^0 \quad (28)$$

where τ is the driving variable. When $\tau = 0$, the system is at the initial state $\mathbf{e}^0$ and when $\tau = 1$, the system is at the theoretical equilibrium $\mathbf{e}^*$. For a given $\mathbf{e}^0$ and $\mathbf{M}_\beta$ (the matrix of all pairwise co-regulation coefficients), the relative concentrations $\mathbf{e}$ are strictly determined by the value of τ . Consistently, results of the simulations showed linear evolutionary trajectories converging towards the theoretical equilibrium (Figures 3A and 3C), with a clear difference between positive and negative co-regulation:

1. Positive co-regulation

When all β_{ij} 's are strictly positive, the relative enzyme concentrations converge towards the theoretical equilibrium $\mathbf{e}^*$ located within the space of relative enzyme concentrations (Figure 3A). At this point, the flux response coefficients are (Supporting Information SI.B.6.1.3):

$$\forall i, \quad R_{E_i}^{J^*} = 1 \quad (29)$$

meaning that every enzyme strictly controls the flux. This is not in contradiction with the previously mentioned evolutionary black hole, because the flux response coefficients concern the *absolute* enzyme concentrations, which increase indefinitely, as does the flux (Figure 3B), while the *relative* enzyme concentrations are stuck at the theoretical equilibrium.

2. Negative co-regulation

If at least one co-regulation coefficient is negative, then at least one enzyme has a negative global regulation coefficient B_i (Supporting Information SI.B.1.2). As concentrations cannot be negative, the theoretical equilibrium $\mathbf{e}^*$ is unattainable because it is outside the space of possible relative enzyme concentrations (Figure 3C). There is another evolutionary equilibrium, which we called the *effective equilibrium* $\tilde{\mathbf{e}}$, defined by (Supporting Information SI.B.6.3):

$$\forall i, \quad R_{E_i}^J = 0 \quad (30)$$

Null response coefficients mean that in this case there is a maximum value of the flux, $\tilde{J}_{\max}$, for the same reason as in the case of competition: the negative co-regulation coefficients prevent the flux from exceeding a certain value because mutations that increase (resp. decrease) the concentration of an enzyme will decrease (resp. increase) the concentration of the other enzyme(s).

The *effective equilibrium* $\tilde{\mathbf{e}}$ could not be found analytically but could be found numerically from another expression of equation 30 (Supporting Information SI.B.6.3):

$$\sum_{j=1}^n \frac{1}{B_j A_j (\tilde{e}_j)^2} = 0 \quad (31)$$

The black curve on Figure 3C shows the set of solutions $\tilde{\mathbf{e}}$ satisfying equation 31 for a given set of A_i and B_i values and a large number of initial enzyme concentrations $\mathbf{e}^0$. All evolutionary trajectories converge towards the theoretical equilibrium, but stop on this curve at their respective effective equilibrium. Figure 3D shows the evolution of the corresponding fluxes and their respective $\tilde{J}_{\max}$.

3.1.4 Case 4: Competition *plus* co-regulation

If there is competition and co-regulation, the relationship between flux and enzyme concentrations is a dome in the multidimensional space of relative enzyme concentrations, but here the relative enzyme concentrations are constrained by the effects of co-regulation to follow line $\mathcal{E}$ (equation 28). Consequently, the flux evolves following the curve defined by the intersection of the dome with the upright plane drawn from line $\mathcal{E}$ in space $(\mathbf{e}, J)$ (Figure 4). There are three remarkable points on this curve: (i) the flux value corresponding to the initial relative enzyme concentrations $\mathbf{e}^0$ (black point); (ii) the flux value corresponding to the theoretical equilibrium of relative enzyme concentrations $\mathbf{e}^*$ (red point); (iii) the maximum of this curve, $\tilde{J}_{\max}$, corresponding to the effective equilibrium of relative enzyme concentrations $\tilde{\mathbf{e}}$ (yellow point). Note that for a given $\mathbf{e}^*$, the orientation of the plane is determined by $\mathbf{e}^0$.

The theoretical equilibrium is the same as in the case of co-regulation alone (Supporting Information SI.B.7.1.1):

$$\forall i, \quad e_i^* = \frac{1}{B_i} \quad (32)$$

and the effective equilibrium is also defined by (Supporting Information SI.B.7.2):

$$\forall i \quad R_{E_i}^J = 0 \quad (33)$$

which leads to the following equation:

$$\sum_{j=1}^n \frac{(1/B_j) - e_j^0}{A_j(\tilde{e}_j)^2} = 0, \quad (34)$$

from which $\tilde{\mathbf{e}}$ can be found numerically. If we consider a large number of possible $\mathbf{e}^0$'s for given $\mathbf{M}_\beta$ and $\mathbf{A}$, we can compute a set of points corresponding to $\tilde{\mathbf{e}}$, which defines a closed curve in the case of positive co-regulation (Figure 5A) and an open curve in the case of negative co-regulation (Figure 5C). As previously, all evolutionary trajectories end on these curves. At $\tilde{\mathbf{e}}$ the flux reaches $\tilde{J}_{\max}$, which depends on $\mathbf{e}^0$, $\mathbf{M}_\beta$, $\mathbf{A}$ and E_{tot}^0 (Figures 5B and 5D).

When there is both competition and positive co-regulation, two possible evolutionary equilibria exist, depending on the initial relative enzyme concentrations $\mathbf{e}^0$ (Figure 4). If the theoretical equilibrium is reached first, the evolutionary black hole, which is also valid in this situation (Supporting Information SI.B.7.1.2), will prevent any flux increase despite the selective pressure. If the effective equilibrium is reached first, selection for high flux will prevent the theoretical equilibrium from being reached because the flux value is higher at the effective equilibrium than at the theoretical equilibrium.

3.1.5 Summary of the results on evolutionary equilibria

In summary, the conclusions drawn from the analytical developments and the simulations are as follows:

- Independence: The flux increases continuously but the relative enzyme concentrations reach an equilibrium value that depends on the pseudo-activities of the enzymes.
- Competition: The flux increases up to a maximum value that depends on both the pseudo-activities of the enzymes and the total enzyme concentration. The relative

enzyme concentrations reach an equilibrium that depends on the pseudo-activities of the enzymes, but which differs from the one obtained in the case of independence.

- Co-regulations without competition: The relative enzyme concentrations tend towards an equilibrium that depends only on the co-regulation coefficients between enzymes. If the co-regulation coefficients are all positive, the flux increases continuously and equilibrium is reached. This equilibrium is an evolutionary black hole because, once it is reached, additional mutations cannot change the *relative* enzyme concentrations. If there is at least one negative co-regulation coefficient, the flux evolves towards a maximum and a different equilibrium is reached, called the effective equilibrium, which depends not only on the co-regulation coefficients and the pseudo-activities of the enzymes, but also on the initial enzyme concentrations.
- Competition *plus* co-regulation: This case is qualitatively similar to the case with negative co-regulation alone, with a flux maximum and a multi-factor effective equilibrium. The difference resides in the shape of the adaptive landscape.

Tables 1 and 2 summarize the evolutionary equilibria and the limit to the flux value in the different situations.

Table 1. Theoretical e^* and effective $\tilde{e}$ equilibrium for relative enzyme concentrations.

	Equilibrium	Without competition	With competition
Without co-regulation	Theoretical	$\forall i, e_i^* = \frac{A_i^{-1/3}}{\sum_{j=1}^n A_j^{-1/3}}$	$\forall i, e_i^* = \frac{A_i^{-1/2}}{\sum_{j=1}^n A_j^{-1/2}}$
Positive co-regulation	Theoretical	$\forall i, e_i^* = 1/B_i$	$\forall i, e_i^* = 1/B_i$
	Effective	—	$\tilde{e} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A})$
Negative co-regulation	Theoretical	$\forall i, e_i^* = 1/B_i$ cannot be reached	$\forall i, e_i^* = 1/B_i$ cannot be reached
	Effective	$\tilde{e} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A})$	$\tilde{e} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A})$

Table 2. Limit to the flux value. The maximum value of the flux corresponds to the effective equilibrium.

	Without competition	With competition
Without co-regulation	$J \rightarrow \infty$	$J_{\max} = f(\mathbf{A}, E_{\text{tot}}^0)$
Positive co-regulation	$J \rightarrow \infty$	$\tilde{J}_{\max} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A}, E_{\text{tot}}^0)$
Negative co-regulation	$\tilde{J}_{\max} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A}, E_{\text{tot}}^0)$	$\tilde{J}_{\max} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A}, E_{\text{tot}}^0)$

3.2 The range of selective neutrality of enzyme concentrations

3.2.1 Selective neutrality at evolutionary equilibrium

In addition to searching for evolutionary equilibria, another key question from an evolutionary point of view is the extent to which mutations affecting enzyme concentrations become neutral under selection, *i.e.* have a negligible effect on the flux, and by extension on fitness. The classical criterion defining the neutral zone for haploid populations is a value of the selection coefficient s_i between $-1/N$ and $+1/N$ (Kimura 1985).

In our model, the selection coefficient is related to δ_i , $R_{E_i}^J$ and E_i following the relation $s_i \approx R_{E_i}^J \frac{\delta_i}{E_i}$ (equation 21). When J is limited (when there is competition, negative co-regulation or both competition and co-regulation), $R_{E_i}^J$ is null at the effective equilibrium of relative enzyme concentrations, and thus mutations are nearly neutral near this point. When J is not limited (when enzymes are independent or when there is positive co-regulation), response coefficients are constant at the theoretical equilibrium and $\frac{\delta_i}{E_i}$ tends towards zero because E_i 's increase indefinitely, and thus mutations are nearly neutral at this point. Therefore, selection for high flux leads to the near neutrality of mutations affecting any enzyme in the pathway in all situations considered.

3.2.2 The range of neutral variation depends on the enzymes and the constraints

Even though the mutations affecting enzyme concentrations become neutral as relative enzyme concentrations approach equilibrium, the size of the neutral zone of mutations is expected to vary depending on the enzyme properties and the constraints applied. The zone of selective neutrality determines the *range of neutral variation* (RNV) of the concentrations. The RNV size is given by $|\delta_i^{\text{sup}} - \delta_i^{\text{inf}}|$, where δ_i^{sup} (resp. δ_i^{inf}) is the effect of a mutation such that $s_i = +\frac{1}{N}$ (resp. $s_i = -\frac{1}{N}$), s_i being the superior (resp. inferior) limit of the neutral zone. Let us examine the four situations.

3.2.2.1 Case 1: Independence From equations 21, 23 and 24, we deduced the selection coefficient at theoretical equilibrium:

$$s_i \approx R_{E_i}^{J*} \frac{\delta_i}{E_{\text{tot}} e_i^*} \approx \frac{A_i^{-2/3}}{\sum_{j=1}^n A_j^{-2/3}} \frac{\delta_i}{\frac{A_i^{-1/3} E_{\text{tot}}}{\sum_{j=1}^n A_j^{-1/3}}}$$

Since $\delta_i = \nu$ when enzyme concentrations are independent, it comes:

$$s_i \approx \frac{\nu}{E_{\text{tot}}} \frac{A_i^{-1/3} \sum_{j=1}^n A_j^{-1/3}}{\sum_{j=1}^n A_j^{-2/3}} \quad (35)$$

We see that, for a given ratio ν/E_{tot} , the pseudo-activities A_j alone determine the s_i 's. For two enzymes i and i' , if $A_i < A_{i'}$, then $|s_i| > |s_{i'}|$, *i.e.* the selection coefficients are inversely related to the pseudo-activities. Therefore, to get out of the neutral zone, a larger mutation effect is required for enzymes with large A_j 's than for those with small A_j 's (Figure S1A). This means that RNV size is positively related to the pseudo-activities

A_i , and hence is inversely related to the theoretical equilibria e^* . Numerical application was performed for a ten-enzyme pathway using equation 22 to compute the limits of the RNVs. Results were fully consistent with the theoretical predictions (Figure 6A).

3.2.2.2 Case 2: Competition We did not find any analytical method to study the RNV size near the equilibrium. Graphically, we observed that an enzyme with low pseudo-activity needs larger mutations to get out of the neutral zone (Figure S1B). RNVs were computed numerically from equation 22. We found that RNVs were *inversely* related to enzyme pseudo-activities, unlike in case 1, and hence were positively related to e^* (Figure 6B). Another striking observation is that, for a given total enzyme concentration, RNVs are one order of magnitude greater in the case of competition than when there is independence.

3.2.2.3 Case 3: Co-regulation From equation 28, the relation between *absolute* enzyme concentration and the driving variable τ when there is co-regulation can be written as (Supporting Information SI.B.8.1):

$$E_i = E_i^0 + \frac{E_{\text{tot}}^0}{B_i} \left(-1 + \frac{1}{1 - \tau} \right) \quad (36)$$

If we consider the superior and inferior limits of the RNV for enzyme i , respectively E_i^{sup} and E_i^{inf} (*i.e.* $E_i + \delta_i^{\text{sup}}$ and $E_i + \delta_i^{\text{inf}}$), we can write:

$$|E_i^{\text{sup}} - E_i^{\text{inf}}| = \left| \frac{E_{\text{tot}}^0}{B_i} \left(\frac{1}{1 - \tau^{\text{sup}}} - \frac{1}{1 - \tau^{\text{inf}}} \right) \right| \quad (37)$$

where τ^{sup} and τ^{inf} are the superior and inferior limits of the driving variable.

We see that RNVs are inversely related to $|B_i|$'s: the enzymes that have the smallest effect on total enzyme concentration *via* co-regulation have the largest RNV (Figures 6C and 6E).

3.2.2.4 Case 4: Competition plus co-regulation The same method as previously was applied when we considered jointly competition and co-regulation. From equation 28, we have (Supporting Information SI.B.8.2):

$$E_i = \tau \left(\frac{E_{\text{tot}}^0}{B_i} - E_i^0 \right) + E_i^0 \quad (38)$$

Considering the superior and inferior limits of the RNV for enzyme i (see above), we get:

$$|E_i^{\text{sup}} - E_i^{\text{inf}}| = E_{\text{tot}}^0 \left| \left(\frac{1}{B_i} - e_i^0 \right) (\tau^{\text{sup}} - \tau^{\text{inf}}) \right| \quad (39)$$

Thus, there is a positive relationship between RNV and $|1/B_i - e_i^0|$, *i.e.* the difference between the theoretical equilibrium ($1/B_i$) and the initial relative concentration (e_i^0) (Figures 6D, 6F), meaning that an enzyme whose concentration is initially far from its theoretical equilibrium will have a large RNV. A geometric representation of this relationship is shown in Figure 7 when there is positive co-regulation.

Table 3 and Figure 6 summarize the factors determining the ranking order of RNVs depending on the type of constraint.

Table 3. Factors determining the ranking order of RNVs

	Without competition	With competition
Without regulation	Same order as the A_i 's	Reverse order of the A_i 's
With regulation	Reverse order of the $ B_i $'s	Same order as the $\left \frac{1}{B_i} - e_i^0 \right $'s

3.2.3 Measuring RNVs from simulations of long-term evolution

We tested these theoretical predictions by simulating the long-term evolution of a three-enzyme pathway. For each enzyme and each situation, we computed the RNV limits at each time step between 60,000 and 120,000 generations (*i.e.* once evolutionary equilibrium is reached), then we computed the mean size of the RNVs. The relationship between mean RNV value and A_i , B_i or $\left| \frac{1}{B_i} - e_i^0 \right|$ was fully consistent with theoretical predictions (Figure S2).

In addition, the simulations led to two observations.

- When the flux is not limited (independence and positive co-regulation), RNV size increases over the course of generations. For example, if $\mathbf{E}^0 \approx (21.9, 30.7, 47.3)$, mean RNV size for enzyme 1 is ≈ 0.13 between 60,000 and 80,000 generations, and ≈ 0.17 between 100,000 and 120,000 generations. The reason is that, over the course of generations, a mutation of a given effect has increasingly less effect on the flux, and therefore on the selection coefficient, because enzyme concentrations increase continuously.
- When there is competition and/or negative co-regulation, the dome-shape of the flux-enzyme relationship can lead to a split of the RNVs. Near the maximum flux, an upper limit of the RNV cannot be defined because mutations are either neutral or deleterious. When the flux moves away from the maximum, advantageous mutations may occur, with the RNV having an upper and a lower limit. But interestingly, another RNV appears on the other side of the dome, due to large mutations that result in flux values that are similar to those caused by mutations of small effect. (Figure S3). The farther the flux is from the maximum, the smaller the two moieties of the RNVs and the stronger the selection to bring the concentrations towards effective equilibrium.

4 Discussion

Metabolic fluxes are commonly used as model traits in quantitative genetics and evolutionary studies, benefitting from the development of elaborate theoretical tools that enable to relate “genotype” (the genetically controlled enzyme properties) to phenotype (the flux). The multidimensional flux-enzyme relationship can be viewed as an adaptive landscape and used to analyze the dynamics of enzyme evolution. In this study, we developed a system of differential equations that describes the evolution of enzyme concentrations in a pathway under directional selection for increased metabolic flux. We considered two realistic constraints, separately or jointly, that prevent enzyme concentrations from varying independently from each other, namely competition for cellular resources and

co-regulation between enzymes. We defined two types of evolutionary equilibria of relative enzyme concentrations: a theoretical equilibrium and an effective equilibrium, the latter being reached when the flux is limited by certain constraint(s) (tables 1 and 2). In both constrained and unconstrained conditions, mutations become near neutral near the evolutionary equilibrium, however the range of neutral variation (RNVs) of enzyme concentrations differs to a large extent depending on the enzyme properties and the constraints applied (table 3). Simulation results of long-term evolution were fully consistent with the analytical predictions.

4.1 Modeling rationale

The effect of an enzyme j on the flux of a linear pathway depends on three factors: the ratio of kinetic parameters $\frac{k_{catj}}{K_{Mj}}$, the abundance E_j and the product of the equilibrium

constants from the first to the $(j - 1)^{th}$ reaction, $K_{1,j-1}$. Both E_j and $\frac{k_{catj}}{K_{Mj}}$ can be genetically variable. Recently, Heckmann et al. (2018) simulated the evolution of turnover numbers k_{catj} and showed that their values can hardly be maximized.

In our model, we distinguished the concentration E_j from the "pseudo-activity" $A_j = \frac{k_{catj}}{K_{Mj}} K_{1,j-1}$ of enzyme j , the latter being a systemic property as its value depends on upstream reactions in the pathway. We focused on enzyme concentrations, which display two features that are relevant from an evolutionary point of view:

(i) The range of natural variability is expected to be noticeably larger for enzyme concentration than for catalytic efficiency $\frac{k_{catj}}{K_{Mj}}$. While genetically controlled changes in kinetic parameter values are caused by substitutions at certain sites of the enzyme-coding gene, changes in protein expression can be caused by mutations at multiple *cis* and *trans* elements affecting the gene regulatory network, the translation machinery, mechanisms of protein turnover and stability, etc. (Landry et al. 2007; Zheng et al. 2011; Gruber et al. 2012). This is consistent with the highly polygenic nature of proteomic variations observed in various species (Damerval et al. 1994; Albert and Kruglyak 2015; Chick et al. 2016). Therefore, proteome variability could play a much larger role in the evolution of pathway efficiency than the variability of kinetic parameters. Thus, in our model, we chose to consider mutations that target enzyme concentrations E_j , and let pseudo-activities $A_j = \frac{k_{catj}}{K_{Mj}} K_{1,j-1}$ be invariable. Note that kinetic parameters and concentration play an identical role in the expression of the flux, but the constraints applied to E do not make sense for k_{cat} and K_M .

(ii) Enzyme synthesis and degradation are energetically costly for the cell; furthermore, space and ribosome availability are limited (Koehn 1991; de Vienne et al. 2001b; Klumpp et al. 2019). Therefore, protein over-expression can result in protein burden effects (Snoep et al. 1995; Eguchi et al. 2018). Continuous selective pressure for optimizing cellular resource utilization imposes a constraint on total enzyme abundance and a trade-off between enzymes (Albertin et al. 2013).

In addition to the constraint on E_{tot} , we also considered the possibility that enzyme concentrations were co-regulated. In the cell, multiple mechanisms can result, directly or indirectly, in dependencies between enzyme quantities, *e.g.* operons in prokaryotes, metabolic gene clusters in eukaryotes (Lee and Sonnhammer 2003; Wong and Wolfe 2005)

regulation cascades (Albert and Kruglyak 2015), *trans* effects due to transcription factors (Guerin et al. 2016; Lambert et al. 2018), transposable elements (Cowley and Oakey 2013) and microRNAs (Wang et al. 2011), etc. To model the evolutionary implications of these dependencies within a single theoretical framework, we considered that a mutation targeting the concentration of an enzyme modifies the concentrations of the other enzymes. To keep things relatively simple, we chose linear and constant co-regulation coefficients, but we assumed that they could take on any value. When coordinated enzyme expression is due to a common *trans* factor or physically clustered genes, the system of differential equations has to be modified slightly because in this case the mutations target the driving variable τ (Supporting Information SI.B.9). This changes the distribution of the mutation effects on the co-regulated enzymes, but does not modify evolutionary equilibria, which do not depend neither on the effects nor on the targets of the mutations. Further developments could rely on more complex models of gene regulatory network functioning and evolution (De Jong 2002; Spirov and Holloway 2013).

The two-step process we used to formalize the mutation effects under competition is not simply theoretically convenient, but may reflect actual events at the molecular level. First, a mutation with a canonical effect ν affects the transcription level of an enzyme-coding gene. Second, transcripts are translated proportionally to their abundance up to the maximum total enzyme concentration allowed by the cell. This results in an actual mutation effect on the concentration of the target enzyme and in pseudo-mutation effects on the other enzymes. The process is similar if there is co-regulation in addition to competition.

As in Serohijos et al. (2012) and Heckmann et al. (2018), our evolution model is based on an adaptive dynamics approach that relies on classical population genetics concepts, such as the selection coefficient s and the fixation probability P_f , with the assumption of constant population size N . Other assumptions were made to render tractable the mathematical developments: small mutation effect relative to enzyme concentration, no fixation of deleterious mutations and large population size. We relaxed these assumptions in the simulations, and this did not affect the predictions of evolutionary equilibria of relative enzyme concentrations and maximum flux values (if applicable).

4.2 Constraints shape the adaptive landscape

4.2.1 Optimal enzyme concentrations depend on the type of constraints

The evolutionary equilibria of relative enzyme concentrations depend on the constraints applied. If enzyme concentrations vary independently, there is an inverse relationship between relative concentrations and pseudo-activities $\mathbf{A}$ at equilibrium, which is intuitively consistent, and both enzyme concentrations and flux can increase indefinitely. This is not biologically realistic because enzyme concentrations are necessarily bounded by limited cellular resources and various regulatory mechanisms. Such constraints shape the adaptive landscape.

When enzymes compete for resources, the adaptive landscape is a dome in the multidimensional space of relative enzyme concentrations; its shape is determined by the enzyme pseudo-activities and its height is determined by the total enzyme concentration. The top of the dome corresponds to the maximum flux reached at equilibrium (Figure 2A). At that point, the relative enzyme concentrations are inversely related to the pseudo-activities $\mathbf{A}$, as when enzyme concentrations vary independently, but they are more dispersed. At present, we have no explanation for this result.

Does selection for high flux result in the optimal distribution of enzyme concentrations when there is competition? The theory of optimality of cost minimization has been developed in detail (Klipp and Heinrich 1999; Liebermeister et al. 2004; Noor et al. 2016). Klipp and Heinrich (1999) addressed the question of the optimal distribution of enzyme concentrations in a pathway when there is competition between enzymes. They considered a fixed flux and searched for the distribution that minimizes the total enzyme concentration, which amounts to maximizing the flux with a fixed total concentration, as we did here. Using the Lagrange multiplier method, they obtained the same equilibrium of relative enzyme concentrations as the one we obtained using an adaptive dynamics approach. Thus, the mutation-selection process for increased flux amounts to optimally minimizing the total enzyme concentration allocated to the pathway.

Co-regulation between enzymes prevents reaching this optimum. With co-regulation, the evolutionary trajectories of the relative enzyme concentrations become linear; the orientation of this line is determined by the initial enzyme concentrations and the co-regulation coefficients. Hence the flux follows the curve defined by the intersection of the dome and the upright plane drawn from the line (Figure 4). The maximal flux value on this curve, $\tilde{J}_{\max}$, is a local maximum corresponding to the effective equilibrium of relative enzyme concentrations.

If there is co-regulation, with or without competition, the theoretical equilibrium of relative enzyme concentrations depends on the co-regulation coefficients, according to the simple relation $e_i^* = 1/B_i$, where the global regulation coefficient B_i quantifies the effect of E_i variation on the total enzyme concentration. This evolutionary equilibrium implies that:

(i) The higher the effect of enzyme concentration variation on the total concentration, the lower the relative concentration of this enzyme at evolutionary equilibrium;

(ii) Pseudo-activities A_i do not appear in this relation. This does not mean that the kinetic parameters and equilibrium constants do not affect the evolutionary trajectory when there is co-regulation. Indeed, A_i 's determine the shape of the adaptive landscape, and hence the flux value at evolutionary equilibrium;

(iii) This equilibrium is an evolutionary black hole: as one gets closer, the effect of mutations on *relative* enzyme concentrations decreases, and is cancelled out at this equilibrium irrespective of the effects of these mutations. This property has an important consequence when there is competition, because it can prevent the flux from reaching its local maximum $\tilde{J}_{\max}$. If the initial flux value is lower than the flux value at theoretical equilibrium *and* both values are on the same side of the curve (*i.e.* the derivatives at these points have the same sign), selection will be unable to bring the flux value above the theoretical equilibrium (Figures 4B and 5A);

(iv) If there is negative co-regulation, at least one B_i is negative and theoretical equilibrium is not achievable. The flux will then reach a maximum corresponding to the effective equilibrium, whether there is competition or not.

In summary, co-regulation can be evolutionarily costly, since it can prevent the flux from reaching the maximum value given the total enzyme concentration allocated to the pathway. Therefore, increasing fitness by increasing the flux should imply having mutations that modify the co-regulation coefficients, *i.e.* that modify the orientation of line $\mathcal{E}$ to bring $\tilde{J}_{\max}$ closer to $J_{\max}$.

4.2.2 The Range of Neutral Variation of enzyme concentrations depends markedly on enzyme properties, constraints and/or initial enzyme concentrations

From visual inspection of the simulation results, we found that relative enzyme concentrations fluctuate close to their evolutionary equilibrium, which is consistent with the mutation-selection-drift balance: selection drives the system towards equilibrium, but near the equilibrium, mutations become nearly neutral and genetic drift dominates (Figures 2B, 3C, 5A, 5C). Near the point of evolutionary equilibrium, mutations have a negligible effect on the flux, and therefore are in the neutral zone. We compared analytically the range of variation of enzyme concentrations in the neutral zone (RNVs) at evolutionary equilibrium under different conditions.

When enzyme concentrations vary independently, the RNVs are positively related to the pseudo-activities; the opposite is found when there is competition (Figures 6A and 6B). Thus, competition results in a positive relationship between RNVs and enzyme concentrations. In addition, competition strongly increases all RNVs for a given E_{tot} .

When there is only co-regulation, enzymes whose variation has the largest effect on total concentration – hence the smallest effective equilibrium of relative enzyme concentrations – have the smallest RNVs. With both competition and co-regulation, the more an enzyme is initially far from the effective equilibrium, the larger its RNV (Figure 7). Therefore, co-regulation abolishes the effect of A on RNVs, whether there is competition or not, as they did for the theoretical equilibrium values.

Interestingly, the existence of a flux maximum can lead to a discontinuous RNV. One part of the RNV corresponds to mutations of small effect, another part to mutations of large effect; the latter push the enzyme concentrations beyond the top of the dome, where flux values are similar to the resident flux. This optimum overshoot is inherent to Fisher's geometric model of adaptive landscapes (Orr 1998), and is more common when phenotypes are close to the optimum (Scott and Queller 2019).

The magnitude of the RNV should be reflected in the level of polymorphism of the genes affecting enzyme abundance, because high-RNV enzymes can accumulate more neutral mutations in these genes than low-RNV enzymes. It would be informative to compare the levels of variability of enzyme abundances within a pathway and investigate how these are related to kinetic parameters and/or known co-regulatory interactions. Quantitative proteomics databases could be exploited for this purpose.

4.3 Conclusion and future prospects

We combined MCT principles, an adaptive dynamics approach and simulations to analyze the evolution of enzyme concentrations in metabolic pathways under selection for increased flux. We showed that the adaptive landscape is shaped by enzyme properties and by constraints applied to the system. When there is both competition for cellular resources and co-regulation that limits the flux, which is likely to be the most common situation in a cell, selective pressures lead to equilibria of enzyme concentrations and to ranges of neutral variation that not only depend on enzyme properties and the type of constraints but also on the enzyme concentrations at the start of the selective process. This evolutionary contingency comes from the effect of co-regulation between enzymes and could be mitigated by mutations modifying certain regulatory parameters. In addition, the assumption that all enzymes of the pathway are co-regulated can be relaxed by considering groups of differently co-regulated enzymes (manuscript in preparation).

Acknowledgements

We warmly thank Dr. Wolfram Liebermeister and Dr. Arnaud Le Rouzic for their careful reading of the manuscript and their very helpful suggestions. We also thank Pr. Amandine Veber for her useful suggestions regarding the differential equation system of enzyme evolution and Hélène Citerne for language corrections of the manuscript.

CC was supported by a PhD thesis grant from the French Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.

The authors declare no conflict of interest.

Author contributions

CD and DdV initiated the project. CD, MLL, GT, CC and DdV developed the model. GT, MLL and CC performed the analysis. CD and MLL wrote the simulation code. CC developed the dedicated R package. CC and DdV wrote the manuscript, with the contribution of CD.

Literature cited

Aguilar-Rodriguez, J. and A. Wagner, 2018. Metabolic Determinants of Enzyme Evolution in a Genome-Scale Bacterial Metabolic Network. *Genome Biol. Evol.* 10:3076–3088.

Albert, F. W. and L. Kruglyak, 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 16:197–212.

Albertin, W., P. Marullo, M. Bely, M. Aigle, A. Bourgaïs, O. Langella, T. Balliau, D. Chevret, B. Valot, T. da Silva, C. Dillmann, D. de Vienne, and D. Sicard, 2013. Linking Post-Translational Modifications and Variation of Phenotypic Traits. *Mol Cell Proteomics* 12:720–735.

Bershtein, S., A. W. Serohijos, and E. I. Shakhnovich, 2017. Bridging the physical scales in evolutionary biology: From protein sequence space to fitness of organisms and populations. *Curr Opin Struct Biol* 42:31–40.

Blein-Nicolas, M., W. Albertin, B. Valot, P. Marullo, D. Sicard, C. Giraud, S. Huet, A. Bourgaïs, C. Dillmann, D. de Vienne, and M. Zivy, 2013. Yeast Proteome Variations Reveal Different Adaptive Responses to Grape Must Fermentation. *Molecular Biology and Evolution* 30:1368–1383.

Brännström, Å., J. Johansson, and N. von Festenberg, 2013. The Hitchhiker's Guide to Adaptive Dynamics. *Games* 4:304–328.

Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill, and S. P. Gygi, 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* 534:500–505.

Colombo, M., H. Laayouni, B. M. Invergo, J. Bertranpetit, and L. Montanucci, 2014. Metabolic flux is a determinant of the evolutionary rates of enzyme-encoding genes. *Evolution* 68:605–613.

Cornish-Bowden, A. and M. L. Cárdenas, 2020. Contrasting theories of life: Historical context, current theories. In search of an ideal theory. *Biosystems* 188:104063.

Cowley, M. and R. J. Oakey, 2013. Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet* 9:e1003234.

Crow, J. F. and M. Kimura, 1970. An introduction to population genetics theory. Harper and Row, New York.

- Damerval, C., A. Maurice, J. M. Josse, and D. de Vienne, 1994. Quantitative Trait Loci Underlying Gene Product Variation: A Novel Perspective for Analyzing Regulation of Genome Expression. *Genetics* 137:289–301.
- 820 De Jong, H., 2002. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* 9:67–103.
- Dieckmann, U. and R. Law, 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biology* 34:579–612.
- D’Souza, G. and C. Kost, 2016. Experimental Evolution of Metabolic Dependency in Bacteria. *PLoS Genet* 12:e1006364.
- 825 Eanes, W., 1999. Analysis of selection on enzyme polymorphism. *Annu. Rev. Ecol.* 30:301–326.
- Eguchi, Y., K. Makanae, T. Hasunuma, Y. Ishibashi, K. Kito, and H. Moriya, 2018. Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. *eLife* 7:e34595.
- 830 Ellis, R., 2001. Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* 26:597–604.
- Flowers, J. M., E. Sezgin, S. Kumagai, D. D. Duvernell, L. M. Matzkin, P. S. Schmidt, and W. F. Eanes, 2007. Adaptive Evolution of Metabolic Pathways in *Drosophila*. *Mol Biol Evol* 24:1347–1354.
- 835 Greenberg, A. J., S. R. Stockwell, and A. G. Clark, 2008. Evolutionary Constraint and Adaptation in the Metabolic Network of *Drosophila*. *Molecular Biology and Evolution* 25:2537–2546.
- Gruber, J. D., K. Vogel, G. Kalay, and P. J. Wittkopp, 2012. Contrasting Properties of Gene-Specific Regulatory, Coding, and Copy Number Mutations in *Saccharomyces cerevisiae*: Frequency, Effects, and Dominance. *PLoS Genet* 8:e1002497.
- 840 Guerin, C., T. Joët, J. Serret, P. Lashermes, V. Vaissayre, M. D. Agbessi, T. Beule, D. Severac, P. Amblard, and J. Tregear, 2016. Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *The Plant Journal* 87:423–441.
- Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985. Limits of Adaptation: The Evolution of Selective Neutrality. *Genetics* 111:655–674.
- 845 Heckmann, D., D. C. Zielinski, and B. O. Palsson, 2018. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. *Nature Communications* 9.
- Heinrich, R. and T. Rapoport, 1974. Linear Steady-State Treatment of Enzymatic Chains - General Properties, Control and Effector Strength. *Eur. J. Biochem.* 42:89–95.
- Heinrich, R., S. Schuster, and H.-G. Holzhütter, 1991. Mathematical analysis of enzymic reaction systems using optimization principles. *European Journal of Biochemistry* 201:1–21.
- 850 Jiang, L.-G., B. Li, S.-X. Liu, H.-W. Wang, C.-P. Li, S.-H. Song, M. Beatty, G. Zastrow-Hayes, X.-H. Yang, and F. Qin, 2019. Characterization of proteome variation during modern maize breeding. *Molecular & Cellular Proteomics* 18:263–276.
- Kacser, H. and R. Beeby, 1984. Evolution of catalytic proteins: On the origin of enzyme species by means of natural selection. *J Mol Evol* 20:38–51.
- 855 Kacser, H. and J. A. Burns, 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27:65–104.
- , 1981. The Molecular Basis of Dominance. *Genetics* 97:639–666.
- Kacser, H., J. A. Burns, H. Kacser, and D. A. Fell, 1995. The control of flux : 21 years on. *Biochemical Society Transactions* 23:341–366.
- 860 Kimura, M., 1985. *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- Klipp, E. and R. Heinrich, 1994. Evolutionary optimization of enzyme kinetic parameters; effect of constraints. *Journal of theoretical biology* 171:309–323.
- , 1999. Competition for enzymes in metabolic pathways:: Implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *Biosystems* 54:1–14.
- 865 Klumpp, S., W. Bode, and P. Puri, 2019. Life in crowded conditions: Molecular crowding and beyond. *Eur. Phys. J. Spec. Top.* 227:2315–2328.
- Koehn, R., 1991. The Cost of Enzyme-Synthesis in the Genetics of Energy-Balance and Physiological Performance. *Biol. J. Linnean Soc.* 44:231–247.

- Kurland, C. G. and H. Dong, 1996. Bacterial growth inhibition by overproduction of protein. *Molecular Microbiology* 21:1–4.
- Lambert, S. A., A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch, 2018. The human transcription factors. *Cell* 172:650–665.
- Landry, C. R., B. Lemos, S. A. Rifkin, W. J. Dickinson, and D. L. Hartl, 2007. Genetic Properties Influencing the Evolvability of Gene Expression. *Science* 317:118–121.
- Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill, Y. Zhang, B. Oliver, and A. G. Clark, 2008. Evolution of protein-coding genes in *Drosophila*. *Trends in Genetics* 24:114–123.
- Lee, D.-H. and B. Ø. Palsson, 2010. Adaptive Evolution of *Escherichia coli* K-12 MG1655 during Growth on a Nonnative Carbon Source, l-1,2-Propanediol. *AEM* 76:4158–4168.
- Lee, J. M. and E. L. Sonnhammer, 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome research* 13:875–882.
- Liebermeister, W., E. Klipp, S. Schuster, and R. Heinrich, 2004. A theory of optimal differential gene expression. *Biosystems* 76:261–278.
- Lion, S., F. Gabriel, B. Bost, J. Fiévet, C. Dillmann, and D. De Vienne, 2004. An extension to the metabolic control theory taking into account correlations between enzyme concentrations. *European Journal of Biochemistry* 271:4375–4391.
- Morrison, E. S. and A. V. Badyaev, 2017. Beyond topology: coevolution of structure and flux in metabolic networks. *J. Evol. Biol.* 30:1796–1809.
- Newton, M. S., V. L. Arcus, M. L. Gerth, and W. M. Patrick, 2018. Enzyme evolution: innovation is easy, optimization is complicated. *Curr. Opin. Struct. Biol.* 48:110–116.
- Noda-Garcia, L., W. Liebermeister, and D. S. Tawfik, 2018. Metabolite–Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks. *Annu. Rev. Biochem.* 87:187–216.
- Noor, E., A. Flamholz, A. Bar-Even, D. Davidi, R. Milo, and W. Liebermeister, 2016. The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLOS Computational Biology* 12:e1005167.
- Orr, H. A., 1998. The Population Genetics of Adaptation: The Distribution of Factors Fixed During Adaptive Evolution. *Evolution* 52:935–949.
- Pál, C., B. Papp, and M. J. Lercher, 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* 37:1372–1375.
- , 2006. An integrated view of protein evolution. *Nat Rev Genet* 7:337–348.
- R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Rausher, M. D., 2013. The Evolution of Genes in Branched Metabolic Pathways. *Evolution* 67:34–48.
- Rausher, M. D., R. E. Miller, and P. Tiffin, 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol* 16:266–274.
- Sambamoorthy, G., H. Sinha, and K. Raman, 2019. Evolutionary design principles in metabolism. *Proc. R. Soc. B.* 286:20190098.
- Scott, T. J. and D. C. Queller, 2019. Long-term evolutionary conflict, Sisyphian arms races, and power in Fisher’s geometric model. *Ecology and Evolution* 9:11243–11253.
- Sellis, D. and M. D. Longo, 2015. Patterns of variation during adaptation in functionally linked loci. *Evolution* 69:75–89.
- Serohijos, A. W. and E. I. Shakhnovich, 2014. Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. *Molecular biology and evolution* 31:165–176.
- Serohijos, A. W. R., Z. Rimas, and E. I. Shakhnovich, 2012. Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly. *Cell Reports* 2:249–256.
- Snoep, J. L., L. P. Yomano, H. V. Westerhoff, and L. O. Ingram, 1995. Protein burden in *Zymomonas mobilis*: negative flux and growth control due to overproduction of glycolytic

enzymes. Microbiology 141:2329–2337.

Spirov, A. and D. Holloway, 2013. Using evolutionary computations to understand the design and evolution of gene and cell regulatory networks. Methods 62:39–55.

de Vienne, D., B. Bost, J. Fiévet, and C. Dillmann, 2001a. Optimisation of Enzyme Concentra-
 925 tions for Unbranched Reaction Chains: The Concept of Combined Response Coefficient. Acta
 Biotheor 49:341–350.

de Vienne, D., B. Bost, J. Fiévet, M. Zivy, and C. Dillmann, 2001b. Genetic variability of
 proteome expression and metabolic control. Plant Physiology and Biochemistry 39:271–283.

Vitkup, D., P. Kharchenko, and A. Wagner, 2006. Influence of metabolic network structure and
 930 function on enzyme evolution. Genome Biol 7:R39.

Wang, Y., X. Li, and H. Hu, 2011. Transcriptional regulation of co-expressed microRNA target
 genes. Genomics 98:445–452.

Wong, S. and K. H. Wolfe, 2005. Birth of a metabolic gene cluster in yeast by adaptive gene
 relocation. Nature genetics 37:777–782.

935 Wright, K. M. and M. D. Rausher, 2010. The Evolution of Control and Distribution of Adaptive
 Mutations in a Metabolic Pathway. Genetics 184:483–502.

Zheng, W., T. A. Gianoulis, K. J. Karczewski, H. Zhao, and M. Snyder, 2011. Regulatory
 Variation Within and Between Species. Annu. Rev. Genom. Hum. Genet. 12:327–346.

5 Figures

Figure 1. Simulation of enzyme evolution in a three-enzyme pathway when enzyme concentrations vary independently. (A) The three-enzyme pathway used in the simulations. Parameters values are: $X = 1$, $A_1 = 1$, $A_2 = 10$, $A_3 = 30$, $E_{\text{tot}}^0 = 100$, $N = 1\,000$. (B) Evolutionary trajectories of relative enzyme concentrations e_1 , e_2 and e_3 represented in a triangular diagram for ten simulations (one color per simulation) with different initial enzyme concentrations (colored squares). The relative enzyme concentrations converge towards the theoretical equilibrium e^* (white diamond). (C) Evolution of the flux over the course of generations from simulations shown in B.

Figure 2. Simulation of enzyme evolution in a three-enzyme pathway when there is competition. (A) The flux-enzyme relationship is dome-shaped because E_{tot} is constant. The purple point is the maximal flux $J_{\max}$, corresponding to the evolutionary equilibrium. (B) and (C) Representations and symbols are as in Figure 1; the grayscale in (B) corresponds to the flux values. When there is competition, the relative enzyme concentrations converge towards the equilibrium (white diamond) and the flux is limited. Parameter values are: $X = 1$, $A_1 = 1$, $A_2 = 10$, $A_3 = 30$, $E_{\text{tot}}^0 = 100$, $N = 1\,000$.

Figure 3. Simulation of enzyme evolution in a three-enzyme pathway when there is co-regulation. Representations and symbols are as in Figure 1. (A) and (B) Positive co-regulation. (C) and (D) Negative co-regulation. When there is co-regulation, the evolutionary trajectories in (A) and (C) are linear. In the case of negative co-regulation, the theoretical equilibrium e^* (white diamond in (C)) is outside the triangle. The black curve in (C) represents the sets of points of the effective equilibrium $\tilde{e}$ determined numerically for a large number of initial concentrations e^0 . The relative enzyme concentrations fluctuate on both sides of this curve due to the mutation-selection-drift balance. (B) and (D) Evolution of the flux over the course of generations from simulations shown in (A) and (C), respectively. Parameter values are: $X = 1$, $A_1 = 1$, $A_2 = 10$, $A_3 = 30$, $E_{\text{tot}}^0 = 100$, $N = 1\,000$, $\beta_{12} = 0.1$, $\beta_{23} = 2$ (A and B) or $\beta_{12} = 0.32$, $\beta_{23} = -0.43$ (C and D).

Figure 4. Relationship between flux and relative enzyme concentrations when there is both competition and positive co-regulation in a three-enzyme pathway. (A) When there is only competition, the relation between flux and relative concentrations of enzymes 1 and 2 is represented by a dome. The purple point indicates the maximum flux $J_{\max}$. The ivory section represents the upright plane drawn from line $\mathcal{E}$ along which enzyme concentrations vary when there is positive co-regulation (green dashed line on the plane). Red point: flux value at theoretical equilibrium e^* . Black point: flux value at the initial enzyme concentrations e^0 . Yellow point: maximum flux value $\tilde{J}_{\max}$ at effective equilibrium $\tilde{e}$. (B) Relationship between J and τ , the driving variable of line $\mathcal{E}$, corresponding to the ivory section of the dome in (A). Solid vertical line: abscissa of the theoretical equilibrium. Dashed vertical line: abscissa of the effective equilibrium corresponding to $\tilde{J}_{\max}$. Colored points: same as in (A). The arrows indicate the variation of flux under selection for different initial enzyme concentrations. Note that the red point (the theoretical equilibrium) is a singular point that cannot be passed if the initial flux is lower than the flux at e^* and if e^* and e^0 are on the same side of the curve. Parameter values are: $X = 1$, $A_1 = 1$, $A_2 = 10$, $A_3 = 30$, $E_1^0 = E_2^0 = E_3^0 = 100/3$, $\beta_{12} = 2.5$, $\beta_{23} = 0.6$, corresponding to $B_1 = 5$, $B_2 = 2$ and $B_3 = 3.33$.

Figure 5. Simulation of enzyme evolution in a three-enzyme pathway when there is competition and co-regulation. Representations and symbols are as in Figure 1. (A) and (B) Competition plus positive co-regulation. (C) and (D) Competition plus negative co-regulation. In (A) and (C), the grayscales correspond to the flux values and the black curves represent the sets of points of the effective equilibrium determined numerically for a large number of initial concentrations e^0 . The evolutionary trajectories converge towards, but do not reach, the theoretical equilibrium e^* (white diamonds). They fluctuate on both sides of the curves due to the mutation-selection-drift balance. (B) and (D) The evolution of the flux over the course of generations from the ten simulations shown in A and C, respectively. Parameter values are: $X = 1$, $A_1 = 10$, $A_2 = 10$, $A_3 = 30$, $E_{\text{tot}}^0 = 100$, $N = 1\,000$, $\beta_{12} = 0.1$, $\beta_{23} = 2$ (A and B) or $\beta_{12} = 0.32$, $\beta_{23} = -0.43$ (C and D).

Figure 6. Constraint-dependent factors impacting the range of neutral variation (RNV) of enzyme concentrations. Numerical applications for a ten-enzyme pathway. If enzymes vary independently (A), the RNV size at evolutionary equilibrium is positively related to the pseudo-activities – hence negatively to the relative enzyme concentrations –, whereas the opposite is found when there is competition (B). For a given total enzyme concentration, the RNV size is one order of magnitude greater when there is competition than when enzyme concentrations vary independently. If there is positive (C) or negative (E) co-regulation, RNV size is inversely related to $|B|$. When there is competition with positive (D) or negative (F) co-regulation, RNV size is linearly related to $|1/B - e_0|$. The pseudo-activities $\mathbf{A}$ were randomly drawn between 0 and 1 000. Initial concentrations $\mathbf{E}^0$ were randomly drawn between 0 and 100, $\mathbf{E}^0$ was then calibrated to have $E_{\text{tot}}^0 = 100$. When there is co-regulation, inverses of B_i 's were randomly drawn between 0 and 100 (and randomly multiplied by +1 or -1 for negative co-regulation), then calibrated to have a sum equal to 1.

Figure 7. Geometric representation of the RNV when there is both competition and positive co-regulation. The triangle is a top view of the dome of Figure 4A. The color scale represents the flux levels. a and b are the ends of the line $\mathcal{E}$. Purple point: top of the dome. Black point: initial enzyme concentrations $\mathbf{e}^0$. Yellow point: effective equilibrium $\tilde{\mathbf{e}}$. Red point: theoretical equilibrium $\mathbf{e}^*$. The red bars outside the triangle represent the difference $\left(\frac{1}{B_i} - e_i^0\right)$ for each of the three enzymes, and the gold bars represent the range of neutral variation (RNV). Note the positive relationship between $\left(\frac{1}{B_i} - e_i^0\right)$ and RNV. The parameters values are as in Figure 4.

Appendix A Glossary of mathematical symbols

Latin symbols

A_i	Pseudo-activity of enzyme i (the parameter includes the kinetic properties of the enzyme and thermodynamic equilibrium of upstream reactions)
$\mathbf{A}$	Vector of pseudo-activities
B_i	Global co-regulation coefficient, representing effect of a mutation of E_i on E_{tot} due to co-regulation
E_i	Concentration of enzyme i
E_i^0	Initial concentration of enzyme i
$\mathbf{E}$	Vector of enzyme concentrations, considered as the genotype
$\mathbf{E}^0$	Vector of initial enzyme concentrations
$\mathbf{E}^{(1)}$	Vector of enzyme concentrations after a mutation
$\mathbf{E}^{(2)}$	Vector of enzyme concentrations after a mutation and the application of a constraint on E_{tot}
$\mathbf{E}^m$	Vector of enzyme concentrations of the mutant
$\mathbf{E}^r$	Vector of enzyme concentrations of the resident
$\mathbf{E}^{(t)}$	Vector of concentrations at time t
E_{tot}	Total enzyme concentration in the metabolic pathway
E_{tot}^0	Initial total enzyme concentration
e_i	Relative concentration of enzyme i
e_i^0	Initial relative concentration of enzyme i
$\mathbf{e}$	Vector of relative enzyme concentrations
$\mathbf{e}^0$	Vector of initial relative enzyme concentrations
$\mathbf{e}^*$	Theoretical equilibrium of relative enzyme concentrations
$\tilde{\mathbf{e}}$	Effective equilibrium of relative enzyme concentrations, corresponding to a maximum of the flux
$\mathcal{E}$	Line in the space of relative enzyme concentrations along which the relative concentrations vary in the case of co-regulation
J	Metabolic flux, considered as a phenotype that is proportional to fitness
J_i^m	Flux resulting from a mutation targeting enzyme i
J^r	Flux of the resident
K_i	Equilibrium constant of reaction i , catalyzed by enzyme i
$K_{1,i}$	Product of the equilibrium constants from the first to the i^{th} reaction: $K_{1,i} = \prod_{k=1}^{k=i} K_k$
$k_{\text{cat},i}$	Catalytic constant of enzyme i
$K_{M,i}$	Michaelis-Menten constant of enzyme i for its substrate
$\mathbf{M}_\beta$	Matrix of co-regulation coefficients
N	Effective population size
n	Number of enzymes in the metabolic pathway
P_f	Fixation probability of a mutation
$R_{E_i}^J$	Flux response coefficient of enzyme i
S_i	Metabolite i , product of reaction i and substrate of reaction $i + 1$
s_i	Selection coefficient of a mutation targeting enzyme i
t	Evolutionary time
w	Fitness
w^m	Mutant fitness
w^r	Resident fitness
X	Constant representing the environment: $X = X_0 - X_n/K_{1,n}$
X_0	Concentration of the input substrate of the pathway
X_n	Concentration of the output product of the pathway

Greek symbols

α_{ij}	Redistribution coefficient, expressing the variation of E_j due to mutation affecting enzyme i
β_{ij}	Co-regulation coefficient of enzyme i on j
γ_{ij}	Competition coefficient, expression of α_{ij} in the case of competition alone
δ_i	Actual effect of a mutation affecting enzyme i
μ	Mutation rate at each time unit
ν	Canonical effect of a mutation
τ	Driving variable of the line $\mathcal{E}$ in the space of relative enzyme concentrations in the case of co-regulation.