The moderately efficient enzyme: evolutionary and physico-chemical trends shaping enzyme parameters

Supplementary Information

Evolution of suboptimal catalytic constants – a mathematical model

The observed trends of the measured catalytic constants call for an evolutionary explanation that accounts both for the physico-chemical nature of enzymes and for the systematic differences between central and secondary metabolism. The evolutionary adaptation of kinetic constants and their co-optimization with enzyme levels has been studied in a series of theoretical works (see, for instance, [1, 2, 3, 4, 5]). A common assumption in these approaches are upper and lower limits for kinetic constants. An optimization for maximal reaction rate, together with trade-offs between kinetic constants, induced for instance by the thermodynamic Haldane relations, will push some of the kinetic constants to their theoretical limits [4].

However, the analysis we perform in the main text suggests that many catalytic constants remain suboptimal and far from their physical limitations. Thus, we developed a mathematical model that can explain these findings. Like earlier models, it describes a metabolic system under selection pressure for high metabolic performance. Since low catalytic constants can be compensated by high enzyme levels, the co-adaptation of enzyme levels is taken into account. The model suggests a relationship between the catalytic constants, the enzyme levels, and their effects on the cellular fitness. This is in line with earlier predictions and with the trends observed in our statistics.

1 Model considerations

Our model is based on the following considerations. An increase in k_{cat} values is generally beneficial because it allows for decreasing the enzyme level and thereby releases resources. However, we observed that k_{cat} values show systematically lower – and therefore, probably sub-optimal – values in certain regions of metabolism, which indicates an evolutionary compromise. Klipp and Heinrich [5] proposed a theoretical model in which enzyme levels are determined by two requirements: a metabolic flux, used as a proxy for cellular fitness, has to be maximized, while the sum of all enzyme concentrations is fixed at a maximal allowed value. The optimization leads to enzyme levels that are proportional to the metabolic control coefficients. Those enzymes that have the biggest impact on the fitness – measured by the relative increase in fitness induces by a relative increase in enzyme level – are also the ones that are most abundant. An extended analysis comprised the co-optimization of kinetic constants with hard upper limit constraints.

Our approach follows the same lines. We consider a metabolic system in a stationary state, which depends on the maximal velocities of all enzymes. The system is required to behave optimally, for instance, to maximize the flux of biomass production. The maximal velocities, in turn, are products of the enzyme levels and the catalytic constants, which both evolve by mutation and selection. For the enzyme levels, we assume a cost term that describes a fitness reduction at increasing enzyme levels. In contrast, large k_{cat} values once they have evolved, do not put an additional burden on the cell. However, while the increase of enzyme levels may be a fairly easy evolutionary task, the evolution of large k_{cat} values is considerably more challenging. In fact, large k_{cat} values may be hard to achieve, not because they are costly, but because fast enzymes have to be found during evolution and to be preserved against mutations. In our model, the difficulty to evolve fast enzymes is described by an apparent evolutionary cost term [1], which is justified below by an underlying stochastic model of mutation and selection.

With the cost term, the common evolution of enzyme levels and catalytic constants can be presented as an optimization problem in which a cellular benefit, depending on both quantities, is balanced by the cost of enzyme production and by the apparent evolutionary cost of the catalytic constants. In contrast to previous models, we do not consider hard upper limits on the catalytic constants, but a soft constraint implemented by the apparent evolutionary cost. Accordingly, our model does not describe an optimization of cellular benefit, but an evolution of k_{cat} values towards values that are reasonably likely to be reached by mutation and selection.

2 Cost-benefit model for catalytic constants

Following [5], we consider a metabolic system in steady state with a metabolic benefit f depending on some of the steady state fluxes and concentrations. For instance, the benefit could simply be given by the rate of biomass production. The rate of each reaction (with index i) follows an enzymatic rate law of the form $v_i(u, c) = u_i\, k_i w_i(c)$ with a maximal velocity $v_i^{\mathrm{max}} = u_i\, k_i$ depending on the enzyme concentration u_i and on the catalytic constant k_i. If the rate law is reversible, forward and backward flux scale proportionally with v_i^{max}. For each enzyme, we assume that an increase in the maximal velocity will increase the metabolic benefit since otherwise, there would be an incentive to down-regulate the enzyme. Stated more technically, each reaction has a positive control coefficient towards the fitness function f. The unscaled control coefficient C_i^{f} describes the sensitivity of the fitness function towards small perturbations of reaction i and is formally defined as $C_i^{\mathrm{f}} = (\partial f/\partial p_i)/(\partial v_i/\partial p_i)$ where p_i is a parameter affecting exclusively the rate of reaction i, $\partial v_i/\partial p_i$ describes the immediate effect of this parameter on the reaction rate, and $\partial f/\partial p_i$ describes its indirect overall effect on the fitness function, evaluated in steady state. The scaled control coefficients, defined by $\hat{C}_i^{\mathrm{f}} = \frac{v_i}{f}\, C_i^{\mathrm{f}}$, are unitless and refer to relative changes of f and v_i.

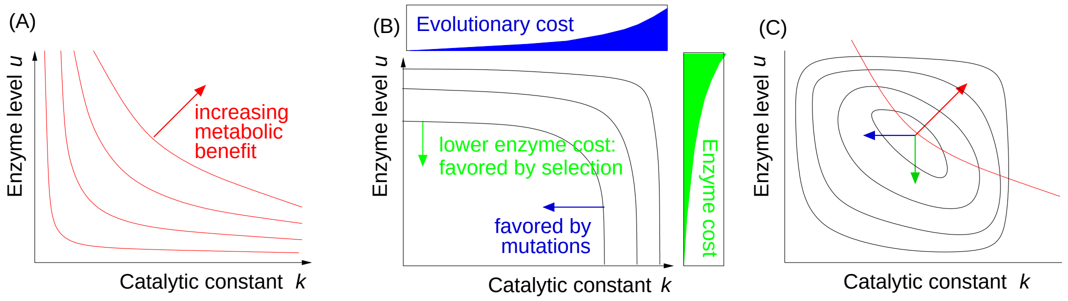

Figure 1: Schematic model for the evolution of catalytic constants. (A) The metabolic benefit – e.g., the stationary flux of biomass production – depends on the maximal velocities $v^{\mathrm{max}} = u\, k$ and therefore on the catalytic constant k (x-axis) and on the enzyme level u (y-axis). The benefit isolines $u\, k = \mathrm{const.}$ are shown as red curves. (B) The catalytic constants and enzyme levels are scored by cost functions (depicted by blue and green areas, respectively). Their sum (show by isolines) describes an evolutionary tendency towards lower enzyme levels and catalytic constants. (C) By subtracting both cost functions from the benefit function, we obtain an apparent fitness function (black isolines). In its maximal point, the gradients of benefit (red arrow) and costs (blue and green arrows) are balanced. For a fixed benefit (red isoline), the position of the optimum results from a balance between the two cost terms.

In the model, the catalytic constants and enzyme levels (in vectors k and u) for a complex metabolic system are supposed to maximize the apparent objective function $F(k, u)$, given by the formula

$$F(k, u) \quad = \quad f(k, u) - g(u) - \sum_i h_i(k_i) \qquad (1)$$

where f, the metabolic benefit, depends on the stationary state and therefore, *via* the maximal velocities $v_i^{\mathrm{max}} = u_i\, k_i$, on all u_i and k_i. The second term, $g(u)$, describes the total enzyme cost, summarizing the investments in protein production, maintenance, and degradation as well as the burden of molecular crowding. The third term, which punishes large k_{cat} values, appears like a cost term, but arises from the fact that high k_{cat} values are less likely to evolve. A more specific interpretation of this cost term is given below. The necessary conditions for maximality of $F(k,u)$, illustrated in Figure 1, read

$$\begin{aligned} 0 &= \nabla_u F = \nabla_u f - \nabla_u g \\ 0 &= \nabla_k F = \nabla_k f - \nabla_k g \end{aligned} \tag{2}$$

Since $\partial v_i/\partial u_i = v_i/u_i$, the derivatives of f can be expressed in terms of the unscaled metabolic control coefficients by $\partial f/\partial u_i = C^{\mathrm{f}}_i \partial v_i/\partial u_i = C^{\mathrm{f}}_i v_i/u_i$ and likewise $\partial f/\partial k_i = C^{\mathrm{f}}_i v_i/k_i$. We obtain the balance equations

$$C^{\mathrm{f}}_i\, v_i = \frac{\partial g}{\partial u_i}\, u_i \tag{3}$$

$$C^{\mathrm{f}}_i\, v_i = \frac{\partial h}{\partial k_i}\, k_i, \tag{4}$$

which contain the unscaled control coefficients. The equations can be recast in terms of the scaled control coefficients as

$$f\,\hat{C}^{\mathrm{f}}_i = \frac{\partial g}{\partial \ln u_i} = \frac{\partial h}{\partial \ln k_i}. \tag{5}$$

The control coefficients describe the quantitative selection pressure on the maximal velocities v_i, which involves both a selection pressure on large enzyme levels and on large catalytic constants. Thus, the equations state a balance between the control of the enzyme on the cellular fitness (first term), its marginal cost (second term), and the difficulty to obtain or maintain a large k_{cat} value during evolution (third term). This suggests, although indirectly, that control coefficients will correlate with enzyme levels and with catalytic constants. Experiments indicate [6] that the cost of enzymes increases more than proportionally with the enzyme level. If we just assume that $g(u)$ increases with each enzyme level u_i and that its second derivative is non-negative, the right-hand side of Eq. (3) – and therefore also the control coefficient – increases with u. For instance, for a cost function $g(u) = y(\sum_i m_i u_i)$ with an increasing function y and individual enzyme cost weights m_i, the derivatives read $\partial g/\partial u_i = y'\, m_i$ where y' is the slope of y. The balance equations imply that scaled control coefficients and cost-weighted enzyme levels are proportional:

$$\hat{C}^{\mathrm{f}}_i \sim m_i\, u_i. \tag{6}$$

If this is true, highly expressed enzymes will also tend to exert a large control on the cellular fitness. For the k_{cat} values, we obtain an analogous result by assuming that $\partial^2 h/\partial {k_i}^2 \geq 0$. For instance, for a power-law function $h_i(k_i) = z_i\, {k_i}^{\gamma_i}$, we obtain

$$\hat{C}^{\mathrm{f}}_i \sim z_i\, \gamma_i\, {k_i}^{\gamma_i} \tag{7}$$

where $z_i > 0$ and $\gamma_i \geq 1$, with larger values for reactions that are difficult to catalyze by enzymes.

3 Model predictions and limitations

In contrast to the approach [5], our model does not push k_{cat} values to their physical upper limits, but towards values that are reasonably likely to arise from mutation and selection. Under the assumptions made, enzyme level should increase until the incentives of increasing the benefit and lowering the cost would cancel each other. Likewise, the catalytic constant should increase on an evolutionary time scale until enzymes with larger k_{cat} values become improbable to find and to preserve against random mutations.

Our model predicts that during evolution, enzyme levels and catalytic constants will increase until both reach a point where the metabolic benefit is balanced by the cost of enzyme production and by a difficulty to evolve higher k_{cat} values. The latter factor may depend on the physico-chemical mechanism, which explains the observed difference between EC classes. Furthermore, reactions exerting a strong control on the metabolic objective of the cell are expected to show both increased enzyme levels and catalytic constants.

Does this imply that enzyme levels and catalytic constants are correlated? The model does not provide a simple general answer to this. On the one hand, both of them correlate with the control coefficient, which may induce a general correlation between them. On the other hand, given a fixed benefit value (hyperbola in Figure 1), we can also expect that low $k_{\rm cat}$ value can be compensated by increased enzyme levels. In a general statistics, both effects will appear in combination.

The main prediction of our model, in line with earlier works in the area, is a statistical correlation between $k_{\rm cat}$ values and the scaled control coefficients towards the cell's metabolic benefit. This is still hard to verify because it would require to determine a large number of quantitative control coefficients, which may strongly differ between organisms and between growth conditions [7] and can be strongly affected by modifications of individual enzymes [8]. Moreover, it would be important to know what are the natural environments under which the metabolic systems evolved and which metabolic state features (other than biomass production) are relevant for the fitness function, e.g. increase of production fluxes, elimination of toxic compounds, and stabilization of metabolite concentrations.

Our evolutionary model has a number of conceptual limitations.

1. For enzymes operating in the linear range, the relevant kinetic parameter determining the speed of the enzyme is not the catalytic constant, but its ratio over the Michaelis constant $K_{\rm M}$ (or the product of Michaelis constants, for a multimolecular reaction). Our evolutionary argument can also be applied to this case, suggesting a selective pressure for higher $k_{\rm cat}/K_{\rm M}$ ratios.

2. Our model assumes a separate enzyme for each individual reaction. In contrast, our statistical analysis showed that enzymes, especially in secondary metabolism, tend to be promiscuous. An adaptation to several alternative substrates is likely to lower the $k_{\rm cat}$ value for each one of them. This effect cannot directly be captured by our model because multifunctional enzymes are disregarded.

3. Our model is based on the assumption of a single metabolic state. However, in reality, a living organism encounters a variety of environmental conditions and metabolic tasks. The actual metabolic network is expected to be selected to be able to respond to these changes under these conditions and to adopt its internal fluxes accordingly.

4. The current enzymes are possibly not an end point of evolution, but still "in the making". If this is true, then the balance between benefits and costs, as predicted by the model, has not yet been reached and further changes in $k_{\rm cat}$ values are expected.

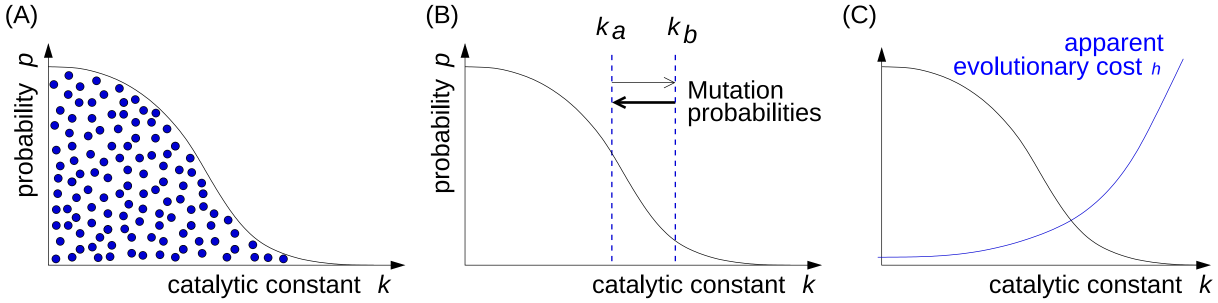

Figure 2: Derivation of the evolutionary cost term. (A) An ensemble of randomly evolved enzymes is sorted by catalytic activity k (x-axis). The probability of finding a certain value k (blue curve) decreases with k. (B) We assume that the distribution $p(k)$ is the stationary distribution of an evolutionary process without selection. Probability ratios determine the relative probabilities of mutations leading from value k_a to k_b and back (indicated by arrows). (C) The same stationary distribution as in (B) would arise from a process with equal jump probabilities and a selection according a hypothetical cost function $h(k)$ (blue curve), proportional to the negative logarithm of $p(k)$. This cost term can act as a proxy for describing the unequal mutation probabilities that make it harder to reach high k values by random mutations.

4 The evolutionary cost can be interpreted in terms of mutation probabilities

In the apparent fitness function, we assume a term $h(k)$ expressing the fact that high $k_{\rm cat}$ values are difficult to obtain and therefore less likely to appear in evolution. To justify this term, we just consider the simple fact that there are many ways to construct slow enzymes, but much fewer possibilities to construct a fast enzyme. As we will see below, this leads to an apparent evolutionary cost term

$$h(k) = -\frac{1}{\beta}\, p^{\circ}_{\rm k}(k) \tag{8}$$

where $p^{\circ}_{\rm k}(k)$ is a probability distribution of $k_{\rm cat}$ values, expected to arise in a random evolution without any selection for the cost or benefit of the enzyme. Since such a neutral evolution is unlikely to yield high $k_{\rm cat}$ values, the function $p^{\rm mut}(k)$ can be assumed to decrease rapidly and accordingly, the apparent cost $h(k)$ will increase with k. The parameter β in the above formula describes the strength of selection.

For a mathematical derivation, we focus on a single enzyme i and consider a stochastic model of evolution by mutation and selection. Our chemical reaction can be catalyzed by a large set of enzyme variants with different $k_{\rm cat}$ values and enzyme evolution will consist of a series of random jumps between the different variants, accompanied by random changes of the catalytic constant. At the same time, mutations in the gene regulation system may lead to random changes of the enzyme level u. Therefore, we obtain a jump process in the space of pairs (k, u).

Let us first consider neutral evolution without selection. We assume that random mutations of the enzyme and its regulation follow a Markov process with a stationary distribution $p^{\circ}(k, u) = p^{\circ}_{\rm k}(k)\, p^{\circ}_{\rm u}(u)$. For simplicity, we assume that the u values are uniformly distributed between zero and some large upper limit. In contrast, the probability density $p^{\circ}_{\rm k}(k)$ (see Figure 2 (A)) is assumed to decrease rapidly since higher catalytic constants are less likely to arise by chance. Given a certain stationary distribution $p^{\circ}(k, u)$, detailed balance requires that conditional jump probabilities for mutations between the pairs (k_a, u_a) and (k_b, u_b) show the ratio

$$\frac{p^{\rm mut}_{a\rightarrow b}}{p^{\rm mut}_{b\rightarrow a}} = \frac{p^{\circ}(k_b, u_b)}{p^{\circ}(k_a, u_a)}. \tag{9}$$

Next, we take into account the effects of selection. We assign to each possible pair (k, u) a fitness function $G(k, u)$; if a mutation involves a fitness change $\Delta G = G(k_b, u_b) - G(k_a, u_a)$, the fixation probability for this mutation is assumed to be given by

$$p^{\rm fix}_{a\rightarrow b} = \gamma_{ab}\, \exp\left(\frac{\beta}{2} \Delta G(k, u)\right) \tag{10}$$

where the same prefactor $\gamma_{ab} = \gamma_{ba}$ holds for mutations in both directions. Therefore, the ratio of conditional jump probabilities is given by

$$\frac{p_{a\rightarrow b}}{p_{b\rightarrow a}} = \frac{p^{\rm mut}_{a\rightarrow b}\, p^{\rm fix}_{a\rightarrow b}}{p^{\rm mut}_{a\rightarrow b}\, p^{\rm fix}_{a\rightarrow b}} = \frac{p^{\circ}(k_b, u_b)}{p^{\circ}(k_a, u_a)} \exp\left(\beta\, \Delta G(k, u)\right) \tag{11}$$

and due to detailed balance, the stationary distribution of pairs (k, u) reads

$$p^{\star}(k, u) = p^{\circ}(k, u) \cdot \exp(\beta\, G(k, u)). \tag{12}$$

During evolution, enzymes will approach the maximum point of this distribution, which can be found by maximizing the apparent fitness function

$$F(k, u) = \frac{1}{\beta}\, \ln p^{\star}(k, u) = G(k, u) + \frac{1}{\beta}\, \ln p^{\circ}_{\rm k}(k) = G(k, u) - h(k). \tag{13}$$

While the term $G(k, u)$ describes the actual fitness of the pair (k, u), our evolutionary cost term $h(k)$ arises from the entropy (the logarithmic probability) of neutrally evolved proteins that realize a certain $k_{\rm cat}$ value. Higher $k_{\rm cat}$ values have lower entropies and are therefore less likely to appear. Accordingly, random mutations will drive the enzyme towards higher "realizability" entropy and therefore towards lower $k_{\rm cat}$ values. Since this entropy term plays a similar role for the apparent fitness as the entropy term that appears in the free energy in thermodynamics, the apparent fitness function could also be called “free fitness”.

The actual form of the distribution $p^{\circ}_{\mathrm{k}}(k)$ is unknown, it will be different for every reaction, and will probably be hard to determine. At least, we may assume that it decreases rapidly with k_{cat}. For example, let $p^{\circ}_{\mathrm{k}}(k)$ be given by a normal distribution $k_{\mathrm{cat}} \sim \mathcal{N}(0, \sigma_{k^2})$

$$\ln p^{\circ}_{\mathrm{k}}(k) = -\frac{k^2}{2\,\sigma_k^2} + \mathrm{const.} \tag{14}$$

Our cost term will read

$$h(k) = -\frac{1}{\beta}\,\ln p^{\circ}_{\mathrm{k}}(k) = \frac{k^2}{2\,\beta\,\sigma_k^2} + \mathrm{const.} \tag{15}$$

where the constant offset can be neglected. Its first and second derivative are positive, and this will also hold for other distributions $p^{\circ}_{\mathrm{k}}(k)$ that decrease more rapidly than a Gaussian.

In our derivation of the evolutionary cost term, we made some simplifying assumptions. First, we assumed that enzyme levels, under a neutral evolution, will show a uniform stationary distribution. If we assume a non-uniform distribution instead, we obtain an additional term that depends on the enzyme level and can formally be included into the cost term $g(u)$. Second, the enzymatic cost g might not only depend on the enzyme level, but also systematically on the k_{cat} value. While we do not see empirical evidence for this, such a dependence could also be included into the model. In this case, g will become dependent on k_{cat}.

Finally, there could be many different justifications for the evolutionary cost term, apart from the one we have discussed above. For example, an enzyme with high k_{cat} value might be more difficult to evolve for other metabolic tasks. This reduction in evolvability can be regarded as an evolutionary cost the organism pays for optimizing the kinetic parameter to a single task. In future works, it will be interesting to explore such alternatives.

References

[1] R. Heinrich and HG. Holzhütter. Efficiency and design of simple metabolic systems. *Biomed Biochim Acta*, 44(6):959–969, 1985.

[2] S. Schuster and R. Heinrich. Time hierarchy in enzymatic reaction chains resulting from optimality principles. *J. Theor. Biol*, 129(2):189–209, 1987.

[3] R. Heinrich, S. Schuster, and HG. Holzhütter. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur J Biochem*, 201(1):1–21, 1991.

[4] R. Heinrich and E. Hoffmann. Kinetic parameters of enzymatic reactions in states of maximal activity. an evolutionary approach. *J. Theor. Biol.*, 151:249–283, 1991.

[5] E. Klipp and R. Heinrich. Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *BioSystems*, 54:1–14, 1999.

[6] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436:588–692, 2005.

[7] Brian J. Koebmann, Hans V. Westerhoff, Jacky L. Snoep, Christian Solem, Martin B. Pedersen, Dan Nilsson, Ole Michelsen, and Peter R. Jensen. The extent to which ATP demand controls the glycolytic flux depends strongly on the organism and conditions for growth. *Molecular Biology Reports*, 29:41–45, 2002. 10.1023/A:1020398117281.

[8] K. Elbing, C. Larsson, RM. Bill, E. Albers, JL. Snoep, E. Boles, S. Hohmann, and L. Gustafsson. Role of hexose transport in control of glycolytic flux in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.*, 70(9):5323–5330, 2004.