A Bird's-Eye View of Enzyme Evolution

A bird's-eye view of enzyme evolution: chemical, physicochemical and physiological considerations

Dan Davidi[1*], Liam M. Longo[2], Jagoda Jabłońska[2], Ron Milo[1] and Dan S. Tawfik[2*]

Departments of Plant and Environmental Sciences[1], and Biomolecular Sciences[2], Weizmann Institute of Science, Rehovot 76100, Israel.

*Corresponding authors: dan.davidi@weizmann.ac.il; dan.tawfik@weizmann.ac.il

Abstract

Enzymes catalyze a vast range of reactions. Their catalytic performances, mechanisms, global folds and active site architectures are also highly diverse, suggesting that enzymes are shaped by an entire range of physiological demands and evolutionary constraints, as well as by chemical and physicochemical constraints. We have attempted to identify signatures of these shaping demands and constraints. To this end, we describe a bird's-eye view of the enzyme space from two angles: evolution and chemistry. We examine various chemical reaction parameters that may have shaped the catalytic performances and active site architectures of enzymes. We test and weigh these considerations against physiological and evolutionary factors. Although the catalytic properties of the 'average' enzyme correlate with cellular metabolic demands and enzyme expression levels, at the level of individual enzymes, a multitude of physiological demands and constraints, combined with the coincidental nature of evolutionary processes, result in a complex picture. Indeed, neither reaction type (a chemical constraint) nor evolutionary origin alone can explain enzyme rates. Nevertheless, chemical constraints are apparent in the convergence of active site architectures in independently-evolved enzymes, although significant variations within an architecture are common.

Table of contents

1. Introduction

The structures and functions of proteins are determined by physicochemical constraints, physiological considerations, and evolutionary processes. The physicochemical constraints on enzyme catalysis include both physical-chemical ones (*i.e.*, constraints related to the thermodynamics and kinetics of the reaction) and chemical ones (*i.e.*, constraints regarding the mechanism of catalysis and active site architecture). By and large, these physicochemical constraints are reasonably well-understood. In contrast, a detailed understanding of evolutionary processes, which reflect physiological necessity as well as chance [1], has been more elusive, in part because physiological and evolutionary constraints are enzyme- and organism-specific.

Here we attempt to identify and quantify various chemical constraints that shape the performances, structures, and mechanisms of the entire repertoire of known enzymes (the 'enzyme space'). These chemical constraints were quantitatively compared to various physiological demands and evolutionary constraints. Our analysis is primarily driven by the observation that – although some enzymes represent a hallmark of kinetic performance, with impressively high catalytic efficiencies and selectivities – the majority of enzymes are far from being highly efficient and/or highly selective catalysts [2]. A chemist may wonder whether the rate of all reactions can be accelerated to the same level, or if there are some fundamental chemical constraints that result in reaction- or mechanism-specific rate limits. From an evolutionary perspective, on the other hand, this 'mediocrity' may reflect limited physiological demand; that is, no fitness benefit for higher enzyme performance. Indeed, the fluxes of metabolic reactions catalyzed by enzymes differ widely [3-4] and demand for high catalytic efficiency should vary accordingly. Evolutionary constraints may also dictate how catalytically efficient an enzyme might become, such as limited population sizes that slow evolutionary adaptations and lead to the persistence of deleterious mutations, or tradeoffs with other functional traits (*e.g.*, selectivity or regulation)[5-6] as discussed further below.

Similar questions arise with respect to active site architecture. The composition of catalytic residues, as well as their distances and angles, relate to the catalytic mechanism and the transition state of the reaction. Accordingly, we note that for a given catalytic mechanism, active site architectures can be nearly identical – even if these architectures are realized in completely different structural contexts (protein folds) [7-9]. Active site convergence, however, need not imply a dearth of catalytic solutions, since coincidence also plays a critical role in shaping the structural and mechanistic landscape of enzyme space. Furthermore, as discussed in Section 7, the same catalytic mechanism is often realized by alternative active site configurations.

In summary, this review brings together a collection of different data with the aim of demonstrating how chemistry dictates certain kinetic and structural properties of enzymes. General principles can be drawn and are outlined here. However, a multitude of physiological demands, combined with the tinkering, coincidental nature of evolutionary processes, result in complex trends observed across the huge variety of kinetic parameters, mechanisms, and structures of natural enzymes.

2. A bird's-eye view of enzyme kinetic parameters

To address the above questions, and obtain a bird's-eye view of contemporary enzymes with respect to kinetic parameters, we have updated our previous analysis of the BRENDA database [10] (**Figure 1**), which now includes over twice as many non-redundant enzyme parameters as before [2]. The BRENDA database provides the two key metrics of enzyme kinetic performance: k_{cat}, a first-order rate constant referred to as the turnover number that reflects the reaction rate under saturating substrate concentrations; and, k_{cat}/K_M, a second-order rate constant referred to as the catalytic efficiency. In the sections below we relate kinetic performance to physiological demands – *i.e.*, to the flux of product molecules per unit time that an enzyme can provide within a living cell. In addition, we have extended our analysis to include evolutionary and structural assignments in an effort to better understand both chemical and evolutionary constraints (Sections 4 to 5 below).

There are obvious caveats associated with a global analysis of kinetic parameters. There is inevitable and considerable variability and noise due to the assays being performed in different laboratories and under different conditions (*e.g.*, pH, temperature, buffers, and cofactor concentrations) as well as inconsistencies between BRENDA and the original reports. However, the general trends seem relevant despite the noise and variability (for a detailed discussion see [2]).

A second caveat associated with the analysis of enzyme rates is that all known parameters were measured *in vitro* and possibly under suboptimal conditions, not just with respect to buffer, ionic and crowding strength, but also unknown allosteric regulators and other activators. Do the *in vitro* measured k_{cat} values represent the rates of enzymes *in vivo*? Using proteomics and metabolomics data (cellular enzyme levels and flux rates) the apparent *in vivo* k_{cat} value can be derived. The correlation between the *in vitro* measured k_{cat} values is relatively high (R^2=0.6 with a root mean square deviation of 0.54 on a log-log plot; ~3.5-fold in linear scale), and accounting for the substrate saturation level gave an even higher correlation (R^2=0.9) [11]. Thus, by and large, it appears that painstakingly collected *in vitro* data, by

thousands of enzymologists throughout the entire 20th century and the beginning of the 21st, do reflect the *in vivo* enzymatic rates.

A third caveat regards bias in enzyme sampling. The enzymes studied so far do not comprise a random sample, let alone an even sample, of the entire enzyme repertoire. For example, core metabolism of fast-growing microbes such as *E. coli* or yeast is overrepresented while specialized metabolism is underrepresented. Enzymes belonging to the former class are significantly faster than the latter (median k_{cat} values ≈30-fold higher). As a wider range of organisms is being explored, core metabolic enzymes with unusually slow rates are also discovered (e.g., [12-13]). The decrease in the median k_{cat} and k_{cat}/K_M values between the previous (ca. 2011; Ref. [2]) and the current analyses may reflect a wider exploration, but the sampling bias is still substantial. It is likely, therefore, that the "average enzyme" is considerably less efficient than the average BRENDA enzyme.

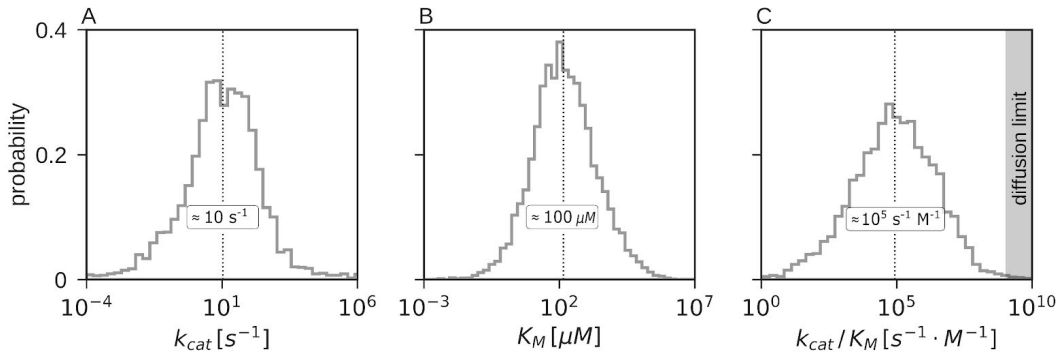

Figure 1: **The distribution of enzymatic kinetic parameters**. Data from BRENDA [10] (ca. April 2017) was filtered for wild-type enzymes catalyzing naturally-occurring reactions (based on the KEGG database [14]) and analyzed as previously described [2]. (**A**) The distribution of k_{cat} values (N = 3,500) has a median rate of 10 s^{-1} (as indicated by the dashed line; previous median 14 s^{-1}; N = 1,942). (**B**) The distribution of K_M values (N = 12,576) has a median K_M value of 140 μM (previously reported median was 130 μM, N = 5,194). (**C**) The distribution of k_{cat}/K_M values (N = 4,451) has a median value of 8 x 10^4 s^{-1} M^{-1} (previous median 13 x 10^4 s^{-1} M^{-1}; N = 1,882). The diffusion rate limit is shaded in gray. A list of all enzymes included in our analysis and their BRENDA reported kinetic parameters is provided in **Supplementary Table S1**.

With these caveats in mind, one can still deduce some interesting conclusions. For example, as discussed in the next Section, "the average enzyme" is far less catalytically efficient than the textbook superstars, perfect enzymes. Rather, the average kinetic parameters seem to

reflect the average metabolic demand. Overall, K_M values and cellular metabolite concentration are correlated [4, 15], supporting the notion that K_M values largely evolved to match the substrates' physiological concentrations. That said, K_M values (**Figure 1B**) and metabolite concentrations are both widely-distributed, and most enzymes seems to operate close to saturation [4, 15].

Does the median turnover number (k_{cat}) value have any meaning? In principle, an enzyme's k_{cat} corresponds to the reaction's rate in the living cell (flux) divided by the enzyme's concentration [11, 16]. As discussed in Section 6 below, even within one organism, say *E. coli*, reaction fluxes and enzyme concentrations vary over several orders of magnitude. However, interestingly, considering *E. coli*'s median reaction flux, and its median enzyme concentration, the approximated *in vivo* k_{cat} value (**Box 1**; ~7 s^{-1}) roughly matches the observed median k_{cat}, 10 s^{-1} (**Figure 1A**; the median k_{cat}, for *E. coli* enzymes, is similar, 11 s^{-1}). This back-of-the-envelope deduction should, of course, be taken with more than a pinch of salt – the median values we applied represent wide, log-normal distributions, and *E. coli* does not represent the typical organism (see the sampling bias caveat above). Nonetheless, Box 1 illustrates the notion that k_{cat} values and cellular enzyme concentrations have largely evolved to meet metabolic flux demands. At a minimum, while avoiding Panglossian reasoning, it appears that k_{cat} values, cellular enzyme concentrations, and metabolic flux rates have coevolved to maximize *E. coli*'s fitness.

Box 1: Averages make sense – the median metabolic reaction rate and cellular enzyme concentration match the median k_{cat}

Median turnover rate of *E. coli*'s enzymatic reactions (V_{max}):

$$V \approx \frac{\text{median reaction flux in } E. coli \text{ (1)} \times \text{dry weight of an } E. coli \text{ cell (2)}}{\text{Volume of an } E. coli \text{ cell (3)}} = \frac{0.014 \text{ mmol/gCDW/hr} \times (3 \times 10^{-13} \text{ gCDW/cell})}{(3600 \text{ s/hr}) \times (10^{-15} \text{ L/cell})} = 1.2 \times 10^{-3} \text{ mmol/L/s} = 1.2 \times 10^{-6} \text{ M/s}$$

(4)
$V_{max} \approx 1.5 \times V = 1.8 \times 10^{-6}$ M/s

$$k_{cat} \approx \frac{V_{max}}{E} \approx \frac{1.8 \times 10^{-6} \text{ M/s}}{0.25 \times 10^{-6} \text{ M}} \approx 7 \text{ s}^{-1}$$

Median enzyme concentration in *E. coli* (E):

$$E \approx \frac{\text{median enzyme abundance in } E. coli \text{ (4)}}{(6.02 \times 10^{23} \text{ copies/mol}) \times \text{Volume of an } E. coli \text{ cell (3)}} = \frac{150 \text{ copies/cell}}{(6.02 \times 10^{23} \text{ copies/mol}) \times (10^{-15} \text{ L/cell})} = 0.25 \times 10^{-6} \text{ M}$$

(1) The median flux across 7 conditions was calculated per reaction as described in Fig. 4A, and the median for all *E. coli* reactions was used here. (2) BioNumbers (http://bionumbers.hms.harvard.edu/) ID: 109837; (3) Ibid., ID: 101788; (4) The median metabolite concentration in *E. coli* (0.2 mM; Ref. 4, 15) is roughly twice the median K_M, and thus, the median V_{max} would be 1.5-fold the median metabolic rate ($V_{max} = V[S/(K_M + S)]$; see also text); (5) The median enzyme abundance across 7 conditions was calculated per enzyme as described in Fig. 4C below, and the median for all *E. coli* enzymes was used here.

3. The diffusion rate limit – a dominant physicochemical constraint?

Among the most well-known physicochemical constraints acting on enzymatic rates is the diffusion rate limit – the point at which collisional frequency alone dictates the rate of substrate conversion into product. But to what degree has this 'kinetic ceiling' shaped the performance of enzymes?

The distribution of k_{cat}/K_M values (**Figure 1C**) suggests a negligible influence of the diffusion limit. True, no known enzyme meaningfully exceeds the diffusion rate limit – consistent with a fundamental physical limitation. However, there is no indication that evolution has 'pushed' enzymatic rates towards the diffusion-limit: (i) The 'average' enzyme exhibits a k_{cat}/K_M value nearly 4 orders of magnitude lower than that predicted for diffusion-limited reactions (≥ 10^9 s^{-1} M^{-1} for low molecular weight substrates; see Ref. [17]). (ii) The log distribution of k_{cat}/K_M values is strikingly symmetric, rather than having a skewed distribution, which would be expected if evolution has pushed enzymes for high efficiency.

Despite their rarity, those enzymes that do approach the diffusion rate limit (so-called 'perfect' enzymes) are important to our understanding of enzymes in general. For example, a perfect

enzyme, by definition, makes no futile encounters with its substrate. By comparison, only 1 out of 10^4 encounters of an 'average' enzyme with its substrate are productive, with the overwhelming majority being futile (resulting in substrate dissociation). The dominance of futile encounters stems from several factors, including conformational heterogeneity [17]. Both enzymes and substrates adopt multiple conformational sub-states, only a subset of which is catalytically competent. Consistent with this interpretation, many of the known perfect enzymes have notably simple substrates, such as carbon dioxide or dihydroxyacetone phosphate. Conversely, minimizing nonproductive conformations, especially in enzymes, is among the chief physicochemical hurdles that evolution must overcome [18-22].

The symmetric distribution of kinetic parameters suggests that across the entire enzyme repertoire there is no strong, consistent evolutionary drive for the emergence of very fast enzymes. Indeed, it seems that perfect enzymes evolved to meet unique physiological needs (Section 6). On the other hand, it may well be that other chemical constraints, dictated, for example, by the nature of the catalyzed reaction, severely limit the rates of many enzymes, thus resulting in the 'average' enzyme being far slower than the diffusion rate limit. This possibility is addressed in Sections 4 and 5 below, with the short answer being no.

4. The relationship between spontaneous and enzymatic reaction rates

The spontaneous rate of enzyme-catalyzed reactions – that is, the reaction rate without the enzyme (at ambient temperature and physiological pH) – span more than 15 orders of magnitude [23], a spread of greater than 90 kJ/mol in activation energy. In contrast, the first-order and second-order rate constants of enzyme-catalyzed reactions span only approximately six orders of magnitude – an impressive compression of rates by over 10^{10}-fold [24]. However, it is unclear if reactions with fast spontaneous rates tend to have fast catalyzed rates as well. Or, have enzymes been primarily shaped by physiological demands such that their catalytic efficiencies bear no correlation to the spontaneous rates of the reactions they catalyze?

To systematically examine this question, we compared the catalytic efficiencies across various hydrolysis reactions. A variety of biochemical bonds are cleaved via hydrolysis, with large differences in inherent reactivity. Using data derived largely from Richard Wolfenden's papers on the spontaneous rates of enzymatic reactions [23-24], and extracting enzymatic k_{cat} values for all enzymes that catalyze these reaction classes from BRENDA, we generated **Figure 2**. As can be seen, the k_{cat} values of hydrolytic enzymes are widely distributed, yet neither the median k_{cat} values, nor the maximum k_{cat} values, within a given reaction class

show a clear relation to typical spontaneous reaction rates (denoted as k_{non} values; **Figure 2**; Refs. [23-24]).

Overall, the median values of k_{cat} are remarkably consistent between different hydrolytic reaction classes, as well as with the full set of enzymes (**Figure 1**). Consequently, evolution has accelerated the rate of phosphodiester hydrolysis by approximately 12 orders of magnitude more than that of lactam hydrolysis. The only class that shows significantly higher k_{cat} values is CO_2 hydration, which is catalyzed by carbonic anhydrases, and has a median k_{cat} of 2 x 10^5 s^{-1}. Although the high spontaneous rate of this reaction may have enabled exceptionally high k_{cat} values, the 'physiological demand' explanation seems more likely: the concentration of CO_2 in most growth media is low, and CO_2 rapidly diffuses through membranes. Thus, to maintain sufficiently high cellular carbonate concentrations, carbonic anhydrases must exhibit exceptionally high rates [25]. Acting in the opposite direction (conversion of carbonate to CO_2), carbonic anhydrases must be fast enough to compete against leakage of CO_2 through membranes, and to enable coupling to enzymes that utilize CO_2.

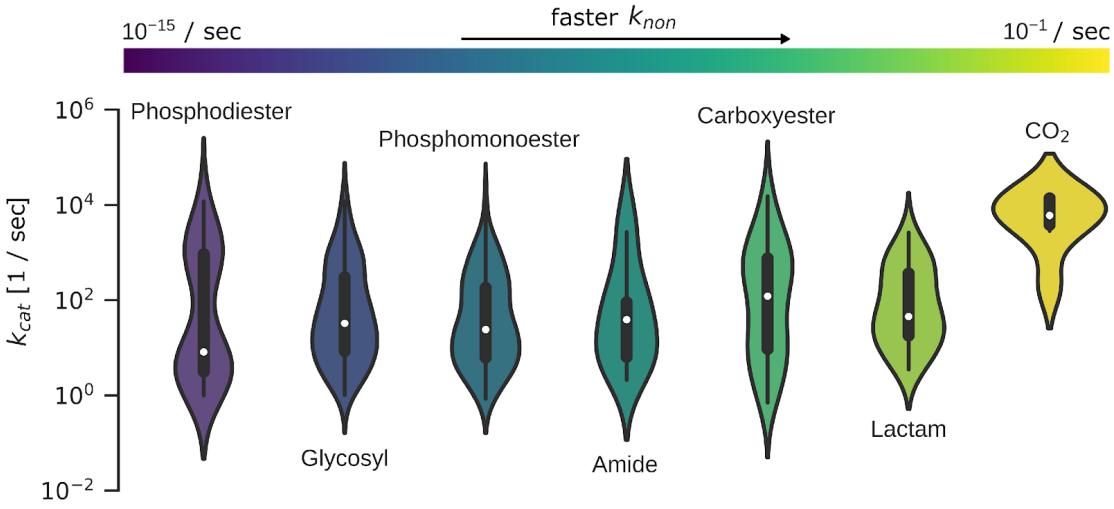

Figure 2: Enzyme rates seem to have evolved with no correlation to how demanding a reaction is. Categories along the x-axis denote different types of hydrolytic reactions, approximately ordered from the most demanding to the most facile in terms activation energy, as indicated by their characteristic spontaneous rates (k_{non} values; extracted mostly from Refs. [23-24]; number of enzymes per category are: CO_2: 7, Lactam: 29, Carboxyester: 39, Amide: 13, Phosphomonoester: 117, Glycosyl: 87, Phosphodiester: 29; for each category, highest and lowest 10% k_{cat} values were removed to exclude outliers but their inclusion does not change the

observed trend). For each reaction class, the distribution of k_{cat} values is rendered as a violin plot; median value (white dot) and interquartile range (thick black line) are shown for each group. Note that CO_2 hydration, catalyzed by carbonic anhydrase (EC 4.2.1.1), is the only group that significantly differs from all other groups (p-value < 10^{-5} by Wilcoxon rank-sum test). A list of the enzymes included in the 7 hydrolytic categories and their kinetic parameters is provided in **Supplementary Table S2**.

Overall it appears that the activation barrier of the reaction itself (in the absence of an enzyme) does not generally constrain enzymatic rates. However, what we do not know is whether, given sufficient physiological incentive, enzymes belonging to other classes could evolve k_{cat} values that are as high as carbonic anhydrases (>10^5 s^{-1}). Further, some reactions are so demanding (in activation energy terms and/or mechanistic complexity) that such rapid turnover rates seem unrealistic; for example, the highest reported k_{cat} value for methane oxidase is 4 s^{-1}, just to compare another enzyme with a gaseous substrate.

5. Catalytic mechanisms and active site architectures do not seem to constrain enzymatic rates

The variability in rates within individual reaction classes is high (**Figure 2**) and in fact, given the different sample sizes, it might be as high as the variability across all enzymes (**Figure 1**). This variability may stem from physiological demands and evolutionary constraints, but may also relate to chemical constraints associated with certain mechanisms of catalysis and/or active site architectures. For many catalyzed reactions, there appear to be multiple catalytic solutions, *i.e.*, fundamentally different mechanisms and corresponding active site architectures. So, do different catalytic solutions dictate different rates? To answer this question, enzymes catalyzing the same reaction were classified as analogs or homologs and their catalytic properties were compared.

Analogous enzymes are enzymes that catalyze the same reaction yet are evolutionary unrelated. Accordingly, they have a different overall structure (fold) and may also apply a different catalytic mechanism that is, in turn, achieved via a completely different active site architecture (*e.g.*, a serine proteinase *versus* a metallo-proteinase). By a conservative estimate, ≥40% of enzyme functions arose independently in more than one fold (estimated with function defined by 4 EC digits [7]; if identical 3 digits are considered as same function, as applied here, the frequency of multiple independent emergences would be about 2-fold higher). Analogs may converge on (*i.e.*, independently evolve) the same mechanism, and even have nearly identical active site architecture as discussed in Section 7 below. However, the likelihood of a shared mechanism is much lower than in homologs (enzymes that diverged from a common ancestor). Indeed, enzymes that catalyze the same reaction are often

evolutionarily related as orthologues (the same enzyme from different organisms) or as paralogues (isoenzymes within the same organism), jointly addressed here as homologs. Homologs, tend to share the same fold, catalytic mechanism, and active site architecture.

If homologous enzymes tend to have more similar k_{cat} values than analogues that catalyze the same reaction, it suggests that evolutionary origin, and by extension, a protein's fold and catalytic mechanism, may have a role in shaping enzyme rates. Conversely, if enzymes that are analogs, and are therefore evolutionarily and structurally unrelated, tend to have k_{cat} values about as similar as homologs, then evolutionary origin, mechanism, and active site architecture have a relatively minor role in shaping enzyme rates.

To examine whether such trends exist, we performed a pair-wise comparison of homologous and analogues enzymes for the entire collection of enzymes in BRENDA (**Figure 3**). A comparison of pairs of homologs to pairs of analogs catalyzing the same reaction shows a mere 1.5-fold average difference in rates for a given reaction (**Figure 3A**). The rate of homologs catalyzing the same reaction compared to homologs catalyzing different reactions shows a larger difference (**Figure 3B**). Innovations in enzymatic functions occur readily [26]. Most changes regard substrate specificity or relatively small changes in the reaction (*e.g.*, amidase to esterase); but profound changes in the reaction (*i.e.*, switches between reaction classes; 1st EC digit) have also occurred throughout evolution [7, 26-27]. However, overall, changes in the catalyzed reaction seem to have a modest effect: a maximum of about a threefold difference compared to the k_{cat} values that range over six orders of magnitude.

As mentioned above, a caveat associated with our analysis is that, due to convergence, evolutionarily unrelated enzymes that catalyze the same reaction (analogous enzymes) do not necessarily employ different catalytic mechanisms. However, the likelihood of different mechanisms is obviously higher in analogues than in evolutionarily related enzymes (homologs). Further, when all enzymes are compared (regardless of whether they catalyze the same reaction or different ones), the cumulative difference is twofold (analysis not shown), *i.e.*, only slightly higher than for enzymes catalyzing the same reaction (~1.5-fold; **Figure 3A**). Enzymes catalyzing different reactions are even more likely to have different mechanisms, further suggesting that evolutionary origin (including the enzyme's fold and active site architecture) and the catalytic mechanism play a relatively minor role in shaping enzyme rates.

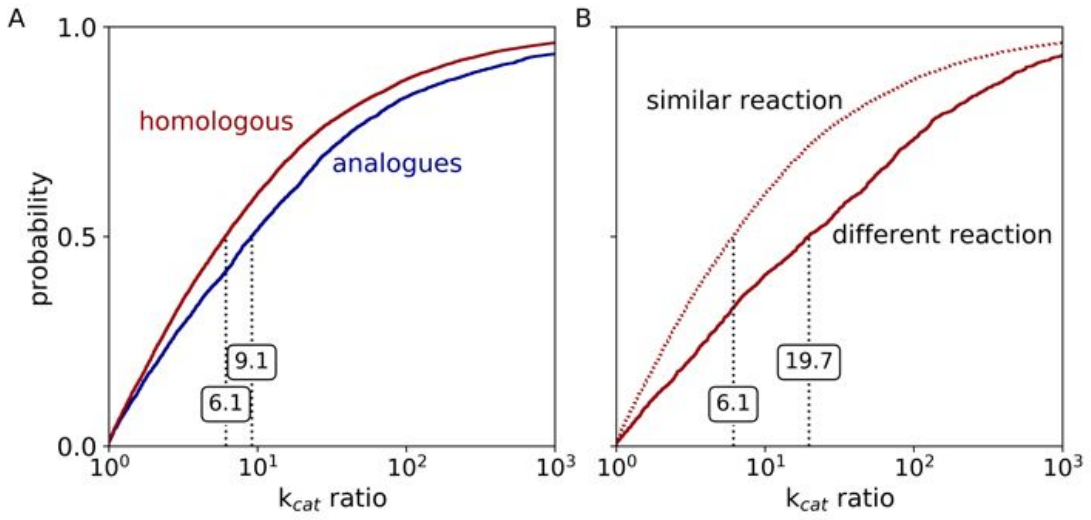

Figure 3: Do different evolutionary origins or different reaction types correlate with k_{cat} value differences? (**A**) Cumulative distributions of k_{cat} ratios (faster enzyme divided by slower one, per each pair) for all pairs of enzymes catalyzing the same reaction (*i.e.*, reactions sharing the first 3 EC digits; derived from all enzymes reported in the BRENDA database; N=2024 pairs) for evolutionary-related enzyme (homologous) versus unrelated enzyme pairs (analogs; N=1269 pairs). An e-value of 10^{-10} was used as the threshold to define sequence homology. The median values, indicated as dashed lines, are 1.5-fold smaller for homologs compared to analogs (p-value < 10^{-10}, Wilcoxon rank-sum test). (**B**) Cumulative distribution of k_{cat} ratios for homologous enzymes that catalyze similar reactions (*i.e.*, reactions sharing the first 3 EC digits; N=1557 pairs) versus homologous enzyme pairs that catalyze different reactions (N=467 pairs); the median k_{cat} ratio for enzymes catalyzing similar reactions is about threefold smaller compared to enzymes that catalyze different reactions (p-value < 10^{-10}). A list of all possible pairs of BRENDA reported enzymes and their categorization as homologs or analogs is provided in **Supplementary Table 3**.

6. Physiological demands and enzyme rates

The sections above suggest that physicochemical constraints play a relatively minor role in shaping the kinetic parameters of enzymes. Might physiological demands be the dominant factor? Can we detect the footprints of such demands in a global analysis of the enzyme repertoire?

In principle, the evolution of enzymatic turnover rates is expected to be driven by the need to maintain an appropriate metabolic flux, *i.e.*, sufficient conversion of substrate to product per unit time, thus maximizing organismal fitness. The apparent cellular k_{cat} values (derived from dividing flux rates by cellular enzyme levels) largely correlate with *in vitro* measured k_{cat} values [11]. Further, the median metabolite concentration in *E. coli* (~0.2 mM) corresponds to the median K_M value, suggesting that, on average, enzymes act *in vivo* at $^2/_3 V_{max}$. Accordingly, the median flux rate in *E. coli* divided by the median cellular enzyme concentration roughly corresponds to the median k_{cat} value (**Box 1**). So can metabolic flux explain the variability in enzymatic rate parameters? In other words, can we find support for the intuitive hypothesis that evolution has shaped the k_{cat} values of individual enzymes to meet reaction rates in the living cell? Apparently not, as the overall correlation between the flux of a given reaction and the k_{cat} of the corresponding enzyme is weak [11, 28]. Indeed, only when flux is normalized by enzyme concentration does a significant correlation with k_{cat} emerge [11].

Moreover, the correlation is weak not only when k_{cat} values of the *E. coli* enzyme are compared [11, 29] but also, as shown here, when a normalized k_{cat} value is considered, denoted here as k_{cat}^{Rel}. k_{cat}^{Rel} was calculated by dividing the k_{cat} value of each *E. coli* enzyme by the highest k_{cat} observed within the same EC class (in any organism, and considering the first three EC digits, and removing highest and lowest 10% values, as in **Figure 2**). The normalized k_{cat}^{Rel} value is therefore a more conservative estimate for the extent to which the turnover rate of a given enzyme has been optimized. This normalization also alleviates a potential bias due to differences in the nature of the reactions catalyzed by these enzymes (although this bias is minor; **Figure 2).** Notably, only about 20% of the enzymes in *E. coli* show a $k_{cat}^{Rel} > 0.5$, indicating that for 80% of the enzymes in *E. coli* there exists at least a twofold faster orthologue. Further, the median k_{cat}^{Rel} for the *E. coli* enzymes analyzed here is ≈0.05 – suggesting that the average *E. coli* enzyme could be replaced with an enzyme that catalyzes the same reaction (an orthologue and/or an analogue) with a 20-times faster turnover number.

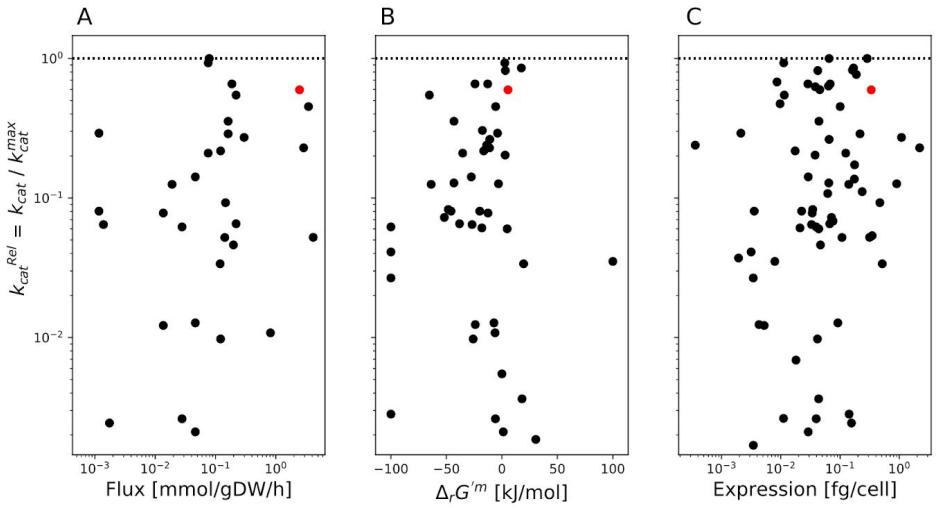

Figure 4: Physiological demands do not correlate with enzyme turnover numbers. Scatter plots showing the k_{cat}^{Rel} values of individual enzymes (k_{cat} of the *E. coli* enzyme divided by the highest k_{cat} observed in the same EC class across all organisms, k_{cat}^{max}) as a function of three different parameters. (**A**) The enzyme-mediated cellular flux weakly correlates with its catalytic efficiency (k_{cat}^{Rel}; Spearman $R^2 = 0.24$, p-value$<10^{-4}$). Flux rates (mmol product per gram dry cell weight per hr) for *E. coli* were derived from metabolic flux analysis measurements in *E. coli* [11, 30] and expanded to the entire metabolic network of *E. coli* using Flux Balance Analysis (FBA; [31-32]). Values used here represent the median flux at exponential growth under seven different carbon sources. (**B**) The reaction's standard free energy does not correlate with k_{cat}^{Rel} (Spearman $R^2 \approx 0$). Plotted values are $\Delta_r G'^m$ (free energy at 1 mM reactants and products, at pH 7.5 and ionic strength of 0.1 M in kJ/mol; obtained from [33-34]). (**C**) Enzyme concentration and k_{cat}^{Rel} are poorly correlated (Spearman R^2=0.19, p-value$<10^{-3}$). Enzyme masses are given in femto (10^{-15}) grams per cell [35]; Similar to the flux values, enzyme expression levels are the median across seven different conditions. The red point designates TIM – triosephosphate isomerase, a "perfect enzyme" [36]. This analysis was done with a curated set of *E. coli* metabolic reactions with which one and only one enzyme (EC number and gene; N=91) is associated (a list of the enzymes included in this analysis and their kinetic and cellular parameters is provided in **Supplementary Table 4**). The complete set of reactions gave a similar distribution yet with even poorer correlation.

The flux through a given metabolic reaction depends, however, not only to the turnover rate of the corresponding enzyme, but also to the thermodynamics of the reaction itself. That is, how far the substrate and product concentrations are from equilibrium. Many, if not most,

metabolic reactions are reversible – in *E. coli*, for example, under normal physiological conditions and metabolite levels, it is thought that two-thirds of the reactions are reversible [37]. As catalysts, enzymes do not change the equilibrium concentrations of products and reactants; they therefore enhance the rates of both the forward and reverse reactions to the very same degree (the Haldane relationships). For reactions that reside near equilibrium, the forward and reverse rates are similar and thus the net forward flux becomes small. As such, the catalytic efficiency (k_{cat}/K_M and/or k_{cat}) and/or enzyme levels required to achieve a significant net forward flux dramatically increase for physiological reactions that occur near equilibrium [38-39]. A classical example for this logic may be manifested in triosephosphate isomerase (TIM) – Knowles' original example of the "perfect enzyme" [36].

TIM is a key enzyme in glycolysis with a typical k_{cat}/K_M of $>10^7$ s^{-1} M^{-1}, and k_{cat}/K_M of $\sim10^{10}$ s^{-1} M^{-1} in some thermophilic orthologues [40]. The drive for TIM's perfection was proposed to be an intense evolutionary pressure, "since for "flight or fight" there is an instant requirement for muscle ATP" [36]. However, other enzymes in the same pathway are notably less efficient. For the enzyme upstream of TIM, fructose bisphosphate aldolase, the highest k_{cat}/K_M values found in BRENDA are on the order of 10^3 $s^{-1}M^{-1}$ – seven orders of magnitude lower than TIM. Similarly, for the enzyme downstream of TIM, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), k_{cat}/K_M is on the order of 10^5 s^{-1} M^{-1}. Why are the k_{cat}/K_M value of TIM so high compared to of its two flanking enzymes? Perhaps it is because TIM catalyzes a reaction that is close to equilibrium (**Figure 4B**). One might hypothesize then that TIM, and other enzymes acting near equilibrium, may be under the highest selection pressure and consequently have evolved toward near-diffusion-limited kinetics (Krebs first proposed the concept of equilibrium enzymes, although in the reverse logic: some enzymes have high rates and levels such that they push the reaction to equilibrium [41-42]). Following this rationale, the selection pressure on the two flanking enzymes to evolve ultra-fast kinetics would be comparatively weak, as they catalyze reactions further from equilibrium, and hence do not comprise flux bottlenecks. Is there consistent evidence for unfavorable reaction thermodynamics dictating high catalytic efficiency?

Following the TIM example (the red dots in **Figure 4**), we examined whether enzymes that support reactions in which the substrate and product concentrations are close to equilibrium tend to have higher catalytic efficiency. However, $\Delta_r G'^m$ (the reaction's free energy at 1 mM reactants and products; $\Delta_r G'^m$ close to zero indicates a reaction near equilibrium) and k_{cat}^{Rel}, do not seem to be anti-correlated as expected (**Figure 4B**). The lack of correlation may relate to the fact that the actual, cellular concentrations of reactants and products are far from 1 mM and may thus result in favorable free energy. However, measurements of the actual concentrations indicate that TIM's reaction, and many other glycolytic reactions, are

near-equilibrium [4]. Nonetheless, most of the enzymes catalyzing these steps exhibit relatively slow turnover rates, as also indicated in **Figure 4B**. Overall, it appears that near-equilibrium thermodynamics are not a global shaping force of turnover rate, and even TIM's near-perfect rate may be unrelated to the catalyzed reaction being near equilibrium. In general, the physiological demands that have shaped perfect or near-perfect enzymes are visible, as is the case with carbonic anhydrase described above, but only in exceptional cases.

Finally, in the living cell, the rate of an enzymatic reaction is proportional to the specific activity (k_{cat}/K_M or k_{cat}) of the enzyme as well as the enzyme's concentration. The latter comprises another physiological constraint, given that both protein synthesis (making amino acids and polymerizing them) and protein maintenance (chaperones, undesirable interactions with other proteins, etc.) come at a cost [43-44]. It is therefore expected that the most highly expressed enzymes have also experienced the strongest selection pressure for maximizing catalytic efficiency. However, such a trend is not seen, at least not when *E. coli* is examined (reliable data, including absolute genome-wide protein levels [35], are available for *E. coli*; **Figure 4C**). Indeed, the reverse argument, and hence an anti-correlation, is equally valid: namely, evolution toward fast rate enables a reduction in expression levels (the same may apply to the free energy consideration; i.e., Krebs' reverse argument regarding enzymes that mediate near-equilibrium reactions).

Overall, while the overall trends and averages make sense (**Box 1**), at the level of the individual enzymes, the effect of the physiological demands examined here (high flux, low $\Delta G^{0'}$, and protein cost) seems to be largely masked by other physiological and/or evolutionary factors. What might these other factors be? Apart from chance, an evolutionary factor that is routinely underestimated, we should keep in mind that rate (k_{cat}/K_M or k_{cat}) is not the only, and possibly not even the primary, enzymatic trait under selection [26, 45]. Selectivity (or accuracy) and regulation are also under selection, and both of these traits often trade off with turnover rate (k_{cat}) and/or catalytic efficiency (k_{cat}/K_M) [5, 46]. For enzymes evolving under such tradeoffs, k_{cat} would be anti-correlated with enzyme levels, as exemplified by Rubisco (Ribulose bisphosphate carboxylase/oxygenase; EC 4.1.1.39) whose rate trades off with CO_2/O_2 selectivity and whose cellular levels tend to be exceedingly high [47]. Tradeoffs between tight regulation and high rate have also been observed (*e.g.*, Ref. [48]). Thus, tradeoffs may explain, at least in part, the lack of overall correlation in **Figure 4C**. Another explanation regards a trait that is potentially under selection: secondary, moonlighting functions. Many metabolic enzymes, for example, also act as transcriptional regulators [49-51] and the latter role may shape their cellular levels. Additionally, most metabolic enzymes are part of complexes, ranging from tight associations with well-defined stoichiometry to transient complexes and 'metabolons' [52-53]

Overall, it appears that many enzymes operate below their catalytic capacity, namely, they are generally expressed at levels higher than needed to support the flux of the reaction they catalyze [28, 45, 54]. Further, it appears that for a given organism many, if not most, enzymes could readily evolve toward higher k_{cat} values (see the k_{cat}^{Rel} analysis above). However, although higher catalytic efficiency may allow reduced cost owing to lower enzyme levels, it might tradeoff with other critical traits such as regulation or secondary functions, and/or might disturb complex stoichiometry [55]. The hypothesis that many natural enzymes possess a potential to evolve higher catalytic efficiency (k_{cat}/K_M, or k_{cat}) can be readily tested by experimental evolution. Numerous enzymes have been evolved in the laboratory toward higher rates, albeit for alternative, promiscuous substrates and/or for non-native reaction conditions. However, in principle, one could attempt to evolve a natural enzyme toward higher rate with its native substrate and under native-like conditions, and then replace the wild-type with the evolved enzyme to examine the potential benefit (reduced expression levels) or cost (loss of regulation, disturbing interaction partners, etc.).

We conclude that while the chemical and physiological factors addressed above are highly relevant, the global variability in kinetic parameters cannot be rationalized by any single factor. We now turn our focus to structural considerations, in an effort to better understand the extent to which chemical constraints shape the evolution of enzyme active sites.

7. Convergence of active site architectures

Up until now, we have examined physicochemical, physiological, and evolutionary constraints in relation to the thermodynamics and kinetics of enzyme-catalyzed reactions. However, chemical constraints dictate other enzyme traits, foremost the mechanism of catalysis of an enzyme, and in turn, the active site architecture needed to exercise this mechanism. Convergent evolution of chemically similar, if not identical, active sites is a common process [56-58]. For example, three carbonic anhydrase families with completely different folds – but with strikingly similar active site architectures – have been identified (**Figure 5A**). These enzyme families most likely arose via completely independent evolutionary origins, although common ancestry at very early stages (parallel evolution [57-59]) is nearly impossible to rule out. The catalytic triad of serine proteinases provides another striking example of active site convergence – the triads of trypsin-like and subtilisin-like proteases exhibit overlapping geometries despite their independent emergence in two completely different folds (**Figure 5B**; trypsin-like in magenta and subtilisin-like in green).

The convergence of active site configurations that execute similar reactions via the same mechanism suggests that, given 20 amino acids and a finite set of available cofactors (organic or inorganic), by and large, there seems to be one relatively simple, readily-accessible, catalytically-competent active site configuration. That convergence of active site architecture is the outcome of chemical constrains is also manifest in independently emerged active sites possessing mirror-image architectures, *e.g.* in metallo-lactonases [57]. Obviously, when a cofactor is the key catalytic element, as is the case with PLP (pyridoxal-5'-phosphate) enzymes or with metallo-enzymes, similar active sites can evolve time and again while converging on the same chemistry. For example, PLP-dependent enzymes have emerged along at least 4 independent lineages [60-61]. In the case of PLP, emergence may be seeded by a single mutation that provides a lysine side-chain that covalently anchors the cofactor [62]. However, in metallo-enzymes, convergence entails multiple residues – four residues in the case of mono-metal sites (*e.g.*, carbonic anhydrases; **Figure 5A**) and even more in the case of bi-metallo enzymes. Nonetheless, bi-metallo active sites have been observed in quorum quenching lactonases that emerged in two different folds, whereby the geometry of as many as 7 residues exhibits mirror-image symmetry [57]. Similarly, a newly identified phosphotriesterase with a β-propeller fold exhibits a nearly identical active site architecture to a previously discovered phosphotriesterase that has a TIM-barrel fold [63].

However, while active site convergence is a dominant phenomenon, it is also clear that active site divergence is as common. Multiple alternative mechanisms typically exist for a given reaction type. For example, for amide hydrolysis (including peptide bond hydrolysis), three completely different mechanisms are known: nucleophilic catalysis (with serine, cysteine, or threonine as the nucleophile); metal catalysis (with either mono- or bi-metal centers); and acid-base catalysis (aspartyl proteinases). In addition, there exist variations within the same mechanism, and accordingly, independently evolved enzyme superfamilies exhibit overlapping yet nonetheless alternative active site architectures. Serine, cysteine, or threonine serves as the nucleophile in various amide hydrolases. There are also variations in the auxiliary residues, *e.g.*, dyads *versus* triads, or lysine replacing histidine as the base (**Figure 5B**) [64].

Variations on the theme also occur in possibly the most striking example of active site convergence – glycosyl hydrolases. In the case of the glycosyl bond hydrolysis, the transition state carries a partial positive charge, thereby excluding catalysis by metal cations. Even when covalent catalysis is applied (*e.g.*, in retaining glycosylases), the nucleophile for glycosyl bond hydrolysis seems to be limited to carboxylate side-chains (*i.e.*, Glu or Asp; **Figure 5C**). Curiously, the same side-chain is preferentially used to activate the water

nucleophile in inverting glycosyl hydrolases that do not act via a covalent intermediate. Overall, Glu/Asp dyads are present in the vast majority of glycosyl hydrolases, including in families that show no detectable sequence identity or even possess a different fold, and are therefore unlikely to have diverged from one another [65] (The CAZy website (http://www.cazy.org/Glycoside-Hydrolases.html) currently lists 17 independent glycoside hydrolases clans that represent 7 completely different folds). However, even within retaining glycosylases, the distance between the carboxylate side-chains of Glu/Asp dyads is more variable than traditionally assumed [65]), and the relative positing of the two carboxylates widely varies (**Figure 5C**). Further, variations is in the second Asp/Glu that acts to protonate the leaving group are also known, foremost Asp-His dyad glycosidases ([66]; for another rare variation, see Ref. [67]). Finally, variations in active-site configurations are observed even within the same superfamily. Thus, despite the same reaction mechanism, and despite shared evolutionary origin, alternative catalytic configurations have evolved. The Aldolase Class I superfamily is just one good example for such divergence (see Figure 8 in [7]).

Overall, it appears that chemical constraints limit the space of evolutionary solutions. However, for the majority of enzyme classes, multiple mechanisms, and accordingly, multiple active site architectures, have emerged independently. Further, even within the same mechanism, there exist variations on the theme that indicate multiple alternative catalytic configurations. This chemical diversity suggests that despite chemical constrains, *de novo* emergences of enzymatic active sites have occurred repetitively and throughout evolutionary time.

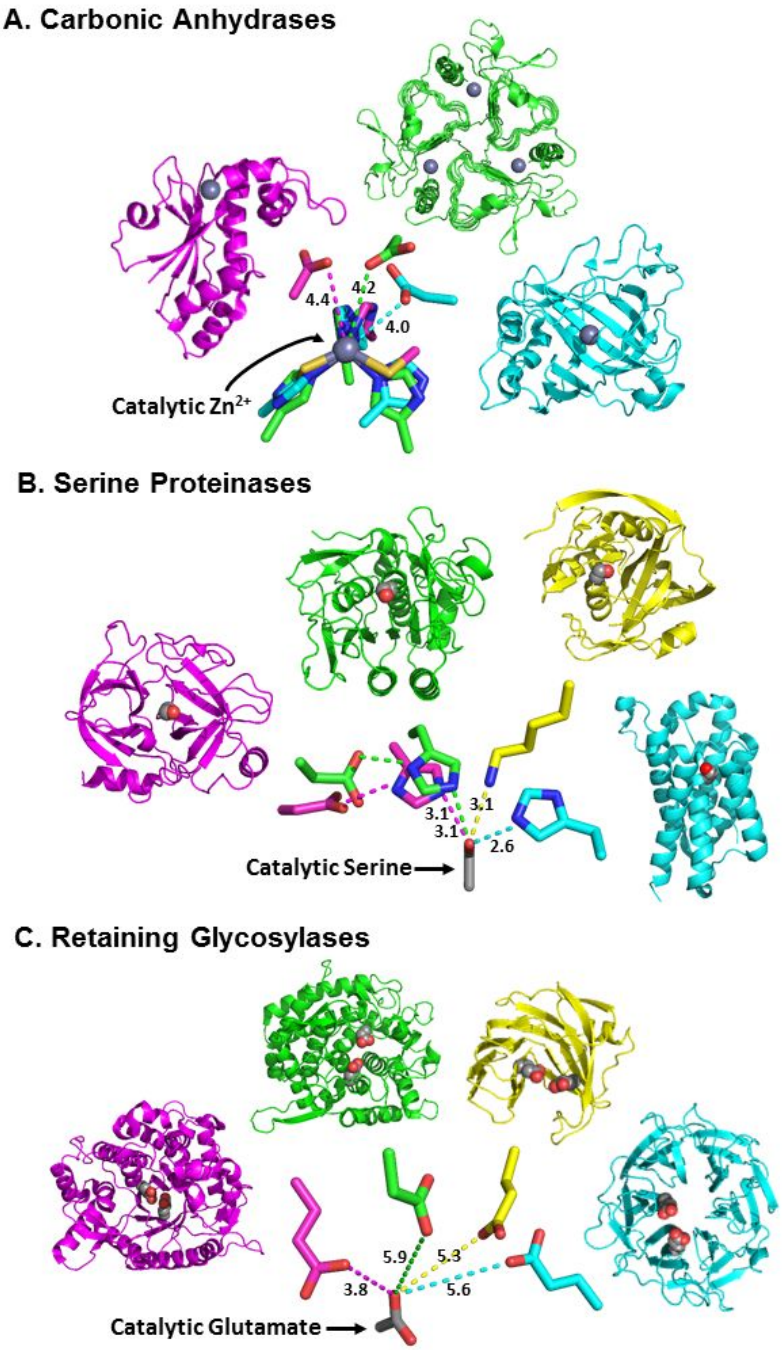

Figure 5: The convergence of active site architectures – considerable variations on the same theme. Enzymes of independent evolutionary origins (assigned by different ECOD X-groups) that share the same catalytic mechanism were analyzed. Distances between key catalytic residues are given in Angstroms. (**A**) Carbonic anhydrases, aligned at the metal-ligating residues: 2w3n (β-carbonic anhydrase-like, magenta); 1thj (Single-stranded left-handed β-helix, green); 3ks3 (Carbonic anhydrase, cyan) (**B**) Serine proteinases aligned at the nucleophilic serine (α-carbon, β-carbon, and γ-oxygen): 4i8g (Cradle loop barrel, magenta); 1gci (Subtilisin-like,

green); 2gef (C-terminal subdomain in Lon-related proteases catalytic domains, yellow); 4njp (Rhomboid-like, cyan). (**C**) Retaining glycosylases aligned at the nucleophilic carboxylate (γ-carbon, δ-carbon, and ϵ-oxygen): 3ahx (TIM β/α-barrel, magenta); 5bx2 (Glycosyl hydrolase domain-like, green); 1xnb (Jelly-roll, yellow); 2w5n (β-propeller-like, cyan).

8. Concluding remarks

Biochemists traditionally approach enzymes from a 'reductionist' perspective, studying the kinetic parameters, structure, mechanism, or physiological roles of an individual enzyme. The accumulation of thousands of such individual case studies, however, enables a 'systems', or bird's-eye view, of 'enzyme space'. By considering the entire repertoire of known natural enzymes, we can better understand the various driving forces and constraints that have shaped natural enzymes. The signatures of chemical constraints are visible, be they thermodynamic and kinetic constraints (*e.g.,* the diffusion rate limit), or constraints acting on catalytic mechanisms, and thereby on active site architectures (**Figure 5**). Nevertheless, evolution seems to have largely overcome hurdles related to high activation barriers and complex reaction mechanisms (**Figure 2**). Further, multiple independent catalytic solutions have emerged for many reactions and, overall, their catalytic efficiencies do not differ much (**Figure 3**).

Along the same vein, the analysis of kinetic parameters suggests that the multitude of physiological and evolutionary demands and constraints, combined with the coincidental nature of evolution, result in a complex picture. The outcome is that the 'average enzyme' is not as impressive of a catalyst as one might expect from looking at the 'perfect enzyme' (**Figure 1**). Nonetheless, given the *in vivo* enzyme concentrations, the 'average enzyme' k_{cat} value seems to match the average metabolic reaction rate (**Box 1**) suggesting that, in most cases, evolution drives enzymes toward 'good enough' rather than 'perfect'. Indeed, TIM, or triosephosphate isomerase – the original 'perfect enzyme' [36] – may exemplify how physiological demands (high glycolytic flux) and unfavorable reaction thermodynamics (acting near equilibrium) have provided an evolutionary incentive towards very high catalytic efficiency. However, analysis of a large number of *E. coli* enzymes indicates that the overall correlation between reaction thermodynamics, flux, and k_{cat}, is poor (**Figure 4**). At least in the case of *E. coli*, the potential to reach a diffusion-limited rate has not been exhausted: Although TIMs with k_{cat}/K_M values of ~10^{10} s^{-1} M^{-1} have been observed in other organisms [40], the *E. coli* TIM exhibits a k_{cat}/K_M value of ~10^8 s^{-1} M^{-1}, which is >10-fold slower than the diffusion limit. Indeed, it seems that in *E. coli* the vast majority of enzymes have the potential to evolve a significantly higher catalytic efficiency, and for enzyme levels to be accordingly

reduced, as suggested from their observed distance from the fastest enzyme in the reaction class (20-fold, on average – median value; **Figure 4**). The potential to evolve higher catalytic efficiency (k_{cat}/K_M, or k_{cat}) likely exists for many enzymes in any given organism, highlighting few noteworthy points with respect to enzyme evolution: (i) selection pressures to maximize the catalytic efficiency of *individual* enzymes, let alone reach the diffusion-rate limit, are relatively rare; (ii) even when such a selection pressure exists, other enzyme properties, such as accuracy (selectivity in discriminating against undesirable, non-cognate substrates), regulation, and/or association in complexes, are as important in shaping enzymes as catalytic efficiency [26, 45]; (iii) mutations generally arise one at a time, and rate improvement in one enzyme is typically insufficient to exert an advantage, or may even be deleterious to the overall metabolic network. Thus, while it could may well be that the vast majority of enzymes have the potential to evolve a significantly higher catalytic efficiency, such a global improvement is evolutionarily inaccessible.

Finally, it appears that chemical constraints dictate the composition and geometry of active site residues (**Figure 5**). Nonetheless, the same active site chemistry, as well as alternative active site chemistries, seem to have emerged time and again. It may well be that some folds and active site architectures are evolutionarily related in the very distant past, and have since diverged beyond recognition. However, a more likely hypothesis is that the *de novo* emergence of a new enzyme may not be as improbable as is generally assumed [26]. Not much is known on how active sites emerge *de novo* in scaffolds devoid of catalytic capabilities. However, two *de novo* emergences of natural enzymes from non-catalytic proteins have been recently unraveled. In both cases, gradual and smooth emergence of catalysis, including stereo- and regio-selectivity, could be reconstructed starting from ligand binding pockets that exhibited no catalysis [68-69].

Acknowledgements

D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry. R.M. is the Charles and Louise Gartner professional chair. Research described here was funded by the Israel Science Foundation (grant 980/14) and the Estate of Mark Scher and the Adelis Foundation (D.S.T.) and the European Research Council (Project NOVCARBFIX 646827; R.M.). D.S.T dedicates this review to the inspiring I.B., to the beauty of enzymes and of those studying them.

Supporting Information

Supplementary Table S1: Enzyme kinetic parameters (k_{cat}, K_M, k_{cat}/K_M) extracted from the BRENDA database.

Supplementary Table S2: The seven hydrolytic enzyme groups and their kinetic parameters.

Supplementary Table S3: Pairwise comparisons of all BRENDA enzymes including the identified pairs of homologs and analogs.

Supplementary Table S4: *E. coli* enzymes – their kinetic parameters, cellular expression levels and mediated cellular fluxes.

References

1. Monod, J., *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. Alfred A. Knopf: New York, 1971.
2. Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D. S.; Milo, R., The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **2011,** *50*, 4402-4410.
3. Metallo, C. M.; Walther, J. L.; Stephanopoulos, G., Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. *J Biotechnol* **2009,** *144*, 167-174.
4. Park, J. O.; Rubin, S. A.; Xu, Y. F.; Amador-Noguez, D.; Fan, J.; Shlomi, T.; Rabinowitz, J. D., Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol* **2016,** *12*, 482-489.
5. Tawfik, D. S., Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr Opin Chem Biol* **2014,** *21*, 73-80.
6. Tokuriki, N.; Jackson, C. J.; Afriat-Jurnou, L.; Wyganowski, K. T.; Tang, R.; Tawfik, D. S., Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun* **2012,** *3*, 1257.
7. Furnham, N.; Dawson, N. L.; Rahman, S. A.; Thornton, J. M.; Orengo, C. A., Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. *J Mol Biol* **2016,** *428*, 253-267.
8. Ribeiro, A. J. M.; Holliday, G. L.; Furnham, N.; Tyzack, J. D.; Ferris, K.; Thornton, J. M., Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res* **2017**, D618-D623.
9. Tyzack, J. D.; Furnham, N.; Sillitoe, I.; Orengo, C. M.; Thornton, J. M., Understanding enzyme function evolution from a computational perspective. *Curr Opin Struct Biol* **2017,** *47*, 131-139.
10. Schomburg, I.; Chang, A.; Ebeling, C.; Gremse, M.; Heldt, C.; Huhn, G.; Schomburg, D., BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* **2004,** *32*, D431-433.
11. Davidi, D.; Noor, E.; Liebermeister, W.; Bar-Even, A.; Flamholz, A.; Tummler, K.; Barenholz, U.; Goldenfeld, M.; Shlomi, T.; Milo, R., Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proc Natl Acad Sci U S A* **2016,** *113*, 3401-3406.
12. Ferla, M. P.; Brewster, J. L.; Hall, K. R.; Evans, G. B.; Patrick, W. M., Primordial-like enzymes from bacteria with reduced genomes. *Mol Microbiol* **2017,** *105*, 508-524.
13. Juarez-Vazquez, A. L.; Edirisinghe, J. N.; Verduzco-Castro, E. A.; Michalska, K.; Wu, C.; Noda-Garcia, L.; Babnigg, G.; Endres, M.; Medina-Ruiz, S.; Santoyo-Flores, J.; Carrillo-Tripp, M.; Ton-That, H.; Joachimiak, A.; Henry, C. S.; Barona-Gomez, F., Evolution of substrate specificity in a retained enzyme driven by gene loss. *Elife* **2017,** *6*.
14. Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **2000,** *28*, 27-30.

15. Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D., Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* **2009,** *5*, 593-599.
16. Valgepea, K.; Adamberg, K.; Seiman, A.; Vilu, R., Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* **2013,** *9*, 2344-2358.
17. Bar-Even, A.; Milo, R.; Noor, E.; Tawfik, D. S., The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry* **2015,** *54*, 4969-4677.
18. Khersonsky, O.; Kiss, G.; Rothlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci U S A* **2012,** *109*, 10358-10363.
19. Elias, M.; Wieczorek, G.; Rosenne, S.; Tawfik, D. S., The universality of enzymatic rate-temperature dependency. *Trends Biochem Sci* **2014,** *39*, 1-7.
20. Audin, M. J.; Wurm, J. P.; Cvetkovic, M. A.; Sprangers, R., The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation. *Nucleic Acids Res* **2016,** *44*, 2962-2673.
21. Hobbs, J. K.; Jiao, W.; Easter, A. D.; Parker, E. J.; Schipper, L. A.; Arcus, V. L., Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates. *ACS Chem Biol* **2013,** *8*, 2388-2393.
22. Romero-Rivera, A.; Garcia-Borras, M.; Osuna, S., Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal* **2017,** *7*, 8524-8532.
23. Wolfenden, R.; Snider, M. J., The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* **2001,** *34*, 938-945.
24. Wolfenden, R., Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu Rev Biochem* **2011,** *80*, 645-667.
25. Merlin, C.; Masters, M.; McAteer, S.; Coulson, A., Why is carbonic anhydrase essential to Escherichia coli? *J Bacteriol* **2003,** *185*, 6415-6424.
26. Newton, M. S.; Arcus, V. L.; Gerth, M. L.; Patrick, W. M., Enzyme evolution: innovation is easy, optimization is complicated. *Curr Opin Struct Biol* **2018,** *48*, 110-116.
27. Martinez Cuesta, S.; Rahman, S. A.; Furnham, N.; Thornton, J. M., The Classification and Evolution of Enzyme Function. *Biophys J* **2015,** *109*, 1082-1086.
28. Davidi, D.; Milo, R., Lessons on enzyme kinetics from quantitative proteomics. *Curr Opin Biotechnol* **2017,** *46*, 81-89.
29. Noor, E.; Flamholz, A.; Bar-Even, A.; Davidi, D.; Milo, R.; Liebermeister, W., The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLoS Comput Biol* **2016,** *12*, e1005167.
30. Gerosa, L.; Haverkorn van Rijsewijk, B. R.; Christodoulou, D.; Kochanowski, K.; Schmidt, T. S.; Noor, E.; Sauer, U., Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Syst* **2015,** *1*, 270-282.
31. Lewis, N. E.; Hixson, K. K.; Conrad, T. M.; Lerman, J. A.; Charusanti, P.; Polpitiya, A. D.; Adkins, J. N.; Schramm, G.; Purvine, S. O.; Lopez-Ferrer, D.; Weitz,

K. K.; Eils, R.; Konig, R.; Smith, R. D.; Palsson, B. O., Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* **2010,** *6*, 390.
32. Orth, J. D.; Thiele, I.; Palsson, B. O., What is flux balance analysis? *Nat Biotechnol* **2010,** *28*, 245-248.
33. Flamholz, A.; Noor, E.; Bar-Even, A.; Milo, R., eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res* **2012,** *40*, D770-D775.
34. Noor, E.; Haraldsdottir, H. S.; Milo, R.; Fleming, R. M., Consistent estimation of Gibbs energy using component contributions. *PLoS Comput Biol* **2013,** *9*, e1003098.
35. Schmidt, A.; Kochanowski, K.; Vedelaar, S.; Ahrne, E.; Volkmer, B.; Callipo, L.; Knoops, K.; Bauer, M.; Aebersold, R.; Heinemann, M., The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol* **2016,** *34*, 104-110.
36. Albery, W. J.; Knowles, J. R., Perfection in Enzyme Catalysis: The Energetics of Triosephosphate Isomerase. *Accounts of Chemical Research* **1977,** *10*, 105-111.
37. Orth, J. D.; Conrad, T. M.; Na, J.; Lerman, J. A.; Nam, H.; Feist, A. M.; Palsson, B. O., A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol Syst Biol* **2011,** *7*, 535.
38. Noor, E.; Bar-Even, A.; Flamholz, A.; Reznik, E.; Liebermeister, W.; Milo, R., Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol* **2014,** *10*, e1003483.
39. Noor, E.; Flamholz, A.; Liebermeister, W.; Bar-Even, A.; Milo, R., A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett* **2013,** *587*, 2772-2777.
40. Sharma, P.; Guptasarma, P., 'Super-perfect' enzymes: Structural stabilities and activities of recombinant triose phosphate isomerases from Pyrococcus furiosus and Thermococcus onnurineus produced in Escherichia coli. *Biochem Biophys Res Commun* **2015,** *460*, 753-758.
41. Engel, P. C., A marriage full of surprises: forty-five years living with glutamate dehydrogenase. *Neurochem Int* **2011,** *59*, 489-494.
42. Williamson, D. H.; Lund, P.; Krebs, H. A., The redox state of free nicotinamide-adenine dinucleotide in the cytoplasm and mitochondria of rat liver. *Biochem J* **1967,** *103*, 514-527.
43. Flamholz, A.; Noor, E.; Bar-Even, A.; Liebermeister, W.; Milo, R., Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci U S A* **2013,** *110*, 10039-10344.
44. Kafri, M.; Metzl-Raz, E.; Jona, G.; Barkai, N., The Cost of Protein Production. *Cell Rep* **2016,** *14*, 22-31.
45. Newton, M. S.; Arcus, V. L.; Patrick, W. M., Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes. *J R Soc Interface* **2015,** *12*.
46. Reznik, E.; Christodoulou, D.; Goldford, J. E.; Briars, E.; Sauer, U.; Segre, D.; Noor, E., Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Rep* **2017,** *20*, 2666-2677.

47. Savir, Y.; Noor, E.; Milo, R.; Tlusty, T., Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc Natl Acad Sci U S A* **2010,** *107*, 3475-3480.
48. Neu, A.; Neu, U.; Fuchs, A. L.; Schlager, B.; Sprangers, R., An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nat Chem Biol* **2015,** *11*, 697-704.
49. Boukouris, A. E.; Zervopoulos, S. D.; Michelakis, E. D., Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. *Trends Biochem Sci* **2016,** *41*, 712-730.
50. Campanini, B.; Benoni, R.; Bettati, S.; Beck, C. M.; Hayes, C. S.; Mozzarelli, A., Moonlighting O-acetylserine sulfhydrylase: New functions for an old protein. *Biochim Biophys Acta* **2015,** *1854*, 1184-1193.
51. Commichau, F. M.; Stulke, J., Trigger Enzymes: Coordination of Metabolism and Virulence Gene Expression. *Microbiol Spectr* **2015,** *3*.
52. Pedley, A. M.; Benkovic, S. J., A New View into the Regulation of Purine Metabolism: The Purinosome. *Trends Biochem Sci* **2017,** *42*, 141-154.
53. Schmitt, D. L.; An, S., Spatial Organization of Metabolic Enzyme Complexes in Cells. *Biochemistry* **2017,** *56*, 3184-3196.
54. O'Brien, E. J.; Utrilla, J.; Palsson, B. O., Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments. *PLoS Comput Biol* **2016,** *12*, e1004998.
55. Bhattacharyya, S.; Bershtein, S.; Yan, J.; Argun, T.; Gilson, A. I.; Trauger, S. A.; Shakhnovich, E. I., Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. *Elife* **2016,** *5*.
56. Gherardini, P. F.; Wass, M. N.; Helmer-Citterich, M.; Sternberg, M. J., Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol* **2007,** *372*, 817-845.
57. Elias, M.; Tawfik, D. S., Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J Biol Chem* **2012,** *287*, 11-20.
58. Galperin, M. Y.; Koonin, E. V., Divergence and convergence in enzyme evolution. *J Biol Chem* **2012,** *287*, 21-28.
59. Laurino, P.; Toth-Petroczy, A.; Meana-Paneda, R.; Lin, W.; Truhlar, D. G.; Tawfik, D. S., An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLoS Biol* **2016,** *14*, e1002396.
60. Catazaro, J.; Caprez, A.; Guru, A.; Swanson, D.; Powers, R., Functional evolution of PLP-dependent enzymes based on active-site structural similarities. *Proteins* **2014,** *82*, 2597-2608.
61. Eliot, A. C.; Kirsch, J. F., Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* **2004,** *73*, 383-415.
62. Vacca, R. A.; Giannattasio, S.; Capitani, G.; Marra, E.; Christen, P., Molecular evolution of B6 enzymes: binding of pyridoxal-5'-phosphate and Lys41Arg substitution turn ribonuclease A into a model B6 protoenzyme. *BMC Biochem* **2008,** *9*, 17.
63. Mabanglo, M. F.; Xiang, D. F.; Bigley, A. N.; Raushel, F. M., Structure of a Novel Phosphotriesterase from Sphingobium sp. TCM1: A Familiar Binuclear

Metal Center Embedded in a Seven-Bladed beta-Propeller Protein Fold. *Biochemistry* **2016,** *55*, 3963-3974.
64. Ekici, O. D.; Paetzel, M.; Dalbey, R. E., Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci* **2008,** *17*, 2023-2037.
65. Davies, G.; Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure* **1995,** *3*, 853-859.
66. Litzinger, S.; Fischer, S.; Polzer, P.; Diederichs, K.; Welte, W.; Mayer, C., Structural and kinetic analysis of Bacillus subtilis N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *J Biol Chem* **2010,** *285*, 35675-35684.
67. Munoz-Munoz, J.; Cartmell, A.; Terrapon, N.; Henrissat, B.; Gilbert, H. J., Unusual active site location and catalytic apparatus in a glycoside hydrolase family. *Proc Natl Acad Sci U S A* **2017,** *114*, 4936-4941.
68. Clifton, B. E.; Kaczmarski, J. A.; Carr, P. D.; Gerth, M. L.; Tokuriki, N.; Jackson, C. J., Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat Chem Biol* **2018,** *14*, 542-547.
69. Kaltenbach, M.; Burke, J. R.; Dindo, M.; Pabis, A.; Munsberg, F. S.; Rabin, A.; Kamerlin, S. C. L.; Noel, J. P.; Tawfik, D. S., Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat Chem Biol* **2018,** *14*, 548-555.