# The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters

Arren Bar-Even,[†] Elad Noor,[†] Yonatan Savir,[‡] Wolfram Liebermeister,[†] Dan Davidi,[†] Dan S. Tawfik,[§] and Ron Milo*,[†]

[†]Department of Plant Sciences, [‡]Department of Physics of Complex Systems, and [§]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

S *Supporting Information*

ABSTRACT: The kinetic parameters of enzymes are key to understanding the rate and specificity of most biological processes. Although specific trends are frequently studied for individual enzymes, global trends are rarely addressed. We performed an analysis of k_{cat} and K_M values of several thousand enzymes collected from the literature. We found that the "average enzyme" exhibits a k_{cat} of ~ 10 s^{-1} and a k_{cat}/K_M of $\sim 10^5$ s^{-1} M^{-1}, much below the diffusion limit and the characteristic textbook portrayal of kinetically superior enzymes. Why do most enzymes exhibit moderate catalytic efficiencies? Maximal rates may not evolve in cases where weaker selection pressures are expected. We find, for example, that enzymes operating in secondary metabolism are, on average, ~ 30-fold slower than those of central metabolism. We also find indications that the physicochemical properties of substrates affect the kinetic parameters. Specifically, low molecular mass and hydrophobicity appear to limit K_M optimization. In accordance, substitution with phosphate, CoA, or other large modifiers considerably lowers the K_M values of enzymes utilizing the substituted substrates. It therefore appears that both evolutionary selection pressures and physicochemical constraints shape the kinetic parameters of enzymes. It also seems likely that the catalytic efficiency of some enzymes toward their natural substrates could be increased in many cases by natural or laboratory evolution.
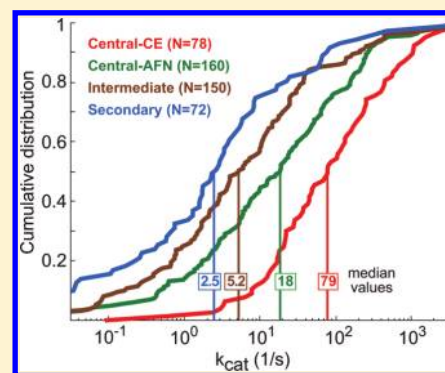
Natural selection shapes the regulation of metabolic enzymes enforcing precise spatial and temporal control. It also plays a central role in the evolution of enzymatic kinetic parameters, including the turnover number (k_{cat}) and the apparent binding constants of substrates (K_M), as both these parameters, together with enzyme concentrations, affect metabolism and growth rates. The literature usually highlights kinetically supreme enzymes such as triosephosphate isomerase, commonly termed a "perfect enzyme".[1,2] What, however, are the kinetic parameters of the "average enzyme"? Are most enzymes similarly "perfect", and if not, why? Here we describe a global view of enzyme parameters with the aim of highlighting the forces that shape the catalytic efficiency of enzymes.

A large body of literature discusses the complex interplay among the various parameters of enzymatic catalysis,[3−9] yet the selective pressures that shaped these parameters remain largely unclear. While traditionally k_{cat}/K_M was thought to be an optimized quantity,[4,8−10] other alternatives were proposed.[1,11,12] For example, K_M values may have evolved to match physiological substrate concentrations,[13] while a substrate-saturated enzyme is expected to maximize k_{cat} and to be insensitive to K_M.[8,14]

There are also several known physicochemical constraints that set boundaries to kinetic parameters.[15,16] For example, theoretical limitations suggest that k_{cat} is unlikely to be higher

than $10^6 - 10^7$ s^{-1}.[6,17] Furthermore, the apparent second-order rate for a diffusion-limited enzyme-catalyzed reaction with a single low-molecular mass substrate (k_{cat}/K_M) cannot exceed $\sim 10^8 - 10^9$ s^{-1} M^{-1}.[18,19] The activation energy of the reaction, as reflected in the uncatalyzed rate, also comprises a barrier: the enzymatic acceleration of an extremely slow reaction, even by many orders of magnitude, may still result in a relatively slow catalyzed rate.[5] The overall thermodynamics of a reaction adds further limits. The Haldane relationship[20] states a dependency between the k_{cat}/K_M of the forward (F) and backward (B) reactions, such that $K_{eq} = (k_{cat}/K_M)_F/(k_{cat}/K_M)_B$, where K_{eq} is the reaction's equilibrium constant. Therefore, even when k_{cat}/K_M in the favorable direction is diffusion-limited, k_{cat}/K_M for the unfavorable direction would be far lower than this limit.

Although there are no obvious limitations on K_M, previous studies have demonstrated that physicochemical properties of ligands and substrates correlate with their binding affinities (K_S) toward receptors and enzymes.[9] One example is the octanol−water partition coefficient (LogP), the ratio of concentrations of a compound partitioning between a hydrophobic

Table 1. Data Extracted for Our Analysis[a]

filter	no. of K_M values	no. of k_{cat} values	K_M and k_{cat}
Brenda, all data	92516	36021	
Brenda, removing mutants	76591	25157	
Brenda, acting on KEGG natural substrates	31162	6530	
unique reactions (enzyme−substrate pairs)	5194	1942	1882
unique enzymes	2588	1176	1128
unique substrates	1612	834	804

[a] Each of the first three rows corresponds to the kinetic parameters downloaded from the Brenda website and filtered as described in the Supporting Information. Briefly, in the first step, parameters referring to mutant enzymes were discarded; in the second step, only parameters that refer to the natural substrate, as indicated by the KEGG database, were included. The last three rows correspond to the amount of unique parameters available after all the filtration steps. The last column gives the number of reactions, enzymes, and substrates for which both K_M and k_{cat} are available.

phase and a water phase. This indicator of hydrophobicity was found to correlate with the binding energy of certain substrates toward their enzymes.[21−24] In addition, previous studies found a correlation between the molecular mass of ligands and their receptor binding affinities.[25,26] However, while binding of ligands to receptors is a relatively simple process, the K_M is a macroscopic value affected by microscopic rates that relate to substrate binding but also to enzymatic catalysis.

Besides the physicochemical considerations, it was also suggested that the evolution of enzyme efficiency is dependent on the enzyme's function within the organism and specific metabolic context.[1] Several cases were found to support this suggestion (e.g., ref 27). On the other hand, it appears that for many enzymes, significant reductions in rate have no effect, and only relatively large reductions in catalytic efficiency hamper organismal fitness (ref 28 and references cited therein). Thus, although it is presumed that both evolutionary selection pressures and physicochemical constrains affect enzymes, the effects of these two factors on enzyme kinetic parameters are currently unclear.

Here we present a global analysis of the kinetic parameters of enzymes. We aim to highlight global factors shaping kinetic parameters of enzymes rather than to discuss specific case studies. To allow such a global analysis, we have mined the Brenda[29] and KEGG[30] databases. The Brenda database contains the kinetic parameters for thousands of enzymes and substrates collected from the literature. Using the KEGG database, we distinguished the parameters for the natural substrates of these enzymes from those of non-native, promiscuous substrates (Supporting Information). In this way, we could separately analyze the kinetic parameters for the natural substrates, avoiding biases from other parameters. Such a distinction has not been made in previously published distributions of kinetic parameters.[29,31] The KEGG database has also allowed us to assign a metabolic pathway for each reaction. Our final data set contained ∼5000 unique reactions, utilizing ∼2500 unique enzymes and ∼1500 unique substrates (Table 1 and Figure S1 of the Supporting Information). The entire data set of the kinetic parameters is given in the Supporting Information.

As expected, such a data set of published kinetic parameters has considerable variability and noise as shown in Table S1 and

discussed in detail in the Supporting Information. This can be attributed to varying measurement conditions, such as pH, temperature, ionic strength, and the concentrations of metals and cofactors, as well as differences in experimental methodologies, and inconsistencies between the values in the original papers and the values entered in the Brenda database (Supporting Information). Further, most parameters in the literature are measured in vitro, which can be considerably different from those measured in vivo.[32,33] These differences can be as great as 3 orders of magnitude.[32] It will be worthwhile to revisit our results as more in vivo data become available. However, it is unlikely that such differences will account for the systematic trends we observe. In addition, temperature significantly affects the kinetic parameters;[34] on average, k_{cat} doubles from 25 to 37 °C (Figure S2 of the Supporting Information). However, we tested such effects and found that none of the trends described below results from a systematic bias in assay temperatures (or pH values).

Overall, the noisy nature of the data collected from the literature limits the resolution of any global analysis and thus limits the study to prominent differences consistently observed in large samples. Moreover, certain subgroups of enzymes might display specific behavior not evident in other groups, thereby masking global trends. Nevertheless, as discussed below, a global analysis does uncover several general trends that underlie kinetic parameters.

■ THE AVERAGE ENZYME IS FAR FROM KINETIC PERFECTION

The median turnover number for the entire data set of enzymes and natural substrates is ∼10 s^{-1} (Figure 1A), where most k_{cat} values (∼60%) are in the range of 1−100 s^{-1}. These rates are orders of magnitude slower than the textbook examples of fast enzymes [e.g., carbonic anhydrase (Figure 1A)], let alone compared to the theoretical limit of 10^6−10^7 s^{-1}.[6,17] The median k_{cat}/K_M is ∼10^5 M^{-1} s^{-1} (Figure 1B), where most k_{cat}/K_M values (∼60%) lie in the range of 10^3−10^6 M^{-1} s^{-1}, orders of magnitude below the diffusion limit. Although pooling all enzymes to one distribution is conceptually problematic, this provides a unique perspective for a comparison of enzymes. For example, Rubisco, the central carbon fixation enzyme, is usually considered to be a slow and inefficient enzyme.[35] However, Figure 1 suggests that it actually possesses rather average kinetic parameters.

When considering the maximal k_{cat}/K_M of each enzyme, the median k_{cat}/K_M is only 5-fold higher (∼5 × 10^5 M^{-1} s^{-1}), suggesting that constraints imposed by the Haldane relationship, and the multisubstrate reactions of enzymes, only partially explain why many enzymes operate far from the theoretical kinetic limits.

We note that previous studies have shown that not all diffusion-controlled reactions operate at k_{cat}/K_M values of 10^8−10^9 M^{-1} s^{-1}. The kinetic parameters for some enzymes could be limited by the rate of substrate−enzyme encounter, as suggested by the viscosity dependence of k_{cat}/K_M, although they display k_{cat}/K_M values in the range of 10^6 M^{-1} s^{-1}.[36−40] In such cases, the low k_{cat}/K_M values may reflect the recruitment of a rare enzyme conformation that mediates catalytically productive binding.[38,40]

The median K_M is ∼100 μM (Figure 1C), where ∼60% of all K_M values are in the range of 10−1000 μM. The distribution of K_M values implies an empirical constraint for the K_M of metabolic
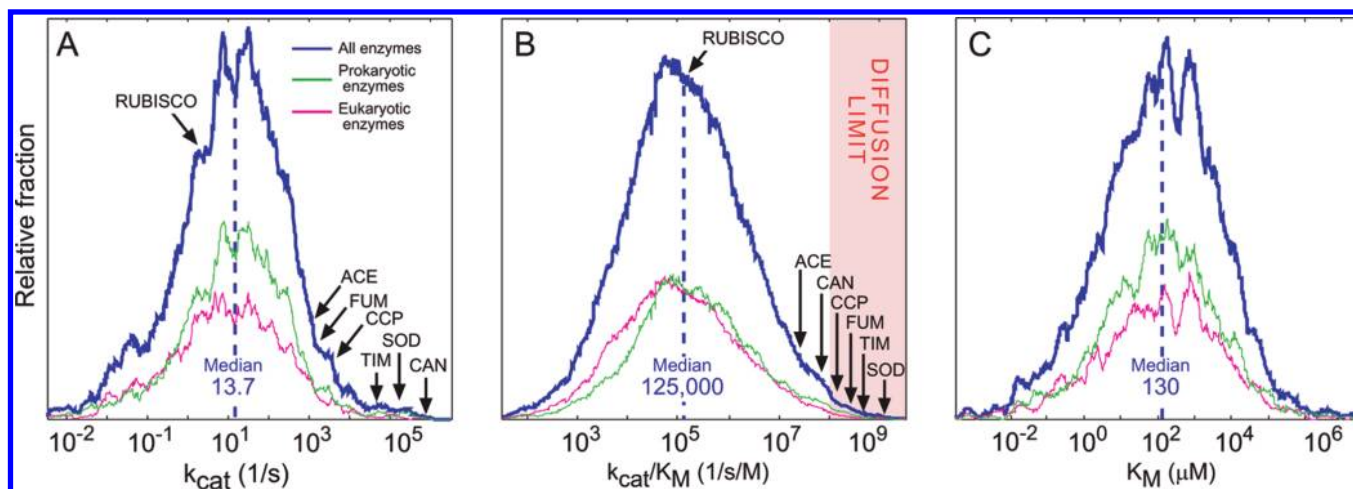
Figure 1. Distributions of kinetic parameters: (A) k_{cat} values ($N = 1942$), (B) k_{cat}/K_M values ($N = 1882$), and (C) K_M values ($N = 5194$). Only values referring to natural substrates were included in the distributions (Supporting Information). Green and magenta lines correspond to the distributions of the kinetic values of prokaryotic and eukaryotic enzymes, respectively. The location of several well-studied enzymes is highlighted: ACE, acetylcholine esterase; CAN, carbonic anhydrase; CCP, cytochrome c peroxidase; FUM, fumarase; Rubisco, ribulose-1,5-bisphosphate carboxylase oxygenase; SOD, superoxide dismutase; TIM, triosephosphate isomerase.

enzymes toward their substrates, where the vast majority of enzymes (>99%) do not exhibit K_M values below 0.1 μM. Such a low K_M characterizes, for example, the glycolysis enzymes glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase utilizing the substrate 1,3-bisphosphoglycerate, the cellular concentration of which is ~0.1 μM.[16]

■ SELECTIVE PRESSURES MOLD ENZYME KINETICS

Why do most enzymes operate far from the theoretical limits of k_{cat} and k_{cat}/K_M? We hypothesize that evolutionary pressures play a key role in shaping enzyme parameters; that is, maximal rates have not evolved in cases where a particular enzyme's rate is not expected to be under stringent selection. To assess the role of selection, we analyzed enzymes that operate at different contexts and examined whether they display significantly different kinetic parameters. We found only a small, yet measureable, effect of the host organism on the kinetics of enzymes (Figure S3 of the Supporting Information). Regardless of the host organism, the metabolic context of an enzyme appears to affect its kinetics considerably as discussed below.

Central Metabolism Enzymes Exhibit Higher k_{cat} and k_{cat}/K_M Values Than Secondary Metabolism Enzymes. We assigned each enzyme—substrate pair to the metabolic modules in which it participates on the basis of the categorization in the KEGG database (Supporting Information). We classified all modules into four primary groups: central-CE (carbohydrate energy) metabolism, involving the main carbon and energy flow; central-AFN (amino acids, fatty acids, and nucleotide) metabolism; intermediate metabolism, including the biosynthesis and degradation of various common cellular components, such as cofactors and coenzymes; and secondary metabolism, related to metabolites that produced in specific cells or tissues, under specific conditions and/or in relatively limited quantities (see Table 2 for the classifications of example metabolic pathways and Table S2 of the Supporting Information for the full list).

As illustrated in Figure 2A, enzyme—substrate pairs associated with central metabolic groups tend to have higher turnover numbers than intermediate and secondary ones. The median

k_{cat} of reactions associated with central-CE metabolism (79 s^{-1}) is ~30-fold higher than that of the reactions associated with secondary metabolism (2.5 s^{-1}). Enzymes associated with central-CE metabolism also exhibit an average k_{cat}/K_M value ~6-fold higher than those of intermediate and secondary metabolism (Figure 2B).

These trends were reproducible even when considering only enzymes from specific functional or structural groups and when analyzing enzymes from *Escherichia coli*, *Saccharomyces cerevisiae*, and humans separately (Figure S4 of the Supporting Information). As the molecular masses of substrates affect their K_M values (discussed below), we also compared substrates within a certain molecular mass range and obtained similar results.

We suggest that these differences may be due to enzymes operating in central metabolism being under stronger selective pressures to increase their rates. A rate maximization of central metabolism enzymes is sensible considering the average high flux these enzymes sustain in vivo. High fluxes enforce the production of many enzyme molecules, a costly process that imposes a significant pressure for improved kinetic parameters that may reduce the amount of enzymes needed and hence alleviate the cost. The selection pressure might be weaker in enzymes functioning in secondary metabolism, either because the overall contribution of these enzymes to organismal fitness is limited or because they operate only under specific conditions or for short periods of time and at relatively low fluxes. Low fluxes can be maintained with relatively low enzyme levels that do not impose a considerable pressure on the kinetic parameters and can be satisfied by moderate k_{cat} or k_{cat}/K_M values. Alternatively, for secondary metabolism enzymes, the selection pressure might not be to increase the catalytic rate but rather to optimize the regulation, control, and localization within a specific context. Finally, it could also be that for a significant fraction of secondary metabolism enzymes, substrates are not correctly assigned, resulting in lower reported parameter values for the so-called natural substrates.

The differences in k_{cat} values between the metabolic groups are more prominent than the differences in the k_{cat}/K_M value. This finding is in line with central metabolism enzymes being under strong selection for higher fluxes and the theoretical view of the evolution of

Table 2. Classification of Several Example Metabolic Pathways[a]

metabolic groups	example of pathways
primary metabolism—CE (carbohydrate and energy)	glycolysis/gluconeogenesis
	pentose phosphate pathway
	citrate cycle (TCA cycle)
	carbon fixation in photosynthetic organisms
	pyruvate metabolism
	glyoxylate and dicarboxylate metabolism
	etc.
primary metabolism—AFN (amino acids, fatty acids, and nucleotides)	glycine, serine, and threonine biosynthesis and metabolism
	phenylalanine, tyrosine, and tryptophan biosynthesis and metabolism
	cysteine and methionine biosynthesis and metabolism
	fatty acid biosynthesis
	purine metabolism
	pyrimidine metabolism
	etc.
intermediate metabolism	pantothenate and CoA biosynthesis
	ubiquinone and other quinone biosynthesis
	glutathione metabolism
	biotin metabolism
	folate biosynthesis
	thiamine metabolism
	etc.
secondary metabolism	flavonoid biosynthesis
	taurine and hypotaurine metabolism
	phenylpropanoid biosynthesis
	bile acid biosynthesis
	caffeine metabolism
	retinol metabolism
	etc.

[a] The full list, containing the classification of 300 metabolic modules, is given in Table 2 of the Supporting Information.

the maximal rate, suggesting that an enzyme, under a constant k_{cat}/K_M, is expected to increase its k_{cat} at the expense of K_M.[9]

■ PROPERTIES OF REACTIONS AND REACTANTS CONSTRAIN ENZYME KINETICS

Evolutionary pressures have their limits and cannot improve the kinetics of an enzyme *ad infinitum*. In the introduction, we reviewed several global constraints imposed on enzymatic catalysis. Here we ask whether we can obtain a deeper understanding of the constraints imposed by the reaction and/or reactants' nature.

While a full consideration of the effects of the reaction characteristics on its kinetics is beyond the scope of this review, we note that different EC classes are characterized by significantly different distributions of k_{cat} values (Figure S5 of the Supporting Information). For example, isomerases (EC 5.*X.X.X*) exhibit a median k_{cat} of 33.5 s^{-1}, an order of magnitude higher than that of ligases (EC 6.*X.X.X*), with a median k_{cat} of 3.7 s^{-1}. It is possible that the different mechanisms and activation energy barriers of different reaction classes may correlate with the
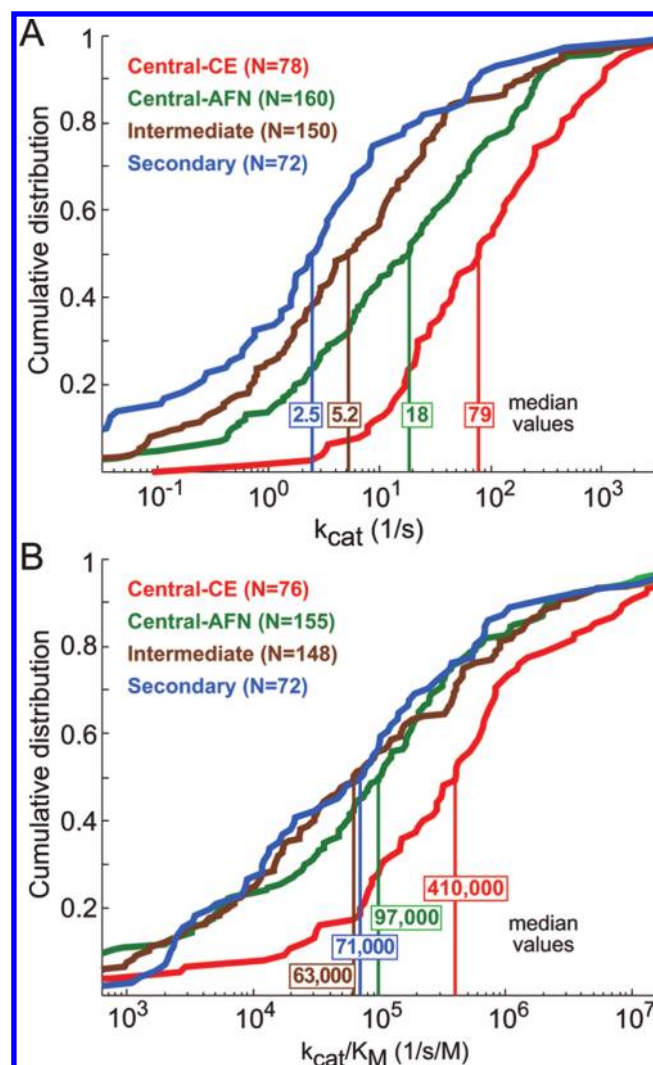

Figure 2. Enzymes operating within different metabolic groups have significantly different k_{cat} and k_{cat}/K_M values. (A) Distribution of k_{cat} values for enzyme−substrate pairs belonging to different metabolic contexts. All distributions are significantly different with a *p* value of <0.0005 (rank-sum test), except for intermediate versus secondary metabolisms (*p* < 0.05). (B) Distribution of k_{cat}/K_M values for enzyme−substrate pairs belonging to different metabolic contexts. Central-CE (carbohydrate and energy) metabolism has significantly higher k_{cat}/K_M values than all other metabolic groups [*p* < 0.0005 (rank-sum test)]. Abbreviations: CE, carbon and energy; AFN, amino acids, fatty acids, and nucleotides. Numbers in parentheses represent the numbers of enzyme−substrate pairs included in each set.

catalyzed rates, but at present, the available data do not allow a systematic exploration of these mechanistic issues. We do observe, however, an effect of the number of substrates involved in a reaction on the K_M for these substrates; the higher the number of substrates, the lower the K_M for each substrate (Figure S6 of the Supporting Information). This observation might result from the fact that as the number of substrates increases, lower K_M values are required to obtain the same concentration of the enzyme−substrate complex; catalysis of a single substrate begins with the formation of a binary E·S complex, of a two-substrate reaction with a ternary complex, etc.

For Small Substrates, K_M Decreases with Increasing Substrate Molecular Mass and Hydrophobicity. The binding of
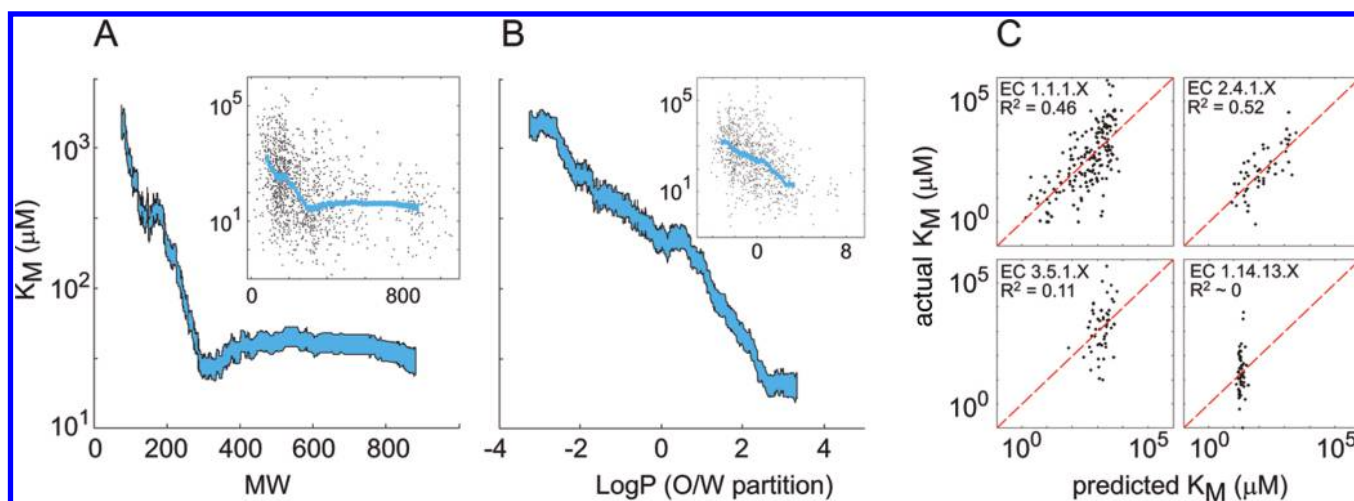
Figure 3. Correlations of K_M and substrate molecular mass and LogP. (A and B) The cyan area corresponds to the standard error around the mean value of a window centered at each X value (Supporting Information). The insets display the distribution of all the points. (A) Two regimes are noted: molecular mass of <350 Da, where the correlation between K_M and molecular mass was found to be negative ($R^2 = 0.13$; $p < 10^{-6}$; slope = -6×10^{-3} Da^{-1}) (Supporting Information), and molecular mass of >350 Da, where K_M values plateau at ~40 μM. (B) K_M vs LogP. Only small substrates (<350 Da) were used for this analysis ($R^2 = 0.24$; $p < 10^{-6}$; slope = $0.25 \times$ LogP^{-1}) (Supporting Information). (C) Correlations between the predicted K_M, according to a substrate molecular mass and LogP, and the reported K_M values for several EC classes: 1.1.1.X ($R^2 = 0.46$), oxidoreductases, acting on the CH−OH group of donors, with NAD$^+$ or NADP$^+$ as the acceptor; 2.4.1.X ($R^2 = 0.52$), hexosyltransferases; 3.5.1.X ($R^2 = 0.11$), hydrolases, acting on carbon−nitrogen bonds, other than peptide bonds, in linear amides; and 1.14.13.X ($R^2 \sim 0$), oxygenases, acting on two donors, where NADH or NADPH is one donor, and incorporating one atom of oxygen into the other donor.

ligands to receptors is usually described by the dissociation constant, K_S, which can be measured and interpreted as a sum of contributions of different chemical bonds.[25,26] In contrast, the interpretation of the apparent dissociation constant of an enzyme−substrate complex, K_M, actually represents a complex function of both the dissociation constant of a substrate (i.e., its binding affinity, or K_S) and its catalytic turnover rate. In reactions of a complex mechanism, K_M is affected by various microscopic rates that do not necessarily relate to the binding of a single substrate. K_M is expected to approximate K_S only for reactions with a very simple mechanism and when the rate of catalysis is significantly slower than the rate of substrate association and dissociation.[9] Still, the existence of a significant correlation between various physicochemical properties of substrates and their K_M values might indicate that these properties are global contributors to the interaction between substrates and enzymes, and their effects are manifested in spite of the mechanistic complexities mentioned above.

We identify two physicochemical properties that significantly correlate with K_M (Table S3 of the Supporting Information): molecular mass and hydrophobicity. As shown in Figure 3A, K_M decreases, on average, by almost 2 orders of magnitude with an increasing molecular mass up to ~350 Da (the molecular mass of AMP). For molecular masses of >350 Da, we observed no correlation between molecular mass and K_M, and the median K_M for >350 Da substrates (~40 μM) is considerably lower than the median for all substrates (~100 μM). Hence, the data indicate a limitation on the binding affinity of small substrates, as observed for the binding of ligands to macromolecular targets.[25,26] We find that affinity increases with an increasing number of non-hydrogen atoms, although with a lower contribution per atom as compared to ligands[25,26] (Supporting Information), and only up to molecular masses of ≤ 350 Da. While the potential for increasing the affinity for large substrates might exist, the evolutionary pressure might not be significant

enough to drive K_M beneath an average value of ~40 μM for most enzymes. Alternatively, a trade-off between K_M and k_{cat} may restrict the improvement of the former.[9]

For small substrates (≤350 Da), the K_M values also decrease ~100-fold with an increasing LogP (Figure 3B). Molecular mass and LogP are not correlated with one another (Table S3 of the Supporting Information), and hence, their correlation with K_M is independent. Indeed, a regression of K_M using both parameters yielded an almost additive correlation ($R^2 = 0.31$). Considering the noisy nature of the data set, and the fact that K_M does not directly represent K_S, we find it noteworthy that almost one-third of the variability in K_M can be explained by two substrate properties. It suggests that molecular mass and hydrophobicity might significantly limit the kinetic parameters and affect the observed parameter distributions.

In line with previous studies, we interpret the correlations between K_M and molecular mass and LogP as contributions from van der Waals forces and the hydrophobic effect.[25] All other substrate properties, such as charge and hydrogen bond donors and acceptors, either did not show an effect on K_M or k_{cat} or were correlated with molecular mass and LogP (Table S3 of the Supporting Information). Therefore, although hydrogen bonding and electrostatic interactions clearly play an important role in substrate binding, their contribution could not be globally detected.

Some EC classes show a strong correlation between K_M and molecular mass and LogP, suggesting an ability to predict parameter values as shown in Figure 3C. For other EC classes, the correlations are very weak (Figure 3C). Although we find only a minor global correlation between k_{cat} and K_M ($R^2 = 0.09$), the correlation between these parameters displays a similar EC dependent contingency (Figure S1 of the Supporting Information). The EC classes that display strong correlations seem to have a rather simple catalytic mechanism that corresponds to the Michaelis−Menten formalism. In such cases, the

apparent effective kinetic constants correspond more directly to the steps in the reaction axis, and K_M largely reflects the substrate binding affinity (K_S).[48] In contrast, EC classes, such as monooxygenases (EC 1.14.13.X), that are characterized by low correlations seem to have complex, multistage mechanisms in which the apparent K_M is highly removed from K_S.[9]

None of the substrate properties we examined correlated with k_{cat} (Table S3 of the Supporting Information).

Large Substrate Modifiers Decrease K_M. It was previously shown that noncovalent interactions between an enzyme and the CoA moiety of a substrate stabilize the reaction's ground state and bring about an approximately 10-fold increase in k_{cat}/K_M.[49] We wanted to test whether this phenomenon is general. We therefore compared the K_M values of enzymes utilizing substituted and unsubstituted substrates. We analyzed modifiers of different sizes for which we were able to obtain sufficient data. The large modifiers analyzed were CoA and NDP (UDP, CDP, etc.), and the small to medium modifiers analyzed were the phosphoryl group, pentoses (ribose, arabinose, etc.), and the N-substituted methyl and acetyl groups.

It appears that substitution with large modifiers, NDP and CoA, significantly decreases K_M, by median factors of 19 and 5, respectively (Figure S7 of the Supporting Information) ($p < 10^{-3}$). Of the small to medium modifiers, only the phosphoryl group significantly decreased K_M, by median factors of almost 4 ($p < 10^{-3}$) (Figure S7).

Thus, apart from chemically activating their substrates, modifiers might have a common functionality in enhancing the affinity of enzymes for the substituted substrates. For example, a micromolar K_M range for acetate (molecular mass of 60 Da) and propionate (molecular mass of 74 Da) might be desirable for certain enzymes, but difficult to achieve because of the lack of a sufficiently large interaction surface for such small a substrate. Indeed, the K_M values for these substrates lie in the range of 1–10 mM, whereas the K_M values for the respective CoA-substituted substrates are in the range of 1–100 μM, indicating a decrease of ~2 orders of magnitude upon CoA substitution. The cellular concentrations of acetate and propionate are in the range of 1 μM to 1 mM.[13] Thus, enzymes that accept these small substrates operate below the K_M and hence significantly below their maximal rate. In contrast, enzymes accepting the acyl-CoA derivatives of these substrates operate within the K_M range and hence are nearly substrate-saturated (the effect of substitution with CoA is further discussed in the Supporting Information).

■ IMPLICATIONS

Most enzymes are characterized by kinetic parameters that are much lower than textbook examples and the commonly presented constraint of the diffusion rate. We propose two possible underlying explanations for this phenomenon, one involving different intensities of selection for rate maximization and the other related to physicochemical constraints. Both explanations impact our understanding of the evolution of enzymatic parameters and suggest that within certain physicochemical constraints, the catalytic efficiency of many enzymes toward their native substrates could potentially be increased by natural or laboratory evolution.

We found indications for both proposals, but laboratory evolution experiments will be needed to test these suggestions in an unambiguous manner. For example, in contrast to very large improvements observed in directed evolution toward promiscuous substrates and reactions (up to 10^6-fold), attempts to improve the catalytic efficiency of Rubisco toward its natural substrate and reaction resulted in relatively small improvements,[50,51] suggesting that this enzyme operates close to the theoretical maximal efficiency.[52] To counter the Rubisco example, it might be of interest to try to improve a secondary metabolism enzyme whose catalytic efficiency with its natural substrate and reaction is suggested to be far from optimized.

The identification of the natural substrate and reaction comprises a challenge for secondary metabolism enzymes. For many of those, substrates with only moderate catalytic efficiencies are identified in vitro, and the validation of their in vivo function remains elusive. For many such enzymes, the real substrate and reaction are yet to be discovered. However, we consider it unlikely that this explanation fully accounts for the trends we observe across hundreds of enzymes. Our global analysis therefore suggests that for secondary metabolism enzymes, k_{cat}/K_M values that are $<10^{-5}$ M^{-1} s^{-1} and k_{cat} values lower than 10 s^{-1} are the norm. Indeed, low k_{cat}/K_M values in the range of 10^3 M^{-1} s^{-1}, or even lower, together with broad substrate specificity, are commonly seen in plant secondary metabolism enzymes.[41–47]

This global view of kinetic parameters has implications with respect to enzyme engineering and design. The benchmark for success in this field has been kinetic parameters and rate accelerations that are comparable to those of natural enzymes,[53] but as one can see in Figure 1, the catalytic efficiencies of natural enzymes span more than 5 orders of magnitude. Designed and/or engineered enzymes begin to reach k_{cat}/K_M values of $\geq 10^4$ (for a recent example, see ref 54). Whereas these values are orders of magnitude from the enzymes' top league ($k_{cat}/K_M \geq 10^8$ M^{-1} s^{-1}), they are not so far from that of the average enzyme ($k_{cat}/K_M \sim 10^5$ M^{-1} s^{-1}). Rate accelerations, and how inherently activated the reaction is, are other important comparison points, but from a biological point of view, catalytic efficiency is the most relevant parameter.

The kinetic parameters analyzed here are also relevant for metabolic engineering. In recent years, this field has advanced from the heterologous expression of individual enzymes to the construction of complex metabolic pathways that combine several enzymes from different origins.[55–58] In many cases, more than one solution for a given metabolic goal exists, and each of these solutions employs a different set of enzymes.[59] The global analysis of kinetic parameters suggests that choosing enzymes that normally operate in central metabolic contexts is more likely to yield higher fluxes, a property that is a key for most metabolic engineering enterprises. Further, it might be beneficial to use pathways that employ substituted substrates (e.g., with phosphoryl or CoA modifiers) to ensure saturation and hence higher rates.

The trends shaping the kinetic parameters can be helpful in establishing and analyzing metabolic models. These trends can further be used to reinterpret the evolutionary objectives assumed by some of these models. For example, optimality analysis previously suggested that under the assumptions of an evolutionary pressure to increase flux, and of an upper limit on the total enzyme concentrations, enzymes with more control over the metabolic flux are expected to be present in higher concentrations compared to those with weaker control.[16,60] It is possible to expand such an analysis to account for the observed trends in the values of k_{cat}. Our observation that k_{cat} values remain far below their theoretical limits and that not all enzyme groups are

optimized to the same extent led us to construct a model with an effective cost term on the increase of k_{cat}, arising from the decreasing probability of finding a mutant with higher k_{cat} values (Supporting Information). Just as enzyme concentrations increase until the benefit and cost balance each other, k_{cat} values increase until the possibility of achieving even larger k_{cat} values becomes improbable. Our model suggests that a strong selective pressure, resulting from strong control over the metabolic flux and from the cost of producing many enzyme copies, can push a k_{cat} value further into the low-probability range of higher values, while a weaker selective pressure would result in lower k_{cat} values. The full model, including a detailed discussion of its strengths and limitations, is given in the Supporting Information.

OPEN QUESTIONS

The global examination of kinetic parameters raises several key questions. First, it is apparent that different EC classes display significantly different behaviors concerning the relationship between kinetic parameters and the physicochemical substrate properties. These differences can be attributed to noise but may also be related to different catalytic mechanisms. In any case, the issue of specific correlations deserves a deeper and more thorough analysis, while considering individual enzyme classes. More specifically, the global data indicate only a minor correlation between k_{cat} and K_M and thus fail to indicate a clear trade-off between these parameters. However, analysis of enzymatic reactions with similar properties might reveal that this trade-off prevails in numerous cases (e.g., ref 52).

A puzzling issue remains regarding the lack of correlation between the substrate's potential for hydrogen bonding (including electrostatic interactions) and its K_M value. This is in spite of the established contribution of these interactions to the binding energy of substrates.[9] A more detailed analysis of smaller groups of substrates possessing similar physicochemical properties might reveal a quantitative contribution of these forces to the binding affinity. It could also be that enzymes evolved to utilize such interactions primarily as the transition state develops, and the substrate's potential for hydrogen bonding therefore contributes less to K_M and more to k_{cat} values.

Another interesting observation on the global scale analyzed here is that regardless of the microscopic mechanisms, the median k_{cat}/K_M observed is $\sim 10^5$ M^{-1} s^{-1}. Given that the diffusion rate for collisions of proteins and low-molecular mass ligands is on the order of 10^8 M^{-1} s^{-1}, it appears that, on average, fewer than one of 1000 substrate molecules that collide with an enzyme undergoes catalytic conversion. Studies of enzyme dynamics, and of single-molecule enzyme reactions, have begun to address this question,[61−63] but we are still far from understanding what determines the fraction of productive collisions and why so few enzymes have evolved such that nearly every collision is catalytically productive.

Finally, it is not clear how the in vitro-measured k_{cat} and K_M really correlate with enzyme activity within living cells. In a highly crowded environment,[64] characterized by numerous interactions that do not occur in the test tube, and by colocalization of substrates and of other enzymes that either generate the target substrate or react with the resulting product, enzymatic activity might display features that could not be captured by the kinetic parameters obtained in vitro.

In spite of its obvious limitations, we suggest that global analyses can shed light on how evolution on the one hand and physicochemical constrains on the other have shaped enzyme catalysis. As more kinetic data are rapidly being gathered, metabolite concentrations within living cells are becoming widely available,[13,65−69] and the means of simulating and measuring metabolic fluxes are developing,[70−72] we are hopeful it will become possible to address some of the open questions highlighted above.

ASSOCIATED CONTENT

Ⓢ **Supporting Information.** The entire data set of kinetic parameters, a detailed description of the data sets used in our analysis and a discussion of the dependencies between the kinetic parameters and of the reasons for the noise in the data sets, the affiliation of metabolic modules to one of the four primary metabolic groups, an analysis of the effect of host organisms on the kinetics of enzymes, the effect of the physicochemical properties on K_M, the energetic contribution of each non-hydrogen atom to binding, and the decrease in K_M upon substitution with large modifiers, the statistical tools used in the analysis, and a mathematical model for the evolution of suboptimal catalytic constants. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: ron.milo@weizmann.ac.il. Phone: +972505714697. Fax: +97289344540.

Author Contributions
A.B.-E. and E.N. contributed equally to this work.

ACKNOWLEDGMENT

ABBREVIATIONS

LogP, logarithm of the octanol−water partition coefficient; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; CE, carbon and energy; AFN, amino acids, fatty acids, and nucleotides; CoA, coenzyme A; NDP, nucleotide diphosphate.

■ REFERENCES

(1) Albery, W. J., and Knowles, J. R. (1976) Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry 15*, 5631–5640.

(2) Radzicka, A., and Wolfenden, R. (1995) A proficient enzyme. *Science 267*, 90–93.

(3) Knowles, J. R. (1991) Enzyme catalysis: Not different, just better. *Nature 350*, 121–124.

(4) Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem. 273*, 27035–27038.

(5) Wolfenden, R., and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res. 34*, 938–945.

(6) Benkovic, S. J., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science 301*, 1196–1202.

(7) Cannon, W. R., and Benkovic, S. J. (1998) Solvation, reorganization energy, and biological catalysis. *J. Biol. Chem. 273*, 26257–26260.

(8) Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. (2004) How enzymes work: Analysis by modern rate theory and computer simulations. *Science 303*, 186–195.

(9) Fersht, A. (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman, New York.

(10) Fersht, A. R. (1974) Catalysis, binding and enzyme-substrate complementarity. *Proc. R. Soc. London, Ser. B 187*, 397–407.

(11) Crowley, P. H. (1975) Natural selection and the Michaelis constant. *J. Theor. Biol. 50*, 461–475.

(12) Kacser, H., and Beeby, R. (1984) Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. *J. Mol. Evol. 20*, 38–51.

(13) Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol. 5*, 593–599.

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

(15) Mavrovountis, M. L., and Stephanopoulos, G. N. (1990) Estimation of upper bounds for the rates of enzymatic reactions. *Chem. Eng. Commun. 93*, 221–236.

(16) Heinrich, R., Schuster, S., and Holzhutter, H. G. (1991) Mathematical analysis of enzymic reaction systems using optimization principles. *Eur. J. Biochem. 201*, 1–21.

(17) Hammes, G. G. (2002) Multiple conformational changes in enzyme catalysis. *Biochemistry 41*, 8221–8228.

(18) Samson, R., and Deutch, J. M. (1978) Diffusion-controlled reaction rate to a buried active site. *J. Chem. Phys. 68*, 285–290.

(19) Schurr, J. M., and Schmitz, K. S. (1976) Orientation constrains and rotational diffusion in bimolecular solution kinetics. A Simplification. *J. Phys. Chem. 80*, 1934–1936.

(20) Haldane, J. B. S. (1930) *Enzymes*, Longmans & Green, London.

(21) Dorovska, V. N., Varfolomeyev, S. D., Kazanskaya, N. F., Klyosov, A. A., and Martinek, K. (1972) The influence of the geometric properties of the active centre on the specificity of α-chymotrypsin catalysis. *FEBS Lett. 23*, 122–124.

(22) Fastrez, J., and Fersht, A. R. (1973) Mechanism of chymotrypsin. Structure, reactivity, and nonproductive binding relationships. *Biochemistry 12*, 1067–1074.

(23) Page, M. I. (1976) Binding energy and enzymic catalysis. *Biochem. Biophys. Res. Commun. 72*, 456–461.

(24) Fersht, A. R., Shindler, J. S., and Tsui, W. C. (1980) Probing the limits of protein-amino acid side chain recognition with the aminoacyl-tRNA synthetases. Discrimination against phenylalanine by tyrosyl-tRNA synthetases. *Biochemistry 19*, 5520–5524.

(25) Kuntz, I. D., Chen, K., Sharp, K. A., and Kollman, P. A. (1999) The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A. 96*, 9997–10002.

(26) Hajduk, P. J. (2006) Fragment-based drug design: How big is too big? *J. Med. Chem. 49*, 6972–6976.

(27) Matagne, A., Dubus, A., Galleni, M., and Frere, J. M. (1999) The β-lactamase cycle: A tale of selective pressure and bacterial ingenuity. *Nat. Prod. Rep. 16*, 1–19.

(28) Wagner, A. (2007) *Robustness and Evolvability in Living Systems*, Princeton University Press, Princeton, NJ.

(29) Pharkya, P., Nikolaev, E. V., and Maranas, C. D. (2003) Review of the BRENDA Database. *Metab. Eng. 5*, 71–73.

(30) Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res. 36*, D480–D484.

(31) Liebermeister, W., and Klipp, E. (2006) Bringing metabolic networks to life: Integration of kinetic, metabolic, and proteomic data. *Theor. Biol. Med. Modell. 3*, 42.

(32) Wright, B. E., Butler, M. H., and Albe, K. R. (1992) Systems analysis of the tricarboxylic acid cycle in *Dictyostelium discoideum*. I. The basis for model construction. *J. Biol. Chem. 267*, 3101–3105.

(33) Ringe, D., and Petsko, G. A. (2008) Biochemistry. How enzymes work. *Science 320*, 1428–1429.

(34) Wolfenden, R., Snider, M., Ridgway, C., and Miller, B. (1999) The Temperature Dependence of Enzyme Rate Enhancements. *J. Am. Chem. Soc. 121*, 7419–7420.

(35) Tcherkez, G. G., Farquhar, G. D., and Andrews, T. J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U.S.A. 103*, 7246–7251.

(36) Brouwer, A. C., and Kirsch, J. F. (1982) Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation. *Biochemistry 21*, 1302–1307.

(37) Steyaert, J., Wyns, L., and Stanssens, P. (1991) Subsite interactions of ribonuclease T1: Viscosity effects indicate that the rate-limiting step of GpN transesterification depends on the nature of N. *Biochemistry 30*, 8661–8665.

(38) Simopoulos, T. T., and Jencks, W. P. (1994) Alkaline phosphatase is an almost perfect enzyme. *Biochemistry 33*, 10375–10380.

(39) Mattei, P., Kast, P., and Hilvert, D. (1999) *Bacillus subtilis* chorismate mutase is partially diffusion-controlled. *Eur. J. Biochem. 261*, 25–32.

(40) Stratton, J. R., Pelton, J. G., and Kirsch, J. F. (2001) A novel engineered subtilisin BPN′ lacking a low-barrier hydrogen bond in the catalytic triad. *Biochemistry 40*, 10411–10416.

(41) Picaud, S., Brodelius, M., and Brodelius, P. E. (2005) Expression, purification and characterization of recombinant (*E*)-β-farnesene synthase from *Artemisia annua*. *Phytochemistry 66*, 961–967.

(42) Picaud, S., Olofsson, L., Brodelius, M., and Brodelius, P. E. (2005) Expression, purification, and characterization of recombinant amorpha-4,11-diene synthase from *Artemisia annua* L. *Arch. Biochem. Biophys. 436*, 215–226.

(43) Greenhagen, B. T., O'Maille, P. E., Noel, J. P., and Chappell, J. (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc. Natl. Acad. Sci. U.S.A. 103*, 9826–9831.

(44) Guterman, I., Masci, T., Chen, X., Negre, F., Pichersky, E., Dudareva, N., Weiss, D., and Vainstein, A. (2006) Generation of phenylpropanoid pathway-derived volatiles in transgenic plants: Rose alcohol acetyltransferase produces phenylethyl acetate and benzyl acetate in petunia flowers. *Plant Mol. Biol. 60*, 555–563.

(45) O'Connor, S. E., and McCoy, E. (2006) Terpene indole alkaloid biosynthesis. In *Recent Advances in Phytochemistry* (Romeo, J. T., Ed.) Springer, Berlin.

(46) Faraldos, J. A., Zhao, Y., O'Maille, P. E., Noel, J. P., and Coates, R. M. (2007) Interception of the enzymatic conversion of farnesyl diphosphate to 5-epi-aristolochene by using a fluoro substrate analogue: 1-Fluorogermacrene A from (2E,6Z)-6-fluorofarnesyl diphosphate. *ChemBioChem 8*, 1826–1833.

(47) Runguphan, W., Qu, X., and O'Connor, S. E. (2010) Integrating carbon−halogen bond formation into medicinal plant metabolism. *Nature 468*, 461–464.

(48) Leskovac, V. (2003) *Comprehensive Enzyme Kinetics*, Kluwer Academic Publishers, New York.

(49) Whitty, A., Fierke, C. A., and Jencks, W. P. (1995) Role of binding energy with coenzyme A in catalysis by 3-oxoacid coenzyme A transferase. *Biochemistry 34*, 11678–11689.

(50) Parikh, M. R., Greene, D. N., Woods, K. K., and Matsumura, I. (2006) Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E. coli*. *Protein Eng., Des. Sel. 19*, 113–119.

(51) Greene, D. N., Whitney, S. M., and Matsumura, I. (2007) Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. *Biochem. J. 404*, 517–524.

(52) Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010) Cross-species analysis traces adaptation of Rubisco towards optimality in a low dimensional landscape. *Proc. Natl. Acad. Sci. U.S.A. 107*, 3475–3480.

(53) Baker, D. (2010) An exciting but challenging road ahead for computational enzyme design. *Protein Sci. 19*, 1817–1819.

(54) Khersonsky, O., Rothlisberger, D., Wollacott, A. M., Murphy, P., Dym, O., Albeck, S., Kiss, G., Houk, K. N., Baker, D., and Tawfik, D. S. (2011) Optimization of the In-Silico-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution. *J. Mol. Biol. 407*, 391–412.

(55) Medema, M. H., Breitling, R., Bovenberg, R., and Takano, E. (2011) Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat. Rev. Microbiol. 9*, 131–137.

(56) Bachmann, B. O. (2010) Biosynthesis: Is it time to go retro? *Nat. Chem. Biol. 6*, 390–393.

(57) Martin, C. H., Nielsen, D. R., Solomon, K. V., and Prather, K. L. (2009) Synthetic metabolism: Engineering biology at the protein and pathway scales. *Chem. Biol. 16*, 277–286.

(58) Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W. H., Mo, J. D., Tidor, B., Stephanopoulos, G., and Prather, K. L. (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. U.S.A. 107*, 13654–13659.

(59) Bar-Even, A., Noor, E., Lewis, N. E., and Milo, R. (2010) Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U.S.A. 107*, 8889–8894.

(60) Klipp, E., and Heinrich, R. (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.

(61) Min, W., English, B. P., Luo, G., Cherayil, B. J., Kou, S. C., and Xie, X. S. (2005) Fluctuating enzymes: lessons from single-molecule studies. *Acc. Chem. Res. 38*, 923–931.

(62) Lerch, H. P., Mikhailov, A. S., and Hess, B. (2002) Conformational-relaxation models of single-enzyme kinetics. *Proc. Natl. Acad. Sci. U. S. A. 99*, 15410–15415.

(63) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol. 5*, 789–796.

(64) Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol. 222*, 599–620.

(65) Kummel, A., Ewald, J. C., Fendt, S. M., Jol, S. J., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., and Heinemann, M. (2010) Differential glucose repression in common yeast strains in response to HXK2 deletion. *FEMS Yeast Res 10*, 322–332.

(66) Wu, L., Mashego, M. R., van Dam, J. C., Proell, A. M., Vinke, J. L., Ras, C., van Winden, W. A., van Gulik, W. M., and Heijnen, J. J. (2005) Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards. *Anal. Biochem. 336*, 164–171.

(67) Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P. Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., and Tomita, M.

(2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science 316*, 593–597.

(68) Ewald, J. C., Heux, S., and Zamboni, N. (2009) High-throughput quantitative metabolomics: Workflow for cultivation, quenching, and analysis of yeast in a multiwell format. *Anal. Chem. 81*, 3623–3629.

(69) Fendt, S. M., Buescher, J. M., Rudroff, F., Picotti, P., Zamboni, N., and Sauer, U. (2010) Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Mol. Syst. Biol. 6*, 356.

(70) Kleijn, R. J., Buescher, J. M., Le Chat, L., Jules, M., Aymerich, S., and Sauer, U. (2010) Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis. *J. Biol. Chem. 285*, 1587–1596.

(71) Tang, Y. J., Martin, H. G., Myers, S., Rodriguez, S., Baidoo, E. E., and Keasling, J. D. (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. *Mass Spectrom. Rev. 28*, 362–375.

(72) Fuhrer, T., Fischer, E., and Sauer, U. (2005) Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol. 187*, 1581–1590.