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

Arren Bar-Even,* Elad Noor,† Yonatan Savir,‡ Wolfram Liebermeister,† Dan Davidi,‡ Dan S. Tawil,§ 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

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 $k_{\text{cat}}$ and $K_M$ values of several thousand enzymes collected from the literature. We found that the “average enzyme” exhibits a $k_{\text{cat}}$ of $\sim 10 \text{ s}^{-1}$ and a $k_{\text{cat}}/K_M$ of $\sim 10^3 \text{ s}^{-1} \text{ M}^{-1}$, 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 ($k_{\text{cat}}$) 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”. 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, yet the selective pressures that shaped these parameters remain largely unclear. While traditionally $k_{\text{cat}}/K_M$ was thought to be an optimized quantity, other alternatives were proposed. For example, $K_M$ values may have evolved to match physiological substrate concentrations, while a substrate-saturated enzyme is expected to maximize $k_{\text{cat}}$ and to be insensitive to $K_M$. There are also several known physicochemical constraints that set boundaries to kinetic parameters. For example, theoretical limitations suggest that $k_{\text{cat}}$ is unlikely to be higher than $10^6 - 10^7 \text{ s}^{-1}$. Furthermore, the apparent second-order rate for a diffusion-limited enzyme-catalyzed reaction with a single low-molecular mass substrate ($k_{\text{cat}}/K_M$) cannot exceed $\sim 10^6 - 10^7 \text{ M}^{-1} \text{ s}^{-1}$. 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. The overall thermodynamics of a reaction adds further limits. The Haldane relationship states a dependency between the $k_{\text{cat}}/K_M$ of the forward (F) and backward (B) reactions, such that $K_{\text{eq}} = (k_{\text{cat}}/K_M)_F/(k_{\text{cat}}/K_M)_B$, where $K_{\text{eq}}$ is the reaction’s equilibrium constant. Therefore, even when $k_{\text{cat}}/K_M$ in the favorable direction is diffusion-limited, $k_{\text{cat}}/K_M$ 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. One example is the octanol–water partition coefficient (LogP), the ratio of concentrations of a compound partitioning between a hydrophobic...
phase and a water phase. This indicator of hydrophobicity was found to correlate with the binding energy of certain substrates toward their enzymes.\textsuperscript{21–24} In addition, previous studies found a correlation between the molecular mass of ligands and their receptor binding affinities.\textsuperscript{25,26} However, while binding of ligands to receptors is a relatively simple process, the efficiency is dependent on 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.\textsuperscript{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\textsuperscript{29} and KEGG\textsuperscript{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.\textsuperscript{29,31} The KEGG database has also allowed us to assign a metabolic pathway for each reaction. Our final data set contained \textasciitilde5000 unique reactions, utilizing \textasciitilde2500 unique enzymes and \textasciitilde1500 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.\textsuperscript{32,33} These differences can be as great as 3 orders of magnitude.\textsuperscript{34} 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;\textsuperscript{34} on average, $k_{\text{cat}}$ 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 \textasciitilde10 s\textsuperscript{−1} (Figure 1A), where most $k_{\text{cat}}$ values (\textasciitilde60%) are in the range of 1–100 s\textsuperscript{−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 $10^{10}$–$10^{12}$ s\textsuperscript{−1}.\textsuperscript{6,17} The median $k_{\text{cat}}/K_M$ is \textasciitilde10$^3$ M\textsuperscript{−1} s\textsuperscript{−1} (Figure 1B), where most $k_{\text{cat}}/K_M$ values (\textasciitilde60%) lie in the range of $10^3$–$10^5$ M\textsuperscript{−1} s\textsuperscript{−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.\textsuperscript{18} However, Figure 1 suggests that it actually possesses rather average kinetic parameters.

When considering the maximal $k_{\text{cat}}/K_M$ of each enzyme, the median $k_{\text{cat}}/K_M$ is only 5-fold higher ($\sim5 \times 10^4$ M\textsuperscript{−1} s\textsuperscript{−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 $k_{\text{cat}}/K_M$ values of $10^9$–$10^{10}$ M\textsuperscript{−1} s\textsuperscript{−1}. The kinetic parameters for some enzymes could be limited by the rate of substrate–enzyme encounter, as suggested by the viscosity dependence of $k_{\text{cat}}/K_M$\textsuperscript{15} although they display $k_{\text{cat}}/K_M$ values in the range of $10^5$–$10^7$ M\textsuperscript{−1} s\textsuperscript{−1}.\textsuperscript{39,40} In such cases, the low $k_{\text{cat}}/K_M$ values may reflect the recruitment of a rare enzyme conformation that mediates catalytically productive binding.\textsuperscript{39,40}

The median $K_M$ is \textasciitilde100 μM (Figure 1C), where \textasciitilde60% 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

### Table 1. Data Extracted for Our Analysis\textsuperscript{a}

<table>
<thead>
<tr>
<th>filter</th>
<th>no. of $K_M$ values</th>
<th>no. of $k_{\text{cat}}$ values</th>
<th>$K_M$ and $k_{\text{cat}}$ values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brenda, all data</td>
<td>92516</td>
<td>36021</td>
<td></td>
</tr>
<tr>
<td>Brenda, removing mutants</td>
<td>76591</td>
<td>25157</td>
<td></td>
</tr>
<tr>
<td>Brenda, acting on KEGG natural substrates</td>
<td>31162</td>
<td>6530</td>
<td></td>
</tr>
<tr>
<td>unique reactions (enzyme–substrate pairs)</td>
<td>5194</td>
<td>1942 1882</td>
<td></td>
</tr>
<tr>
<td>unique enzymes</td>
<td>2588</td>
<td>1176 1128</td>
<td></td>
</tr>
<tr>
<td>unique substrates</td>
<td>1612</td>
<td>834 804</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{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 $k_{\text{cat}}$ are available.
enzymes toward their substrates, where the vast majority of enzymes (>99%) do not exhibit $K_M$ values below 0.1 $\mu$M. 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 $\sim0.1$ $\mu$M.\textsuperscript{16}

**FIGURE 1.** Distributions of kinetic parameters: (A) $k_{\text{cat}}$ values ($N=1942$), (B) $k_{\text{cat}}/K_M$ 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.

## SELECTIVE PRESSURES MOLD ENZYME KINETICS

Why do most enzymes operate far from the theoretical limits of $k_{\text{cat}}$ and $k_{\text{cat}}/K_M$? 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 measurable, 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 $k_{\text{cat}}$ and $k_{\text{cat}}/K_M$ 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 are 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 $k_{\text{cat}}$ of reactions associated with central-CE metabolism ($79$ s$^{-1}$) is $\sim30$-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 $k_{\text{cat}}/K_M$ value $\sim6$-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 $k_{\text{cat}}$ or $k_{\text{cat}}/K_M$ 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 $k_{\text{cat}}$ values between the metabolic groups are more prominent than the differences in the $k_{\text{cat}}/K_M$ 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

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

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 $K_M$.

■ 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 S.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 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...
ligands to receptors is usually described by the dissociation constant, $K_d$, 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 ligands25,26 (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

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 ($R^2 = 0.13; p < 10^{-3}$, 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 ($R^2 = 0.24; p < 10^{-5}$, slope = 0.25 × 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 1.14.13.1X ($R^2 = 0.10$), oxygenases, acting on two donors, where NADH or NADPH is one donor, and incorporating one atom of oxygen into the other donor.
apparent effective kinetic constants correspond more directly to
the steps in the reaction axis, and \( K_M \) largely reflects the substrate
binding affinity (\( K_S \)). In contrast, EC classes, such as mono-
oxigenases (EC 1.14.13.X), that are characterized by low corre-
lations seem to have complex, multistage mechanisms in which
the apparent \( K_M \) is highly removed from \( K_C \).

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 \( k_{cat}/K_M \).
We wanted to test whether this phenomenon is general. We
therefore compared the \( K_M \) values of enzymes utilizing substi-
tuted 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, modi-
fiers might have a common functionality in enhancing the affinity
of enzymes for the substituted substrates. For example, a micro-
molar \( K_M \) 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 \( K_S \) values for the respective CoA-
substituted substrates are in the range of 1–100 \( \mu \)M, indicating
a decrease of ∼2 orders of magnitude upon CoA substitution.
The cellular concentrations of acetate and propionate are in the
range of 1 \( \mu \)M to 1 mM. 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 pre-
sented 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 para-
meters and suggest that within certain physicochemical con-
straints, 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 improve-
ments, suggesting that this enzyme operates close to the
theoretical maximal efficiency. To counter the Rubisco exam-
ple, 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 efficien-
cies 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 en-
zymes, \( k_{cat}/K_M \) values that are <10^-3 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.

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,
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 inher-
ently 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
different enzymes from different origins. In many cases, more
than one solution for a given metabolic goal exists, and each of
these solutions employs a different set of enzymes. 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 phos-
phoryl 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 as-
sumed by some of these models. For example, optimality analysis
previously suggested that under the assumptions of an evolu-
tionary 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 concentra-
tions compared to those with weaker control. 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 $k_{\text{cat}}$, arising from the decreasing probability of finding a mutant with higher $k_{\text{cat}}$ values (Supporting Information). Just as enzyme concentrations increase until the benefit and cost balance each other, $k_{\text{cat}}$ values increase until the possibility of achieving even larger $k_{\text{cat}}$ 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 $k_{\text{cat}}$ value further into the low-probability range of higher values, while a weaker selective pressure would result in lower $k_{\text{cat}}$ 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 $k_{\text{cat}}$ 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 $k_{\text{cat}}$ values.

Another interesting observation on the global scale analyzed here is that regardless of the microscopic mechanisms, the median $k_{\text{cat}}/K_M$ observed is $\sim 10^8$ 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 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 $k_{\text{cat}}$ 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,70–75 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.

### Funding Sources

A.B.-E. is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. E.N. is supported by the Azrieli Foundation and the Kahn Center for System Biology. R.M. is supported by research grants from the ERC (260392 - SYMPAC), Israel Science Foundation (Grant 750/09), and the Mr. and Mrs. Yossie Hollander and Miel de Botton Aynsley funds. R.M. is the incumbent of the Anna and Maurice Boukstein career development chair, and D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry. D.S.T. is supported by the EU BioModular H2 network.

### ACKNOWLEDGMENT

We thank Asap Aharoni for help with the metabolic modules affiliation and Naama Barkai and her lab members, Pedro Bordalo, Avigdor Eldar, Deborah Fass, Avi Flamholz, John Gerlt, Yuval Hart, Amnon Horovitz, Roy Kishony, Ronny Neumann, Irlienka Nobeli, Guy Polturak, Irit Sagi, Gideon Schreiber, and Tomer Shlomi for helpful discussions. We thank Steve Brudz and the ChemBack team, Peter Erth, John Irwin, and Vladimir Sobolev for help with the analysis of the substrates’ physical properties.

### 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

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


