# Enzyme engineering: reaching the maximal catalytic efficiency peak

Moshe Goldsmith and Dan S Tawfik

The practical need for highly efficient enzymes presents new challenges in enzyme engineering, in particular, the need to improve catalytic turnover ($k_{cat}$) or efficiency ($k_{cat}/K_M$) by several orders of magnitude. However, optimizing catalysis demands navigation through complex and rugged fitness landscapes, with optimization trajectories often leading to strong diminishing returns and dead-ends. When no further improvements are observed in library screens or selections, it remains unclear whether the maximal catalytic efficiency of the enzyme (the catalytic 'fitness peak') has been reached; or perhaps, an alternative combination of mutations exists that could yield additional improvements. Here, we discuss fundamental aspects of the process of catalytic optimization, and offer practical solutions with respect to overcoming optimization plateaus.

**Address**
Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Corresponding author: Goldsmith, Moshe (moshe.goldsmith@weizmann.ac.il)

## Introduction

It is commonly assumed that all enzymes were born poor catalysts and were subsequently optimized by evolution. Laboratory experiments that mimic this process have now become a matter of routine. In some cases, these experiments attempt to reproduce the evolutionary emergence of a natural enzyme from its putative ancestor. However, in most cases, the target of enzyme optimization is primarily applicative — to create a highly active and stable enzyme that can catalyze the target reaction in a non-biological setting. Natural enzymes present two common limitations: first, with few exceptions, natural enzymes exhibit low catalytic efficiency with non-cognate substrates that are typically application-relevant (i.e. substrates that differ from the

enzyme's natural substrate yet are promiscuously transformed by it). Second, natural enzymes exhibit low protein stability especially under applicative conditions. The feasibility and cost of application depend on the catalytic turnover of the enzymes being high. To this end, directed evolution and other protein engineering methods such as computational design are used to optimize enzymes for a variety of practical applications spanning from organic synthesis to therapeutics (for recent reviews, see Refs [1–5]).

Here, we summarize several key aspects regarding the laboratory optimization of enzymatic traits, and specifically of catalytic efficiency ($k_{cat}/K_M$, or $k_{cat}$ for enzymes working under substrate saturation [6]). The last decade has seen a leap in the understanding of how enzymes evolve. However, several key questions still prevail, especially with respect to how enzymes can be optimized toward high, let alone maximal catalytic efficiency (in fact, what 'high' or 'maximal' means is a complex issue, as discussed below). We focus on optimization of catalytic efficiency, primarily by directed evolution. There are fundamental differences between evolution in nature and in the laboratory. Nonetheless, lessons from natural evolution can be implemented in enzyme engineering, and the latter also teaches us about the former.

## The optimization challenge

The catalytic efficiencies of natural enzymes with non-cognate substrates can be extremely low; reported $k_{cat}/K_M$ values as low as $1 \ M^{-1} \ s^{-1}$ are not an exception (e.g. [7•,8–10]). In contrast, the average catalytic efficiency value ($k_{cat}/K_M$) of natural enzymes with their cognate substrates is $\sim 10^5 \ M^{-1} \ s^{-1}$, and some enzymes approach $10^9 \ M^{-1} \ s^{-1}$ [11]. Thus, the gap between the catalytic efficiency with a cognate substrate versus a promiscuous, non-cognate substrate may span several orders of magnitude. Natural evolution can readily bridge such gaps. For example, xenobiotics such as the pesticide metabolite paraoxon are promiscuously hydrolyzed by many natural enzymes, although at very different rates. Bacterial lactonases known as PLLs exhibit promiscuous paraoxonase activities with $k_{cat}/K_M$ values that span over 4 orders-of-magnitude (from 0.5 up to $4 \times 10^3 \ M^{-1} \ s^{-1}$) [12]. A natural paraoxonase named PTE (phosphotriesterase), had diverged from an unidentified PLL. This enzyme evolved in soil bacteria, in a matter of few decades, to hydrolyze paraoxon with a rate that approaches diffusion limit ($k_{cat}/K_M \sim 10^8 \ M^{-1} \ s^{-1}$) [13].

Directed evolution and computational redesign of natural enzymes have proven capable of bridging equally large gaps; possibly even greater ones when applying starting points that exhibit no detectable activity with the target substrate (e.g. Refs [14•,15•]). However, most publications describe only modest improvements of up to two orders of magnitude in $k_{cat}/K_M$ (Figure 1). Large improvements are rare, and increases in $k_{cat}/K_M$ of $\geq 10^4$-fold comprise only ∼5% of our literature sampling (Figure 1). Improving an enzyme's catalytic efficiency with a non-cognate substrate by an order of magnitude or two typically requires only a few rounds of directed evolution, especially when the initial catalytic efficiency is relatively low. In contrast, efforts to bridge large gaps by directed evolution are likely to encounter diminishing returns and optimization plateaus, and thus require many rounds of directed evolution with no guarantee of success. The correlation between fold-improvement and the number of introduced mutations is not strictly linear, but improvements of above 1000-fold typically demand at least 10 mutations (Figure 2).

## The landscape of evolutionary optimizations

The search for a combination of beneficial mutations that would yield an optimized enzyme is usually depicted as a sequential advancement in protein sequence-space, where each step is associated with a change in fitness. An in-depth discussion of this topic is beyond the scope of this review, but for the purposes of this discussion, fitness landscapes can be illustrated by a simplified 3D space (Box 1). In the simplest scenario, there exists a trajectory that is both continuous and gradual — an uphill climb to the 'fitness peak' that comprises a series of mutations, each of which provides a distinct fitness advantage (Box 1, trajectory A).

The fitness peak represents the maximal possible catalytic efficiency of an enzyme for a particular reaction and substrate in a given region of sequence space. In theory, an unlimited exploration of protein sequence space will reveal the 'global fitness peak', which represents the maximal $k_{cat}/K_M$ value possible for a given enzyme, reaction and substrate. However, the number and heights of local fitness peaks that exist for a particular enzyme and substrate are unknown, and at present are also impossible to predict. Given the vastness of the theoretical sequence space, only a minute portion of it is accessible to natural evolution, let alone to laboratory evolution. The size of sequence-space explored by natural evolution is restricted by factors such as the frequency of mutations, their types and the size of the evolving population. Similarly, in laboratory evolution, the diversity of the applied gene library (i.e. the number of mutations per gene, their types, the number of variants per library), and the throughput of the screening/selection method and its stringency, dictate what fraction of sequence space will be explored per round.

**Figure 1**
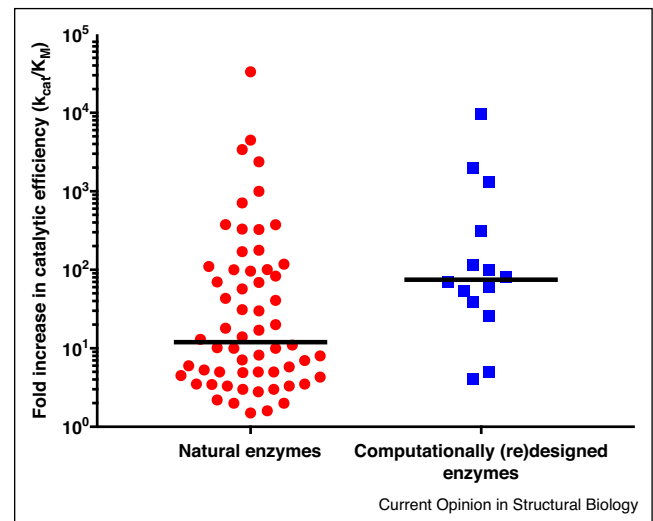


Laboratory optimizations of catalytic efficiency. Shown are the fold-improvements in catalytic efficiency ($k_{cat}/K_M$) reported in the period of 2012-2016 for 60 natural enzymes optimized by directed evolution (red circles) [22•,43•,45,46•,73–78,79•,80–109,110•,111–128], and for 14 enzymes that were computationally designed or redesigned and further optimized by directed evolution (2009–2016; blue squares) [7•,44,50,129–139]. With few exceptions, the data describes the optimization of different enzymes. Indicated are the improvements for the most catalytically efficient variant compared to its starting point. Only studies that reported catalytic efficiency values of purified proteins, and described the number of directed evolution rounds and incorporated mutations, were included. The black, horizontal bars indicate the median fold improvement.

**Figure 2**
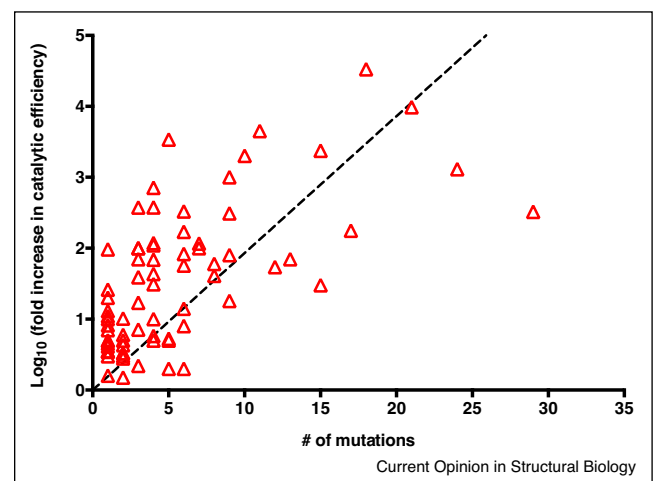


Large increases in catalytic efficiency demand multiple mutations. Plotted are the $\log_{10}$ values of the fold-increases in $k_{cat}/K_M$ of evolved enzymes versus the number of non-synonymous mutations in these variants (dataset of Figure 1). The line represents a linear fit ($y = 0.1931X \pm 0.01461$) suggesting that large improvements (>$10^3$-fold) demand on average 5 mutations per order-of-magnitude improvement in catalytic efficiency.