Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Enzyme evolution: innovation is easy, optimization is complicated

Matilda S .Newton[1], Vickery L Arcus[2], Monica L Gerth[3] and Wayne M Patrick[3]

Enzymes have been evolving to catalyze new chemical reactions for billions of years, and will continue to do so for billions more. Here, we review examples in which evolutionary biochemists have used big data and high-throughput experimental tools to shed new light on the enormous functional diversity of extant enzymes, and the evolutionary processes that gave rise to it. We discuss the role that gene loss has played in enzyme evolution, as well as the more familiar processes of gene duplication and divergence. We also review insightful studies that relate not only catalytic activity, but also a host of other biophysical and cellular parameters, to organismal fitness. Finally, we provide an updated perspective on protein engineering, based on our new-found appreciation that most enzymes are sloppy and mediocre.

Addresses
[1] BioTechnology Institute, University of Minnesota, St. Paul, MN, United States
[2] School of Science, University of Waikato, Hamilton, New Zealand
[3] Department of Biochemistry, University of Otago, Dunedin, New Zealand

Corresponding author: Patrick, Wayne M (wayne.patrick@otago.ac.nz)

Current Opinion in Structural Biology 2018, **48**:110–116

This review comes from a themed issue on **Proteins**

Edited by **Birte Hocker** and **Jakob Winther**

https://doi.org/10.1016/j.sbi.2017.11.007

Introduction

New enzymes have been evolving on Earth for at least four billion years, and will continue to do so for another two billion or so — at which point the expanding sun will sterilize our planet [1]. The goal of this article is to review recent studies that shed new light on enzyme evolution, with a focus on work published since 2015.

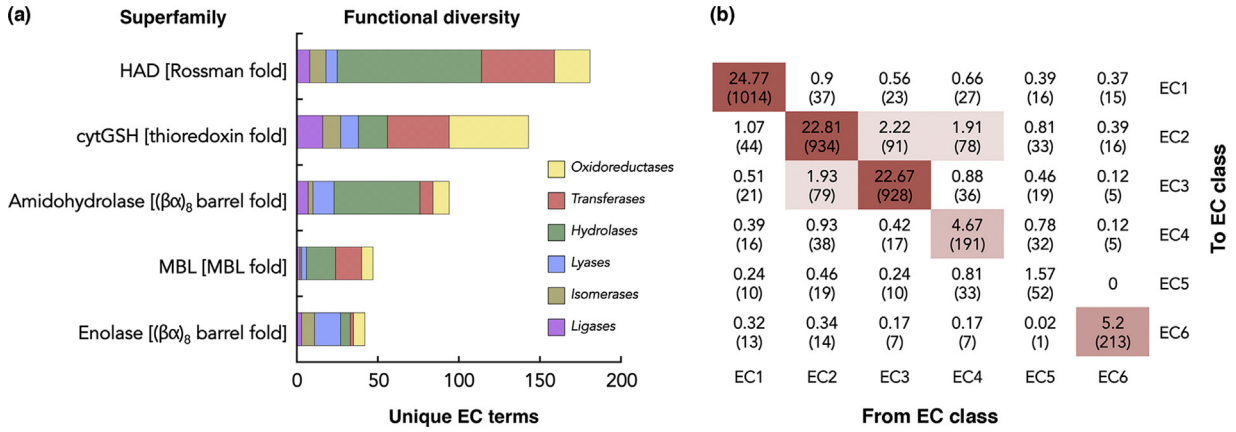
Innovation is easy

A general model for the evolution of enzymes with new functions was articulated by Yčas and Jensen, independently, in the mid-1970s [2,3]. Each proposed that ancestral enzymes were multifunctional generalists, with the ability to catalyze broad classes of reactions on a range of substrates. From this low-fidelity starting point, gene duplication and divergence would have given rise to more specialized enzymes with higher activities towards their preferred substrates.

The Yčas–Jensen model had two important implications. First, divergent evolution of new enzymes was most likely to be enabled, and constrained, by catalytic chemistry. Gerlt and Babbitt were among the first to emphasize the importance of 'chemistry driven' evolution from multifunctional ancestors, giving rise to superfamilies of homologous enzymes. As originally defined, the members of these superfamilies share the same fold and either catalyze the same reaction with different substrate specificities, or catalyze different overall reactions that share a common mechanistic feature such as a partial reaction, an intermediate or a transition state [4]. The second implication of the Yčas–Jensen model was that the promiscuous (secondary, non-physiological) activities of existing enzymes remain important starting points for the evolution of new functions, because today's enzymes are tomorrow's ancestors. It is now well accepted that most — and probably all — extant enzymes are, in fact, promiscuous [5,6].

Recent large-scale studies, both computational and experimental, have opened our eyes to the enormous functional diversity among existing enzyme superfamilies, the vastness of 'promiscuity space,' and therefore the seemingly limitless potential for future evolutionary innovation. Baier *et al.* surveyed the functional diversity, as represented by Enzyme Commission (EC) numbers, in five common superfamilies [7•]. Each superfamily contained enzymes from all six of the EC classes (Figure 1a). Furnham *et al.* went further and used a phylogenetic approach [8] to reconstruct the evolutionary histories of 379 superfamilies from the Class, Architecture, Topology, Homology (CATH) database, and to ask how often a change in EC number was observed over the course of their evolution [9•]. While 81% of the functional changes were within an EC class, every possible change between EC classes was also observed (Figure 1b), with the exception of a change from a ligase (EC class 6) to an isomerase (EC class 5). These bioinformatics studies emphasize that there is little, if anything, that constrains particular catalytic chemistries to particular folds.

Figure 1



Functional diversity in enzyme superfamilies. **(a)** Five abundant superfamilies, each with 13 000–91 000 representatives in the sequence databases, are HAD (haloalkanoate dehalogenase), cytGST (cytosolic glutathione transferase), amidohydrolase, MBL (metallo-β-lactamase), and enolase. Each of these superfamilies contains homologous enzymes that fall into all six EC classes. Adapted with permission from [7•]. Copyright (2016) American Chemical Society. **(b)** An EC exchange matrix, counting the changes from one EC number to another during the evolution of 379 different superfamilies. Counts are expressed as a percentage of the total number observed, with the raw numbers of exchanges in parentheses. Colouring is on a red intensity scale. Reproduced with modifications from [9•].

Four high-throughput experimental studies (reviewed in detail elsewhere [7•]) have reached a similar conclusion. Dozens of enzymes from within the cytosolic glutathione transferase [10], β-keto acid cleavage enzyme [11], metallo-β-lactamase [12], and haloalkanoate dehalogenase [13••] superfamilies were each tested for activity towards a range of different substrates. In each case, many enzymes were found to have multiple functions *in vitro*. In the most comprehensive study, 217 members of the haloalkanoate dehalogenase superfamily were expressed, purified, and screened for phosphatase or phosphonatase activity towards 167 substrates (most of which were naturally occurring metabolites). The authors discovered breathtakingly broad substrate specificities. A median of 15.5 substrates were recognized by each enzyme, 50 of the enzymes could utilize 40 or more substrates, and remarkably, one enzyme could utilize 143 [13••].

Together, these computational and experimental studies highlight the genuine risk in using homology to assign physiological function(s) to uncharacterized proteins in databases [14]. A further caveat with *in vitro* experiments is that it can be difficult to elucidate which activities are physiological (being maintained by selection) and which are promiscuous. Even in cases where an enzyme appears to have a clear-cut physiological role, it is theoretically possible that one or more of its weak side activities may be contributing to the fitness of the organism — either by contributing to the metabolite pool, or by inducing a regulatory effect. Regardless, the old idea of 'one enzyme, one substrate' is now shown to be quaint and outdated. The leading evolutionary biochemist, Prof Shelley Copley, has made the entirely reasonable estimate that an average enzyme may have 10 promiscuous activities [6]. Thus, even the simplest bacteria are likely to harbour 10 000–20 000 promiscuous activities, any one of which may be the starting point for the evolution of a new enzyme. Not only that, but two upcoming studies have retraced the evolution of enzymes from their non-catalytic ancestors, via a small number of key mutations in each case [15,16]. When non-enzymatic scaffolds are also considered, there is certainly no shortage of possibilities for future evolutionary innovation!

Gene loss drives functional innovation, too

The Yčas–Jensen model and its descendants (e.g. [17,18]) are centred on the importance of gene duplication and divergence as the evolutionary route to new enzymes. However, genome reduction is also a pervasive force in evolution. Every lineage, apart from that tiny fraction leading to extant animals and plants, appears to undergo rapid bursts of genomic complexification, followed by much longer periods in which genetic material is slowly lost [19]. Two recent studies have combined phylogenomics and biochemistry to examine how gene loss can shape enzyme evolution.

Juárez-Vázquez *et al.* [20••] continued their groundbreaking research into the evolution of PriA, a bifunctional isomerase that catalyzes the HisA and TrpF reactions (in histidine and tryptophan biosynthesis, respectively) in some bacteria [21–23]. Extensive phylogenomic analysis and the construction of genome-scale metabolic models for 33 bacteria led to the identification of PriA homologues that were predicted to fulfil different roles (bifunctional, HisA-only, or TrpF-only), depending on the pattern of gene loss in the host organism (Figure 2a).