ADVANCED DATABASE MINING INTEGRATING SEQUENCE AND STRUCTURE BIOINFORMATICS WITH MICROFLUIDICS CHALLENGES ENZYME ENGINEERING

Zbynek Prokop, Masaryk University, Czech Republic
zbynek@chemi.muni.c
Michal Vasina, Pavel Vanacek, David Kovar, Antonin Kunka, David Bednar, Jiri Damborsky, Masaryk University; St. Ann’s Hospital, Czech Republic
Jiri Hon, St. Ann’s Hospital, Czech Republic
Hana Faldynova, Stanislav Mazurenko, Masaryk University, Brno, Czech Republic
Tomas Buryska, Stavros Stavrakis, Andrew deMello, ETH Zürich, Switzerland
Christoffel P. S. Badenhorst, Uwe T. Bornscheuer, Greifswald University, Germany

Keywords: bioinformatics, microfluidics, enzyme, catalysis,

The growing popularity of industrial enzyme applications increases the demand for discovering new biocatalysts with industrially useful properties. The fundamental question is how to get better biocatalysts? Shall we explore the natural sequences or improve available enzymes with protein engineering tools? Outstanding progress has taken place thanks to the genomics revolution. The avalanche of protein sequences, which now fills the databases at a rocket pace, represents a vast potential, bringing new challenges to its practical utilization. Only a negligible fraction of gene sequences deposited in databases has been experimentally characterized.

Incorrect automatic annotations are common and tend to percolate, leading to error accumulation in the databanks [1]. Without advanced bioinformatic expertise, relying on database annotations, many projects dedicated to finding new biocatalysts do not succeed, even after large investments and the application of high-throughput screening campaigns [2]. This can lead to overlooking or underestimations of the potential of natural diversity hidden in sequence databases.

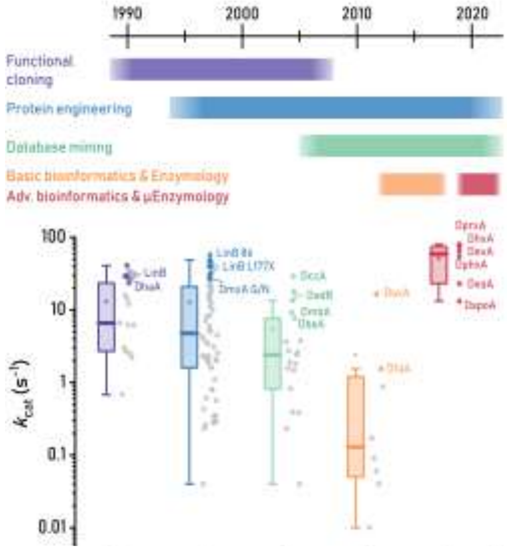

Figure 1. Strategies for biocatalyst discovery illustrated with the HLD family. Functional cloning (purple), protein engineering (blue), database mining (green), bioinformatics & enzymology (orange), advanced bioinformatics & microfluidic enzymology (red). Box charts compare turnover numbers for enzyme variants obtained by individual strategies. The box shows median (line), mean (small square), quartiles, minima, and maxima.

Herein, we will present advanced database mining of novel biocatalysts by combining automated sequence and structural bioinformatics [3] and microfluidic enzymology methods for efficient experimental characterization [4]. In a single run of this workflow, we doubled the number of experimentally characterized members of a model enzyme family, haloalkane dehalogenases (HLDs), and obtained biocatalysts catalytically surpassing the previously known variants whether discovered or engineered.

We compared the biocatalytic effectivity of variants obtained by this advanced database mining with enzymes previously isolated by classical enzymological approaches, as well as with the variants systematically constructed for more than 20 years by various protein engineering strategies (Figure 1), including optimization of active sites, redesigning access tunnels, engineering dynamical protein loops or resurrecting ancestral sequences. Our study provides an interesting conceptual view of current approaches used in biocatalyst development, and the substantial potential of natural sequence diversity.

References:

1. Gilks W. R., et al. 2005. Percolation of annotation errors through hierarchically structured protein sequence databases. *Mathematical Biosciences* 193, 223–234.
2. Schnoes A. M., et. al. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLOS Computational Biology* 5, e1000605.
3. Hon J., et al. 2020. EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities. *Nucleic Acids Research* 48 (W1), W104–W109.
4. Vasina M., et al. 2020. Exploration of enzyme diversity: High-throughput techniques for protein production and microscale biochemical characterization. In Methods in Enzymology Enzyme Engineering and Evolution: General Methods., D. S. Tawfik, ed., Academic Press, pp. 51–85.