NEW TECHNOLOGIES FOR ENZYME ENGINEERING: COMBINING COMPUTATIONAL PREDICTIONS AND AUTOMATED EXPERIMENTAL FEEDBACK

Uwe Jandt, Hamburg University of Technology, Germany
uwe.jandt@tuhh.de
Cornelius Jacobi, Hamburg University of Technology, Germany
Sibel Ilhan, Hamburg University of Technology, Germany
Sukanya Sekar, Hamburg University of Technology, Germany
Wael Sabra, Hamburg University of Technology, Germany
An-Ping Zeng, Hamburg University of Technology, Germany

Key Words: simulation-experimental feedback loop, machine learning, fourth wave of enzyme engineering

The targeted design and optimization of novel enzymes and enzymatic reaction cascades increasingly demands a close connection between rational design, computational prediction and experimental feedback. In recent years, lots of effort have been put on increasing the throughput of experimental results, however, this approach frequently tends to stick in local minima and unsatisfying performance improvement despite considerable screening efforts. Contrary, model-based computational predictions, despite increasing available computation power, need to introduce severe simplifications and therefore will continue to lack accuracy and perfect predictability in the foreseeable future. The interplay of thorough model-based understanding, automated experimental feedback and, based on the latter, refinement of model predictions using for example machine learning methods, will in the near future become an important approach to combine the best of the two worlds. Ultimately, this provides potential to boost highly efficient automated or semi-automated design of new enzymatic properties in the scope of a “fourth wave” of enzyme engineering.

We present a new integrated directed evolution framework to achieve this simulation-experimental feedback loop, called “Feedback Guided Enzyme Optimization” (FEO). The implementation includes the setup of a suitable simulation back-end, robot-based experimental generation of mutants and evaluation of their performance [1], and finally feedback to the simulation in order to close the loop and verify and refine the quality of the predictions.Focus is laid on thorough statistical analysis of both prediction and experimental results, in order to tune false positive vs. false negative error rate, depending on experimental conditions: This includes, e.g., availability of time, ingredients, parallel workflows and distortions (random noise and potential systematic deviations) in both experimental and simulation setups.

The framework is being implemented in an automated robotic setup. We demonstrate results on three exemplary enzymatic systems: Firstly, GFP is employed as a simple role model to demonstrate the looping principle. The second example, aspartokinase III (AK3), is a key enzyme for the biosynthetic production of amino acids and derivatives thereof. Its activity is naturally limited by its own downstream products, e.g., lysine. Simulated predictions of the sensitivity of AK3 towards lysine have been compared to experimental data. This allowed a significant ($p<0.05$) simulation-based discrimination of highly resistant versus non-resistant variants. Determination of new lysine resistant mutants by multiple point mutations is performed within few dozen of iterations. The obtained candidates were validated, showing that new Lys-resistant variants can be obtained using the new workflow without special a priori knowledge or extensive (random) screening.

The third and most sophisticated enzyme system is the pyruvate dehydrogenase complex (PDC) which involves interesting features like shielding of reaction intermediates, renewal of co-factors, self-assembly, modularity and others. Based on recently published models of PDC by our group [2-3] and in collaborations [4], we demonstrate how the dynamic self-assembly of mutants of PDC and structurally similar enzymes complexes can be predicted, iteratively refined and in the future used for the creation of new enzyme cascades.

This presented framework is expected to have large impact on design and evolution of novel biomolecules and biosystems.

[1] Wurm, M., Ilhan, S., Jandt, U., Zeng A.-P. (2019). Direct and Highly Sensitive Measurement of Fluorescent Molecules in Bulk Solutions using Flow Cytometry. Analyt. Biochem. 570, 32-42.

[2] Hezaveh, S., Zeng, A. P., Jandt, U. (2018). Enzyme Interaction in Human Pyruvate Dehydrogenase Complex: Full Complex Simulation. Journal of Chemical Information and Modeling, 58(2), 362-369.

[3] Hezaveh, S., Zeng, A. P., Jandt, U. (2017). Investigation of Core Structure and Stability of Human Pyruvate Dehydrogenase Complex: A Coarse-Grained Approach. ACS Omega, 2(3), 1134-1145.

[4] Depta, P.N., Jandt, U., Dosta, M., Zeng, A.-P., Heinrich, S. (2019). Toward Multiscale Modeling of Proteins and Bioagglomerates: An Orientation-Sensitive Diffusion Model for the Integration of Molecular Dynamics and the Discrete Element Method. J. Chem. Inf. Model. 59(1), 386-398.