COMPUTATIONAL PROTEIN DESIGN TO ACCELERATE THE CONCEPTION OF FINE-TUNED BIOCATALYSTS

Sophie Barbe, LISBP, INSA, CNRS, INRA, Université de Toulouse, Toulouse, France
sophie.barbe@insa-toulouse.fr
Simoncini David, LISBP, INSA, CNRS, INRA, Université de Toulouse, Toulouse, France
Clément Viricel, LISBP, INSA, CNRS, INRA, Université de Toulouse, Toulouse, France
Antoine Charpentier, LISBP, INSA, CNRS, INRA, Université de Toulouse, Toulouse, France
Seydou Traoré, LISBP, INSA, CNRS, INRA, Université de Toulouse, Toulouse, France
Juan Cortés, CNRS LAAS, avenue du Colonel Roche BP 54200 Toulouse 31031 Cedex 4, France
Isabelle André, LISBP, INSA, CNRS, INRA, Université de Toulouse, Toulouse, France
David Allouche, INRA MIAT, UR 875, Castanet-Tolosan, 31326 Cedex, France
Thomas Schiex, INRA MIAT, UR 875, Castanet-Tolosan, 31326 Cedex, France

Key Words: Computational enzyme design, In silico rational enzyme engineering, Artificial Intelligence, Combinatorial optimization algorithms, Large enzyme sequence-conformation search space.

The remarkable properties of enzymes (high catalytic efficiency, regio- and stereo-selectivity) have been recognized and largely exploited in biocatalysis. Accordingly, enzyme-driven processes should play an increasing role in the next decades, potentially substituting chemical processes and contributing to the raise of bioeconomy. However, to foresee a viable future to biocatalysis, advances in R&D are required to accelerate the delivery of fine-tuned enzymes displaying high chemical specificity on non-cognate substrates, high efficiency and better stability in reaction conditions. To this end, structure-based Computational Protein Design (CPD) is a promising strategy to fully rationalize and speed-up the conception of new enzymes while reducing associated human and financial costs.

By combining physico-chemical models governing relations between protein amino-acid composition and their 3D structure with optimization algorithms, CPD seeks to identify sequences that fold into a given 3D-scaffold and possess the targeted biochemical properties. Starting from a huge search space, the protein sequence-conformation space, this in silico pre-screening aims to considerably narrow down the number of mutants tested at experimental level while substantially increasing the chances of reaching the desired enzyme. While CPD is still a very young and rapidly evolving field, success stories of computationally designed proteins highlight future prospects of this field. Nonetheless, despite landmark achievements, the success rate of the current computational approaches remains low, and designed enzymes are often way less efficient than their natural counterparts. Therefore, several limitations of the CPD still need to be addressed to improve its efficiency, predictability and reliability.

Herein, we present our methodological advances in the CPD field that enabled overcoming technological bottlenecks and hence propose innovative CPD methods to explore large sequence-conformation spaces while providing more accuracy and robustness than classical approaches. Our CPD methods speed-up search across vast sequence-conformation spaces by several orders of magnitude, find the minimum energy enzyme design and generate exhaustive lists of near-optimal sequences, defining small mutant libraries. These new methods, in rupture with classical approaches are based on efficient algorithms issued from recent research in artificial intelligence. The performance and accuracy of our computer-aided enzyme design methods have been evaluated and validated on various types of protein design problems.

This work was partially funded by INRA/Région Midi-Pyrénées, IDEX Toulouse, Agreenskills and the French National Research Agency (PROTICAD, ANR-12-MONU-0015-03).