

SUCCESSFUL EXAMPLES OF THE APPLICATION OF NOVEL ITERATIVE TRAINABLE ALGORITHMS TO GUIDE RATIONAL MUTATION STRATEGIES FOR ENZYME ENGINEERING: FROM PREDICTION TO LAB TESTING TO ALGORITHM RETRAINING

Alvaro Olivera-Nappa, Centre for Biochemical Engineering and Biotechnology – CeBiB, Department of Chemical Engineering and Biotechnology, University of Chile
aolivera@ing.uchile.cl

Barbara A. Andrews, Centre for Biochemical Engineering and Biotechnology – CeBiB, Department of Chemical Engineering and Biotechnology, University of Chile

Juan A. Asenjo, Centre for Biochemical Engineering and Biotechnology – CeBiB, Department of Chemical Engineering and Biotechnology, University of Chile

Key Words: Rational protein design, protein engineering, rational protein evolution, mutation selection, bioinformatics algorithms.

Both natural mutations occurring in a homologous enzyme family and mutations engineered in a given protein can have a tremendous impact in the activity and binding behavior of the enzyme towards substrates or other molecules. Binding and catalytic properties can be modified by rationally mutating selected amino acids in a protein. For instance, new specificity properties can be engineered into existing enzymes, which can be applied to the rational design of mutations to alter its catalysis. Although this approach has been largely used, the modifications introduced in the target protein have not been exempt of deleterious effects on protein function, binding or physicochemical properties. Much finer tuned modifications should be designed in order to alter the desired catalytic or binding properties of a protein and simultaneously not affecting other protein properties or functions. These engineered mutations usually require a thorough knowledge of the relevant structure-function relationships in the protein molecule. If no precise structure-function information is available for a protein, the amount of possible amino acid mutations to be tested precludes a direct search. Furthermore, in many cases a directed evolution strategy cannot be successfully used to achieve the desired results due to the unavailability of suitable screening tests.

In the last years, we have developed new and powerful *in silico* methodologies to automatically propose, test and redesign mutagenesis strategies for a target protein, based only on evolutionarily conserved physicochemical properties of amino acids in a protein family where the target protein belongs, and on structural properties, including calculation of vibrational entropies, if available, with no need of explicit structure-function relationships. This methodology identifies amino acid positions that are putatively responsible for function, specificity, stability or binding interactions in a family of proteins and calculates amino acid propensity and distributions at each position. Not only conserved amino acid positions in a protein family can be labelled as functionally relevant, but also non-conserved amino acid positions can be identified to have a meaningful functional effect, and even amino acid substitutions that are unobserved in nature. These results can be used to predict if a given mutation can have a functional implication and which mutation is most likely to be functionally silent for a protein. Through several rounds of mutation suggestions, laboratory testing of the mutants and feedback of results to retrain the algorithms, our methodology can be used to rapidly and automatically discard any irrelevant mutation and guide the research focus toward functionally significant mutations.

In this work, we will show how we have successfully used our publicly available methods to guide mutant design in enzyme engineering applied to xylanases (producing an improved octuple mutant in a single mutagenesis round), proteases, glucanases, ubiquitin ligases and other enzymes, to alter protein function, stability or thermodynamic properties independently of their catalytic properties *in vitro* and *in vivo*. We will also show how the predictions of these methods have been employed to shift chromatographic elution profiles of xylanases and ferritin nanocages for better purification without affecting their activity and to obtain ferritin variants with better properties to be used in nanotechnological applications, including modifications to the external and internal surface of the protein to change its interaction properties, improve its recombinant production, alter the characteristics of nanoparticles within or change its organic molecule carrier capacity. Finally, we will show how a similar approach has been integrated in an artificial intelligence classification scheme to identify somatic mutations in the human VHL gene that are related to renal clear-cell cancer and to predict the clinical outcome and prognosis of pVHL mutation and malfunction in humans, based on specific disruption of interactions with VHL binding partners.

Clearly, our techniques show promising performance as a valuable and powerful bioinformatics tool to aid in the computer-aided design of engineered enzyme variants and in the understanding of function-structure, binding and affinity relationships in enzymes and other proteins.