VALIDATION AND STABILIZATION OF A PROPHAGE LYSIN OF *CLOSTRIDIUM PERFRINGENS* BY YEAST SURFACE DISPLAY AND CO-EVOLUTIONARY MODELS

Ben Hackel, University of Minnesota, USA
hackel@umn.edu
Seth Ritter, University of Minnesota, USA

Key Words: antimicrobial proteins, protein engineering, co-evolutionary models, homolog

Bacteriophage lysins exhibit high specificity and activity towards host bacteria with which the phage co-evolved. These properties of lysins make them attractive for use as antimicrobials. Though there has been significant effort to develop platforms for rapid lysin engineering, there have been numerous shortcomings when pursuing the ultra-high throughput necessary for discovery of rare combinations of mutations to improve performance. In particular, the biotechnological utility and evolvability of lysins would be aided by elevated stability. Lysin catalytic domains, which evolved as modular entities distinct from cell wall binding domains, can be classified into one of several families with highly conserved structure and function, many of which contain thousands of annotated homologous sequences. Motivated by the quality of this evolutionary data, the performance of generative protein models incorporating co-evolutionary information was analyzed to predict the stability of variants in a collection of 9,749 multi-mutants across 10 libraries diversified at different regions of a putative lysin from a prophage region of a *Clostridium perfringens* genome. Protein stability was assessed via a yeast surface display assay with accompanying high-throughput sequencing. Statistical fitness of mutant sequences, derived from second-order Potts models inferred with different levels of sequence homolog information, was predictive of experimental stability with AUCs ranging from 0.78 to 0.85. To extract an experimentally derived model of stability, a logistic model with site-wise score contributions was regressed on the collection of multi-mutants. This achieved a cross-validated classification performance of 0.95. Using this experimentally derived model, 5 designs incorporating 5 or 6 mutations from multiple libraries were constructed. All designs retained enzymatic activity with 4 of 5 increasing melting temperature, with the highest performing design achieving an improvement of +4 °C. In addition to validation of a putative lysin and stabilization thereof, the experimental and computational methods presented herein offer a new avenue for improving protein stability and is easily scalable to analysis of tens of millions of mutations in single experiments.

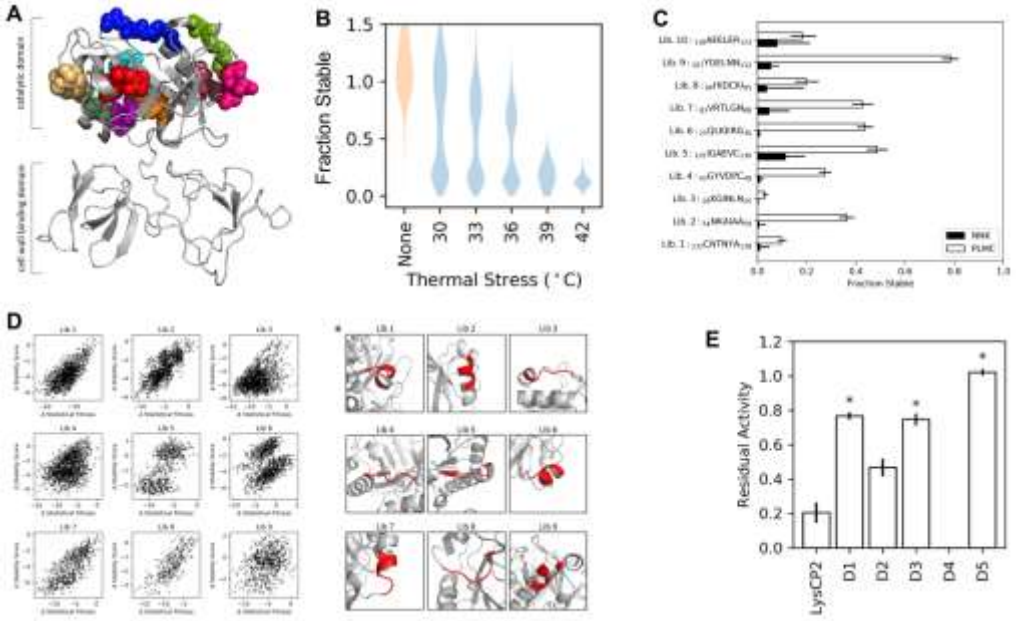

Figure 1. (A) Bacteriophage lysin LysCP2 homology model with diversified sites highlighted. (B) LysCP2 libraries exhibit a broad distribution of stability; high and low stability variants were selected by flow cytometry. (C) The co-evolutionary model yielded more stable lysin variants than a random control library. (D) The predictive quality of the experimentally-derived mutant model was strong at select locations (sub-libraries). (E) Second-generation designed mutants exhibit substantially improved activity after thermal stress relative to wild-type.