Title: Mutations for Improved Enzyme Functionality at High Temperatures and Low pH

Authors: Ambuj Kumar[1*] (ambuj@iastate.edu), Kejue Jia,[1] Pranav M. Khade,[1] Onyeka Onyenemezu,[1] Jetendra K. Roy,[1] Robert L. Jernigan,[1] **Laura Jarboe**[1]

Institutions: [1]Iowa State University

Project Goals: The initial goal is to computationally screen protein mutants to identify ones that will be efficient at high temperatures and low pH for greater fermentation efficiency. This will be followed by experimental selection for the best of these.

Abstract Text: Amino acid mutations can introduce functional changes ranging from loss of function to adaptation to extreme environmental conditions. Mutations have frequently been used to design thermostable proteins to retain function at high temperatures. Furthermore, predictions of damaging protein mutations have been widely applied in cancer research and drug discovery. Here we are designing a new machine learning classifier with the xgboost algorithm [1] to estimate mutation-induced changes in protein stability (both stabilizing and destabilizing) for a wide range of temperatures and pHs. A total of 4302 mutation induced protein stability changes were collected from Chen et al, 20019 [2], which was split into 80 percent for training and 20 percent for testing. Our results yield 91.1% accuracy and a 0.69 Mathews correlation coefficient, which is relatively higher than other sequence-based methods such as I-Mutant 2.0 [3], Mupro [4], iPTREE-STAB [5], and iStable 2.0 [2]. Since our model is based on information extracted from sequences alone, it can even be used on proteins where structure information is not available. We will continue developing this approach to include additional protein information. This initial approach informs us about how to take the next steps on our project "Novel Systems Approach for Rational Engineering of Robust Microbial Metabolic Pathways"

References

1. T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13-17-August-2016 (2016) 785–794. https://doi.org/10.1145/2939672.2939785.

2. C.W. Chen, M.H. Lin, C.C. Liao, H.P. Chang, Y.W. Chu, iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules, Comput. Struct. Biotechnol. J. 18 (2020) 622–630. https://doi.org/10.1016/j.csbj.2020.02.021.

3. E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005). https://doi.org/10.1093/nar/gki375.

4. J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins Struct. Funct. Genet. 62 (2006) 1125–1132. https://doi.org/10.1002/prot.20810.

5. L.T. Huang, M.M. Gromiha, S.Y. Ho, iPTREE-STAB: Interpretable decision tree based method for predicting protein stability changes upon mutations, Bioinformatics. 23 (2007) 1292–1293. https://doi.org/10.1093/bioinformatics/btm100.

Funding Statement: This project is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), DE-SC0022090