ENGINEERING THE SUBSTRATE SPECIFICITY OF TOLUENE DEGRADING ENZYME XYLM USING BIOSENSOR XYLS AND MACHINE LEARNING

Yuki Ogawa, The University of Tokyo, Japan
yuki.ogawa@riken.jp
Yutaka Saito, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST) ; AIST-Waseda University Computational Bio Big-Data Open Innovation La-boratory (CBBD-OIL); The University of Tokyo, Japan
yutaka.saito@aist.go.jp
Hideki Yamaguchi, The University of Tokyo, Japan
Yohei Katsuyama, The University of Tokyo, Japan
Yasuo Ohnishi, The University of Tokyo, Japan

Key Words: XylM, enzyme engineering, directed evolution, machine learning, biosensor

With the recent development of machine learning, enzyme engineering using this technology has rapidly been widespread. However, to obtain a large amount of data on enzyme activities for training data, it is necessary to develop a high-throughput and accurate method for evaluating enzyme activities. Here, we examined whether a biosensor-based enzyme engineering method can be applied to machine learning. As a model experiment, we aimed to modify the substrate specificity of XylM, a rate-determining enzyme in a multistep oxidation reaction catalyzed by XylMABC in *Pseudomonas putida*. XylMABC are known to convert toluene and xylene to benzoic acid and toluic acid, respectively. We aimed to engineer XylM to improve its conversion efficiency to a non-native substrate, 2,6-xylenol. Wild-type XylMABC slightly converted 2,6-xylenol to 3-methylsalicylic acid, which is the ligand of the transcriptional regulator XylS in *P. putida*. By locating a fluorescent protein gene under the control of the Pm promoter to which XylS binds, a XylS-producing *Escherichia coli* strain showed higher fluorescence intensity in a 3-methylsalicylic acid concentration-dependent manner. We evaluated the 3-methylsalicylic acid productivity of 126 XylM variants using the fluorescence intensity of the sensor strain as an indicator. The obtained data provided the training data for machine learning for the directed evolution of XylM. After the first round of machine learning-assisted directed evolution, five XylM variants with higher productivity than wild-type XylM were acquired. XylM-XylM-D140E-F243L-N244S (XylM-PEVLS) showed the highest productivity with 10 times higher productivity than wild-type XylM[1]. Then, the relative fluorescence intensity activities of the additional 50 XylM variants were added to the training data of machine learning to update the model. After the second round of machine learning-assisted directed evolution, three XylM variants with higher productivity than XylM-PEVLS were acquired. XylM-D140E-V144K-F243L-N244S showed the highest productivity with 15 times higher productivity than wild-type XylM[1]. These results demonstrate that an indirect enzyme activity evaluation method using biosensors is sufficiently quantitative and high-throughput to be used as training data for machine learning. The findings expand the versatility of machine learning in enzyme engineering.

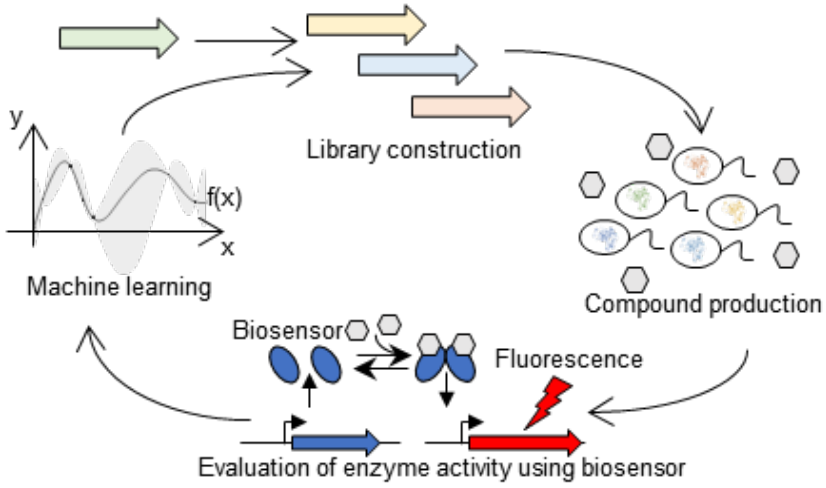

Figure 1 – Enzyme engineering using biosensor and machine learning

[Reference] [1]Ogawa, Y., Saito, Y., Yamaguchi, H., Katsuyama, Y. & Ohnishi, Y. Engineering the Substrate Specificity of Toluene Degrading Enzyme XylM Using Biosensor XylS and Machine Learning. ACS Synth. Biol. 12, 572–582 (2023)