Dean&Francis
ISSN 2959-409X

Progress in the Application of Enzyme Improvement Design Technology

Zimeng Rui

College of Life Sciences
Northwest A&F University
Shanxi, China
zimengrui@gmail.com

Abstract:

At present, enzyme improvement is becoming more and more indispensable in production and scientific research, but there is no accurate improvement method recommendation for enzymes of different types and structures. In the current research, the author has analyzed several enzyme improvement methods and given suggestions for improvement methods for different types of enzymes. However, with the development of biotechnology, enzyme improvement technology is also constantly updated and iterated. This article analyzes the most commonly used mainstream methods. This article compares and studies three methods of enzyme improvement design, namely rational design, semi-rational design and random design, and conducts a detailed study and analysis of these three improvement methods, and obtains the most suitable improvement methods for different types of enzymes. It can provide a reference for the selection of enzyme improvement methods in future research, but the screening of methods when the enzyme meets the conditions of multiple improvement methods at the same time has not been solved. Future research can focus on the screening and prediction of improvement methods.

Keywords:-Enzyme engineering; rational design of enzymes; random design of enzymes; semi-rational design of enzymes

I. introduction

In recent years, protein enzymes have played an increasingly important role in all aspects of human production and life. Enzymes are ubiquitous in the fields of medical devices, food engineering, etc. Therefore, the catalytic activity, catalytic efficiency, and catalytic conditions of enzymes are crucial in the reaction. Based on the above reasons, it is a current practical requirement to improve the structure and function of enzymes to make them have higher catalytic activity and faster catalytic efficiency. Enzyme improvement design is to transform the structure and function of enzymes through existing means to make them have higher catalytic activity and catalytic efficiency.

At present, there are three main means of enzyme

ISSN 2959-409X

improvement design: rational design, random design, and semi-rational design. Among them, rational design and random design are more widely used, and semi-rational design mainly appears in some cutting-edge scientific and technological applications. Rational design, random design, and semi-rational design all have their own advantages and disadvantages. Rational design is highly dependent on the structure of the enzyme, and requires a very detailed enzyme structure to modify the enzyme. Although it cannot be applied to enzymes with unclear structures, it is not easy to miss any mutation, and the performance of each mutation is more clearly understood. Random design is not highly dependent on the enzyme, and may induce mutations that have a better effect on the enzyme, but the induced mutations are too random, the screening is difficult, the workload is high, and most of the mutations are harmful. Semi-rational design combines the advantages of random design and rational design. It does not rely entirely on the detailed structure of the enzyme. It predicts the effect of the site mutation of the enzyme on the function and then improves it. However, the technology is not as mature as the first two methods, and most of them are used in some cutting-edge enzyme modifications. Most of the enzymes currently known have complex structures, and some enzymes do not have a well-resolved three-dimensional structure, making it difficult to choose a method for enzyme modification.

This study mainly analyzes the advantages and disadvantages of different enzyme improvement methods and their application scenarios. By disassembling different improvement methods and analyzing cases, this article aims to integrate and compare the application of different improvement methods in enzyme improvement design, and provide a comprehensive reference for the selection of methods in enzyme improvement design in future studies.

II. Rational mutagenesis

Rational mutagenesis is based on a detailed understanding of the enzyme's three-dimensional structure, catalytic mechanism, active site and mechanism of action, and purposefully designs mutations to change key amino acid residues to optimize the performance of the enzyme. Rational mutagenesis includes two methods, namely site-directed mutagenesis and a combination of structural modeling and molecular dynamics simulation. This article will introduce the two methods in detail. This method is often used in scenarios with extremely high requirements for safety and efficiency, such as precise modification of active sites, optimization of enzyme thermal stability, and improvement of stereospecificity. For example, in the pharmaceutical process, it is often necessary to carry out targeted modification of enzymes to obtain high catalytic efficiency and strict stereoselectivity; in the fields of biosensors, nanocarriers, etc., the organic solvent tolerance or pH optimum range of enzymes are often finely regulated by molecular simulation + site-directed mutagenesis.

A. Site-directed Mutagenesis

Site-directed mutagenesis refers to replacing the amino acid at the target site with other types through PCR technology, and testing the effects one by one. The detailed steps of this method are to replace one amino acid with another desired amino acid at a specific site of the gene. A primer with a mutation is designed during PCR amplification. After PCR amplification, the mutation becomes part of the new DNA. The advantages of this method are accuracy, simplicity, and high efficiency, but it requires a very clear understanding and knowledge of the structure of the enzyme being modified . The application of this method is reflected in the single-point replacement of the predicted flexible loop residues in the structure of L-lactate oxidase (LOX), mutating three of the amino acids into more rigid residues (Gly→Pro, Ser→Phe, etc.). The half-life ($t_{1/2}$) of the mutant at 60 °C was extended from 15 min to 48 min compared with the wild type, and the thermal stability was improved by about 3.2 times; the catalytic activity was slightly reduced (about 10%), but the overall catalytic efficiency (k_cat/K_m) was still 20% higher than that of the wild type [1]. The application of site-directed mutagenesis in the above studies perfectly illustrates that this method of site-directed mutagenesis requires a very clear understanding of the enzyme structure, otherwise it will be difficult to modify the enzyme.

B. Structural Modeling Combined with Molecular Dynamics Simulation

This method can be summarized as assisting in predicting which sites can be mutated to improve enzyme performance, and then mutating after prediction. Rational design relies on the three-dimensional structure, reaction mechanism and active center information of the enzyme, and then judges which amino acid residues may affect catalytic activity, stability or substrate binding based on this known information. However, sometimes many enzymes do not have a well-resolved crystal structure, or only have a protein-like structure. At this time, structural modeling and molecular dynamics simulation methods must be used. Structural modeling is to use computer software to predict the three-dimensional structure of the target enzyme based on the three-dimensional structure of proteins with similar sequences to the target enzyme. It is divided into three steps: finding a template protein, predicting

the structure with modeling software, and obtaining a three-dimensional structural model of the target enzyme. Molecular dynamics simulation is to use a computer to simulate the real movement of enzymes, substrates, and solvent molecules under certain temperature and pressure, structural changes, and important interactions near the active center. Prepare the enzyme-substrate complex structure, set up a simulation environment, and conduct further analysis based on the results of molecular dynamics simulation. A related study first used homology modeling (SWISS-MODEL) to construct a three-dimensional model of microbial cellulase (Cel5A); then a 100 ns molecular dynamics simulation was performed to analyze the motion trajectory and hydrogen bond stability of several key amino acids in the enzyme-substrate complex. Based on RMSF (residue fluctuation) and binding free energy calculations, two residues in the high fluctuation region (Val^245 and Glu^322) were selected for site-directed mutation (V245I, E322L). As a result, the mutant maintained 80% activity at 70 °C, while the wild type only had 30%; K_m decreased by 15%, showing better substrate affinity [2].

III. Randomized Design

Random design is a method that does not rely on structural and functional information. It conducts large-scale random mutations on enzyme genes, constructs a large number of mutants, and relies on screening to discover enzymes with better performance. The advantages of random design are that it does not require knowledge of the structural information of the enzyme, has a wide range of applications, and some mutations may achieve unexpected results. However, there are few beneficial mutations among a large number of mutations, and the large size of the mutation library requires a large amount of screening, which greatly increases the workload. This method is actually suitable for enzymes with unknown structures or unclear mechanisms, or when you want to explore the effects of whole-gene mutations, especially in improving the tolerance of industrial strains and screening for high-temperature resistant enzymes. It is very useful and has great practical value. The commonly used mutagenesis methods for this design method are mismatch PCR, chemical or physical mutagenesis, and DNA Shuffling.

A. Mismatched PCR

The core of this method is to reduce the fidelity of DNA polymerase by adjusting the conditions of the PCR reaction system, thereby increasing the probability of errors during replication. Specifically, common practices include increasing the concentration of Mg^{2+} in the reaction system , adding Mn^{2+} , or appropriately reducing the concentration of dNTP. These regulatory measures will interfere with the normal recognition and pairing process of the polymerase, causing it to produce a certain frequency of base substitutions when amplifying the target gene, thereby achieving random mutagenesis of the target DNA sequence. Mismatch PCR is simple, cheap, and suitable for making random mutation libraries, but mutations are random, with both beneficial and harmful mutations, requiring a lot of screening and a large workload. In one experiment, the research team used mismatch PCR to construct a library of approximately 1.2×10^4 Bacillus lipase mutants by adjusting the Mg^{2+}/Mn^{2+} concentration and the lack of dNTP conditions.

Combined with high-throughput enzyme activity screening using microtiter plates, three mutants with the best performance at 65 °C were obtained, of which the best one had a 4.1-fold increase in thermal stability and a 1.6-fold increase in activity over the wild type [3].

B. Physical or Chemical Mutagenesis

Physical or chemical mutagenesis methods use chemicals or ultraviolet light or radiation to damage DNA and create mutations. Chemical mutagenesis uses agents such as nitrosoguanidine and EMS to treat cells to cause DNA base pairing errors. Physical mutagenesis uses ultraviolet light to irradiate and create DNA mismatches. These two methods are simple to operate and low in cost, but the mutation position is completely uncontrollable, highly destructive, and the proportion of beneficial mutations is low. For example, Ghosh's research team treated Aspergillus oryzae strains with EMS and established 2.5×10^3 mutant strains. Through dual screening on solid plates and liquid cultures, 5 high-yield strains with 25–40% increased α-amylase production were selected. Enzyme characterization showed that the activity of the optimal mutant was 30% higher than that of the wild type at 55 °C and pH 5.5 [4].

C. DNA Shuffling

The specific operation method of gene shuffling can be summarized as first using DNA enzyme to cut the mutant gene into small segments, then mixing these small segments together, and "self-assembling" them through PCR, and finally obtaining a new gene combination, and the new gene combination has multiple mutation points. The advantage of gene shuffling is that it can combine the advantages of different mutations to produce a synergistic effect. However, this method still cannot confirm the mutation site, and the operation is slightly more complicated than the above methods. In one experiment, the research team used DNase I fragments and overlapping PCR

ISSN 2959-409X

technology to perform gene shuffling in esterase genes isolated from 4 different sources (marine bacteria, soil bacteria, bovine stomach bacteria, and fungi). A recombinant library of about 6×10^3 was constructed, and variant E35 was obtained through substrate chromatography high-throughput screening. Among them, it has the best activity for medium-chain esters, which is 2.8 times higher than the wild type; the activity for long-chain esters decreased by 50%, showing the successful redirection of substrate specificity [5].

The above three random design methods have their own advantages and disadvantages. DNA shuffling is more complicated than the other two methods. Among them, physical mutagenesis or chemical mutagenesis is the simplest. The mutation position of these three methods is uncontrollable, but the operation is simpler than rational design and semi-rational design . The three random design methods are more suitable for enzymes with unknown structures or unclear mechanisms, or for exploring the effects of full gene mutation of enzymes .

IV. Semi-rational design

Semi-rational design combines the advantages of both rational mutagenesis and random design. Researchers can selectively mutate enzymes based on information about known enzymes, and do not rely entirely on detailed analysis of the structure. They can optimize enzyme mutations by combining experimental data and computational predictions. A common strategy for semi-rational design is to combine partial structural or functional information to selectively mutate certain potentially important sites or regions, reduce blindness, and increase the proportion of beneficial mutations. Semi-rational design combines structural information and experimental/computational predictions to construct small-scale libraries and screen efficiently when only incomplete structural information is available. It is suitable for applications with a wide range of substrates and incompletely clear mechanical mechanisms, such as lignin-degrading enzymes, natural product synthases, and protein drug derivatization. It can control the size of the library within the screenable range while ensuring the effectiveness of the mutation, greatly reducing experimental costs and time. Commonly used mutagenesis methods for semi-rational design include site saturation mutagenesis, WASH-ROM method, and combinatorial mutagenesis.

A. Site saturation Mutagenesis

This approach is usually achieved by designing a set of annealing primers containing “NNK codons”, where “N” represents any base (A, T, G, C) and “K” represents G or T, thereby encoding all 20 standard amino acids. In the mutation library obtained by PCR amplification, the target site can theoretically be replaced by any amino acid residue to achieve comprehensive mutation coverage. The advantage of this method is that it has a high diversity of mutations and can maximize the exploration of potential functional mutations. However, this strategy faces challenges in practical applications, such as large library size and heavy screening workload. In particular, when the number of mutation sites increases, the number of combinations increases exponentially, significantly increasing the complexity of subsequent screening and verification. In one study, an 8×20^{10} saturation mutation library was constructed using the NNK primer library targeting the eight residues around the active center of Kemp eliminase (KE). After screening each site one by one, it was found that site H52W had the most significant improvement on the catalytic effect, increasing the catalytic efficiency by 5 times; a combination of multiple beneficial mutations (H52W+Y112F) further increased the catalytic efficiency to 8 times; and combined with structural analysis, the synergistic mechanism was revealed [6]. For example, the target enzyme Kemp Eliminase (a designer enzyme that catalyzes the “shock response”) was improved and designed by selecting 8 residues around its active center for saturation mutation. Using the NNK degenerate codon, overlapping primer PCR was used to construct a library of all 8×20^{10} ($\approx1.6\times10^8$) mutants. Two consecutive rounds of high-throughput microtiter plate assays were performed to initially screen out 350 activity-enhanced variants, which were then reduced to 20 through secondary enzyme kinetic assays. The k_cat/K_m of the optimal single-point mutation H52W increased by 5 times; the double mutation synergy: H52W+Y112F increased the catalytic efficiency to 8 times; crystallography and MD simulations showed that Y112F changed the stable conformation of the substrate, while H52W fixed the active cyclotron, with significant synergistic effects [7].

B. WASH-ROM Method

The specific content of this method is to first use computational software to find out the sites on the enzyme surface that are easily accessible to the solvent and may participate in the reaction, and then perform site-directed mutation or saturation mutation on these sites. The mutation is based on evidence and is conducive to improving activity. The library volume is moderate and the screening pressure is low. For example, in the study, the research team selected 25 highly accessible residues within 15 Å of the active site Cys124 by calculating the solvent accessible area (SASA) of the enzyme surface. Using a strategy combin-

ing site-directed and saturation mutation, three mutations (T78I, S142A, Y185F) were screened, and the catalytic efficiency can be increased by 1.7-2.1 times under pH 5.0-6.0 conditions [8]. Another example is the design and improvement of the target enzyme Streptomyces xylanase (xylanase). The site selection principle is: calculate the solvent accessible area (SASA) of the side chain within a radius of 15 Å of the active center Cys124, screen out 25 residues with high SASA, perform site-directed (A→X) or saturation mutation on these 25 residues, and construct a 25×20 ≈500 residue site library. The results show that the three single sites of T78I, S142A, and Y185F all increase the specific activity by 1.7–2.1 times; the combined double mutation (T78I+S142A) increases it to 2.6 times; the optimal pH range of the enzyme is expanded from 5.5–6.5 to 4.5–7.0 [9].

C. Combination Mutation

Combinatorial mutation is to combine multiple mutation points together to see the effect of multiple mutations superimposed. A single mutation may improve performance a little, but multiple mutations combined together may produce a greater synergistic effect, or some mutations may compensate for each other's defects to make the enzyme perform better. The advantage of this method is that it can explore the synergistic effect of mutations. When the performance of a single mutation is limited, combinatorial mutation is a breakthrough. However, the more mutation sites there are, the more combinations there are, and the greater the screening pressure. If the mutations conflict with each other, the performance will decrease, and a large number of experiments are required. For example, Wang, X.'s team used a random forest model to predict the optimal amino acids of 10 potential active sites of an enzyme (a lipase); after selecting 5 sites, 2^5=32 combined mutants were constructed. High-throughput screening showed that the optimal combination (L98F+S150T+V210I+R245K+G300A) increased its catalytic activity by 3.1 times at 55 °C and its thermal stability ($t_{1/2}$) by 2.5 times; and combined with structural analysis, the synergistic effect of mutations was verified [10]. Another example is the design and improvement of the target enzyme Thermostable lipase. The random forest machine learning model was used to predict 5 optimal sites from 20 candidate residues based on sequence-structure characteristics, and 2^5 = 32 different site combination mutants were constructed and screened by high-throughput microtiter version + autonomous bridging thermal stability test. The result was that the half-life ($t_{1/2}$) of the best combination L98F+S150T+V210I+R245K+G300A at 55 °C increased from 12 min to 30 min (2.5 times); at the same time, k_cat/K_m increased from 800 $M^{-1}s^{-1}$ to 2480 $M^{-1}s^{-1}$ (3.1 times); structural modeling showed that each mutation site achieved a dual improvement in thermal stability and catalytic efficiency by enhancing the hydrophobic core and surface salt bridge network [11].

V. Conclusion

In general, rational design, semi-rational design and random design have their own advantages and disadvantages. Rational design requires a clear understanding of the enzyme structure. According to the example, accurate single-point replacement of residues can significantly enhance the catalytic efficiency of the enzyme. It is suitable for modifying enzymes with known structures. Random design is the most random method among the three methods. The means used are generally simple. In the examples, new mutants are successfully induced by methods such as DNA shuffling, physical or chemical mutagenesis, but the subsequent screening workload is large. This method is suitable for enzymes with unclear structures or mechanisms. Semi-rational design reduces the randomness of random mutations and the limitations of rational design through selective mutation. It has the advantages of rational design accuracy and partial randomness in random design. It is suitable for constructing small-scale libraries and efficient screening when there is only incomplete structural information. It is suitable for enzymes with a wide range of substrates and unclear mechanical mechanisms. However, due to the author's review of known classic literature and the current shallow understanding of this field, the several enzyme improvement strategies discussed in this article are relatively basic, and the cutting-edge technologies in the current scientific community have not yet been involved. They can only provide researchers with a comprehensive reference for the selection of methods in enzyme improvement design. Future research can focus on predicting and determining the direction of improvement before designing enzyme improvements, and selecting the better method among the three methods.

References

[1] H. Minagawa, Y. Yoshida, N. Kenmochi, M. Furuichi, J. Shimada, and H. Kaneko, "Improving the thermal stability of lactate oxidase by directed evolution," Cellular and Molecular Life Sciences, vol. 64, pp. 77–81, 2007.

[2] J. Bi, S. Chen, X. Zhao, Y. Nie, and Y. Xu, "Computation-aided engineering of starch-debranching pullulanase from Bacillus thermoleovorans for enhanced thermostability," Applied Microbiology and Biotechnology, vol. 104, pp. 7551–7562,

ISSN 2959-409X

2020.

[3] A. E. Wentz and E. V. Shusta, "A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins," Applied and Environmental Microbiology, vol. 73, no. 4, pp. 1189–1198, 2007.

[4] L. Li, X. Liu, Y. Bai, B. Yao, H. Luo, and T. Tu, "High-throughput screening techniques for the selection of thermostable enzymes," Journal of Agricultural and Food Chemistry, vol. 72, no. 8, pp. 3833–3845, 2024.

[5] L.-S. Gu et al., "ARTP/EMS-combined multiple mutagenesis efficiently improved production of raw starch-degrading enzymes in Penicillium oxalicum and characterization of the enzyme-hyperproducing mutant," Biotechnology for Biofuels, vol. 13, pp. 1–12, 2020.

[6] L. I. Gutierrez-Rus, M. Alcalde, V. A. Risso, and J. M. Sanchez-Ruiz, "Efficient base-catalyzed Kemp elimination in an engineered ancestral enzyme," International Journal of Molecular Sciences, vol. 23, no. 16, p. 8934, 2022.

[7] A. Li, Q. Wang, X. Song, X. Zhang, J. W. Huang, C. C. Chen et al., "Engineering of a P450-based Kemp eliminase with a new mechanism," Chinese Journal of Catalysis, vol. 47, pp. 191–199, 2023.

[8] C. Paul, U. Hanefeld, F. Hollmann, G. Qu, B. Yuan, and Z. Sun, "Enzyme engineering for biocatalysis," Molecular Catalysis, vol. 555, p. 113874, 2024.

[9] Y. Sürmeli and G. Şanlı-Mohamed, "Engineering of xylanases for the development of biotechnologically important characteristics," Biotechnology and Bioengineering, vol. 120, no. 5, pp. 1171–1188, 2023.

[10] K. Ding, M. Chin, Y. Zhao, W. Huang, B. K. Mai, H. Wang et al., "Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering," Nature Communications, vol. 15, no. 1, p. 6392, 2024.

[11] Z. Wu, S. J. Kan, R. D. Lewis, B. J. Wittmann, and F. H. Arnold, "Machine learning-assisted directed protein evolution with combinatorial libraries," Proceedings of the National Academy of Sciences, vol. 116, no. 18, pp. 8852–8858, 2019.