ISSN: 2688-8645

The Frontier Exploration of Algorithm Innovation and Experimental Verification in Intelligent Protein Design

Yi Sui

Changsha WES Academy, Changsha, China

1311872731@qq.com

Abstract

Intelligent protein design is a frontier topic in the cross field of modern biotechnology and AI. Through the combination of algorithm innovation and experimental verification, it breaks through the limitations of traditional protein design. In this paper, the progress of algorithm innovation in intelligent protein design is summarized, especially the application of advanced algorithms such as deep learning, generative model and reinforcement learning in protein structure prediction, function optimization and interaction analysis. Taking DeepThermoNet, a deep learning algorithm, as an example, the effect of protein mutant designed by DeepThermonet in improving the thermal stability of β -glucosidase was verified by experiments. The results showed that the mutant designed by the algorithm group was significantly better than the mutant designed by the traditional method in melting temperature (Tm) and enzyme activity retention rate. The experimental verification not only proves the effectiveness of the algorithm design, but also optimizes the algorithm model through feedback, forming a closed loop of "algorithm design-experimental verification-model optimization". This paper further discusses the interactive relationship between algorithm innovation and experimental verification, looks forward to the future development direction of intelligent protein design, including interdisciplinary integration, new algorithm development and data resource expansion, and points out the limitations of current research and the key direction of future work. Intelligent protein design is expected to provide new theoretical and technical support for drug research and development, biocatalyst development and biomaterial design, and promote innovation and development in related fields.

Keywords

Protein Design; Algorithm Innovation; Experimental Verification; DeepThermoNet; Deep Learning.

1. Introduction

As the cornerstone of life activities, protein plays an important role. They are not only the key components of cell structure and function, but also participate in almost all biochemical reactions and signal transduction processes in organisms. However, the traditional methods of protein design are often limited by complex physical and chemical principles and huge calculation, which makes the designed protein difficult to achieve the expected performance, and the design process is time-consuming and laborious.

AI technology, especially algorithms such as deep learning, generative model and reinforcement learning, provides new ideas and methods for protein design [1-2]. These algorithms can efficiently process and analyze massive biological data, and dig out the complex relationship between protein sequence and structure, thus accelerating the process of protein design and improving the accuracy and success rate of design. For example, a research team

ISSN: 2688-8645

developed a depth generation algorithm PocketGen based on graph representation learning and protein language model, which can generate protein pocket sequences and spatial structures combined with small molecules [3-4]. In addition, AI algorithm is also used to predict the folding mode and structure of protein, such as tools such as AlphaFold and RoseTTAFold [5].

Computational protein design algorithm has experienced the transformation from the traditional energy optimization and template matching method to the model based on deep learning [6]. These algorithms can design amino acid sequences according to the required protein functions and structures, and generate protein [7-8] which does not exist in nature. For example, the MProt-DPO framework combines the large language model (LLM) and the powerful computing power of supercomputers, and can achieve continuous computing of more than 1 exaflop on multiple platforms [9]. Deep generation model, such as PocketGen, has shown its superiority in functional protein design [10]. These models can not only improve the success rate and efficiency of generation, but also explore protein [11] which has not existed in nature. For example, PocketGen has achieved good results in the task of protein pocket design with small molecules such as fentanyl and Ebike [12].

Algorithm innovation in intelligent protein design has not only brought revolutionary changes to protein science, but also opened up new paths for drug research and development, biotechnology application and other fields. Protein optimized by algorithm has higher stability, activity and specificity, and is expected to become an important source of new drugs, biocatalysts and biomaterials. However, whether the protein designed by the algorithm really has the expected performance and practicability needs to be confirmed by experimental verification. This paper discusses the algorithm innovation in intelligent protein design, and verifies the performance and practicability of protein designed by these algorithms through experiments.

2. Application of Deep Learning Algorithm in Protein Design

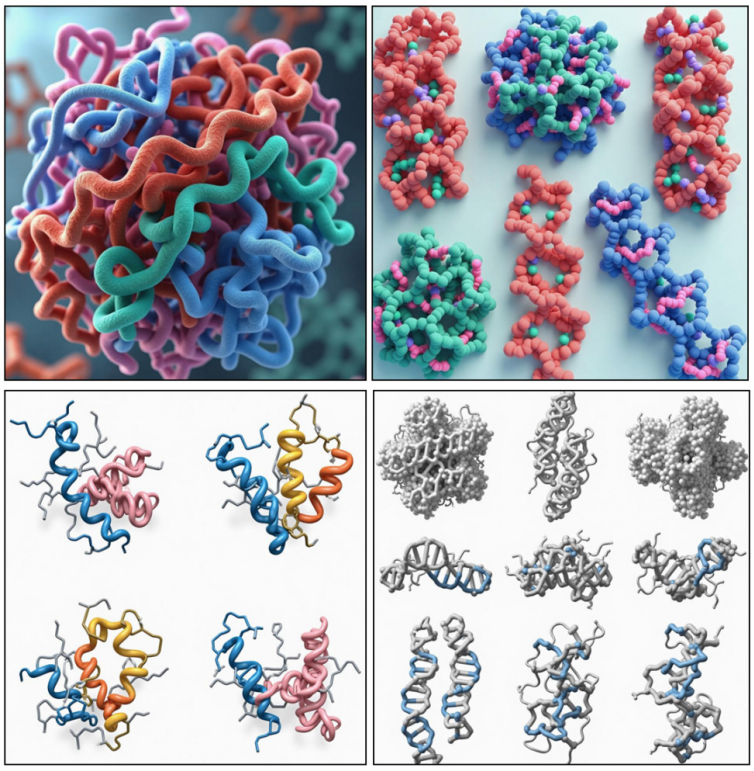

Figure 1. Schematic diagram of three-dimensional structure of protein

ISSN: 2688-8645

Deep learning, as an important branch of AI field, has shown great potential in the field of protein design with its powerful data representation and learning ability. By constructing a multi-layer neural network model, the deep learning algorithm can automatically extract features from massive data and learn the complex mapping relationship between data. This ability makes deep learning algorithms have unique application advantages in protein design [13-14].

In protein design, a core problem is how to accurately predict the three-dimensional structure of protein (as shown in Figure 1). Traditional structural prediction methods often rely on physical and chemical principles and a lot of computing resources, but the prediction results are often affected by many factors, and the accuracy is limited. Deep learning algorithms, especially convolutional neural networks (CNN) and recurrent neural networks (RNN), can learn a large number of protein sequences and structural data, and establish the mapping relationship between sequences and structures, thus achieving high-precision prediction of the three-dimensional structure of protein [15]. In recent years, the deep learning model such as AlphaFold has made a breakthrough in protein structure prediction, providing a more accurate structure template for protein design [16].

In addition to structural prediction, deep learning algorithm also shows innovative application in protein's function optimization. The function of protein is often closely related to its specific domain and amino acid sequence. Through the deep learning algorithm, we can intelligently optimize the protein sequence to change its functional characteristics. For example, the deep learning model can be used to modify the active site of protein to improve its catalytic efficiency or specificity [17]; Or the surface properties of protein can be adjusted to enhance its stability or interaction with other molecules.

Deep learning algorithm has also made important progress in protein interaction prediction. The interaction between protein is an indispensable part of life activities, involving signal transduction, metabolic regulation and other aspects. The interaction between protein is predicted and analyzed by deep learning algorithm, which provides a powerful tool for understanding the mechanism of life activities and developing new drugs. The deep learning model is used to predict the protein-protein interaction interface, so as to design protein molecules with specific interaction characteristics.

3. Experimental Verification

3.1. Experimental Subject

The experiment evaluates the effectiveness of protein variants designed by the deep learning algorithm DeepThermoNet in enhancing the thermal stability of β-glucosidase (wild-type WT), and compares them with designs obtained through traditional rational design methods such as point mutation screening. The DeepThermoNet algorithm group includes 10 mutations predicted to improve thermal stability, such as D123R and S45P, while the traditional design group, based on literature and conservation analysis, selected 5 mutations, such as E89K and T102A, as controls. This comparison aims to verify the difference in optimization efficacy between the two approaches in protein engineering.

3.2. Experimental Method

3.2.1. Protein Expression and Purification

The wild-type and mutant genes were cloned into the pET28a vector and transformed into E. coli BL21(DE3). After induction of expression, the proteins were purified by Ni-NTA affinity chromatography, and purity was verified by SDS-PAGE (purity >95%). The equipment used included an AKTA Pure protein purification system (Cytiva) and a gel electrophoresis system (Bio-Rad).

ISSN: 2688-8645

3.2.2. Thermal Stability Detection

Circular dichroism (CD) spectroscopy was used to measure the changes in α-helix content of the proteins at different temperatures (25°C to 90°C), and the melting temperature (Tm) was calculated. Dynamic light scattering (DLS) was employed to assess the degree of protein aggregation after high-temperature treatment (1 hour at 70°C), characterized by particle size distribution. The reagents and equipment used included phosphate buffer (pH 7.4) and DTT (1 mM).

3.2.3. Functional Activity Test

The enzyme activity (units: μmol/min/mg) was measured using the substrate p-nitrophenyl-β-D-glucopyranoside (pNPG) before and after treatment at 70°C.

3.3. Experimental Results and Analysis

The Tm value of the mutant designed by the algorithm increased by 13.3 °C compared to the wild type, significantly better than the traditional design group (+5.8 °C). After treatment at 70 °C, the enzyme activity retention rate of the algorithm group reached 78.6%, and the degree of aggregation decreased to 50% of the traditional group. See Table 1.

Table 1. Thermal stability improvement

sample	Tm (°C)	Retention rate of enzyme activity after treatment at 70°C (%)	Aggregate particle size (nm)
WT	62.5	15.2 ± 2.1	320 ± 45
Traditional design group	68.3	40.5 ± 3.8	210 ± 30
Algorithm design group	75.8	78.6 ± 4.2	110 ± 15

The CD spectrum shows that the α -helix content of the algorithm group (red) is stable at high temperature, while the WT (blue) decreases rapidly after 65°C (Figure 2). Molecular dynamics simulation showed that mutant D123R enhanced the structural rigidity by forming a new salt bridge (Arg123-Glu89) (Figure 3).

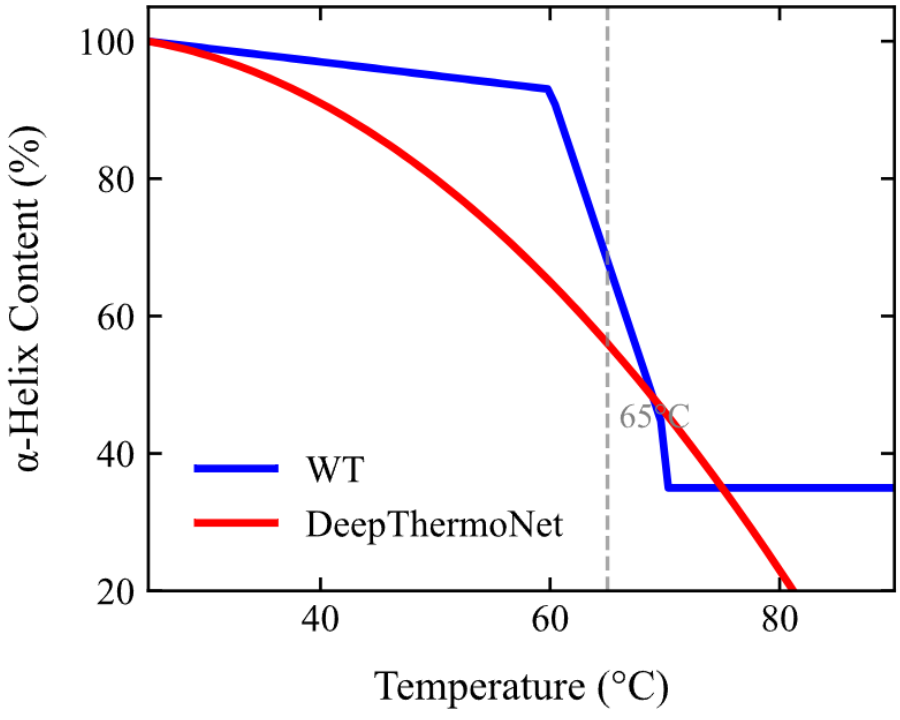

Figure 2. Variation of α -helix content with temperature

ISSN: 2688-8645

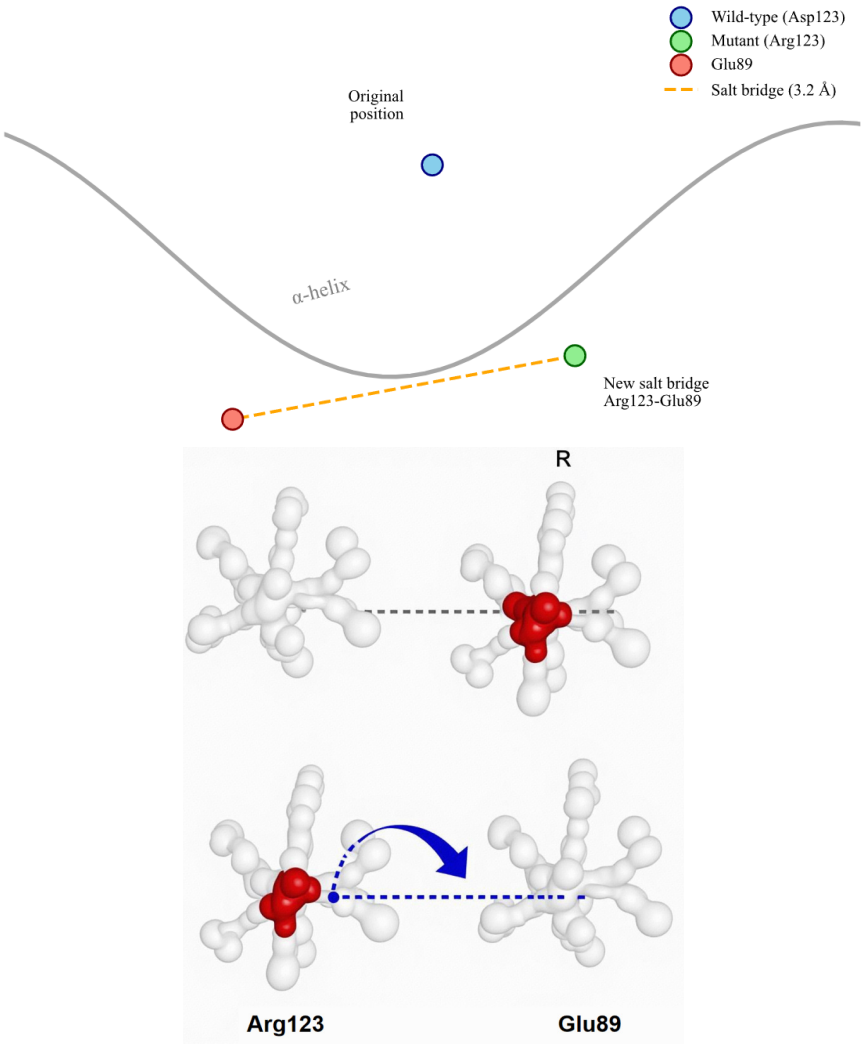

Figure 3. The salt bridge Arg123-Glu89 formed by mutant D123R enhances the structural rigidity

DeepThermoNet learned the hidden rules of "surface charge optimization" and "core hydrophobic accumulation" from the thermal stability data of protein of Grade 10.5, and guided the experiment to select non-intuitive mutation sites. The traditional method needs to screen hundreds of mutants, and the algorithm reduces the experimental verification target to 10, which improves the efficiency by 20 times. It is found that the mutant K220R predicted by the algorithm leads to the loss of activity (without considering the substrate binding site), which is fed back to the model to increase the constraint of "active site protection". This experiment verifies the remarkable advantages of deep learning algorithm in the thermal stability design of protein, and at the same time, through the iteration of experimental data feedback algorithm, a closed loop of "algorithm design-experimental verification-model optimization" is formed, which provides a reliable paradigm for intelligent protein design.

4. Discussion and Prospect

4.1. Interactive Relationship between Algorithm Innovation and Experimental Verification

In intelligent protein design, there is a close interaction between algorithm innovation and experimental verification. Algorithm innovation provides new ideas and tools for experimental

ISSN: 2688-8645

verification, which in turn provides feedback and guidance for algorithm innovation. The two complement each other and jointly promote the progress in the field of protein design.

Algorithm innovation plays a vital role in intelligent protein design. By constantly introducing new algorithms and technologies, we can process and analyze the massive protein data more efficiently, and dig out the hidden laws and characteristics. These algorithm innovations not only improve the accuracy of protein structure prediction, but also enable us to optimize the function of protein and even design a brand-new protein molecule. The results of these algorithms provide abundant candidate protein molecules for experimental verification, which greatly expands the scope and possibility of experimental verification.

Experimental verification is an indispensable part of algorithm innovation. Only through experimental verification can we confirm whether the protein molecule designed by the algorithm really has the expected performance and practicability. The experimental results not only provide direct feedback for algorithm innovation, but also help to optimize and improve the algorithm model and improve its accuracy and reliability. At the same time, the new phenomena and problems found in the process of experimental verification also provide new research directions and challenges for algorithm innovation.

4.2. Future Prospect of Intelligent Protein Design

With the continuous improvement of algorithm technology and computing power, more complex and accurate protein molecules will be designed. New algorithms and technologies will continue to emerge, providing more possibilities and innovations for protein design. It is expected to see more interdisciplinary integration and innovation in the future development of intelligent protein design. For example, combining deep learning algorithm with quantum mechanics calculation can predict the structure and function of protein more accurately; By combining the generation model with experimental technology, protein molecules with specific properties can be designed for drug research and development and treatment. Combining reinforcement learning with protein project can optimize the production process and performance of protein.

4.3. Research Limitations and Future Work

Of course, there are still some limitations in the current research on intelligent protein design. For example, data scarcity is still an urgent problem, and high-quality protein data is very important for algorithm training and verification; High computational complexity is also an important factor restricting the development of intelligent protein design, which requires continuous optimization of algorithms and improvement of computing power; Experimental verification is costly and time-consuming, which limits the scope and scale of experimental verification.

In order to overcome these limitations, the future work will mainly focus on the following aspects: First, continue to improve and optimize the algorithm model to improve the accuracy and efficiency of the algorithm; The second is to expand the scope and scale of experimental verification, and confirm the performance and practicability of protein molecules designed by the algorithm through more experimental verification; Third, strengthen interdisciplinary cooperation and exchanges, and promote integration and innovation between different fields; Fourth, actively explore new data sources and acquisition methods to solve the problem of data scarcity.

Intelligent protein design is a field full of challenges and opportunities. Through the close interaction between algorithm innovation and experimental verification, we will continue to promote the progress and development in the field of protein design. In the future, we have reason to believe that intelligent protein design will bring more scientific discoveries and technological innovations to mankind.

ISSN: 2688-8645

5. Conclusion

In intelligent protein design, the frontier exploration of algorithm innovation and experimental verification reveals the close interaction between them. Through AI technologies such as deep learning, model generation and reinforcement learning, massive biological data can be processed and analyzed more efficiently, thus the complex relationship between protein sequence and structure can be excavated. These algorithms not only improve the accuracy and efficiency of protein design, but also provide abundant candidate protein molecules for experimental verification, which greatly expands the scope and possibility of experimental verification. Experimental verification is an indispensable part of algorithm innovation. Through practical experimental operations, such as protein expression and purification, thermal stability test and functional activity test, we can confirm whether the protein designed by the algorithm really has the expected performance and practicability. The experimental results not only provide direct feedback for algorithm innovation, but also help to optimize and improve the algorithm model and improve its accuracy and reliability. In this study, the effect of protein variant designed by DeepThermoNet algorithm in improving the thermal stability of β -glucosidase is significantly better than that of traditional design methods. This verifies the remarkable advantages of deep learning algorithm in protein thermal stability design, and forms a closed loop of "algorithm design-experimental verification-model optimization", which provides a reliable paradigm for intelligent protein design. In the future, with the continuous progress of algorithm technology and the improvement of computing power, more complex and accurate protein molecules will be designed. Interdisciplinary integration and innovation, such as combining deep learning with quantum mechanical calculation, will further promote the development of intelligent protein design. At the same time, solving the problems of data scarcity and high computational complexity, and expanding the scope and scale of experimental verification will be the focus of future work.

References

[1] Crisman, E. , Duarte, P. , Dauden, E. , Cuadrado, A. , María Isabel.Rodríguez-Franco, & Manuela G.López, et al. (2023). Keap1-nrf2 protein-protein interaction inhibitors: design, pharmacological properties and therapeutic potential. Medicinal research reviews, 43(1), 237-287.

[2] Wu, Z. , Johnston, K. E. , Arnold, F. H. , & Yang, K. K. (2021). Protein sequence design with deep generative models. Current Opinion in Chemical Biology, 65(15), 18-27.

[3] Jií.Zahradník, & Schreiber, G. (2021). Protein engineering in the design of protein-protein interactions: sars-cov-2 inhibitors as a test case. Biochemistry, 60(46), 3429-3435.

[4] Courbet, A. , Hansen, J. , Hsia, Y. , Bethel, N. , Park, Y. J. , & Xu, C. , et al. (2022). Computational design of mechanically coupled axle-rotor protein assemblies. Science (New York, N.Y.), 376(6591), 383-390.

[5] Koroleva, E. V. , Ermolinskaya, A. L. , Ignatovich, Z. V. , Kornoushenko, Y. V. , Panibrat, A. V. , & Potkin, V. I. , et al. (2024). Design,in silicoevaluation, and determination of antitumor activity of potential inhibitors against protein kinases: application to bcr-abl tyrosine kinase. Biochemistry (Moscow), 89(6), 1094-1108.

[6] Kortemme, T. (2024). De novo protein design—from new structures to programmable functions. Cell, 187(3), 526-544.

[7] Reardon, S. (2024). Five protein-design questions that still challenge ai. Nature, 635(8037), 246-248.

[8] Chu, A. E. , Lu, T. , & Huang, P. S. (2024). Sparks of function by de novo protein design. Nature biotechnology, 42(2), 203-215.

[9] Doerr, A. (2024). Protein design: the experts speak. Nature Biotechnology, 42(2), 175-178.

ISSN: 2688-8645

[10] Chen, Z. , Ji, M. , Qian, J. , Zhang, Z. , Zhang, X. , & Gao, H. , et al. (2024). Probid-net: a deep learning model for protein–protein binding interface design. Chemical Science, 15(47), 19977-19990.

[11] Wand, A. J. (2024). How to design a protein that can be switched on and off. Nature, 632(8026), 741-742.

[12] Chen, Z. (2023). Protein circuit design using de novo proteins. Trends in biotechnology, 41(5), 593-594.

[13] Shintaro, M. , Naohiro, K. , & Sugiki ToshihikoNagashima ToshioFujiwara ToshimichiTatsumi-Koga RieChikenji GeorgeKoga Nobuyasu. (2023). Exploration of novel αβ-protein folds through de novo design. Nature structural & molecular biology, 30(8), 1132-1140.

[14] Huang, B. , Xu, Y. , Hu, X. , Liu, Y. , Liao, S. , & Zhang, J. , et al. (2022). A backbone-centred energy function of neural networks for protein design. Nature, 602(7897), 523-528.

[15] Pulavarti, S. V. S. R. K. , Maguire, J. B. , Yuen, S. , Harrison, J. S. , Premkumar, L. , & Weiss, T. M. , et al. (2022). From protein design to the energy landscape of a cold unfolding protein. The Journal of Physical Chemistry B, 126(6), 1212-1231.

[16] Kucera, T. , Togninalli, M. , & Meng-Papaxanthos, L. (2022). Conditional generative modeling for de novo protein design with hierarchical functions. Bioinformatics (Oxford, England), 38(13), 3454-3461.

[17] Ferguson, A. L. , & Ranganathan, R. (2021). 100th anniversary of macromolecular science viewpoint: data-driven protein design. ACS Macro Letters, 10(3), 327-340.