Dean&Francis
ISSN 2959-409X

Principles and Methods of De Novo Protein Design and Applications

Ziping Tian[1,*]

[1]Beijing Haidian Kaiwen Academy, Beijing, China
*Corresponding author: 20200705002@hd.kaiwenacademy.cn

Abstract:

De novo protein design strategy enables to create novel protein sequence that are able to fold in stable and functional structure, while breaking through the limitation of natural evolution. This method, which is based on the idea that the amino acid sequence of a protein dictates its three-dimensional structure, identifies a particular sequence with a low-energy state by combining computational methods, molecular interaction modeling, and thermodynamic optimization. Stabilizing interactions between molecules are precisely engineered to ensure the protein is folded into desired structure, while maintaining conformational stability. Recent advanced methods in de novo protein design includes flexible and fixed backbone modeling and deep learning-guided assembly. These methods have improved accuracy of modeling and broadened available space for protein sequence. De novo protein design can be applied across multiple biomedical domains. De novo protein design achieves notable success in vaccine innovation, where stabilized epitopes are presented to enhance the immune targeting, and in immunotherapy, where synthetic cytokine mimetics achieve high selectivity with reduced toxicity. These developments demonstrate how de novo protein design is transitioning from a conceptual idea to a robust platform for advanced therapeutics.

Keywords: De novo protein design; prediction of protein structure; vaccine innovation.

1. Introduction

De novo protein design is now highly revolutionizing field in molecular biology today with the future possibility of designing proteins that have never been seen in nature both in terms of structure and also function. Last 20 years has witnessed the maturing of the field from a theoretical concept into a mature scientific discipline with state-of-the-art computational protocols, high-resolution techniques in structural biology and powerful tools for experimental validation. This has enabled researchers to design proteins with remarkable precision, leading to breakthroughs in medicine, synthetic biology and materials science.

This advancement has been enabled by major scientific advances of predicting folding-structures with greater stability by energy-based modeling methods, deep-learning algorithms to predict complex structural patterns, and large protein structure databases that vastly increase the reachable design space. provide access to the vast and mostly unexplored sequence space, that in turn leads to the design of proteins not only with high level of structural stability but with custom biochemical functionality, from catalysis to molecular recognition. Intersection of these tools has transformed de novo protein design from a merely conceptually appealing idea to fundamentally sound and reproducible science.

David Baker and his team at the University of Washington were some of these pioneers who helped progress computational de novo protein applications. The development of computational platforms such as Rosetta-tools widely used in de novo protein design. The Baker group's accomplishments include generation of new enzymes with improved catalytic efficiency, structure-based vaccine designing targeting complex viral epitopes, and in-silico designed therapeutic proteins with enhanced stability and reduced immunogenicity. These achievements not only indicate that the field of de novo protein design has matured but also in so doing serve as a reminder of the broad challenges facing medicine, biotechnology and materials science which can be met by these proteins of our own design.

2. De Novo Protein Design

2.1 Fundamental Concept and Scope

Designing completely new protein sequences and structures from scratch is known as de novo protein design. Meanwhile, implementation of designed proteins into a particular, stable and also functional three-dimensional structure. Unlike protein engineering where you change natural proteins that are already there, de novo design seems to imitate natural evolution at the level of natural proteins, accessing the available protein sequences that have not been explored yet by natural proteins. Christian Anfinsen's 1963 theory, which states that a protein's three-dimensional structure is determined by its amino acid sequence, is the foundation for the prospect of de novo protein creation [1]. This principle enables one to predict and design new sequences which fold to prescribed structures and perform predefined functions.

2.2 Physical Principle Underlying De Novo Protein Design

2.2.1 Thermodynamic Determinants of Structural Stability

In the de novo protein design process, the role of considering thermodynamics is to help predict which protein conformation will be the most stable state by adjusting it into the lowest-energy state. In such a conformation, the structure becomes more resistant to perturbation. As shown in Figure 1, any deviation from this state demands additional energy to break these stabilizing interactions. Since the input of extra energy is required for change, spontaneous alterations are thermodynamically unfavorable. This energy requirement acts as a "barrier", making the protein's folded structure resistant to spontaneous shape changes. Therefore, minimizing energy is equivalent to maximizing stability [1].

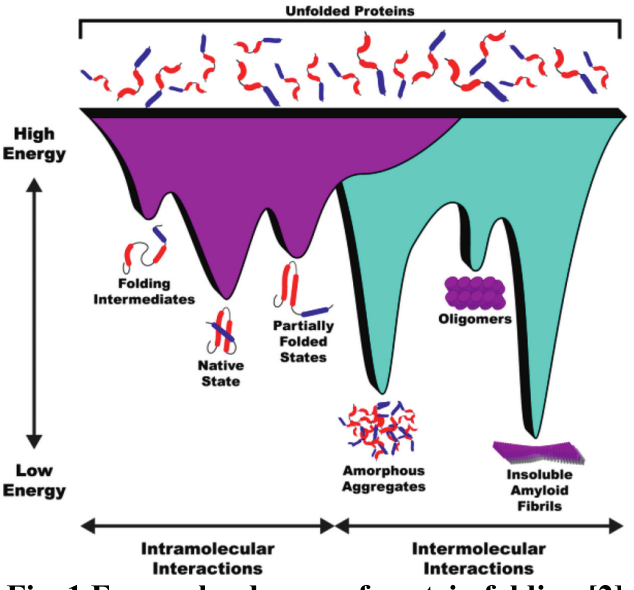

Fig. 1 Energy landscape of protein folding [2]

This process is thermodynamically regulated by ΔG (Gibbs free energy). A protein will spontaneously fold to its native structure only if the ΔG of folding is negative. That is to say the fold is energetically preferred over the unfolded state. Therefore, in protein design seeking sequences which minimize the Gibbs free energy ensures spontaneous folding and structural stability in physiological conditions [3].

2.2.2 Role of Molecular Interactions in Fold Stabilization

The designed sequence is folded into its target state and the fold is stable under normal conditions because of positive stabilizing molecular interactions. As shown in Figure 2, hydrogen bonds, disulphide bridges and hydrophobic effects are the major interactions. For example, hydro-

ISSN 2959-409X

gen bonds play a crucial role in stabilising α-helices and β-sheets by bridging backbone atoms [4]. While charged residues on the protein's surface provide it solubility and specificity, hydrophobic interactions bury non-polar residues in the protein's interior. The proper trade-off between these interactions is approximated in a perfect protein design.

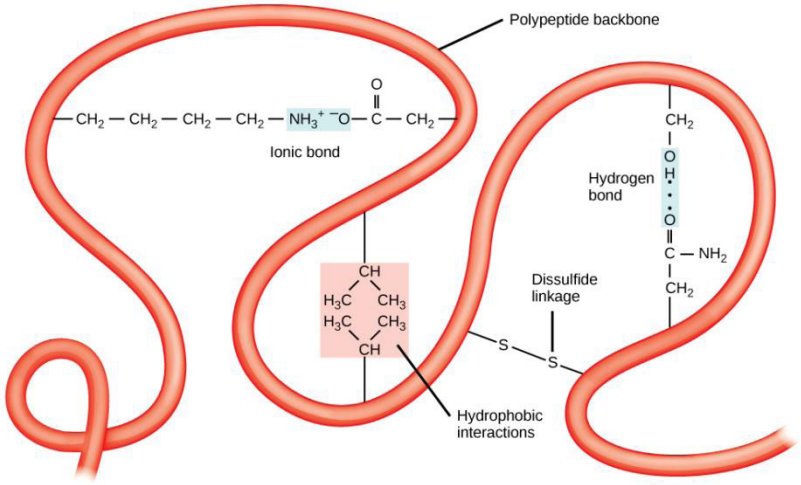

Fig. 2 molecular interactions used to stabilize protein folding structure [5]

2.3 Computational strategies and Tools in De Novo Protein Design

2.3.1 Software and Algorithms

Because the candidate sequences it creates amounts to a lot of computation, the de novo design results depend greatly on computational tools such as Rosetta, RFdiffusion, and AlphaFold2, which could help operate protein folding, optimize sequences, and predict protein's structures [6]. These computational techniques are essential to the advancement of modern medicine. The modeling of side-chain arrangement, its optimization and selection of the best energy state are performed with many algorithms.

One deterministic algorithm used in protein design, dead-end elimination (DEE), is highly efficient for pruning the rotamer conformations that cannot exist within the most stable sequences of proteins. The basic idea behind DEE is: if one rotamer at a particular position consistently has higher energy than the other, there exists no minimum energy set of conformations that contains the rotamer with higher energy. It can be unambiguously eliminated. When DEE is applied iteratively across all residue positions, the number of conformations to evaluate is reduced significantly, while it remains guaranteed that the lowest energy state is within the pruned solution space [7].

The Self-Consistent Mean Field (SCMF) approach is an application of mean field theory to the optimization of side-chain conformations in protein design. SCMF uses the average environment to represent every side chain, instead of all combinations of rotamers over residues. Thus, it calculates each rotamer's energy contribution iteratively under such averaged conditions, and it updates these probabilities until convergence. Although this approach has lower accuracy than a full search or deterministic approach such as DEE, it is much faster to compute- particularly in large design spaces-and produces near-optimal results in practice [8].

Monte Carlo (MC) Simulation sampling, in a stochastic process designed for protein design, efficiently searches sequence and conformational space. It makes a random change, either in the backbone or sidechain rotamers; then it evaluates this with an energy function. If the change decreases the total energy, it will be accepted. With greater energy changes there is a diminishing chance of acceptance. This also makes it possible for the algorithm to escape from any local minimum in which it has become and thus succeed in finding almost optimal solutions MC sampling forms the basis of protein design software like Rosetta.

2.3.2 Modeling of Backbone Flexibility

The amino acid sequence that will fold into the intended three-dimensional structure is determined by de novo protein design. This design process is commonly divided into fixed-backbone and flexible-backbone strategies [3]. Figure 3 shows, in a fixed-backbone design, the protein's backbone is kept rigid and unchanged, and only the side chains are redesigned to find the best-fitting sequence for the given shape. Fewer changes lower the amount of

available sequence, which decreases the computational complexity of selection. However, this method may miss better-fitting sequences that require backbone adjustments. In a flexible backbone design, parts of the backbone are allowed to adjust slightly during the design process, which more accurately reflects how real proteins can shift to achieve better packing and stability. Flexible-backbone design has enabled more ambitious redesigns, such as modifying loop regions to reshape substrate channels or binding pockets for better performance [9].

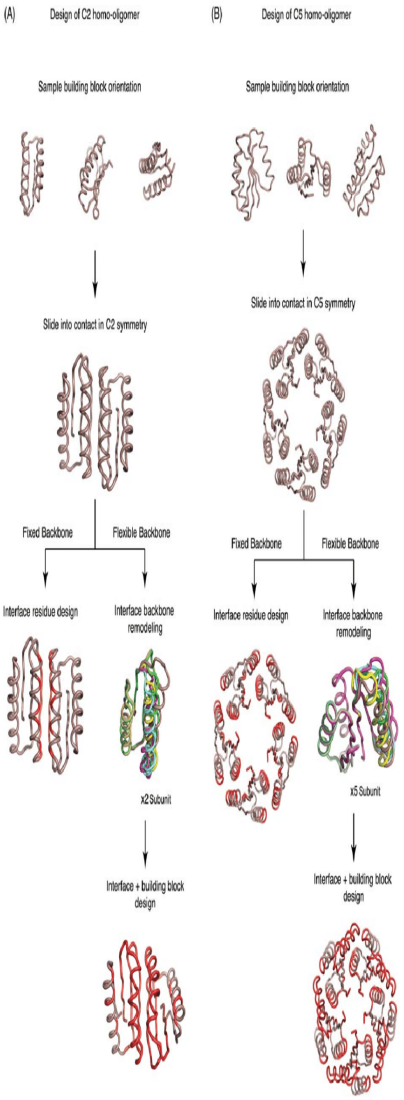

Fig. 3 Comparison of Fixed and Flexible backbone design strategies in protein Oligomer Assembly of C5 homo Oligomer [10]

2.3.3 Stepwise Workflow of De Novo Protein Design

Before a stable, functioning protein can be produced, the de novo protein design process usually entails a number of computational and experimental phases. First, the process starts by building a backbone template library out of well-known secondary structures and desired hydrogen bond networks or binding pockets. Afterwards, deep learning software such as RFdiffusion or TopoBuilder stitches together fragments into scaffolds, which can then identify specific motifs or functional sites [6]. Then, a targeting peptide segment or binding site is put into the template library and the best matching segment is found, followed by optimizations in backbone flexibility and side-chain orientation using algorithms like RFdiffusion or AlphaFold2 for a better structural fit and specificity [9]. After that, computational means are utilized to calculate the energy function for candidates large enough to launch bid apropos of perspective and decide optimal placement and most stable candidate. Having picked out a likely sequence, the corresponding gene is cloned and expressed in a host for experimental confirmation, then the protein is folded and its ability to bind to its target tested through experiments that confirm success of its design. he practical application of De novo protein design.

3. Application of De Novo Protein Design in Biomedicine

3.1 Vaccine innovation through Epitope Stabilization

De novo protein design is transforming the innovation of vaccines by increasing the possibility of creating precise immunogens that copy complex viral targets. The design of DS-Cav1, a stabilized prefusion form of the respiratory syncytial virus (RSV) fusion (F) protein, shown in Figure 4, is a notable example. Site Ø, a critical neutralizing site in the prefusion conformation of RSV F, is the main element that triggers an antibody response. The natural form of this neutralizing site is unstable on its own, which significantly affects the feasibility of a successful antibody response. McLellan et al. created mutations that lock the fusion protein in its prefusion conformation in order to get around this restriction. This stabilizes the native epitope and makes it more stable than the protein's natural form [11]. As shown in Figure 5, for stabilizing RSVF with a cocktail vaccine, de novo epitope scaffolding is employed: using computational design aided by TopoBuilder for de novo template building and 3D assembly to generate stable immunogens. These immunogens are then formulated into a synthetic cocktail vaccine, and when administered, they enable multi-site focusing in non-human primates (NHPs), eliciting defined antibody specificities to stabilize RSVF, which shows the effectiveness of de novo protein design.

ISSN 2959-409X

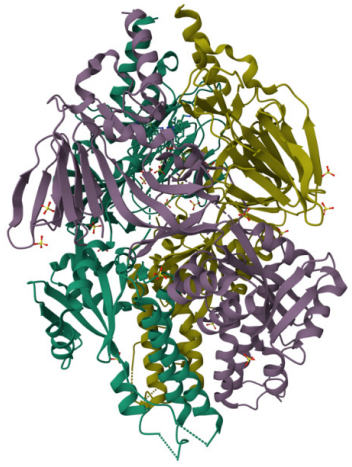

Fig. 4 3D structure of DS-Cav1 [12]

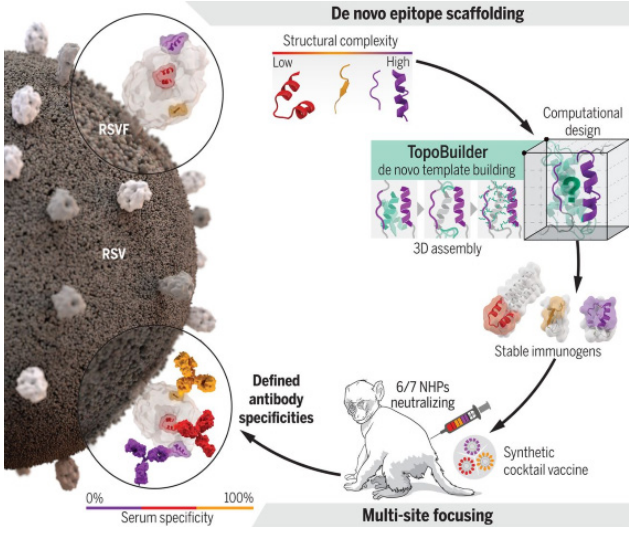

Fig. 5 The process of designing cocktail vaccine to stabilize RSVF protein using de novo protein design [13]

Based on this foundation, Boyoglu-Barnum et al. leveraged de novo protein design to present DS-Cav1 on a computationally designed nanoparticle scaffold, I53_dn5 [14]. This approach enhances the activation of B cells by mimicking dense epitope presentation on viruses, showing how the stabilization of a fragile epitope using de novo-designed multivalent scaffold can enhance vaccine efficacy against recalcitrant pathogens. This study shows how de novo design not only enhances antigen stability but also allows for structural precision in epitope presentation, providing potent solutions for targeting pathogens that are not practical to use conventional vaccine approaches.

3.2 Immunotherapy

De Novo protein design has the potential to improve immunotherapy. Stable cytokine mimics synthetically produce proteins that can function as E. coli-produced human interleukin-12, are able to be biomedically screened for clinical trials. Using computational design, the protein produced frequently has a higher receptor affinity and greater structural stability than its natural counterparts. For example, designed analogs of interleukin-2 (IL-2) can bind to specific receptors selectively, enhancing efficacy and minimizing off-target effects [15]. Unlike natural cytokines, which often have short half-life or undesirable immune activation functions, the proteins Design reduces immunogenicity and improves pharmacokinetics. De novo protein design also allows safer and more effective treatments to be devised for cancer or other diseases. By concentrating the immune response on a specific target.

4. Conclusion

Theoretical principles like Anfinsen's principle, combined with thermodynamic analysis and careful control over molecular interactions, make it possible to predict stable folds and design proteins that function in a biological environment. Such deterministic success is the justification for the development of protein design as a general field. As ways are found to quantify the role played in folding by energetic and entropic changes, continuing increases in sophistication seem likely. Through methodological improvements in computational tools these could be used to engineer proteins whose hydrogen bonds, packing of hydrophobic groups, or surface charges are optimally balanced. These advances have brought real results: stabilized, structure-guided immunogen design has proven helpful in vaccine development, as for hitherto refractory pathogens which conventional methods failed to handle; and immunotherapy is extended by synthetic cytokine mimetics into well targeted hormones with fewer side effects In future, AI-accelerating modeling, high-throughput screening and structural validation will come together to speed the generation of individually designed proteins for medicine, biotechnology and other uses. De novo protein design will become a critical technology for future biomedical innovation.

References

[1] Anfinsen, C. B. The genetic control of tertiary protein structure. Cold Spring Harbor Symposia on Quantitative Biology, 1963, 28: 439–449.

[2] Agbas, A.Baki. Trends of Protein Aggregation in Neurodegenerative Diseases, 2018, 10.5772/intechopen.81224.

[3] Huang, P.-S., Boyken, S. E., & Baker, D. The coming of age of de novo protein design. Nature, 2016, 537(7620): 320–327.

[4] Leaver-Fay, A., et al. Rosetta3: An object-oriented software

suite for the simulation and design of macromolecules. Methods in Enzymology, 2011, 487: 545–574.
[5] Douglas, M., Choi, J., & Mary Ann Clark. Biology 2e. Rice University, 2018.
[6] Watson, J. L., Juergens, D., Bennett, N. R., et al. De novo design of protein structure and function with RFdiffusion. Nature, 2023, 620:1089–1100.
[7] Desmet, J., De Maeyer, M., Hazes, B., & Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature, 1992, 356(6369): 539–542.
[8] Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., & Baker, D. Design of a novel globular protein fold with atomic-level accuracy. Science, 2003, 302(5649): 1364–1368.
[9] Huang, P.-S., Ban, Y.-E. A., et al. RosettaRemodel: A generalized framework for flexible backbone protein design. PLoS ONE, 2011, 6(8): e24109.
[10] Lin, Y.-R., Koga, N., Vorobiev, S. M., & Baker, D. Cyclic oligomer design with de novo αβ-proteins. Protein Science, 2017, 26(11), 2187–2194.
[11] McLellan, J. S., Chen, M., Joyce, M. G., Sastry, M., & Graham, B. S. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. Science, 2013 ,342(6158): 592–598.
[12] McLellan, J.S. RCSB PDB - 7KQD: Prefusion RSV F Bound to RV521. Rcsb.org, 2020.
[13] Fabian Sesterhenn et al., De novo protein design enables the precise induction of RSV-neutralizing antibodies. Science368, eaay5051, 2020.
[14] Boyoglu-Barnum, S., et al. Quadrivalent influenza nanoparticle vaccines induce broad protection. Nature, 2020, 592(7855): 623–628.
[15] Silva, D.-A., Yu, S., Ulge, U. Y., et al. De novo design of potent and selective mimics of IL-2 and IL-15. Nature, 2019, 565(7738): 186–191.