Biotechnology Advances 83 (2025) 108656

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Advances of computational protein design: Principles, strategies and applications in nutrition and health

Ziling Zhao [a,b,1], Qiyang Qu [a,1], Fuwei Sun [c], Jiachen Zang [a], Bowen Zheng [a], Tuo Zhang [a], Guanghua Zhao [a], Chenyan Lv [a,*], Zhongjiang Wang [c,*]

[a] College of Food Science & Nutritional Engineering, Beijing Key Laboratory of Functional Food from Plant Resources, China Agricultural University, Beijing, China
[b] School of Automation, Beijing Institute of Technology, Beijing, China,
[c] College of Food Science, Northeast Agricultural University, Haerbin, Heilongjiang, China

ARTICLE INFO

Keywords:
Computational protein design
Physics-mediated
AI-mediated
De novo design

ABSTRACT

Computational methods and AI technology have had a profound impact on protein design, significantly enhancing the ability to predict protein structures and create proteins with custom-tailored functions. With the help of computational methods, traditional protein design strategies such as directed evolution, fusion protein, and key subunit interface redesign show unprecedented progress in the design of various protein biomaterials such as nanocages, nanocarriers, antibodies, biocatalytic enzymes and inhibitory peptides. Strategies include physics-mediated design, which leverages the physical principles underlying protein structure and dynamics, and AI-mediated design, which employs machine learning techniques to generate and optimize protein configurations. Together, these approaches represent the cutting-edge methodologies in the rational design of novel proteins with desired functions. By using these approaches, novel protein molecules, assemblies, antibodies and responsive nanofibrils were constructed, which can be further applied in the field of nutrition and health.

1. Introduction

Proteins play essential roles in the fundamental processes of living organisms and carry out a variety of functions such as immune defense, neurotransmitter transmission, and signal transduction. These functions are often related to the structure of the proteins. The composition of their multidimensional structures has a crucial impact on their functionality. Due to the development of protein science, computational protein design is increasingly applied to address a number of key challenges in nutrition, biomedicine and biological engineering.

Herein, we summarize the principles of protein design and the main forces driving protein folding. More importantly, we review key strategies in computational protein design along with their emerging applications in the fields of nutrition and health, which aim to provide intuitive insights for researchers engaged in protein engineering.

2. Basic principles of protein design

An ideal protein backbone generally involves a combination of constituent α-helix and β-strand. To idealize this type of backbone, the lengths, folding arrangement, and the order of the backbone component are first to be specified. Then, low-energy amino acid sequence of each backbone is identified by combinatorial rotamer optimization. Lastly, protein structure prediction system is used to test whether the sequence is corresponded to the initial ideal backbone (King et al., 2015; Koga et al., 2012; Lin et al., 2015). In most cases, backbone design involves several rules that describe tertiary motifs of adjacent secondary structure components.

For an instance of RosettaDesign methodology, the orientation of βα- and αβ-unite, parallel (P) or antiparallel (A), and the chirality of ββ-unite, left handed (L) or right handed (R), required to be defined (Fig. 1). The vector of Cα-to-Cβ, $\overrightarrow{C_\alpha C_\beta}$, is defined as the vector of the strand residue which is preceding or following the linking loop. $\vec{u}$ is defined as the vector along the first strand and $\vec{v}$ as the vector from the kernel of the first strand to the kernel of the second strand. The chirality of ββ-unite follows the basic rules: if the value of $(\vec{u} \times \vec{v}) \bullet \overrightarrow{C_\alpha C_\beta}$ is positive, the chirality of ββ-unite is R; if the value of $(\vec{u} \times \vec{v}) \bullet \overrightarrow{C_\alpha C_\beta}$ is

* Corresponding authors.
E-mail addresses: 2019023@cau.edu.cn (C. Lv), wzjname@neau.edu.cn (Z. Wang).
[1] Ziling Zhao and Qiyang Qu contributing equally to this work. They both are listed as co-first author.

https://doi.org/10.1016/j.biotechadv.2025.108656
Received 9 January 2025; Received in revised form 6 July 2025; Accepted 20 July 2025
Available online 24 July 2025

negative, the unite is L. The orientation of βα- and αβ-unite generally follows such rules: one βα unite is P/A if the $\overrightarrow{strand\ to\ helix}$ is parallel/antiparallel to the $\overrightarrow{C_\alpha C_\beta}$ of the last residue in strand; one αβ-unite is P/A if the $\overrightarrow{C_\alpha C_\beta}$ of the first residue in strand is parallel/antiparallel to the vector $\overrightarrow{helix\ to\ strand}$. Moreover, the loop length of one strand also has effect on βα- and αβ-unite orientation. For βα-unite, the loop length affects its backbone bendability which decides the angle between adjacent secondary structure elements, and the angle is about ≤60° between the helix and strand. For αβ-unite, despite its loop length is from 2 to 4 residues, the preferred orientation is P as hydrogen-bonded helix capping provides longer loop length, which can stabilize the helix. Generally, these simulations with difference in loop length strongly consistent with the structure distribution of the native proteins. It should be noted that these rules not rely on the amino acid sequence of the secondary structure elements or connecting loops, but on their intrinsic characteristics (Koga et al., 2012).

3. Main forces for protein folding

It is widely accepted that native structures of proteins have the lowest free energy (Anfinsen, 1973; Fändrich et al., 2001). Actually, whatever proteins will be designed, the first factor we should take into account is their stability.

Globular proteins always contain hydrophobic core constituted by non-polar amino acids, while those polar amino acids are situated on the surface of the sphere protein (Fig. 2). It is such a structure that segregates non-polar residues from solvent to achieve the most stable state of energy. Moreover, the hydrophobic effect mainly stems from the entropy reduction effect induced by disruption of hydrogen bond network between water molecules (Kauzmann, 1959).

Besides hydrophobic effect, hydrogen bonding also plays a critical role in protein folding, particularly in stabilizing protein structures (Fig. 2). When hydrophobic residues are packaged into the core, N—H and C═O polar groups from hydrophobic residues backbone that ought to have formed hydrogen bonds with solution molecules are away from solution, triggering the disruption of the hydrogen bonds and thus causing the increase of free energy. To shape more stable protein architectures, the backbone polar groups from protein interior must interact with each other to form new hydrogen bonds to compensate, otherwise stripping water induces the large energy cost will disfavor folding (Huang et al., 2016a).

When the residues with opposite charges contact with each other (the distance <5 Å), electrostatic interaction and salt bridge will be formed (Fig. 2). For globular protein in water solution, the residues with charges tend to the exterior polar environment. However, the residues with charges can as well enter the interior hydrophobic environment during protein folding. To stabilize protein structure, these charged residues in protein core will interact with other charged residues to form salt bridge (Kumar and Nussinov, 2002). Therefore, if the decreased free energy is able to compensate the desolvation cost, the protein architecture will be more stable.

Additionally, when atoms of protein contact extremely close to each other, the van der Waals force contributes to protein stability significantly (Fig. 2). Nevertheless, van der Waals couple hydrogen bonds in most cases. On one hand, van der Waals force causes N—H and C═O group more close that enhance hydrogen bonds; on the other hand, hydrogen bonds shorten the distance between atoms so that form and thus increase van der Waals force (Chen et al., 2000; Nelson et al., 2005). Hence, the rationale of protein folding is the guidance for computational protein custom-design.

4. Computational protein design strategy

Proteins can be designed with specific topology and functional features. Protein design, which follows the process of specifying a desired function, designing a structure, executing this function, and finding a sequence that folds into this structure (Chu et al., 2024), has tremendously increased the number of available viable nanomaterials by exploring virtually the totality of the possible sequence and structural space, beyond that populated by natural proteins (Hamley, 2019). Traditional methods of protein design have largely relied on laborious and costly trial-and-error processes in the laboratory, constrained by experimental conditions and technological limitations. In recent years, advancements in computational technologies and algorithms have led to the emergence of computational protein design as a powerful approach in protein engineering. Computational protein design leverages sophisticated computational models and techniques to predict and generate protein sequences and structures with desired properties. This approach not only accelerates the design process but also expands the possibilities for creating novel proteins that might be challenging or impossible to develop through conventional methods.

Computational protein design approaches can be broadly categorized into physics and geometry mediated methods, and AI-mediated strategies. The former relies on fundamental physical principles and structural constraints to model protein folding and stability, while the latter leverages data-driven algorithms and machine learning to generate and optimize novel protein sequences and structures. In this part, all the strategies were illustrated and the possible applications were highlighted. Normally, the methods will be combined to overcome

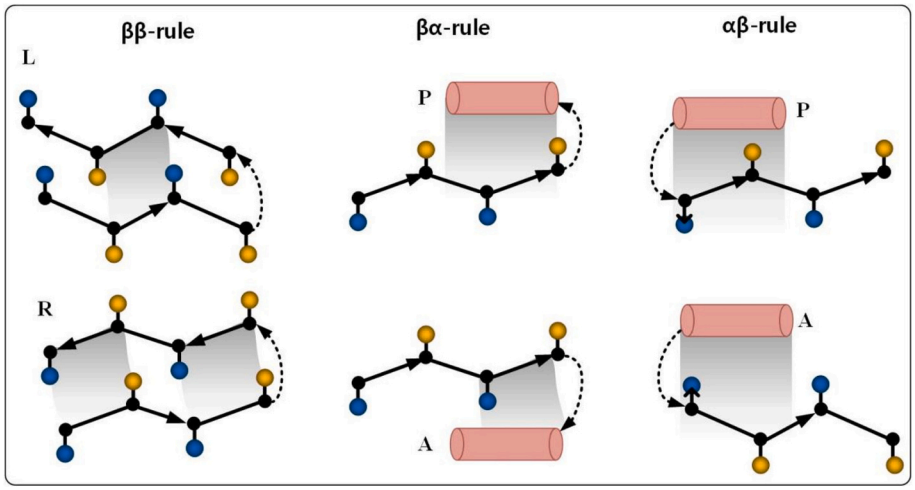

Fig. 1. RosettaDesign methodology. Reprinted with permission from Ref. (Koga et al., 2012). Copyright © 2012, Nature Publishing Group.

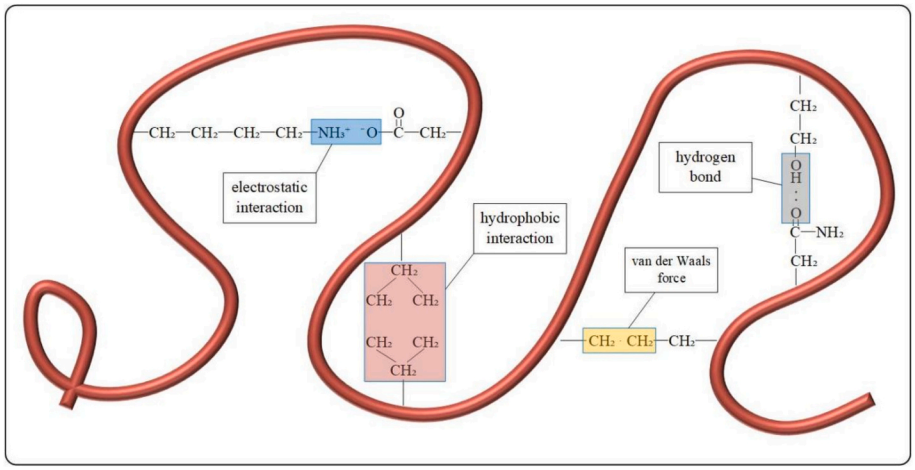

Fig. 2. Main forces for protein folding.

the challenges and to advance the design of functional protein.

4.1. Definition and challenges of computational protein design

Computational protein design is most fundamentally formulated as an optimization problem. Within this framework, the design of a desired structure (or function) can be achieved by identifying the optimal arrangement of amino acids to ensure the protein fold reaches its best possible state in terms of functionality and stability.

A key challenge in protein design is the immense complexity of the sequence and structure space. For 20 naturally occurring amino acids, each residue in a protein sequence has 20 possible choices. For a protein consisting of 100 residues, the number of possible sequences is 20^100, an astronomically large number that far exceeds the total number of atoms in the universe.

Due to the vast search space of proteins, where functional proteins occupy only a small fraction, scoring functions are necessary to evaluate whether the designed protein sequences are successful or not. Common scoring functions include energy-based functions (Gordon et al., 1999) typically used in physics and geometry mediated design, and knowledge-based functions (Reynolds et al., 2013) typically used in AI-mediated design. Energy-based functions estimate protein stability by considering detailed molecular mechanics and simulating physical interactions at the atomic level, such as atomic packing interactions, hydrogen bonding, electrostatic interactions, and solvation terms. Knowledge-based functions utilize AI techniques and statistical information derived from known protein structures to evaluate the likelihood and stability of designed proteins. Scoring functions quantify various physical and chemical properties of proteins, providing comparisons and rankings of different protein sequences and conformations.

The extensive search space inherent to proteins presents a substantial challenge for scoring functions, necessitating a delicate balance between computational efficiency and accuracy. As for energy-based function, achieving this balance often involves approximations on certain terms. However, this approximation has proven difficult, as fast and highly approximate functions tend to have poor correlation with the true free energy of proteins (Kortemme, 2024). On the other hand, knowledge-based functions, which leverage machine learning approaches, do not rely on these approximations, and thus show promising prospects for handling large-scale sequence evaluations more effectively.

4.2. Physics and geometry mediated protein design strategy

4.2.1. Strategies and essential procedures

Based on the above difficulties and challenges, most physics-based protein design approaches simplify the process into two steps: structure generation and sequence optimization. This two-step method effectively addresses the complexities involved in designing functional proteins by breaking down the problem into manageable parts. The procedures of physics and geometry mediated protein design are shown in Fig. 3.

(1) *Structure generation*

The first step involves designing the protein backbone, which includes determining the overall 3D structure of the protein at the atomic level. This step focuses on the spatial arrangement of the main chain atoms and the secondary structure elements (alpha-helices, beta-sheets, *etc.*). It is imperative that the designed backbone is feasible, meaning it must be capable of being realized by at least one amino acid sequence that can stably fold into the intended structure. The most straightforward approach to ensure the feasibility of the design is to utilize natural protein backbones. This method centers on redesigning or optimizing existing sequences based on the framework of existing protein structures to enhance current functionalities. For example, researchers have utilized current backbones to redesign enzyme substrate specificity (Ollikainen et al., 2015), enhance protein thermostability (Wijma et al., 2014), and optimize enzyme catalytic efficiency (Risso et al., 2013).

As for *de novo* design of backbone structure, helical bundles were firstly solved by utilizing Crick's parameterization due to their inherent regularity (Crick, 1953; Hill et al., 2000). This systematic sampling method and the high stability of helical bundles make them excellent models for designing a variety of functions, including ligand binding (Polizzi and DeGrado, 2020), ion transport (Joh et al., 2014), and switches (Langan et al., 2019). Extensive efforts in design and experimental validation have culminated in the development of a comprehensive "periodic table" of coiled-coil architectures (Moutevelis and Woolfson, 2009). The applicability of Crick's parameterization extends beyond coiled-coils, encompassing a wide range of helical bundle architectures that exhibit exceptional thermostability when subjected to laboratory testing (Grigoryan and DeGrado, 2011; Huang et al., 2014). Moreover, the inherent regularity in the geometry of helices permits the precise alignment and fusion of helices from distinct proteins, thereby enabling the generation of structurally diverse and larger helical assemblies (Huang et al., 2016a; Jacobs et al., 2016; Sahtoe et al., 2022). The wide array of designable all-helical structures continues to form the foundation for numerous successful applications in *de novo* protein design (Kortemme, 2024). Despite the substantial progress in designing α-helical proteins, achieving more complex functions may necessitate the incorporation of structures that diverge from canonical helical geometries. For further insights into the progress of coiled-coil designs, Woolfson's review provides additional details (Woolfson, 2017).

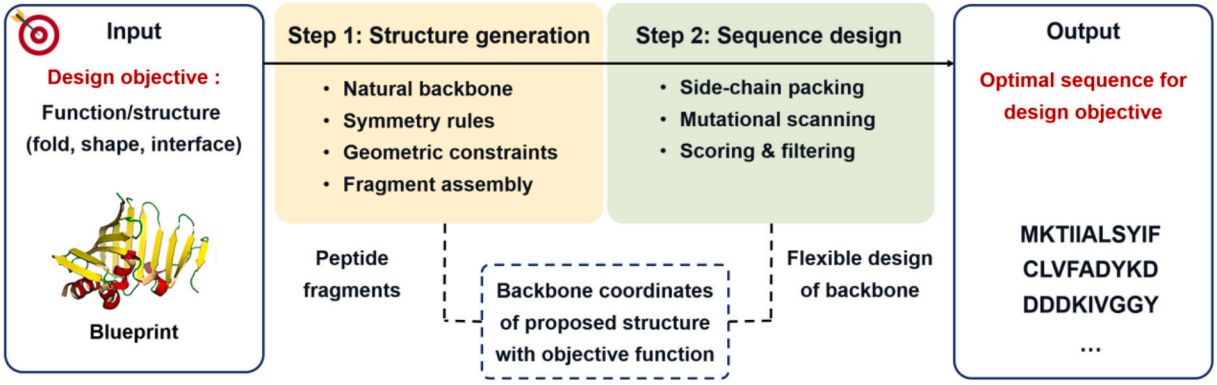

Fig. 3. Procedures of physics and geometry mediated protein design.

For more general proteins containing both α-helices and β-strands, the design process typically begins with the creation of a blueprint. At this stage, designers specify the arrangement and interconnections of secondary structure elements, essentially creating a roadmap for protein architectures. The next phase is achieved through the strategic assembly of peptide fragments, which are extracted from the vast repository of the Protein Data Bank (PDB). A prime example is the creation of Top7 (Kuhlman et al., 2003), a pioneering protein designed using fragment assembly techniques, featuring a fold topology previously unseen in nature. In certain scenarios, the design process necessitates the application of specific rules derived from natural protein structures. A notable example is the creation of symmetrical TIM-barrel proteins (Huang et al., 2016b), where precise hydrogen bonding between side chains and the backbone is crucial for maintaining the proper alignment of barrel segments. Alternative assembly strategies for backbone design without defining a blueprint have also emerged. One such method involves the use of modular leucine-rich repeat units to create proteins with controllable curvatures (Park et al., 2015). Another innovative approach is the Structure Extension with Native-substructure Graphs (SEWING) technique. This method combines helical building blocks, either continuous or discontinuous, sourced from existing proteins (Jacobs et al., 2016).

(2) *Sequence optimization*

The second step focuses on designing the amino acid sequence that will fold into the predetermined backbone structure. This step involves selecting specific amino acids that will form stable interactions and ensure the desired functionality of the protein. The sequence design must consider various factors such as hydrophobicity, charge distribution, and potential interaction sites.

In the sequence optimization process, considering backbone flexibility has become a crucial factor. Early side-chain design methods typically assumed a fixed protein backbone, but this assumption has limitations. Due to the Lennard-Jones potential term in scoring functions being highly sensitive to distance changes (usually inversely proportional to the 12th power of distance), even minor adjustments to the backbone structure can lead to significant energy changes. To address this issue, modern side-chain design methods simultaneously consider sampling of side-chain rotamers and local backbone conformations (Ollikainen et al., 2015; Georgiev et al., 2008; Keedy et al., 2012; Georgiev and Donald, 2007). The importance of considering backbone flexibility has been demonstrated in practical protein design cases. For example, the designing process of Top7 (Kuhlman et al., 2003) used the Rosetta (Das and Baker, 2008) program to first generate a new protein backbone, followed by iterative cycles of (1) selecting amino acid sequences that best fit a given fixed backbone conformation, and (2) slightly adjusting the backbone geometry to better accommodate the newly selected sequence. This example highlights a key concept: protein backbones are not fixed but change, albeit often slightly, with sequence changes during design or as proteins perform their functions. To account for this backbone flexibility in the design process, researchers have developed various methods: (1) inserting backbone minimization steps between fixed backbone designs, as in the Top7 (Kuhlman et al., 2003) example; (2) sampling small backbone adjustments during the design process (Georgiev et al., 2008; Ollikainen et al., 2013); or (3) pre-generating backbone conformation ensembles on which sequences are then designed and scored (Davey and Chica, 2012; Friedland and Kortemme, 2010). The development of these methods has greatly improved the accuracy and reliability of protein design, making it possible to design more complex and functional proteins.

4.2.2. *Applications in artificial protein design*

(1) Design of amyloidogenic peptide-binding proteins.

Amyloidogenic peptides, such as amyloid-β (Aβ), play a significant role in the formation of amyloid fibrils, which are associated with diseases like Alzheimer's. Designing protein binders that can specifically capture and inhibit these peptides is a promising approach to preventing fibril formation (Sahtoe et al., 2024).

Based on the physics and geometry mediated protein design strategies, researchers employed Rosetta software for blueprint-based backbone building, a computational design approach used to generate a scaffold with a β-sheet structure and a deep peptide-binding cleft. The scaffold was specifically designed to capture amyloidogenic peptides that form β-strands. The design process involved starting from a pre-existing scaffold topology (designed by FoldIt (Koepnick et al., 2019) players) and extending the β-sheet using Rosetta to create a binding interface capable of stabilizing the β-strand conformation of the peptide through hydrogen bonding. To further reinforce the structure, additional α-helices were introduced to support the β-sheet. After generating the backbone, the scaffold and peptide sequences were optimized by Rosetta to ensure high-affinity binding, while minimizing unsatisfied polar atoms in the buried interface. The scaffold was further refined by introducing hydrophobic interaction pairs in the solvent-exposed areas of the interface to enhance stability and specificity. These designed amyloid-β (Aβ) peptides can prevent their aggregation into toxic fibrils, which are key contributors to Alzheimer's disease. These protein binders offer potential therapeutic and research applications by inhibiting fibril formation and serving as tools for studying amyloid-related diseases.

(2) Computational design of pH-responsive protein assemblies.

Recent advances in computational protein design have enabled the construction of synthetic protein assemblies that respond sharply and reversibly to environmental pH changes. Two representative studies demonstrate distinct but convergent strategies for designing pH-

responsive protein nanomaterials with precise control over assembly-disassembly dynamics.

In one study, researchers (Shen et al., 2024) designed helical protein filaments by transforming pH-sensitive trimeric coiled-coils containing buried histidine-mediated hydrogen bond networks into monomeric units *via* loop fusion (Fig. 4). These monomers were computationally docked into a wide range of helical filament arrangements using rigid-body sampling, followed by interface redesign to stabilize specific filament architectures. The resulting filaments exhibited highly cooperative and reversible pH-dependent disassembly over narrow pH ranges (as small as 0.3 pH units), tunable by adjusting the number and location of buried histidines. RosettaDesign was used to optimize inter-subunit interface sequences and tune histidine protonation thresholds, achieving sharp responsiveness and structural stability. Cryo-EM confirmed the fidelity of the designed architecture, supporting applications in controlled release and smart biomaterials.

In a parallel study, researchers (Yang et al., 2024) developed pH-responsive, non-porous antibody nanoparticles through modular design. An octahedral nanoparticle framework was constructed from independently expressed tetramers (C_4 axis) and Fc domains (C_2 axis), while the unoccupied C_3 symmetry axes were sealed with engineered trimeric "plug" subunits (Fig. 4). To achieve pH sensitivity, the trimeric plugs were designed to include multiple histidine residues at the interface, enabling protonation-dependent weakening of hydrogen bond and packing interactions. These plugs were computationally docked using RPXDock (Sheffler et al., 2023) and extended *via* WORMS helical fusion (Hsia et al., 2021), and Rosetta was used to optimize the sequences at the plug–core interface. By varying the number and positioning of histidine residues, the researchers fine-tuned the disassembly pH thresholds, achieving plug release across a tunable pH range. Cryo-EM reconstructions validated the structural assembly, and functional assays confirmed effective encapsulation and pH-triggered release of protein or RNA cargo.

Despite their architectural differences, both systems leverage histidine-mediated hydrogen bond interactions to encode pH sensitivity into protein assemblies. Sequence optimization *via* Rosetta allows tuning of disassembly thresholds and structural integrity. These studies exemplify how computational design can program environmentally responsive behavior into synthetic protein assemblies, opening avenues for precision drug delivery, cargo protection, and dynamic biomaterials.

4.2.3. *Applications in key interface protein redesign*

Proteins interact with each other through protein interfaces, which greatly enhance our understanding of biological functions (Keskin et al., 2008). By analyzing the three-dimensional structures, information on protein interfaces will be obtained. It is well known that protein-protein interactions at protein interfaces are the chief contributors to construct the diversified protein nanostructures (Jones and Thornton, 1996; Kortemme and Baker, 2004; Schreiber and Fleishman, 2013). Redesign of these interfaces will benefit the development of protein design. Among all of the interfaces, key interface of protein controls the size, shape and function of protein or protein assemblies. A variety of interactions including non-covalent interactions such as hydrogen bonding, electrostatic interactions and hydrophobic interactions, and covalent interactions including disulfide bonds and metal coordination bonds contribute to the key inter-subunit interfaces. In 2016, our group has described an engineering strategy termed key subunit interface redesign (KSIR), which has been widely used for fabrication of non-native multisubunit protein architectures (Zhang et al., 2016). This strategy includes three steps: determining key subunit interfaces of a target symmetric protein architecture; identifying amino acid residues that not involved in interfacial interactions but located at the key subunit interface; deleting the amino acid residues not involved in interfacial interaction and redesign protein (Fig. 5). Although this strategy was firstly used to redesign protein nanocage according to our experiences in ferritin, computational aided method has been combined with key interface redesign strategy, and tried in symmetrical or non-symmetrical multisubunit proteins (Liu et al., 2022; Zhang et al., 2021).

By using this strategy, our group was able to facilitate the conversion of dimeric ferritin to 24-meric ferritin (Liu et al., 2022). Similarly, our research group (2019) presented an efficient disulfide-mediated approach to construct 3 protein variants based on an 8-mer bowl-like protein architecture (NF-8). This NF-8 protein was converted into a 24-mer ferritin-like nanocage or a 16-mer lenticular nanocage by

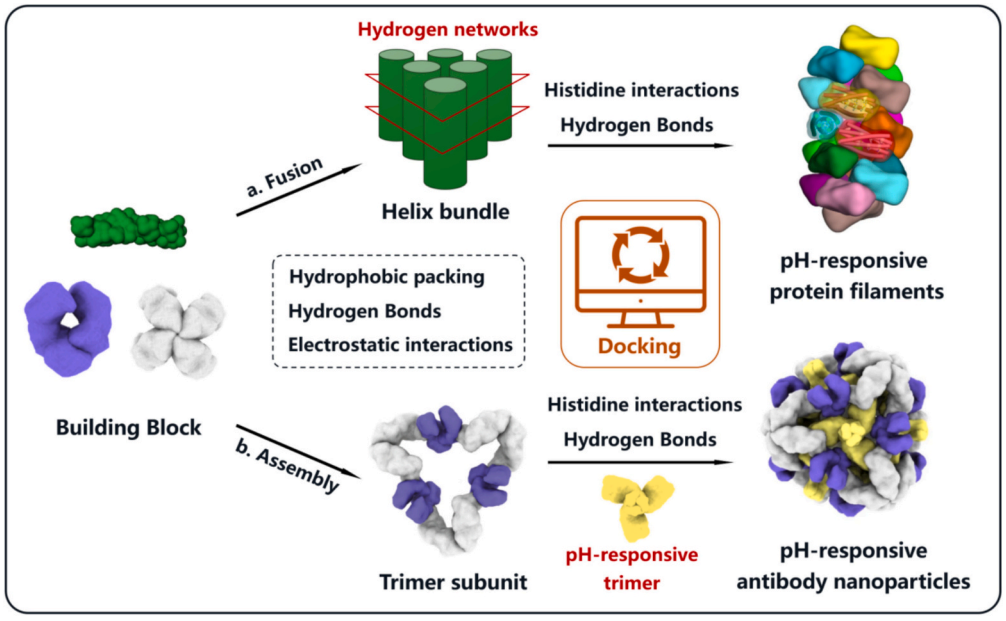

Fig. 4. Design of two representative pH-responsive protein assemblies. (a) Computational design of pH-responsive helical protein filaments. Coiled-coil bundles containing buried histidine residues are fused into monomers, forming hydrogen bond networks that drive disassembly upon protonation. (b) Design of pH-responsive antibody nanoparticles. Tetramers, Fc domains, and histidine-rich trimeric plugs are modularly assembled along symmetry axes. Protonation of interfacial histidines under acidic conditions enables controlled disassembly and cargo release. Reprinted with permission from Ref. (Shen et al., 2024) and Ref. (Yang et al., 2024). Copyright © 2024, Nature Publishing Group.

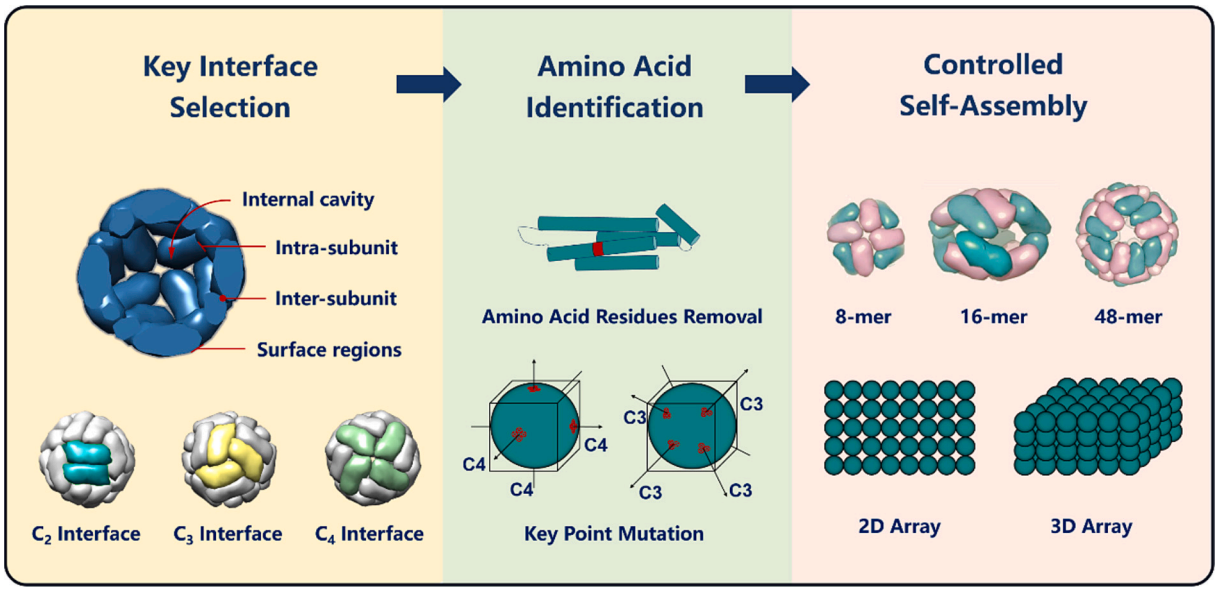

Fig. 5. Schematic illustration of the key subunit interface redesign strategy (KSIR) for engineering protein nanocages. Left: Key interface selection, based on structural analysis of subunit–subunit interactions (C_2, C_3, and C_4 interfaces) and spatial features such as internal cavities and interfacial regions. Middle: Identification and removal of non-contributory amino acid residues at the selected interface regions, followed by rational point mutation of silent residues that do not participate in interfacial stabilization. Right: Controlled self-assembly of the redesigned protein building blocks leads to the formation of discrete nanocage architectures with distinct oligomeric states (*e.g.*, 8-mer, 16-mer, and 48-mer), as well as higher-order 2D and 3D arrays. Reprinted with permission from Ref. (Zhang et al., 2016). Copyright © 2016, American Chemical Society.

selective deletion of an inherent intra-subunit disulfide bond or selective insertion of an inter-subunit disulfide bond respectively. Assembly of 48-mer protein nanocage happened when the above-mentioned 2 modification scenarios were combined onto NF-8 (Zang et al., 2019).

4.2.4. *Applications in hierarchical protein design*

Hierarchical protein design involves engineering proteins in a manner that imitates natural protein architecture to ensure each level of the hierarchy (primary, secondary, tertiary, and quaternary) is designed to achieve a desired shape or function. In hierarchical protein design, the theories behind are mainly based on the principles of (i) protein structure-function relationship; (ii) protein folding thermodynamics; (iii) molecular Interactions. The protein structure and function dedicate to explore the influences of each protein structural level (primary to quaternary) to its functionality, serving as the foundation for discovering potential applications. This hierarchy often involves designing or modifying the interface between proteins or protein subunits to achieve a desired architecture or function, for example designing the bioreactor or nanocage for drug delivery.

In recent years, the typical methods for hierarchical design of synthetic proteins or protein complex are various, such as functional peptides, alpha-helical coiled coils, self-assembling fusion proteins, and metal-mediated protein architectures. Zaccai et al. (2011) designed and characterized the first stand-alone coiled-coil hexamer (CC-Hex) based on the principles of sequence-to-structure relationship of coiled-coil domains stems and knobs-into-holes (KIH) packing (Fig. 6). Coiled-coil sadomains are structural motifs generally found in proteins which are characterized by more than two alpha helixes interacting with each other to form a supercoil (*e.g.*, coiled-coil dimer, coiled-coil trimer and coiled-coil tetramer). These alpha helixes are typically featured as a repeating pattern of seven amino acids called heptad repeat (HPPHPPP)n, and denoted as adcdefg. The H and P in the pattern represent hydrophobic and polar amino acid residues respectively. The knobs-into-holes (KIH) packing is the typical interaction between side chains of these amphiphilic α helixes, where knobs are large hydrophobic residues like leucine and isoleucine fit into holes (smaller residues such as glycine and alanine) between the neighboring helix. In this study, the researchers started from a fully *de novo* designed coiled-coil tetramer (CC-Tet), where position abcdefg were leucine (L), alanine (A), alanine (A), isoleucine (I), lysine (K), Glutamine (Q), and glutamic acid (E). Then, they attempted to expand the hydrophobic interfaces between the adjacent helices of coiled-coil tetramer (CC-Tet) to make higher-order coiled-coil oligomer (*e.g.*, CC-Hex) by altering the peripheral KIH interactions provided by amino acids at position *e.g.*, b and c. With this principle in mind, they exchanged all of the lysine (K) at position e into alanine (A), and successfully built a CC-Hex with an inner channel diameter around 6 Å which is permeable to water. Their findings of CC-Hex provide a foundation of the *de novo* design based on the α-helical coiled coils for a wide range of biological, therapeutic and material science applications (Zaccai et al., 2011).

Inspired by the former study, Thomas et al. (2016) posted modularity assembly strategies of peptide nanotubes by using hexameric α-helical barrels as building blocks (Fig. 7). They started with a well-developed hexameric α-helical barrels, and premutated them at the N-termini to expose hydrophobic residues, leaving complementary positive and negative charged residues at N-termini and C- termini. These mutants (CC-Hex-T) were further proved to be associated longitudinally and laterally, giving a broadened multilayer peptide nanotubes assembly at neutral pH. Besides, these assemblies were pH-sensitive where they can be reversed into identical single hexameric peptide nanotubes (3–4 nm) at acidic pH below 5.6. For further making a single longitudinally extended peptide nanotubes without lateral association at neutral pH, they redesigned the CC-Hex-T by mutating the f position glutamine on the heptad sequence repeats into lysine, resulting an enhanced positive charge of +3 per peptide and + 18 per hexameric building block (CC-Hex-T+) at neutral pH. These CC-Hex-T+ were further proved to be assembled into fibrils with length around 1 mm and 3–4 nm diameters at neutral pH, meaning no lateral association was happened, only longitudinal extension.

In addition to alpha-helical coiled coils based assembly, the symmetry-based protein assembly concept is one of the most basic and widely used strategies for constructing artificial protein nanocages or modifying natural protein nanocages for the desired application since a symmetric structure can lead to energetically favorable interactions

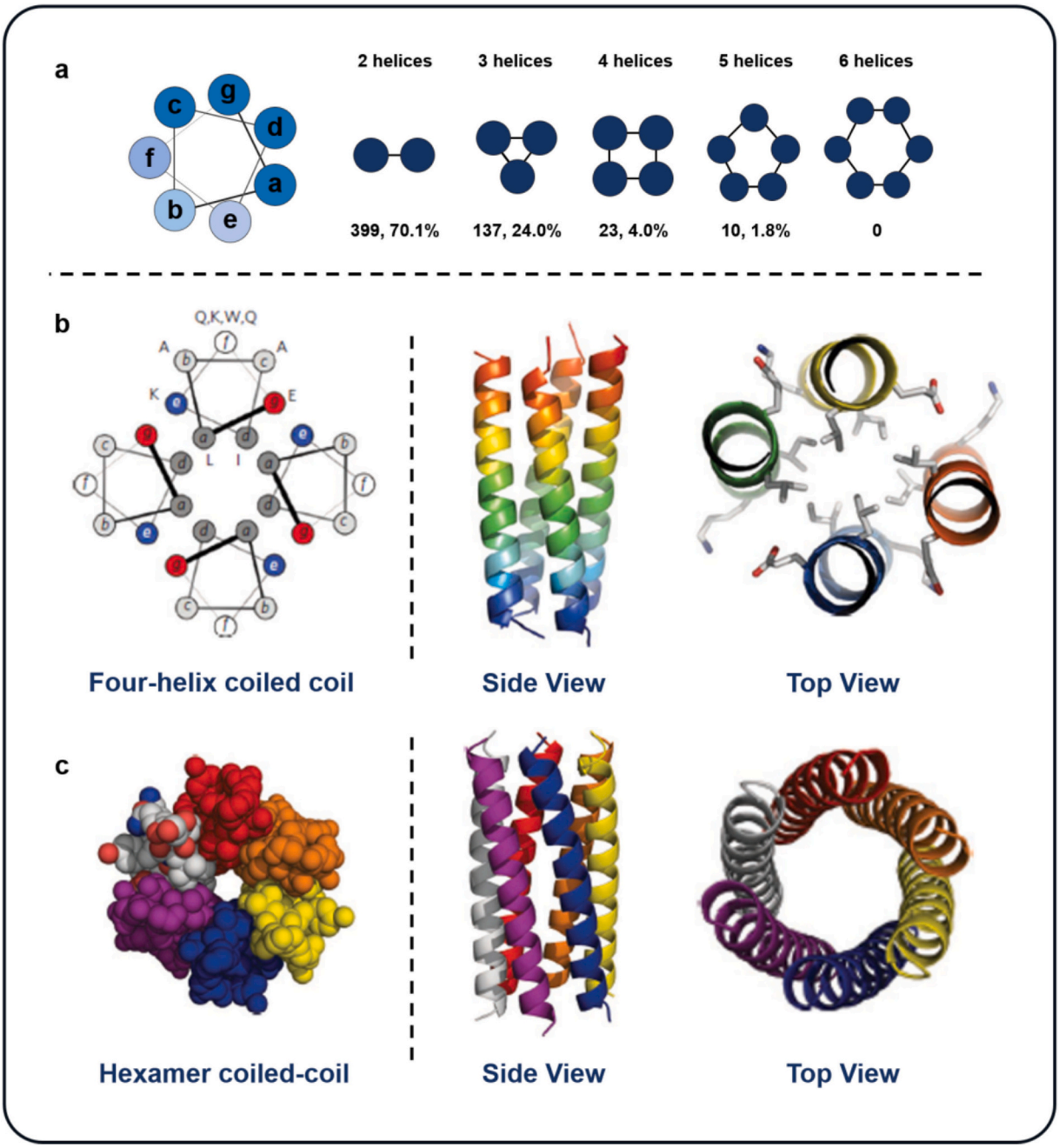

Fig. 6. Schematic representation of coiled-coil helical wheel and possible assemblies based the alpha helices as building block. (a) Helical wheel consisted of a heptad repeat residues abcdefg; (b) Structure of a four-helix coiled coil (CC-Tet); (c) Structure of a parallel coiled-coil hexamer (CC-Hex). Reprinted with permission from Ref. (Zaccai et al., 2011). Copyright © 2011, Nature Publishing Group.

under evolutionary stress (Chen et al., 2021). Padilla and co-workers constructed protein nanocages following the principle of symmetry (Fig. 8). They conducted 2 oligomer proteins (a dimer and a trimer) connected each other into a single larger molecule called a fusion protein by genetic manipulation. This fusion protein hence brings the characteristic of the selected dimer and trimer, where each dimer and trimer on this fusion protein has a strong tendency to interact with the other copies of itself. Through this process, many protein nanocages self-assembled by these fusion proteins were produced (Padilla et al., 2001).

Biswas et al. (2009) designed a metal-ion-induced tubular biocontainer by using the mutant chaperonin GroEL. GroEL is a large, oligomeric complex composed of 14 identical subunits, arranged in two stacked heptameric rings. In their study, the K311 and L314 amino acid residues in GroEL were mutated to cystine, and the mutant GroEL which bears 14 cystine residues in each entrance region was used to react with spirobenzopyran-appended maleimide (SPMI) to produce the GroELSP/MC. These genetically and chemically modified GroELSP/MC can simultaneously assemble into micro-level nanotubes with the presence of $MgCl_2$, and can be cut into short-chain nanotubes by adding EDTA. The mutation sites and the supramolecular assembly interfaces were rationally selected based on prior structural modeling, and the photochromic conjugation units were designed using computational simulations to ensure predictable responsiveness and structural compatibility. Moreover, this nanocage system was proven to successfully trap denatured R-lactalbumin, indicating its potential for serving as a novel protein nanocage for bio-therapy (Biswas et al., 2009). As a continuation of the above research, Shuvendu and coworkers functionalized the surface of the protein nanocage by using boronic acid derivative, providing

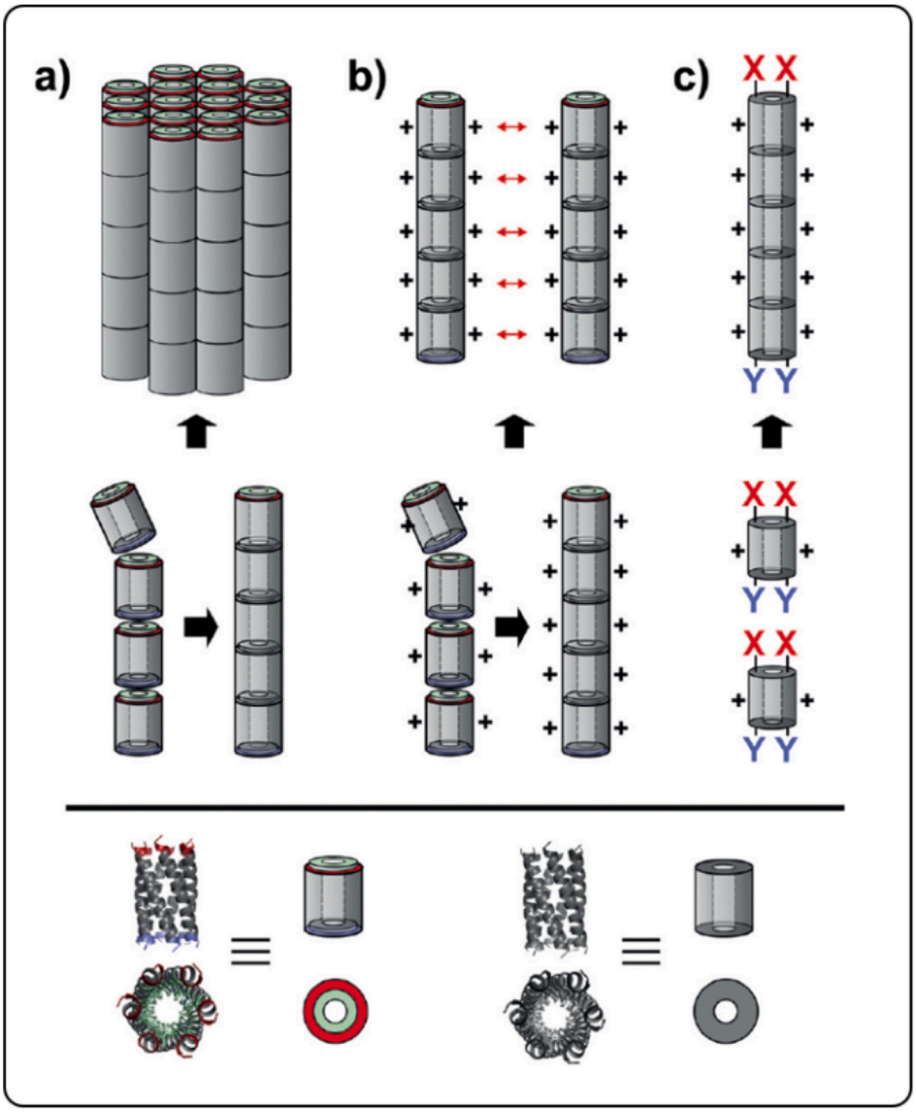

Fig. 7. Schematic representation of Peptide nanotubes assembly. (a) Longitudinally and laterally assembly of CC-Hex-T; (b) Single longitudinally assembled CC-Hex-T+; (c) Chemical ligation stabilized peptide nanotube fibrils assembled by CC-Hex-T + co; Red: positively charged N-terminus; Blue: negatively charged C-terminus. Reprinted with permission from Ref. (Thomas et al., 2016). Copyright © 2016, Royal Society of Chemistry. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

them the ability to penetrate the cells. Besides, they developed an approach where the intracelluar adenosine-5′-triphosphate (ATP) was induced as the key to disassemble the protein nanocage. The hydrolysis of the intracellular ATP into ADP generates a mechanical force leading to the conformational changes of the chaperonin units, which subsequently release the guest molecules trapped in the protein nanocages. Moreover, the release kinetics are sigmoidal depending on ATP concentration, where a dramatic disassembly of the protein nanocage occurs at around 45 μM ATP concentration, meaning these protein nanocages can selectively release the cargo based on the intracellular ATP concentration. Such ATP-triggered behavior was further validated and optimized through *in silico* simulations of intracellular ATP distributions and nanocage dynamics. These are promising results for tumor therapy since the concentration of the extracelluar ATP in tumors is around 700 μM compared with a normal tissue (< 5 μM) (Biswas et al., 2013). From theory to practice, researchers (Ashley et al., 2011) proposed a way to modify MS2 virus-like particles (VLPs) by peptide (S94) to achieve the purpose of selectively delivering cargos such as chemotherapeutic drugs, siRNA cocktails and protein toxins to human tumor cells. The peptide SP94 (SFSIIHTPILPL) is a specific sequence designed computationally to target hepatocellular carcinoma which is a type of liver cancer. In this study, the modified MS2 VLPs exhibit a 10^4-fold higher avidity for HCC compared to normal cells (*e.g.*, endothelial cells, monocytes, lymphocytes and hepatocytes). Moreover, when ricin toxin A-chain (RTA) loaded MS2 VLPs were simultaneously modified by SP94 targeting peptide and a histidine-rich fusogenic peptide (H5WYG), they kill the entire population of Hep3B without altering the activity of control cells.

In a physics-based rational design framework, researchers (Yang et al., 2025) engineered high-affinity binders for several immunomodulatory receptors using concave helical scaffolds. These scaffolds were manually modeled to complement the convex binding surfaces of targets such as PD-L1, CTLA-4, and TGFβRII. The interface regions were optimized *via* Rosetta FastDesign and iterative mutational scanning, guided by energy-based scoring functions. The designed binders demonstrated sub-nanomolar affinities and high functional activity, as confirmed by co-crystal structures and cell-based immune assays. These binders function as immune checkpoint modulators, offering potential therapeutic applications in cancer immunotherapy.

Meanwhile, the designed hierarchical protein or protein assemblies

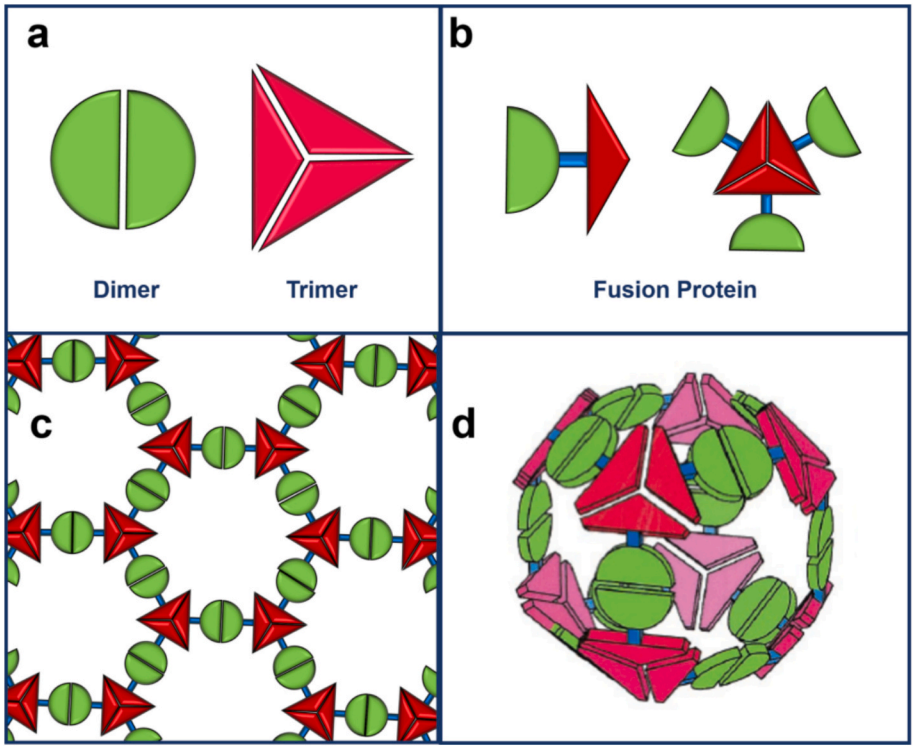

Fig. 8. Schematic representation of a general strategy for fusion protein design. (a) A green dimeric domain with C2 symmetry and a red trimeric domain with C3 symmetry. (b) Construction of fusion proteins by genetically connecting dimeric and trimeric domains through rigid α-helical linkers (blue), enabling specific geometric orientation. (c) Two-dimensional layer formed by self-assembly of the fusion protein when dimeric and trimeric symmetry axes are parallel and non-intersecting, yielding hexagonal symmetry. (d) Three-dimensional protein cage formed by fusion proteins assembling into a tetrahedral nanostructure, following the design rules for dimer-trimer-based architectures. Reprinted with permission from Ref. (Padilla et al., 2001). Copyright © 2001, National Academy of Sciences, U. S.A. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

can be applied in cargo delivery. Our research group developed a dual-compartment system using 3D ferritin-based nanocage arrays, of which can precisely realize spatiotemporal control over its self-assembly based on two designed switches (pH switch and metal switch) (Fig. 9). Through mutating the leu160 which protrudes into the C_4 channels on the previously designed T158HMjFer to histidine, they were able to achieve disassembly of ferritin (H158/H160MjFer) at a higher acidic pH (pH 3.5 *versus* pH 2.0 in conventional protocols), and reassembly of ferritin at mildly basic conditions (pH 8.0). This modification prevents acidic denaturation of ferritin, resulting in a significantly higher protein recovery rate (81.4 % *versus* 40 % from the previous studies) during the drug encapsulation process. Moreover, the introduced His motifs can function as ligands, coordinating with metal ions such as Ni^{2+} to connect four subunits through the C_4 rotational channels, which subsequently allows ferritin nanocages to assemble into 3D protein lattices where the interstitial spaces of the lattice can act as an additional cavity to encapsulate drugs alongside the individual ferritin cavities. Consequently, this system demonstrated the potential hierarchical encapsulation and release of two distinct cargo molecules (R6G and BHQ-2) through stepwise pH and metal switch, offering promising applications in hierarchical protein design for medical therapies (Zhang et al., 2022).

4.3. AI-mediated protein design

In recent years, artificial intelligence (AI) has emerged as a transformative force across various scientific domains, and protein design is no exception. The ability of deep learning to process vast amounts of data and identify complex patterns has opened a new pathway for designing proteins with high precision and efficiency. AI-mediated protein design leverages machine learning and deep learning algorithms to tackle challenges that were previously difficult to overcome. From generating protein backbones with defined topologies to designing amino acid sequences that encode desired functionalities, AI has reshaped the landscape of protein engineering.

Following the design process, structural validation is typically conducted to assess the folding stability and plausibility of the designed constructs. The latest versions of RosettaFold (All-Atom) and AlphaFold (version 3) have become essential tools in this stage. RosettaFold All-Atom (Krishna et al., 2024) is suitable for backbone refinement and symmetric assemblies, whereas AlphaFold 3 (Abramson et al., 2024) enables high-resolution modeling of complex systems involving ligands, nucleic acids, and protein–protein interfaces. These tools are commonly incorporated into post-design workflows to support structural verification prior to experimental validation.

4.3.1. Strategies and essential procedures

AI-based protein design methods, similar to physics-based approaches, are primarily divided into two key steps: structure generation and sequence optimization. It is noteworthy that AI methods capable of efficiently performing both structure generation and sequence optimization simultaneously are still relatively rare, representing a significant challenge and research direction in this field (Kortemme, 2024).

(1) *Structure generation*

AI-based approaches to protein structure generation represent a significant departure from traditional parametric or blueprint-based methods. A key distinguishing feature is that these AI methods do not necessarily require *a priori* definition of the desired protein structure or fold class. This flexibility opens up new possibilities for exploring the vast space of potential protein structures. Recent advancements in AI-based protein design have introduced diffusion models (Anand and

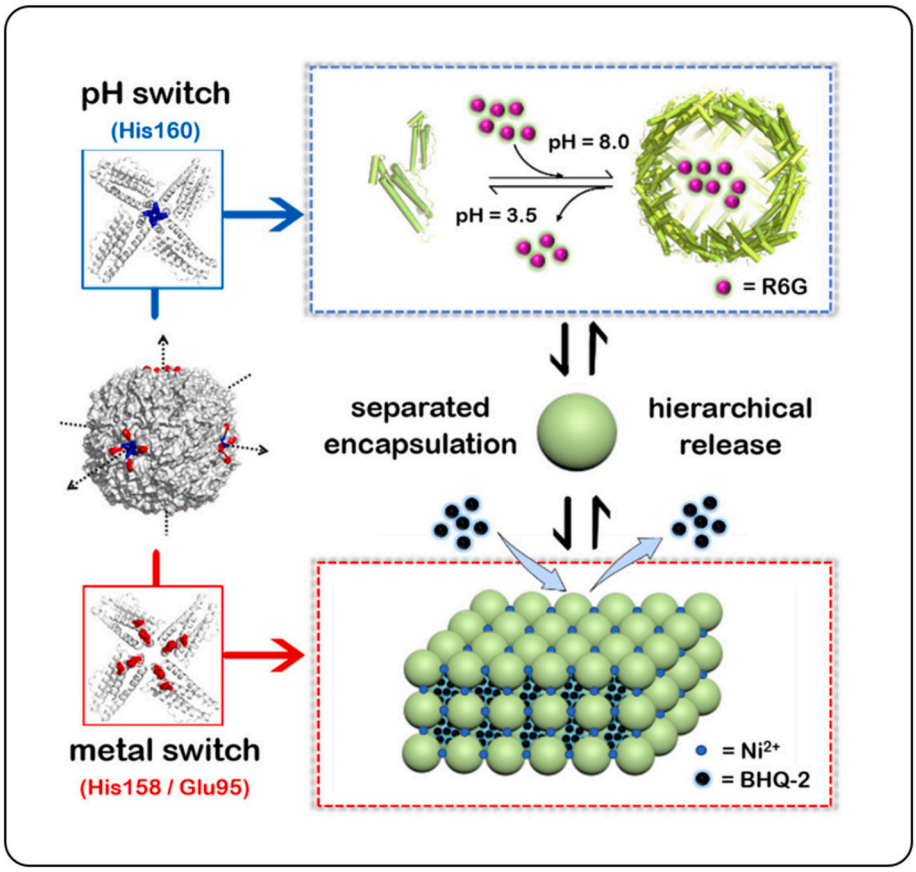

Fig. 9. Schematic representation of the dual-compartment system 3D ferritin-based nanocage arrays for stepwise encapsulation and release of two distinct cargo molecules (R6G and BHQ-2), where R6G was a synthetic dyes and BHQ-2 served as the quencher of R6G. Reprinted with permission from Ref. (Zhang et al., 2022). Copyright © 2022, Royal Society of Chemistry.

Achim, 2022; Watson et al., 2023; Ingraham et al., 2023), inspired by techniques used in image generation. These models start with protein structures, progressively add noise to the protein coordinates, and then train a network to recover the original structures from the noised samples. For design purposes, the process begins with random noise, and the denoising process generates protein structures with properties resembling typical proteins. Notable examples of this approach include RFdiffusion (Watson et al., 2023) and Chroma (Ingraham et al., 2023), which have been used to generate experimentally validated protein monomers, symmetrical assemblies, and protein binders.

A particularly promising aspect of diffusion models is their ability to be conditioned in various ways. This flexibility allows for generating specific fold topologies or preserving designated functional sites, opening up exciting possibilities for the *de novo* design of proteins with targeted molecular functions.

(2) *Sequence optimization*

AI has revolutionized protein sequence optimization by leveraging the vast amount of information available in protein sequence databases. Unlike traditional methods that rely heavily on structural information, AI-based approaches can extract valuable insights from sequence data alone, including sequences without known structures. Several machine learning models have been developed for protein sequence design, with large language models (LLMs) emerging as a particularly promising approach. These models, such as ProtGPT2 (Ferruz et al., 2022), ESM-2 (Lin et al., 2023), and ProGen (Madani et al., 2023), are trained to predict missing amino acids in protein sequences, similar to how language models predict missing words in sentences. Once trained, these models can generate entirely new protein sequences.

For example, ESM-2 (Lin et al., 2023), a language model trained solely on sequence data, has successfully designed new proteins that exhibit stability and monomeric behavior when tested experimentally. Notably, these designed proteins are predicted to have diverse structures, including some that differ significantly from naturally occurring proteins. This suggests that the model has learned fundamental principles of protein sequence-structure relationships that extend beyond its training examples. Another LLM, ProGen (Madani et al., 2023), was trained on sequences from over 19,000 protein families, including functional property labels. When fine-tuned on specific enzyme families, ProGen generated designed variants with catalytic parameters comparable to natural proteins, even for sequences with low similarity to those in the training set.

In addition to LLMs, structure-conditioned models such as ProteinMPNN (Dauparas et al., 2022) have shown strong performance in sequence design tasks by leveraging backbone geometry as input. ProteinMPNN employs a graph neural network (GNN) to generate sequences with high native recovery from backbone structures. By modeling amino acid residues and their spatial relationships as graph nodes and edges, GNNs can effectively capture the geometric and chemical context required for accurate sequence prediction. This allows the model to generalize across diverse topologies and structural motifs. Researchers (Sumida et al., 2024) enhanced this approach by fine-tuning the model with experimental data on expression and stability, leading to better success rates for complex targets such as membrane proteins. This refined model has been successfully applied in automated protein engineering pipelines, facilitating rapid screening and selection of functional variants at scale. More recently, LigandMPNN (Dauparas et al., 2025) extended this framework to include atomic-level context of ligands and cofactors, enabling accurate interface design and significantly

improved binding affinity in redesigned complexes.

These AI-based methods offer several advantages over traditional approaches. They can design functional sequences without relying on sequence alignments, which is particularly useful for highly diverse protein families like nanobodies. Additionally, they have shown promise in model-guided affinity maturation of antibodies. However, it is important to note that experimental validation remains crucial in determining the true success of these AI-designed sequences. While metrics like native sequence recovery are commonly used to evaluate model performance, they don't always correlate perfectly with experimental success. Even a single incorrectly predicted amino acid in a critical position can lead to experimental failure, despite only causing a small decrease in native sequence recovery.

4.3.2. Applications in protein design based on deep-learning neural networks

Deep-learning neural networks have recently revolutionized protein design by enabling the generation of novel protein structures and functional sequences beyond those found in nature. Unlike traditional structure prediction tools like AlphaFold and trRosetta—which focus on inferring structure from known sequences—deep learning models such as hallucination networks and transformer-based generators are actively used to design new proteins with defined folds, binding interfaces, and catalytic sites. These approaches leverage large-scale protein databases and structural patterns to learn complex sequence–function relationships, thereby facilitating the creation of synthetic proteins for therapeutic, diagnostic, and industrial applications.

A notable example demonstrating the integration of deep learning in protein design involves a multi-stage pipeline combining several neural network models (Lauko et al., 2025). Starting from random residue distributions and minimal active-site geometries, RFdiffusion generated protein backbones that scaffold complex catalytic triads with atomic precision, while LigandMPNN was used to assign functional sequences. To evaluate catalytic preorganization throughout the reaction cycle, the authors developed PLACER, a neural network that generates structural ensembles for each reaction intermediate and assesses key geometric features such as Ser-His hydrogen bonds and oxyanion hole formation. The resulting designs not only displayed catalytic efficiencies up to 2.2 $\times 10^5\ M^{-1}\ s^{-1}$ but also adopted novel folds not found in nature. Crystal structures of selected designs revealed root mean square deviations (RMSDs) of <1 Å from the computational models, underscoring the atomic accuracy achievable with AI-guided workflows. The study illustrates how deep learning frameworks can be integrated throughout the entire design cycle—from backbone generation to functional validation—to create enzymes with both novel structures and catalytic activity.

To validate and refine the structural fidelity of designed proteins, structure prediction tools such as RoseTTAFold have been widely employed. RoseTTAFold was first introduced in 2021 by David Baker and his research group (Fig. 10). This network was designed specifically to predict protein structures by deep learning, which is a subset of machine learning based on a three-track neural network (1D sequence alignment track, 2D distance matrix track, and 3D coordinate track). To be specific, the 1D sequence track dealt with protein amino acid sequence to predict protein structural properties; the 2D distance track process the pairwise distance between two amino acid residues to reveal spatial relation in different part of protein; the 3D-coordinate track refines the 3D protein structure prediction by integrating the information from the 1D and 2D tracks iteratively. This integration enhances the accuracy of the protein structure prediction, allowing the model to treat types of information simultaneously.

To test the prediction accuracy of RoseTTAFold from different aspects, four crystallographic datasets (a bacterial surface layer protein, a secreted protein Lrbp, a glycine N-acyltransferase, and a bacterial oxidoreductase), which were difficult to identify by molecular-replacement based models in the protein data bank, were reanalyzed by the constructed model. All cases showed high structural similarity to the real protein structures which can be estimated by molecular replacement. Besides, the modeling of G protein-coupled receptors (GPCRs) with unknown structure by RoseTTAFold provided high model accuracy where the prediction of their active and inactive state was quite resemble to their corresponding native GPCRs and the closest homolog of known structure. Lastly, accurate protein-protein complex models were directly generated from sequence information, where the template modeling score (TM-score) of predicting proteins of known structures was greater than 80 (*e.g.*, NADH-quinone oxidoreductase = 97, tryptophan synthase = 92, and tRNA-dependent amidotransferase = 89), meaning the predicted protein models were very resembling to the actual structures. This development of RoseTTAFold was a vital advancement in aspect of accurate protein structure prediction,

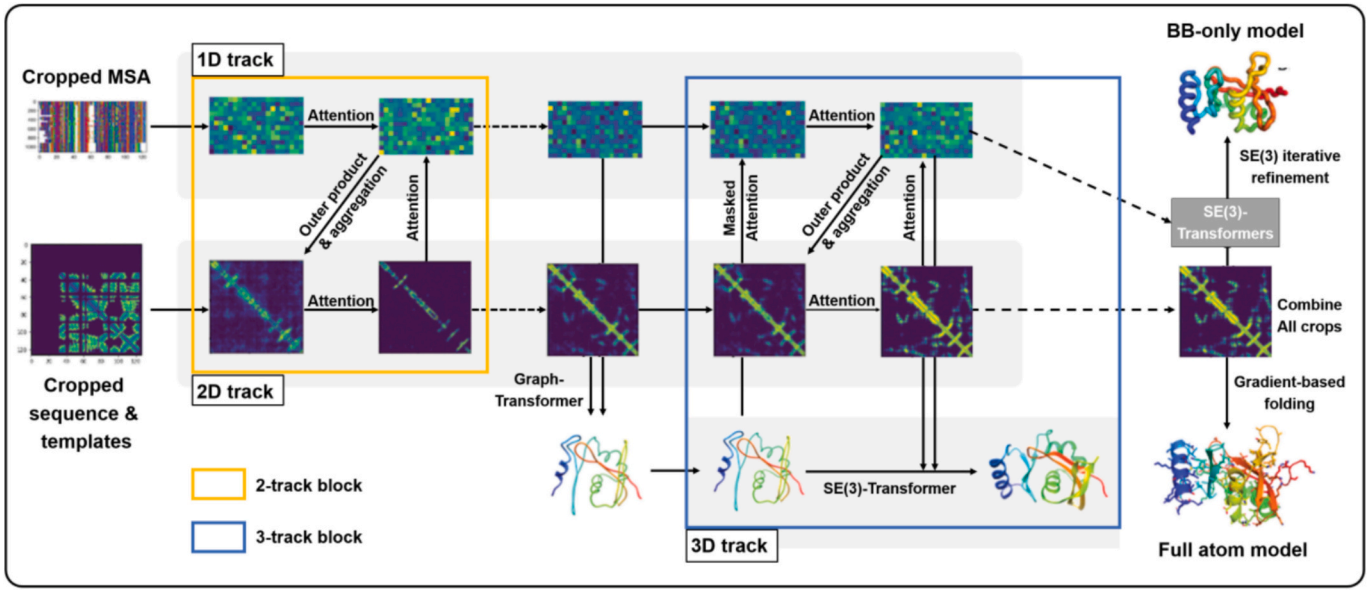

Fig. 10. Architecture of RoseTTAFold with 1D, 2D and 3D attention tracks, in which the tracks are iteratively connected together to check and balance the relationships between sequences, distances, and coordinates. Reprinted with permission from Ref. (Baek et al., 2021). Copyright © 2021, American Association for the Advancement of Science.

providing assistance of solving arduous x-ray crystallography and cryo-EM modeling problems as well as insights into protein function without experimentally determined structures. These insights are crucial for a protein complex or protein binder design for therapeutic usage (Baek et al., 2021).

In addition, Baker' group generates large numbers of protein scaffolds containing diverse pockets with ideal shape to hold the synthetic luciferin (*e.g.*, diphenylterazine, DTZ and 2-deoxycoelenterazine, CTZ) based on the deep learning-based hallucination approach. They started from designing a shape-complementary active site which stabilize the substrates (*e.g.*, anionic state of DTZ). Then, the remaining sequence of the protein scaffolds was designed using RosettaDesign or ProteinMPNN to predict and optimize protein sequences that can fold into desired three-dimensional structures, where 7648 and 46 designs were selected for further experimental screening of luciferases specific to DTZ and h-CTZ, respectively. After a colony-based screening method on the 7648 designs and an expression of the 46 designs, they got 3 active designs within the 7648 and 2 active designs within the 46. The most active luciferase is a small (13.9KDa) and thermostable (melting temperature > 95 °C) enzyme which has a higher substrate specify to DTZ and comparable catalytic efficiency ($K_{cat}/K_m = 10^6\ M^{-1}\ s^{-1}$) to native luciferases. Their results facilitate the development of computational enzyme design, where the lack of available scaffolds of desired enzymatic function can be eliminated with the help of deep learning neural network (Yeh et al., 2023).

One representative study highlights the AI-driven design of orally active miniproteins targeting Th17-related cytokines (Berger et al., 2024). The authors employed a deep generative model to construct initial protein scaffolds capable of binding IL-17 A and IL-23R. These scaffolds were then structurally refined and validated using AlphaFold2 to ensure the formation of stable folds and correct binding topologies. Subsequent interface redesign using Rosetta enabled the enhancement of binding affinity, protease resistance, and oral bioavailability. The resulting candidates showed picomolar binding to targets and outperformed therapeutic antibodies in murine models of colitis, establishing a proof-of-concept for deep learning-guided functional binder development. These designed proteins act as Th17 antagonists and hold promise for treating autoimmune inflammatory diseases such as inflammatory bowel disease.

In addition, Researchers generated signal peptides (SPs) using an attention based neural network called the transformer encoder-decoder model (Wu et al., 2020). The generated signal peptide displayed competitive secreted enzyme activities compared with the natural signal peptides. Furthermore, the identity of these generated signal peptides was various, sharing as little as 58 % sequence identity (73 % on average) to the closest signal peptides in Swiss-prot. Their success of applying attention-based neural network to generate synthetic signal peptides with desired function as the natural SPs stands for a significant progress in the field of computational biology. Despite the enormous progress of deep-learning methods, a general deep-learning framework for *de novo* binder design and design of higher-order symmetric architectures, has yet to be described.

4.3.3. Applications in protein design based on de novo strategy

De novo protein design removes the dependence on naturally evolved scaffolds, and has the potential for a deeper understanding of the contribution that every side chain makes toward the structure, stability and function of *de novo* proteins. In this part, the *de novo* designed protein was truly *de novo* proteins rather than those achieved through protein engineering or redesign. *De novo* protein design is an interdisciplinary field which combines principles from biology, chemistry, physics and computer science to create new proteins with desired functions and structures. With the help of computational design, massive research progress has been achieved in terms of *de novo* protein design, ligand design, and protein modification for therapeutic usage in recent decades.

In the last century, Bogan and coworkers (1998) proposed the concept of hot spots which refer to specific regions on the protein interface that prominently contribute to the binding energy. They found certain amino acid residues contribute the bulk of the binding energy and an irregular distribution of energetic contributions of individual amino acid residues across each interface. Besides, hot spots were found near the center of the interfaces and were surrounded by a shell of less energetic important residues, which isolated them from the bulk solvent. Moreover, they concluded three typical amino acids (tryptophan, arginine and tyrosine) with the highest frequency appeared in hot spots, since these amino acids are capable of providing multiple types of interactions for binding such as hydrophobic interaction, aromatic π-π interaction and hydrogen bonding. These important findings provided insightful guidance for the further development of computational protein design (Bogan and Thorn, 1998).

With these concepts in mind, Kuhlman and coworkers designed the first *de novo* design protein called Top7 which is a 93-residue alpha/beta protein. Top 7 was found to be folded and structurally stable, as well as the x-ray structure of Top 7 is similar to the design model (Kuhlman et al., 2003). Top7 stands for a milestone in the field of protein design, demonstrating the feasibility of creating new protein from scratch using computational methods. This breakthrough has laid the foundation for ongoing advances in protein engineering with applications across biotechnology, medicine, and materials science. Progressively, Fletcher and coworkers constructed unilamellar sphere cages with a diameter of around 100 nm by using their designed coiled-coil bundle peptides. These peptides were joined back-to-back to form 2 types of complementary peptide complexes hub A and hub B, which can render hexagonal networks and subsequently turn into cages (Fig. 11) (Fletcher et al., 2013).

Kaltofen et al. (2015) developed a *de novo* design protocol based on the crystal structure of the seven-residue amyloid-like peptide (GNNQQNY) from the yeast protein sup35. This peptide sequence can form a β-strand and was decorated to connect with a loop-helix-loop segment as the primary building block. These protein building blocks were simultaneously assembled into fibers with a most frequent width of 5 nm at the rational condition, which was almost the same as the prediction in the structural model fibril. These results proved the feasibility of precise engineering of new biomolecules by computational design (Kaltofen et al., 2015).

Besides, Park et al. (2015) proposed a strategy for controlling of the curvature of repeat-protein scaffolds using RosettaRemodel. By introducing a diverse set of leucine-rich repeat protein modules as building-blocks and junction modules (Fig. 12), they built 12 new proteins with different curvatures, which shows the feasibility of controlling the repeat-protein curvature with atomic-level accuracy (Park et al., 2015). Moreover, Shane Gonen and coworkers (2016) successfully designed two-dimensional protein arrays mediated by zipper-like hydrophobic packing.

In addition to the *de novo* protein design, progress has also been made in terms of customizing antibody and protein nanocarriers for encapsulation and controllable drug release. Poosarla et al. (2017) generated 5 types of *de novo* designed single-chain antibody (scFv) with reference to antibody 2D10. These 5 *de novo* designed single-chain antibodies shared less than 75 % sequence identity compared to all existing natural antibody sequences, and 3 of them show nano-molar binding affinities to the dodecapeptide. Their results demonstrate the *de novo* designed antibody can exhibit thermally and conformationally stable characteristics as well as high binding affinity to antigens, providing an insightful reference for the development of synthetic antibodies (Poosarla et al., 2017).

Similarly, miniproteins were designed that antagonize cytokine storm inducers by *de novo* strategies (Huang et al., 2024). Specifically, RosettaDesign was used to successfully create miniproteins that can efficiently bind and block the IL-6 receptor, GP130 co-receptor, and IL-1 receptor. These designed proteins demonstrated high binding affinity in

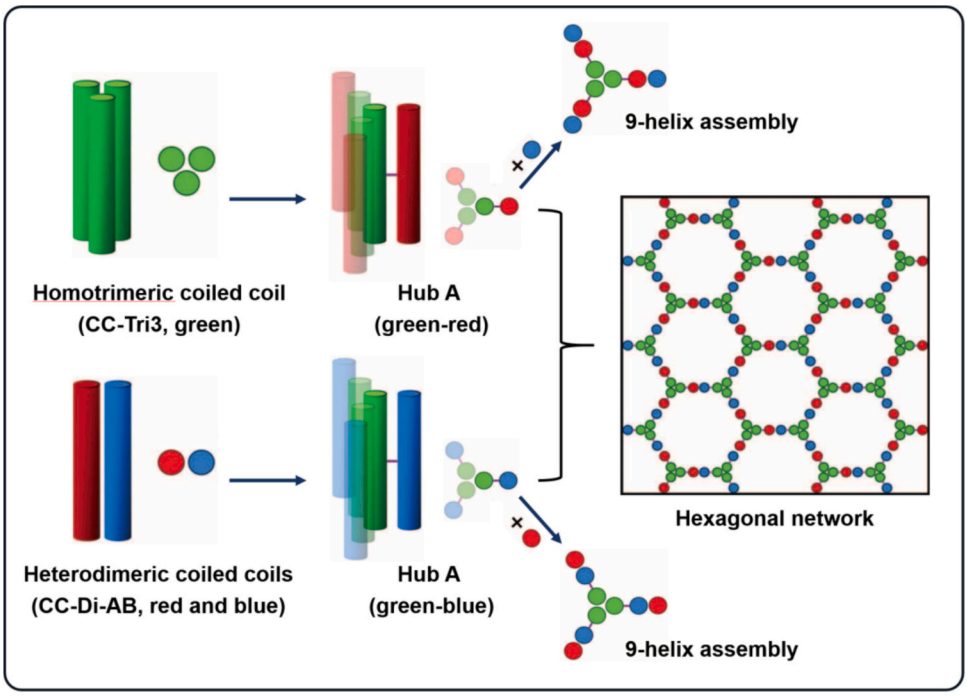

Fig. 11. Schematics for the design and self-assembly of peptide-based cages. Reprinted with permission from Ref. (Fletcher et al., 2013). Copyright © 2013, Association for the Advancement of Science.

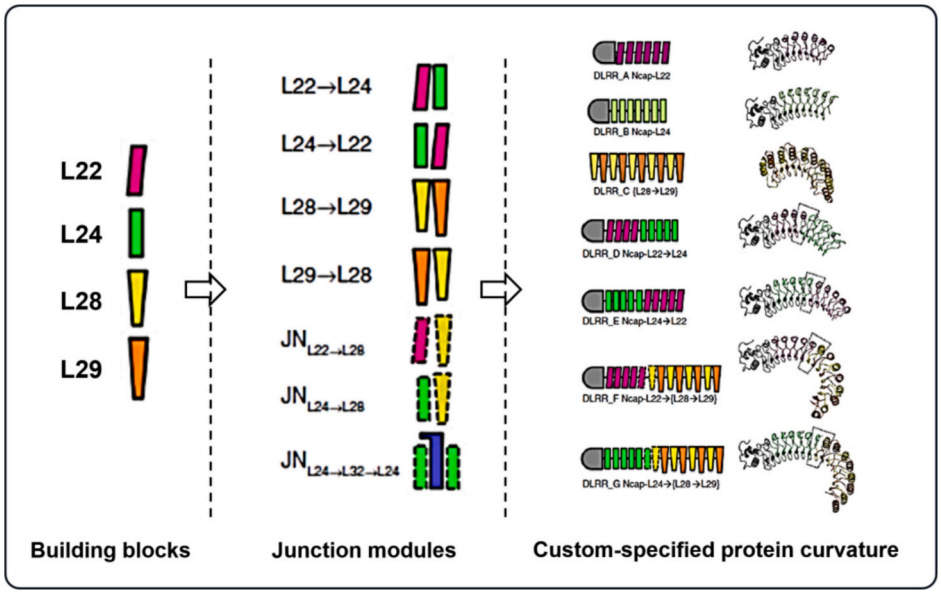

Fig. 12. leucine-rich repeat protein modules for building Custom-specified protein curvature. Reprinted with permission from Ref. (Park et al., 2015). Copyright © 2015, Nature Publishing Group.

the picomolar to nanomolar range, with their structures closely matching the design models, as confirmed by X-ray crystallography. In human heart organoid disease models, these IL-1 receptor antagonists showed protective effects against inflammation and heart damage. The significant potential of computational protein design in developing novel biologics was highlighted, particularly in addressing acute and localized inflammatory responses.

5. Outlook and challenges

The field of computational protein design is poised for a transformative phase, driven by rapid advancements in AI and big data technologies. These developments are making it increasingly feasible to design proteins with high precision and specific functions, marking a significant leap forward from traditional approaches. AI-mediated design, in particular, has enabled the creation of versatile protein structures and large assemblies with applications in areas such as vaccine development, where structural accuracy is paramount. The ability to design *de novo* proteins that tightly bind to target surfaces has broad implications, ranging from basic cell biology research to the development of novel therapeutics. While challenges remain in achieving more complex design goals, such as engineering proteins with advanced

functions like conformational changes and allosteric regulation, these objectives are gradually becoming attainable. The ongoing integration of computational strategies with biological systems also opens up new possibilities for controlling cellular processes in ways that were previously unimaginable.

Despite progress, key challenges remain. Predicting protein behavior—including binding affinities and dynamics—requires deep learning advances and high-quality datasets. Designing proteins with composite functions demands a deeper understanding of behavioral principles. There is also potential in *de novo* design of tunable, controllable, modular proteins, which could interact with biological systems in novel ways. Such innovations are particularly promising in the field of nutrition and health, where computationally designed proteins with controllable activity and site-specific targeting capabilities could enable precise modulation of physiological processes. Potential applications include targeted nutrient delivery, programmable enzyme supplementation, and the development of bioactive compounds that interact selectively with gut microbiota or metabolic pathways. These advances may drive a paradigm shift toward personalized, proactive, and precision-based nutritional interventions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Nos. 32372320).

Data availability

No data was used for the research described in the article.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Anand, N., Achim, T., 2022. Protein Structure and Sequence Generation With Equivariant Denoising Diffusion Probabilistic Models. arXiv preprint. https://doi.org/10.48550/arXiv.2205.15019.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science 181, 223–230. https://doi.org/10.1126/science.181.4096.223.

Ashley, C.E., Carnes, E.C., Phillips, G.K., Durfee, P.N., Buley, M.D., Lino, C.A., Padilla, D. P., Phillips, B., Carter, M.B., Willman, C.L., Brinker, C.J., Caldeira, J.D.C., Chackerian, B., Wharton, W., Peabody, D.S., 2011. Cell-specific delivery of diverse cargos by bacteriophage MS2 virus-like particles. ACS Nano 5, 5729–5745. https://doi.org/10.1021/nn201397z.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Dustin Schaeffer, R., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., Van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Christopher Garcia, K., Grishin, N.V., Adams, P.D., Read, R.J., Baker, D., 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science 373, 871–876. https://doi.org/10.1126/science.abj8754.

Berger, S., Seeger, F., Yu, T.-Y., Aydin, M., Yang, H., Rosenblum, D., Guenin-Macé, L., Glassman, C., Arguinchona, L., Sniezek, C., Blackstone, A., Carter, L., Ravichandran, R., Ahlrichs, M., Murphy, M., Pultz, I.S., Kang, A., Bera, A.K., Stewart, L., Garcia, K.C., Naik, S., Spangler, J.B., Beigel, F., Siebeck, M., Gropp, R., Baker, D., 2024. Preclinical proof of principle for orally delivered Th17 antagonist miniproteins. Cell 187, 4305–4317.e18. https://doi.org/10.1016/j.cell.2024.05.052.

Biswas, S., Kinbara, K., Oya, N., Ishii, N., Taguchi, H., Aida, T., 2009. A tubular biocontainer: metal ion-induced 1D assembly of a molecularly engineered chaperonin. J. Am. Chem. Soc. 131, 7556–7557. https://doi.org/10.1021/ja902696q.

Biswas, S., Kinbara, K., Niwa, T., Taguchi, H., Ishii, N., Watanabe, S., Miyata, K., Kataoka, K., Aida, T., 2013. Biomolecular robotics for chemomechanically driven guest delivery fuelled by intracellular ATP. Nature. Chem. 5, 613–620. https://doi.org/10.1038/nchem.1681.

Bogan, A.A., Thorn, K.S., 1998. Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280, 1–9. https://doi.org/10.1006/jmbi.1998.1843.

Chen, J., Lu, Z., Sakon, J., Stites, W.E., 2000. Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. J. Mol. Biol. 303, 125–130. https://doi.org/10.1006/jmbi.2000.4140.

Chen, H., Tan, X., Fu, Y., Dai, H., Wang, H., Zhao, G., Zhang, Y., 2021. The development of natural and designed protein nanocages for encapsulation and delivery of active compounds. Food Hydrocoll. 121, 107004. https://doi.org/10.1016/j.foodhyd.2021.107004.

Chu, A.E., Lu, T., Huang, P.-S., 2024. Sparks of function by de novo protein design. Nat. Biotechnol. 42, 203–215. https://doi.org/10.1038/s41587-024-02133-2.

Crick, F.H., 1953. The Fourier transform of a coiled-coil. Acta Crystallogr. 6, 685–689. https://doi.org/10.1107/S0365110X53001952.

Das, R., Baker, D., 2008. Macromolecular modeling with Rosetta. Annu. Rev. Biochem. 77, 363–382. https://doi.org/10.1146/annurev.biochem.77.062906.171838.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I. M., Courbet, A., de Haas, R.J., Bethel, N., Leung, P.J.Y., Huddy, T.F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A.K., King, N.P., Baker, D., 2022. Robust deep learning–based protein sequence design using ProteinMPNN. Science 378, 49–56. https://doi.org/10.1126/science.add2187.

Dauparas, J., Lee, G.R., Pecoraro, R., An, L., Anishchenko, I., Glasscock, C., Baker, D., 2025. Atomic context-conditioned protein sequence design using LigandMPNN. Nat. Methods 22, 717–723. https://doi.org/10.1038/s41592-025-02626-1.

Davey, J.A., Chica, R.A., 2012. Multistate approaches in computational protein design. Protein Sci. 21, 1241–1252. https://doi.org/10.1002/pro.2128.

Fändrich, M., Fletcher, M.A., Dobson, C.M., 2001. Amyloid fibrils from muscle myoglobin. Nature 410, 165–166. https://doi.org/10.1038/35065514.

Ferruz, N., Schmidt, S., Höcker, B., 2022. ProtGPT2 is a deep unsupervised language model for protein design. Nat. Commun. 13, 4348. https://doi.org/10.1038/s41467-022-32007-7.

Fletcher, J.M., Harniman, R.L., Barnes, F.R.H., Boyle, A.L., Collins, A., Mantell, J., Sharp, T.H., Antognozzi, M., Booth, P.J., Linden, N., Miles, M.J., Sessions, R.B., Verkade, P., Woolfson, D.N., 2013. Self-assembling cages from coiled-coil peptide modules Europe PMC funders group. Science 340, 595–599. https://doi.org/10.1126/science.1233936.Self-Assembling.

Friedland, G.D., Kortemme, T., 2010. Designing ensembles in conformational and sequence space to characterize and engineer proteins. Curr. Opin. Struct. Biol. 20, 377–384. https://doi.org/10.1016/j.sbi.2010.02.004.

Georgiev, I., Donald, B.R., 2007. Dead-end elimination with backbone flexibility. Bioinformatics 23, i185–i194. https://doi.org/10.1093/bioinformatics/btm197.

Georgiev, I., Keedy, D., Richardson, J.S., Richardson, D.C., Donald, B.R., 2008. Algorithm for backrub motions in protein design. Bioinformatics 24, i196–i204. https://doi.org/10.1093/bioinformatics/btn169.

Gordon, D.B., Marshall, S.A., Mayo, S.L., 1999. Energy functions for protein design. Curr. Opin. Struct. Biol. 9, 509–513. https://doi.org/10.1016/s0959-440x(99)80072-4.

Grigoryan, G., DeGrado, W.F., 2011. Probing designability via a generalized model of helical bundle geometry. J. Mol. Biol. 405, 1079–1100. https://doi.org/10.1016/j.jmb.2010.08.058.

Hamley, I.W., 2019. Protein assemblies: nature-inspired and designed nanostructures. Biomacromolecules 20, 1829–1848. https://doi.org/10.1021/acs.biomac.9b00228.

Hill, R.B., Raleigh, D.P., Lombardi, A., DeGrado, W.F., 2000. *De novo* Design of Helical Bundles as models for understanding protein folding and function. Acc. Chem. Res. 33, 745–754. https://doi.org/10.1021/ar970004h.

Hsia, Y., Mout, R., Sheffler, W., Edman, N.I., Vulovic, I., Park, Y.-J., Redler, R.L., Bick, M. J., Bera, A.K., Courbet, A., Kang, A., Brunette, T.J., Nattermann, U., Tsai, E., Saleem, A., Chow, C.M., Ekiert, D., Bhabha, G., Veesler, D., Baker, D., 2021. Design of multi-scale protein complexes by hierarchical building block fusion. Nat. Commun. 12, 2294. https://doi.org/10.1038/s41467-021-22276-z.

Huang, P.-S., Oberdorfer, G., Xu, C., Pei, X.Y., Nannenga, B.L., Rogers, J.M., DiMaio, F., Gonen, T., Luisi, B., Baker, D., 2014. High thermodynamic stability of parametrically designed helical bundles. Science 346, 481–485. https://doi.org/10.1126/science.1257481.

Huang, P.-S., Boyken, S.E., Baker, D., 2016a. The coming of age of de novo protein design. Nature 537, 320–327. https://doi.org/10.1038/nature19946.

Huang, P.-S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D.A., Höcker, B., Baker, D., 2016b. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. 12, 29–34. https://doi.org/10.1038/nchembio.1966.

Huang, B., Coventry, B., Borowska, M.T., Arhontoulis, D.C., Exposit, M., Abedi, M., Jude, K.M., Halabiya, S.F., Allen, A., Cordray, C., Goreshnik, I., Ahlrichs, M., Chan, S., Tunggal, H., DeWitt, M., Hyams, N., Carter, L., Stewart, L., Fuller, D.H., Mei, Y., Garcia, K.C., Baker, D., 2024. De novo design of miniprotein antagonists of cytokine storm inducers. Nat. Commun. 15, 7064. https://doi.org/10.1038/s41467-024-50919-4.

Ingraham, J.B., Baranov, M., Costello, Z., Barber, K.W., Wang, W., Ismail, A., Frappier, V., Lord, D.M., Ng-Thow-Hing, C., Van Vlack, E.R., Tie, S., Xue, V., Cowles, S.C., Leung, A., Rodrigues, J.V., Morales-Perez, C.L., Ayoub, A.M., Green, R., Puentes, K., Oplinger, F., Panwar, N.V., Obermeyer, F., Root, A.R., Beam, A.L.,

Poelwijk, F.J., Grigoryan, G., 2023. Illuminating protein space with a programmable generative model. Nature 623, 1070–1078. https://doi.org/10.1038/s41586-023-06728-8.
Jacobs, T.M., Williams, B., Williams, T., Xu, X., Eletsky, A., Federizon, J.F., Szyperski, T., Kuhlman, B., 2016. Design of structurally distinct proteins using strategies inspired by evolution. Science 352, 687–690. https://doi.org/10.1126/science.aad8036.
Joh, N.H., Wang, T., Bhate, M.P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G., DeGrado, W.F., 2014. De novo design of a transmembrane Zn $^{2+}$−transporting four-helix bundle. Science 346, 1520–1524. https://doi.org/10.1126/science.1261172.
Jones, S., Thornton, J.M., 1996. Principles of protein-protein interactions. Proc. Natl. Acad. Sci. U. S. A. 93, 13–20. https://doi.org/10.1073/pnas.93.1.13.
Kaltofen, S., Li, C., Huang, P.-S., Serpell, L.C., Barth, A., André, I., 2015. Computational de novo design of a self-assembling peptide with predefined structure. J. Mol. Biol. 427, 550–562. https://doi.org/10.1016/j.jmb.2014.12.002.
Kauzmann, W., 1959. Some factors in the interpretation of protein denaturation. In: Advances in Protein Chemistry. Academic Press, pp. 1–63. https://doi.org/10.1016/S0065-3233(08)60608-7.
Keedy, D.A., Georgiev, I., Triplett, E.B., Donald, B.R., Richardson, D.C., Richardson, J.S., 2012. The role of local backrub motions in evolved and designed mutations. PLoS Comput. Biol. 8, e1002629. https://doi.org/10.1371/journal.pcbi.1002629.
Keskin, O., Tuncbag, N., Gursoy, A., 2008. Characterization and prediction of protein interfaces to infer protein-protein interaction networks. Curr. Pharm. Biotechnol. 9, 67–76. https://doi.org/10.2174/138920108783955191.
King, I.C., Gleixner, J., Doyle, L., Kuzin, A., Hunt, J.F., Xiao, R., Montelione, G.T., Stoddard, B.L., DiMaio, F., Baker, D., 2015. Precise assembly of complex beta sheet topologies from de novo designed building blocks. Elife 4, e11012. https://doi.org/10.7554/eLife.11012.
Koepnick, B., Flatten, J., Husain, T., Ford, A., Silva, D.-A., Bick, M.J., Bauer, A., Liu, G., Ishida, Y., Boykov, A., Estep, R.D., Kleinfelter, S., Nørgård-Solano, T., Wei, L., Players, F., Montelione, G.T., DiMaio, F., Popović, Z., Khatib, F., Cooper, S., Baker, D., 2019. De novo protein design by citizen scientists. Nature 570, 390–394. https://doi.org/10.1038/s41586-019-1274-4.
Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., Baker, D., 2012. Principles for designing ideal protein structures. Nature 491, 222–227. https://doi.org/10.1038/nature11600.
Kortemme, T., 2024. De novo protein design—from new structures to programmable functions. Cell 187, 526–544. https://doi.org/10.1016/j.cell.2023.12.028.
Kortemme, T., Baker, D., 2004. Computational design of protein–protein interactions. Curr. Opin. Chem. Biol. 8, 91–97. https://doi.org/10.1016/j.cbpa.2003.12.008.
Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., McHugh, R., Vafeados, D., Li, X., Sutherland, G.A., Hitchcock, A., Hunter, C.N., Kang, A., Brackenbrough, E., Bera, A.K., Baek, M., DiMaio, F., Baker, D., 2024. Generalized biomolecular modeling and design with RoseTTAFold all-atom. Science 384, eadl2528. https://doi.org/10.1126/science.adl2528.
Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., Baker, D., 2003. Design of a Novel Globular protein fold with atomic-level accuracy. Science 302, 1364–1368. https://doi.org/10.1126/science.1089427.
Kumar, S., Nussinov, R., 2002. Close-range electrostatic interactions in proteins. ChemBioChem 3, 604. https://doi.org/10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X.
Langan, R.A., Boyken, S.E., Ng, A.H., Samson, J.A., Dods, G., Westbrook, A.M., Nguyen, T.H., Lajoie, M.J., Chen, Z., Berger, S., 2019. De novo design of bioactive protein switches. Nature 572, 205–210. https://doi.org/10.1038/s41586-019-1432-8.
Lauko, A., Pellock, S.J., Sumida, K.H., Anishchenko, I., Juergens, D., Ahern, W., Jeung, J., Shida, A.F., Hunt, A., Kalvet, I., Norn, C., Humphreys, I.R., Jamieson, C., Krishna, R., Kipnis, Y., Kang, A., Brackenbrough, E., Bera, A.K., Sankaran, B., Houk, K.N., Baker, D., 2025. Computational design of serine hydrolases. Science. https://doi.org/10.1126/science.adu2454.
Lin, Y.-R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A.F., Montelione, G.T., Baker, D., 2015. Control over overall shape and size in de novo designed proteins. Proc. Natl. Acad. Sci. U. S. A. 112, E5478–E5485. https://doi.org/10.1073/pnas.1509508112.
Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., Dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A., 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379, 1123–1130. https://doi.org/10.1126/science.ade2574.
Liu, Y., Zang, J., Leng, X., Zhao, G., 2022. A short helix regulates conversion of dimeric and 24-meric ferritin architectures. Int. J. Biol. Macromol. 203, 535–542. https://doi.org/10.1016/j.ijbiomac.2022.01.174.
Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos, J.L., Xiong, C., Sun, Z.Z., Socher, R., 2023. Large language models generate functional protein sequences across diverse families. Nat. Biotechnol. 41, 1099–1106.
Moutevelis, E., Woolfson, D.N., 2009. A periodic table of coiled-coil protein structures. J. Mol. Biol. 385, 726–732. https://doi.org/10.1016/j.jmb.2008.11.028.
Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.Ø., Riekel, C., Grothe, R., Eisenberg, D., 2005. Structure of the cross-β spine of amyloid-like fibrils. Nature 435, 773–778. https://doi.org/10.1038/nature03680.
Ollikainen, N., Smith, C.A., Fraser, J.S., Kortemme, T., 2013. Flexible backbone sampling methods to model and design protein alternative conformations. In: Methods in Enzymology. Elsevier, pp. 61–85. https://doi.org/10.1016/B978-0-12-394292-0.00004-7.
Ollikainen, N., de Jong, R.M., Kortemme, T., 2015. Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. PLoS Comput. Biol. 11, e1004335. https://doi.org/10.1371/journal.pcbi.1004335.
Padilla, J.E., Colovos, C., Yeates, T.O., 2001. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc. Natl. Acad. Sci. U. S. A. 98, 2217–2221. https://doi.org/10.1073/pnas.041614998.
Park, K., Shen, B.W., Parmeggiani, F., Huang, P.-S., Stoddard, B.L., Baker, D., 2015. Control of repeat-protein curvature by computational protein design. Nat. Struct. Mol. Biol. 22, 167–174.
Polizzi, N.F., DeGrado, W.F., 2020. A defined structural unit enables de novo design of small-molecule–binding proteins. Science 369, 1227–1233. https://doi.org/10.1126/science.abb8330.
Poosarla, V.G., Li, T., Goh, B.C., Schulten, K., Wood, T.K., Maranas, C.D., 2017. Computational de novo design of antibodies binding to a peptide with high affinity. Biotechnol. Bioeng. 114, 1331–1342. https://doi.org/10.1002/bit.26244.
Reynolds, K.A., Russ, W.P., Socolich, M., Ranganathan, R., 2013. Evolution-based design of proteins. In: Methods in Enzymology. Elsevier, pp. 213–235. https://doi.org/10.1016/B978-0-12-394292-0.00010-2.
Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., Sanchez-Ruiz, J.M., 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. J. Am. Chem. Soc. 135, 2899–2902. https://doi.org/10.1021/ja311630a.
Sahtoe, D.D., Praetorius, F., Courbet, A., Hsia, Y., Wicky, B.I.M., Edman, N.I., Miller, L. M., Timmermans, B.J.R., Decarreau, J., Morris, H.M., Kang, A., Bera, A.K., Baker, D., 2022. Reconfigurable asymmetric protein assemblies through implicit negative design. Science 375, eabj7662. https://doi.org/10.1126/science.abj7662.
Sahtoe, D.D., Andrzejewska, E.A., Han, H.L., Rennella, E., Schneider, M.M., Meisl, G., Ahlrichs, M., Decarreau, J., Nguyen, H., Kang, A., Levine, P., Lamb, M., Li, X., Bera, A.K., Kay, L.E., Knowles, T.P.J., Baker, D., 2024. Design of amyloidogenic peptide traps. Nat. Chem. Biol. 20, 981–990. https://doi.org/10.1038/s41589-024-01578-5.
Schreiber, G., Fleishman, S.J., 2013. Computational design of protein–protein interactions. Curr. Opin. Struct. Biol. 23, 903–910. https://doi.org/10.1016/j.sbi.2013.08.003.
Sheffler, W., Yang, E.C., Dowling, Q., Hsia, Y., Fries, C.N., Stanislaw, J., Langowski, M.D., Brandys, M., Li, Z., Skotheim, R., Borst, A.J., Khmelinskaia, A., King, N.P., Baker, D., 2023. Fast and versatile sequence-independent protein docking for nanomaterials design using RPXDock. PLoS Comput. Biol. 19, e1010680. https://doi.org/10.1371/journal.pcbi.1010680.
Shen, H., Lynch, E.M., Akkineni, S., Watson, J.L., Decarreau, J., Bethel, N.P., Benna, I., Sheffler, W., Farrell, D., DiMaio, F., Derivery, E., De Yoreo, J.J., Kollman, J., Baker, D., 2024. De novo design of pH-responsive self-assembling helical protein filaments. Nat. Nanotechnol. 19, 1016–1021. https://doi.org/10.1038/s41565-024-01641-1.
Sumida, K.H., Núñez-Franco, R., Kalvet, I., Pellock, S.J., Wicky, B.I.M., Milles, L.F., Dauparas, J., Wang, J., Kipnis, Y., Jameson, N., Kang, A., De La Cruz, J., Sankaran, B., Bera, A.K., Jiménez-Osés, G., Baker, D., 2024. Improving protein expression, stability, and function with ProteinMPNN. J. Am. Chem. Soc. 146, 2054–2061. https://doi.org/10.1021/jacs.3c10941.
Thomas, F., Burgess, N.C., Thomson, A.R., Woolfson, D.N., 2016. Controlling the assembly of coiled–coil peptide nanotubes. Angew. Chem. Int. Ed. 55, 987–991. https://doi.org/10.1002/anie.201509304.
Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Hanikel, N., Pellock, S.J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S.V., Lauko, A., De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T.S., DiMaio, F., Baek, M., Baker, D., 2023. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100. https://doi.org/10.1038/s41586-023-06415-8.
Wijma, H.J., Floor, R.J., Jekel, P.A., Baker, D., Marrink, S.J., Janssen, D.B., 2014. Computationally designed libraries for rapid enzyme stabilization. Protein Eng. Des. Sel. 27, 49–58. https://doi.org/10.1093/protein/gzt061.
Woolfson, D.N., 2017. Coiled-coil design: Updated and upgraded. In: Parry, D.A.D., Squire, J.M. (Eds.), Fibrous Proteins: Structures and Mechanisms, Subcellular Biochemistry. Springer International Publishing, Cham, pp. 35–61. https://doi.org/10.1007/978-3-319-49674-0_2.
Wu, Z., Yang, K.K., Liszka, M.J., Lee, A., Batzilla, A., Wernick, D., Weiner, D.P., Arnold, F. H., 2020. Signal peptides generated by attention-based neural networks. ACS Synth. Biol. 9, 2154–2161. https://doi.org/10.1021/acssynbio.0c00219.
Yang, E.C., Divine, R., Miranda, M.C., Borst, A.J., Sheffler, W., Zhang, J.Z., Decarreau, J., Saragovi, A., Abedi, M., Goldbach, N., Ahlrichs, M., Dobbins, C., Hand, A., Cheng, S., Lamb, M., Levine, P.M., Chan, S., Skotheim, R., Fallas, J., Ueda, G., Lubner, J., Somiya, M., Khmelinskaia, A., King, N.P., Baker, D., 2024. Computational design of non-porous pH-responsive antibody nanoparticles. Nat. Struct. Mol. Biol. 31, 1404–1412. https://doi.org/10.1038/s41594-024-01288-5.
Yang, W., Hicks, D.R., Ghosh, A., Schwartze, T.A., Conventry, B., Goreshnik, I., Allen, A., Halabiya, S.F., Kim, C.J., Hinck, C.S., Lee, D.S., Bera, A.K., Li, Z., Wang, Y., Schlichthaerle, T., Cao, L., Huang, B., Garrett, S., Gerben, S.R., Rettie, S., Heine, P., Murray, A., Edman, N., Carter, L., Stewart, L., Almo, S.C., Hinck, A.P., Baker, D., 2025. Design of high-affinity binders to immune modulating receptors for cancer immunotherapy. Nat. Commun. 16, 2001. https://doi.org/10.1038/s41467-025-57192-z.
Yeh, A.H.-W., Norn, C., Kipnis, Y., Tischer, D., Pellock, S.J., Evans, D., Ma, P., Lee, G.R., Zhang, J.Z., Anishchenko, I., Coventry, B., Cao, L., Dauparas, J., Halabiya, S., DeWitt, M., Carter, L., Houk, K.N., Baker, D., 2023. De novo design of luciferases

using deep learning. Nature 614, 774–780. https://doi.org/10.1038/s41586-023-05696-3.

Zaccai, N.R., Chi, B., Thomson, A.R., Boyle, A.L., Bartlett, G.J., Bruning, M., Linden, N., Sessions, R.B., Booth, P.J., Brady, R.L., Woolfson, D.N., 2011. A de novo peptide hexamer with a mutable channel. Nat. Chem. Biol. 7, 935–941. https://doi.org/10.1038/nchembio.692.

Zang, J., Chen, H., Zhang, X., Zhang, C., Guo, J., Du, M., Zhao, G., 2019. Disulfide-mediated conversion of 8-mer bowl-like protein architecture into three different nanocages. Nat. Commun. 10, 778. https://doi.org/10.1038/s41467-019-08788-9.

Zhang, S., Zang, J., Zhang, X., Chen, H., Mikami, B., Zhao, G., 2016. "Silent" amino acid residues at key subunit interfaces regulate the geometry of protein Nanocages. ACS Nano 10, 10382–10388. https://doi.org/10.1021/acsnano.6b06235.

Zhang, X., Liu, Y., Zheng, B., Zang, J., Lv, C., Zhang, T., Wang, H., Zhao, G., 2021. Protein interface redesign facilitates the transformation of nanocage building blocks to 1D and 2D nanomaterials. Nat. Commun. 12, 4849. https://doi.org/10.1038/s41467-021-25199-x.

Zhang, X., Zeng, R., Zhang, T., Lv, C., Zang, J., Zhao, G., 2022. Spatiotemporal control over 3D protein nanocage superlattices for the hierarchical encapsulation and release of different cargo molecules. J. Mater. Chem. B 10, 9968–9973. https://doi.org/10.1039/D2TB01961B.