

Reshaping Protein-Based Nanoparticles: Innovative Artificial Intelligence-Driven Strategies for Structural Design and Applications

Mohammad Mahmoudi Gomari, Mehdi Alidadi, Neda Rostami, and Sidi A. Bencherif*

Nanoparticles (NPs) have become a pivotal technology in biomedical research due to their unique physicochemical properties and nanoscale size, allowing for targeted applications. Among NP materials, proteins and their derivatives stand out for their biocompatibility, engineering flexibility, and inherent biological functions, making them especially attractive for NP design. However, the structural and biochemical complexity of proteins has historically presented challenges in NP development. Recent advancements in artificial intelligence (AI) have transformed this field. Neural network models such as AlphaFold, ProteinMPNN, and RoseTTAFold, along with protein language models like evolutionary scale modeling, enable the design of protein-based NPs (PNPs) with diverse symmetries, shapes, and functionalities. These AI-driven approaches address traditional limitations, unlocking new possibilities in nanomedicine. This review explores the transformative role of AI in PNP design, emphasizing its potential to broaden applications, solve challenges, and drive innovative solutions in biotechnology and medical research.

and enhance their physicochemical properties.^[2] NPs range in size from 1 to 1000 nm, but those smaller than 100 nm are commonly utilized in medicine for research in drug delivery, imaging, and tissue engineering. NPs can be used as a vehicle to transport various types of molecules such as nucleic acid derivatives, protein derivatives, and drug molecules, through surface attachment or encapsulation. These constructs prevent degradation and rapid clearance of therapeutics, improve biodistribution and solubility, and increase compound concentrations in target tissues, resulting in enhanced effectiveness.^[3]

Materials for NP fabrication range from synthetic to natural sources, with proteins offering particular advantages. Their modular structure, biodegradability, and low toxicity make proteins ideal for NP design.

1. Introduction

The use of nanoparticles (NPs) dates back centuries, with our ancestors incorporating these particles into artworks and everyday life. However, modern nanotechnology was first introduced by Richard Feynman in 1959.^[1] Since then, the applications of NPs have expanded across various fields of science, including medicine, prompting numerous efforts to facilitate NP design

They are easier to manufacture than other colloidal carriers and can be produced through eco-friendly processes without toxic chemicals. PNPs can enhance drug delivery efficiency by controlling size, shape, and surface charge, making them a promising alternative to synthetic polymers.^[4] This type of NPs can be used for different purposes (Figure 1). Their versatility allows for the design of PNPs with specific functionalities, such as targeting cells, responding to environmental cues, and enabling controlled

M. M. Gomari
Ali-Asghar Clinical Research Development Center
Department of Pediatrics
School of Medicine
Iran University of Medical Sciences
Tehran 1449614535, Iran

M. Alidadi
Department of Anatomy
School of Medicine
Tehran University of Medical Sciences
Tehran 1461884513, Iran

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/anbr.202500017>.

© 2025 The Author(s). Advanced NanoBiomed Research published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/anbr.202500017

N. Rostami
Department of Chemical Engineering
Arak University
Arak 3848177584, Iran

S. A. Bencherif
Department of Chemical Engineering
Northeastern University
Boston, Massachusetts 02115, USA
E-mail: s.bencherif@northeastern.edu, sidi.bencherif@cncr.fr

S. A. Bencherif
Polymers, Biopolymers
Surfaces Laboratory (PBS, UMR CNRS 6270)
University of Rouen Normandy
76130 Mont-Saint-Aignan, France

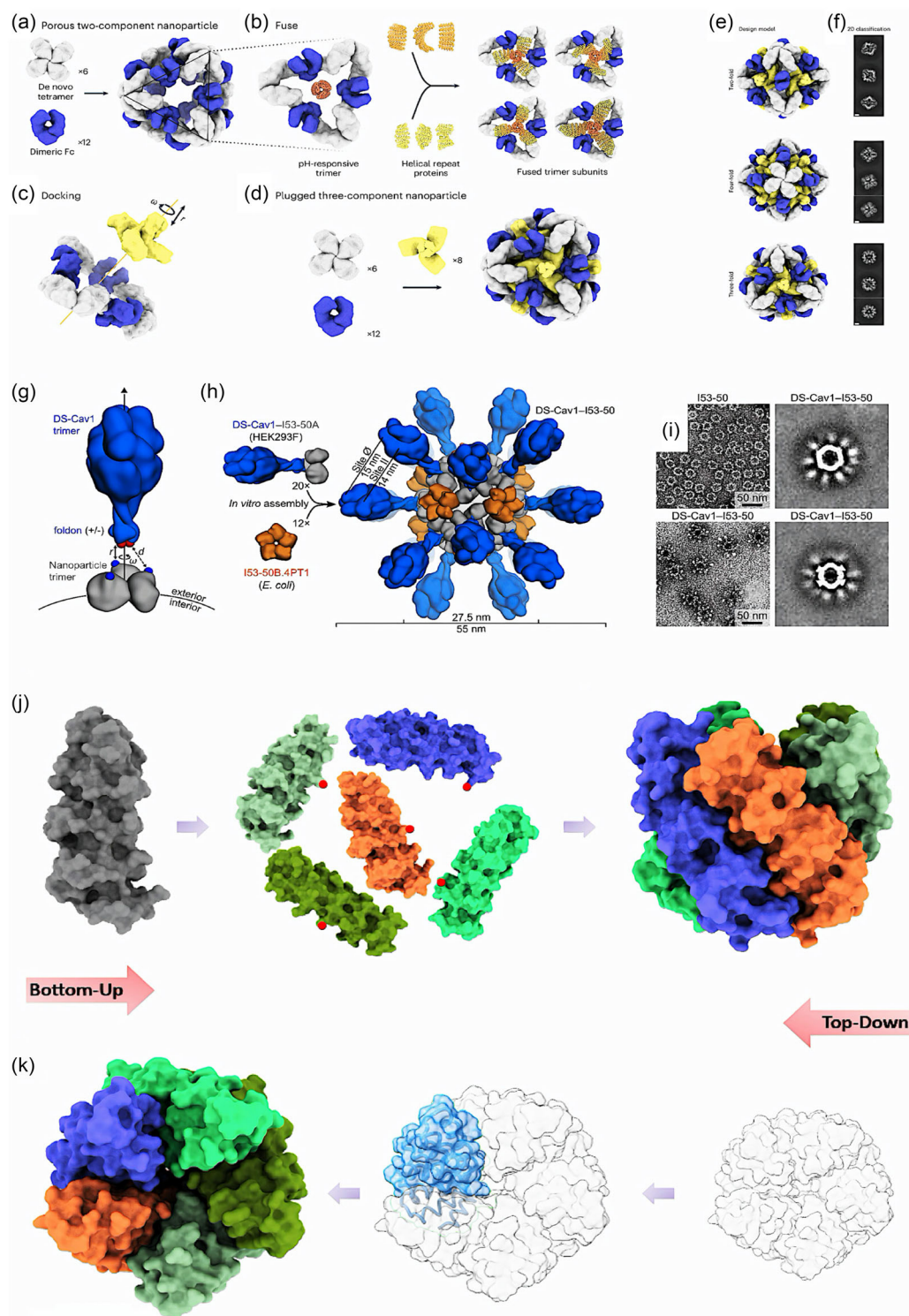


Figure 1. Design strategies and diverse applications of PNPs. One application of PNPs is the design of Ab NPs with specific shapes and conformations. a–f) The design and characterization of a pH-responsive Ab NPs. These PNPs can encapsulate proteins, nucleic acids, and drugs; target specific cells through Ab-mediated mechanisms; and undergo controlled disassembly in acidic environments (pH 5.9–6.7). Reproduced (Adapted) with permission.^[9] Copyright 2024, Springer Nature. g–i) Another application of PNPs is the development of multivalent antigens. Compared to conventional vaccines, these constructs can elicit a more robust immune response. Reproduced (Adapted) with permission.^[10] Copyright 2019, Elsevier. j,k) Bottom-up and top-down strategies. In bottom-up approaches, AI technologies are used to design monomers, predict structures, and optimize interface residues to create PNPs with specific symmetries. Conversely, top-down strategies utilize AI models to define the desired symmetry and design subunits for assembly.

therapeutic release. Additionally, proteins offer numerous binding sites and can fold into complex structures, facilitating the incorporation of various active compounds.^[5]

Despite their many advantages, PNPs face several challenges in their development and clinical translation, particularly in controlling immunogenicity, which can hinder therapeutic efficacy. Strategies like modifying surface glycosylation or using polyethylene glycol (PEG) aim to reduce immune responses, but issues such as the presence of anti-PEG antibodies (Abs) persist.^[6] Additionally, translating in vitro success to in vivo results is difficult due to biological barriers and degradation. Manufacturing PNPs with precision and consistency for clinical use also presents significant hurdles.^[7] Nevertheless, the future of PNPs in nanomedicine is promising. Advances in protein engineering, computational biology, and analytical techniques are leading to more effective PNPs. As researchers refine design and manufacturing processes, PNPs hold significant potential for revolutionizing targeted drug delivery and other therapies.^[8]

Innovative strategies, including artificial intelligence (AI) approaches, will be crucial for overcoming current obstacles and realizing the full potential of PNPs in clinical settings.^[9,10] Over the past decade, knowledge about the structure and sequence co-evolution of proteins has significantly increased. This advancement has made protein design more feasible than ever, particularly through AI strategies.^[11,12] Current AI approaches to designing self-assembling protein-based PNPs (PNPs) include top-down and bottom-up design (Figure 1j,k). AI systems are poised to have a significant impact on addressing the challenges of existing methods for self-assembling PNPs. Developed AI-based algorithms for working with biomolecules like proteins enable the design of PNPs with tetrahedral, octahedral, and icosahedral symmetry. This review examines the current advancements in AI for PNP design in nanomedicine, along with the exciting opportunities and challenges that lie ahead.

2. Language Models Offer Fast and Efficient Solutions

Given the importance of proteins in cellular function and their role as well-known therapeutic candidates for various diseases, many AI models have been developed to predict their structure and function, as well as to design new scaffolds for these biomolecules. On such type of AI model used for working with proteins is protein language models (PLMs). Many PLM-based tools have been developed, including OmegaFold,^[13] SPIRED,^[14] and evolutionary scale modeling (ESM).^[15] Among these, ESM has gained particular attention for its application to multi-subunit protein structures.^[15] ESM-2, a powerful LM trained on millions of protein sequences, significantly advances protein structure prediction. Using masked LM, ESM-2 correlates perplexity with contact map and structure prediction accuracy. Its structure prediction machine, ESM-Fold (Figure 2a), achieves speeds up to 60 times faster than conventional methods and excels in predicting multi-subunit protein structures, which are critical for PNPs. ESM-Fold enables direct, multiple sequence alignment (MSA)-free structure prediction, achieving high confidence for over 225 million structures out of 617 million analyzed metagenomic

proteins, thus offering unprecedented speed and accuracy in protein research.^[16]

Another PLM is ESM-3, a large-scale multimodal generative LM model capable of reasoning about protein behavior. Researchers have demonstrated that ESM-3 can generate functional proteins significantly different from known natural ones. The model uses discrete tokens to represent protein sequences, structures, and functions, enabling it to explore diverse protein designs. Notably, ESM-3 successfully generated a novel fluorescent protein called esmGFP, illustrating its ability to generate functional proteins in previously unexplored regions of protein space. This achievement represents a level of protein design that would otherwise require hundreds of millions of years of natural evolution.^[17] Additionally, since ESM-3 can decode multimers, it shows potential for handling PNPs. However, further studies are required to validate its applicability in this area.

One of the remarkable capabilities of LMs is their application in inverse protein folding. ESM-IF1 is one such model designed for protein inverse folding and protein design. It is a sequence-to-sequence transformer model equipped with invariant geometric input processing layers. The model was trained using both experimental and predicted structures. When evaluated on tasks such as fixed-backbone design, protein complex design, and partially masked structure prediction, ESM-IF1 demonstrated high sequence recovery, showcasing its effectiveness.^[18] Given the abundance of natural PNPs, including viruses, ferritin, and encapsulins, inverse folding presents an exciting avenue for designing novel PNPs. This method involves generating new protein sequences that can fold into scaffolds resembling these naturally occurring proteins. It holds promise for both bottom-up and top-down PNP design strategies; however, extensive evaluations are necessary to confirm its potential in these applications.

A key limitation of ESM models is their restriction to natural amino acids, which excludes proteins that interact with small molecules. To address this, Zheng et al. developed ESM all-atom (ESM-AA), a model pretrained on multiscale code-switch sequences. Using multiscale position encoding, ESM-AA maps residue-atom relationships by unzipping protein residues into their constituent atoms and assigning coordinates, thereby enabling the analysis of functionalized proteins. This advancement is particularly significant for the design and modification of PNPs, as it enhances their stability and functionality.^[20] ESM-AA opens new possibilities for creating advanced PNP carriers and tailoring their surfaces to improve performance in biomedical and nanotechnology applications. Although ESMs^[21] and other LM-based algorithms, such as PLMGraph-Inter,^[19] have been implemented for various protein types, including multimers, they have not yet been applied to PNPs. Exploring the potential of these models for PNP design in the future could be both promising and essential.

3. NN Models Present a Precise Solution

The main advantage of PLMs, such as the ESM models, is their high computational speed as they typically bypass the need for MSAs in structure prediction. In contrast, when working on protein structures using neural network (NN) models like RF2,^[22] RF All-Atom,^[23] and AF3,^[24] along with their derivatives

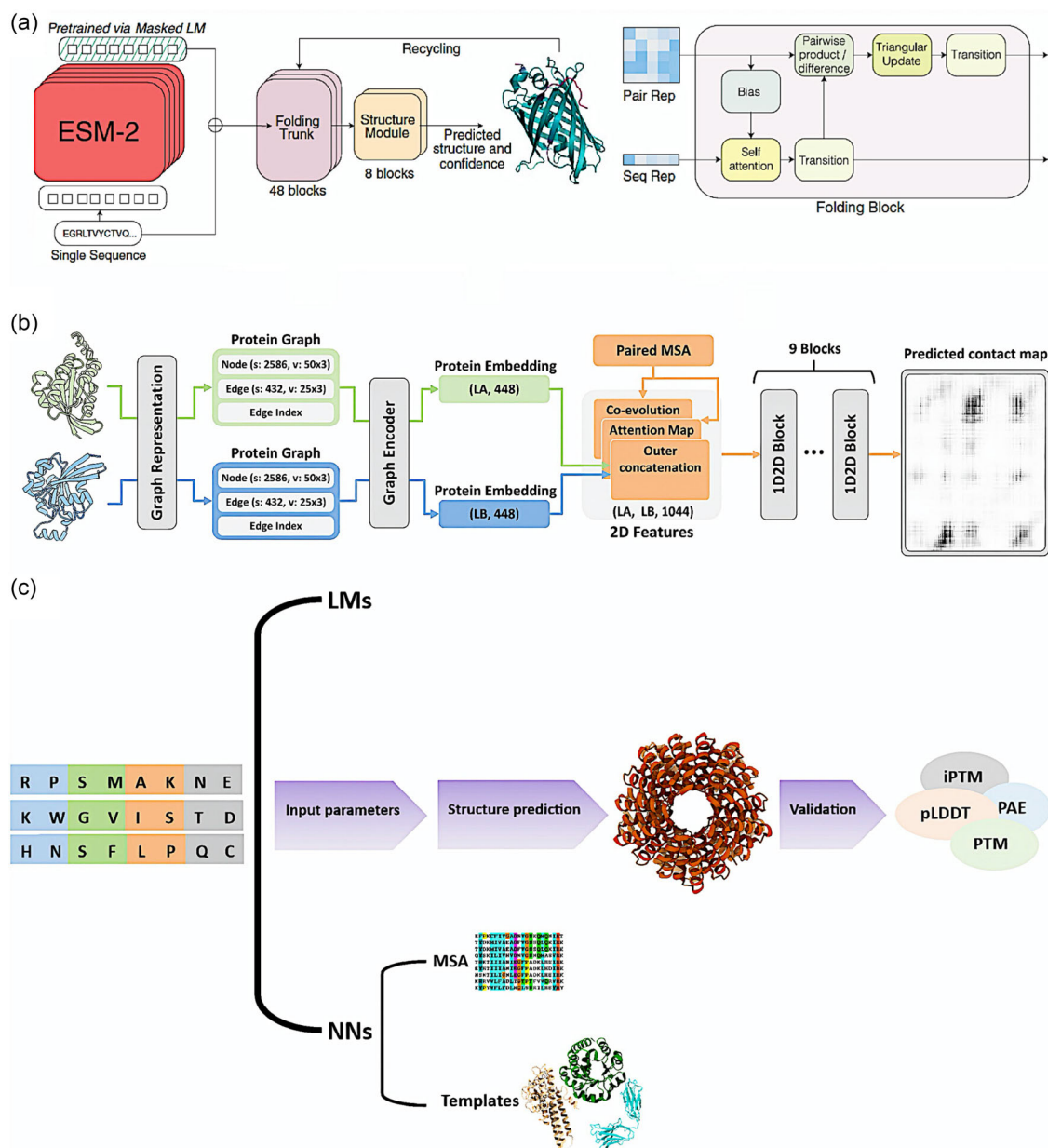


Figure 2. Architecture of selected LMs and their comparison with NN models. a) The architecture of ESM-Fold, an end-to-end protein structure predictor built on ESM-2, featuring folding blocks and recycling steps to enhance prediction accuracy. Reproduced (Adapted) with permission.^[16] Copyright 2023, AAAS. b) The architecture of PLMGraph-Inter, a LM designed for analyzing protein-protein interactions. Reproduced (Adapted) with permission.^[19] Copyright 2023, eLife. Finally, c) compares NNs and LMs. NNs rely on an input MSA file and optional templates to predict the structure of multichain proteins. In contrast, LMs predict structures directly from input sequences, largely eliminating the reliance on MSAs and templates. While LMs offer faster predictions, NNs generally produce structures of higher quality.

such as ColabFold,^[25] an MSA file is necessary. A structural template can optionally be used to further improve the performance of these models. Although NN models are slower compared to LMs, they excel in terms of quality.^[26,27] RoseTTAFold (RF) and AlphaFold (AF) models consistently achieve high scores in the critical assessment of protein structure prediction (CASP) competition and demonstrate significant efficiency in research applications. As a result, many advanced tools for protein structure

prediction and design are built on these architectures. Examples include AfDesign,^[28] ProteinMPNN, and RFdiffusion, which exemplify this trend (Figure 3c).^[29–31]

Traditional score function-based methods have been widely used to predict protein-protein interactions in self-assembling NPs, significantly improving design speed and efficiency.^[32–35] However, these methods face limitations, including high computational demands, a restricted design space, and challenges

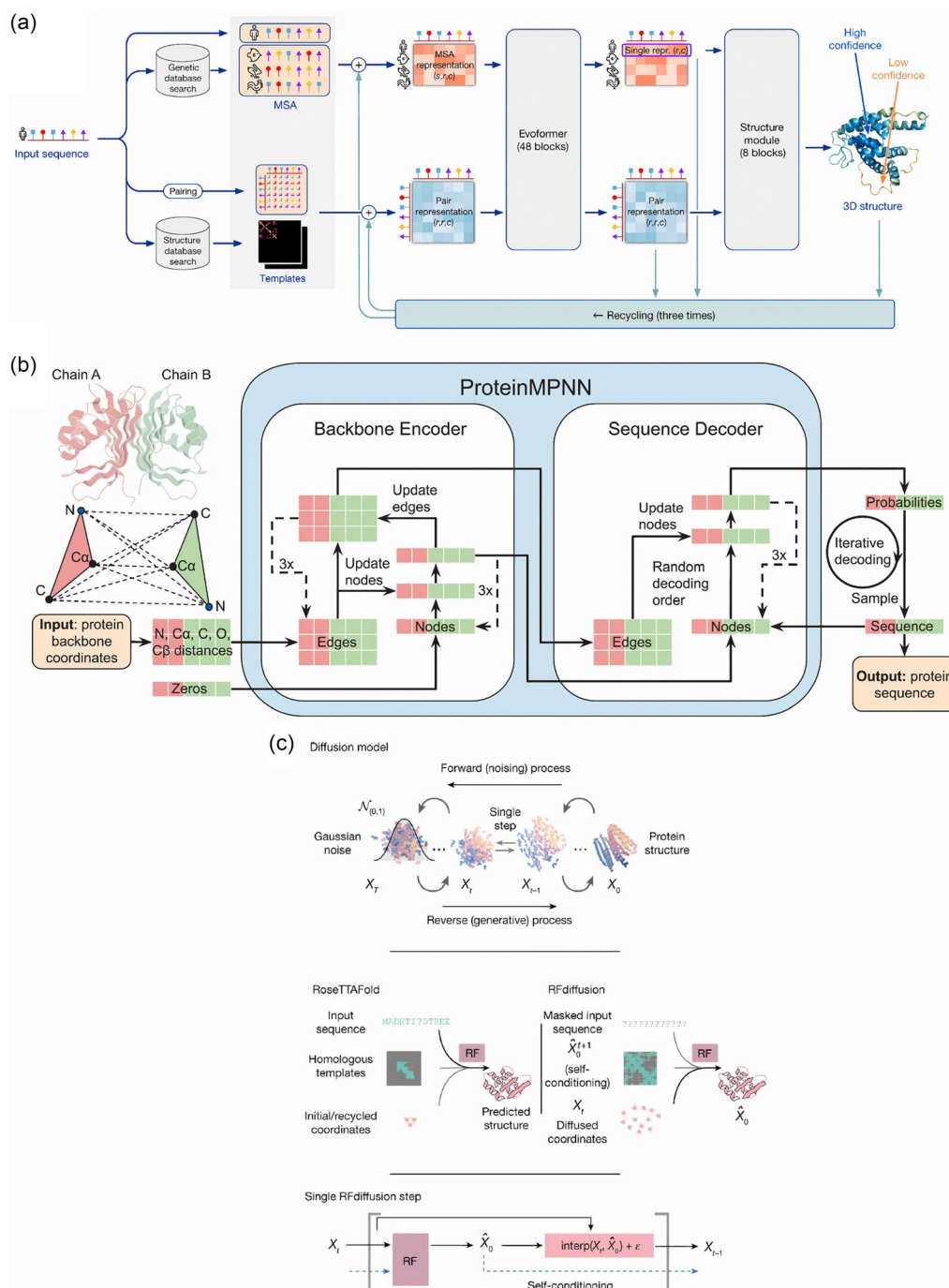


Figure 3. NN architectures for protein structure prediction and design. a) The architecture of AF2, one of the most widely used NN models for protein structure prediction. Reproduced (Adapted) with permission.^[30] Copyright 2021, Springer Nature. b) The protein multimeric protein NN (ProteinMPNN) architecture, an NN-based model designed to optimize protein sequences. Reproduced (Adapted) with permission.^[31] Copyright 2022, AAAS. Finally, c) highlights the RFdiffusion architecture, an algorithm specifically developed for designing and generating novel proteins with desired rearrangements. Reproduced (Adapted) with permission.^[29] Copyright 2023, Springer Nature.

associated with certain symmetries. NNs, such as ProteinMPNN, address these issues by enabling the design of protein sequences with desired folding properties. ProteinMPNN employs a message-passing NN architecture that leverages features like interatomic distances and random decoding orders, achieving a

sequence recovery rate of 52.4%, which outperforms traditional methods like Rosetta (32.9%). ProteinMPNN excels in designing sequences for a variety of protein structures, including monomers, oligomers, and protein–protein interfaces, with recovery rates of 52% for monomers, 55% for homomers, and 51% for

heteromers. In nanomaterials, ProteinMPNN has successfully designed tetrahedral NP backbones and protein assemblies critical for applications such as structure-based vaccine design. These designs demonstrated high success rates in forming stable assemblies, outperforming manual and traditional computational approaches. The integration of ProteinMPNN and AF2 further enhanced the design process. While AF2 is known for its accurate structure predictions based on sequences with co-evolutionary information, it struggles with single sequences. However, sequences generated by ProteinMPNN were more confidently predicted by AF2 to adopt the target structures than native sequences. This synergy enabled iterative validation and refinement, ensuring stable and properly folded proteins. Together, ProteinMPNN and AF2 constitute a powerful toolkit for designing PNPs with broad applications in biotechnology and medicine.^[31]

In a study investigating the role of structural flexibility in the assembly of computationally designed PNPs, researchers found that while most designs assembled as intended, some deviated from the target models. Detailed characterization of these cases, using AF2 and Rosetta software, revealed that flexibility within the protein building blocks contributed to the formation of alternative architectures. These findings suggest that controlled flexibility, rather than rigid symmetry, may be essential for designing adaptable protein assemblies. This approach opens new possibilities for creating custom nanomaterials, such as those used for cargo encapsulation, by exploring quasisymmetric architectures that were previously inaccessible through conventional methods. This integration of AI-based techniques further enhances these capabilities.^[36]

A recent study highlighted significant advancements in protein design, particularly in creating multimeric protein cages for applications in nanotechnology, biotechnology, and medicine (**Figure 4**). A major challenge in this field is designing stable subunit interfaces, which was addressed by integrating AF2 and ProteinMPNN. AF2 was employed to predict accurate structural models of individual subunits, which were then docked to identify potential interfaces. ProteinMPNN was used to explore sequence space and design sequences that reliably folded into the desired structures, enabling the creation of novel proteins with improved structural and functional properties. The method achieved a native sequence recovery rate of at least 45%, and 16% of the designs successfully formed complete assemblies, as verified by transmission electron microscopy. The study highlighted the superiority of AI-driven approaches over traditional methods, emphasizing the importance of backbone flexibility during the design process. This flexibility was especially critical for facilitating interactions involving non-helical motifs, thereby broadening the range of achievable designs.^[37]

In a study conducted by Bethel et al. nanofibers were created from extendable proteins (helical repeat proteins). The researchers presented a systematic approach to designing modular repeat protein oligomers with coincident cyclic (C2 to C8) and superhelical symmetry axes. The process involves generating repeat protein monomers using Rosetta's fragment assembly, computationally assembling them into cyclic structures, and assigning sequences. Initially, designs were generated using Rosetta's FastDesign; however, better results were achieved with ProteinMPNN. Out of 67 ProteinMPNN-designed multiplexes expressed in *E. coli*, 11 were

successfully validated experimentally. High-resolution structures of five designs (C2 to C6 symmetry) were determined by X-ray crystallography and cryoEM, closely matching their design models. AF2 and AF Multimer were used to verify the designs prior to experimental testing. These building blocks can form unbounded fibers with tunable geometry by introducing hydrophilic surface patches. This approach opens new possibilities in protein nanomaterial design, with applications in biomineralization, ligand binding, and hydrogels. The mechanical properties can be modulated by adjusting the pore size. Furthermore, the designs' robust thermostability and high soluble yields enable large-scale manufacturing using standard procedures.^[38]

A separate study compared ProteinMPNN's performance with that of Rosetta in creating two-component tetrahedral protein nanomaterials. ProteinMPNN matched Rosetta's success rate by producing 13 new experimentally validated assemblies but achieved this with significantly less computational effort and without the need for manual refinement. High-resolution crystal structures confirmed the accuracy of the ProteinMPNN designs. ProteinMPNN efficiently redesigned 27 tetrahedral protein assemblies initially created by Rosetta, generating each sequence in about 1 s compared to several minutes required by Rosetta. Despite comparable performance metrics, ProteinMPNN's designs featured a higher fraction of polar residues, enhancing biophysical properties and assembly success. Six out of six ProteinMPNN-designed nanomaterials assembled effectively in vitro, whereas only one Rosetta-designed assembly did. These results highlight ProteinMPNN's advantages in speed, simplicity, and design quality, making it a valuable tool for developing protein-based technologies and vaccines.^[39]

In a related study addressing the challenges of constructing large protein assemblies, researchers developed a modular system of twistless helix repeat (THR) protein blocks. These blocks, categorized into linear, curved, and turn types, offered versatile geometric configurations. ProteinMPNN played a key role in designing sequences for these blocks, ensuring uniformity and structural integrity in symmetrical structures. Its computational efficiency and high success rate made it ideal for designing repeat proteins and homomeric oligomers. Additionally, AF2 was used for structural validation, predicting protein structures with high accuracy. The designed THR blocks were validated through metrics such as predicted local distance difference test (pLDDT) and predicted template modelling (pTM) scores, demonstrating their structural accuracy. The results highlighted the potential of THR blocks to create complex structures, including concentric rings and polyhedral nanocages. This modular approach advances the programmability of protein nanomaterials, paving the way for customizable applications in PNP design.^[40]

AI methodologies have been effectively applied to top-down protein design, focusing on desired properties such as symmetry and porosity rather than individual monomer stability. This approach has enabled the systematic construction of protein subunits tailored to meet these specifications. In one study, ProteinMPNN was used to design RC_L1-H9 and RC_L1-H11, which successfully formed the intended symmetric capsids. Validation was performed using AF2 predictions and Rosetta metrics to ensure compatibility between the designed sequences and target structures. AF2 played a critical role in refining designs by evaluating various factors, including binding energy, interface quality, and structural stability.

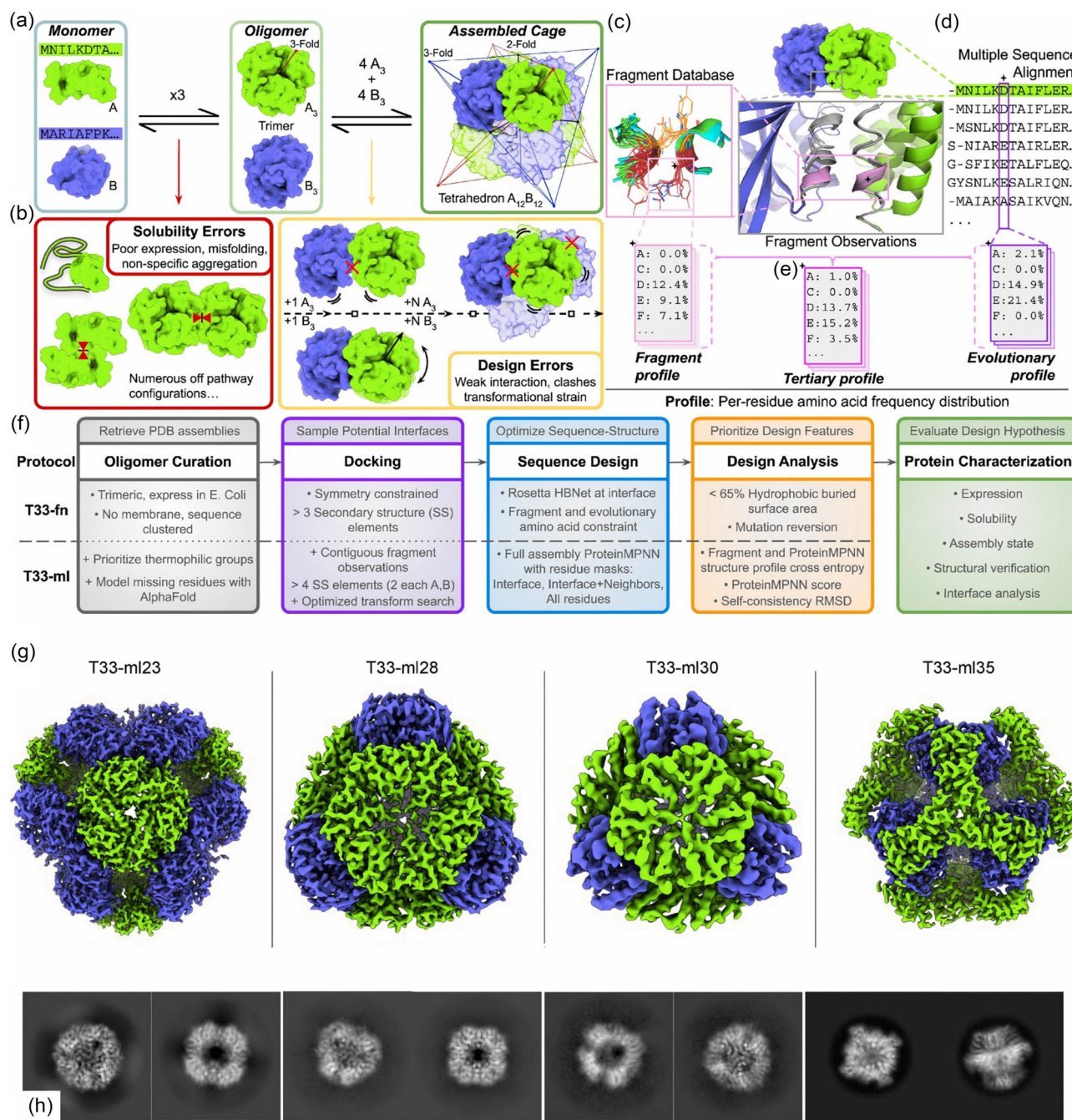


Figure 4. Designing protein cages for biotechnology and medical applications: a–f) illustrate the frameworks used to design protein cages, while g,h) depict the structural shapes of several tetrahedral cages created in the study. Reproduced (Adapted) with permission.^[37] Copyright 2024, Elsevier.

The iterative optimization process, which combined ProteinMPNN and AF2, improved both the sequences and the backbones, achieving high structural fidelity. Experimental validation, including cryo-EM analysis, confirmed the accuracy of the designs, showing close alignment between the predicted models and the observed structures.^[41]

In a recent study, researchers successfully created symmetry-broken protein cages using steric occlusion, introducing a novel

approach in the field. The main challenge was designing cages that exceeded the size limitations of purely icosahedral symmetry, which is typically achieved through complex protein–protein interfaces. Using AF2 for structural predictions, the team engineered a modified subunit with an inward-facing domain, ensuring that only one such subunit could fit within the cage. This design was further refined by adding an outward-facing DARPin domain, creating unique external attachment points.

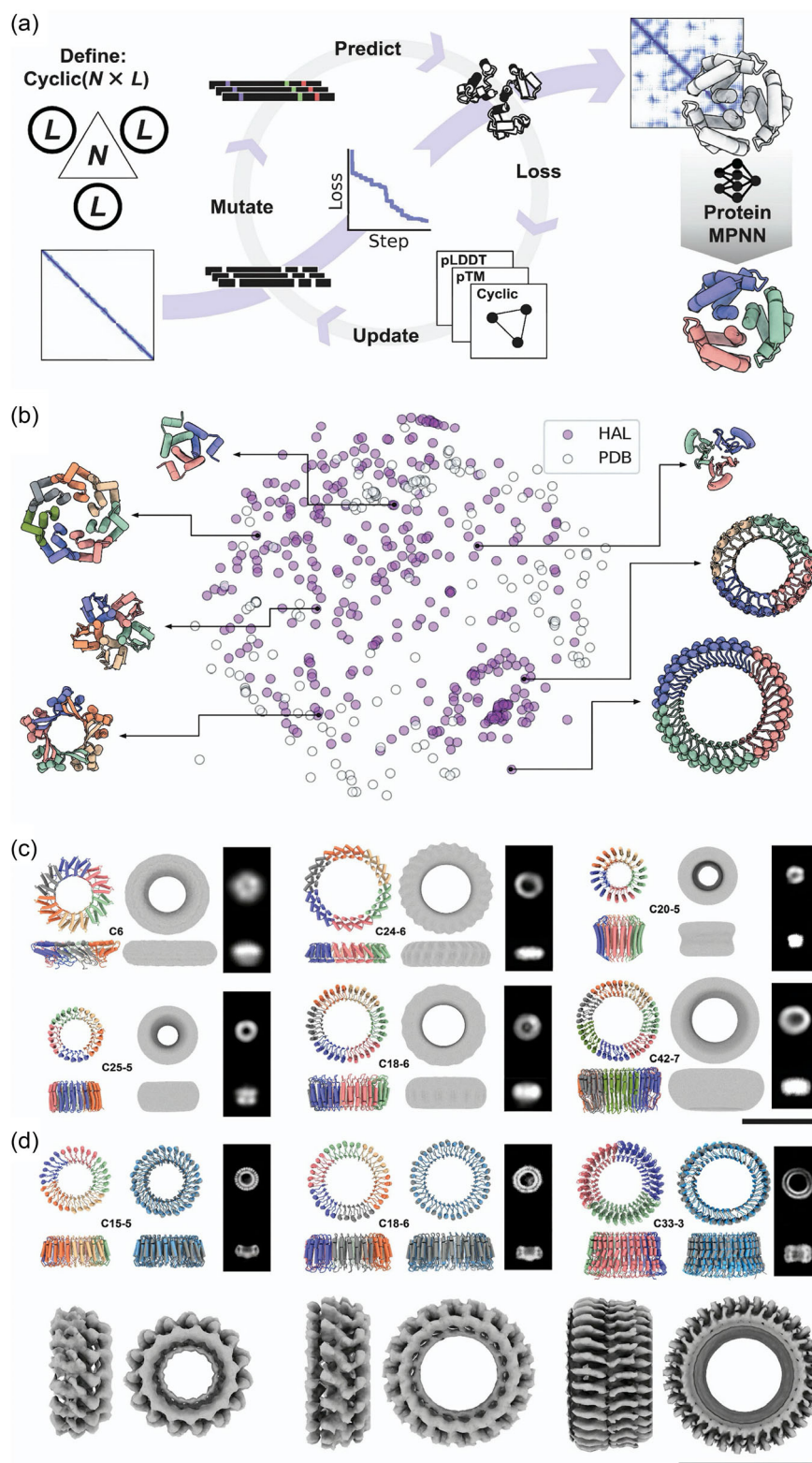


Figure 5. Designing symmetrical PNPs for advanced applications: a) outlines the process, where cyclic symmetry and protein length are selected, followed by MCMC optimization using AF2 and sequence redesign via ProteinMPNN; b) shows multidimensional scaling of similarities between experimentally validated protomers and all de novo cyclic oligomers in the PDB; c,d) present validation of large assemblies using Cryo-EM and negative stain electron microscopy techniques. Reproduced (Adapted) with permission.^[43] Copyright 2022, AAAS.

The researchers selected the SUMO protein and a flexible GGSSG linker to correctly position the internal fusion. This method enabled the precise presentation of a single outward-facing binding domain, simplifying the creation of symmetry-broken features.^[42]

In a study utilizing the hallucination method for PNP design (Figure 5), researchers explored sequence space to create proteins with desired symmetries and properties without relying on predefined structures. This approach combined AF2 and ProteinMPNN for the design and optimization of cyclic protein oligomers. AF2 predicted oligomeric structures by jointly optimizing protomer and oligomer configurations through a Markov Chain Monte Carlo (MCMC) search in sequence space. The search was guided by a composite loss function, incorporating AF2's confidence metrics (pLDDT and pTM) and a cyclic symmetry term. After generating initial structures, ProteinMPNN was used to redesign the backbone sequences, producing multiple variants for each AF2-generated structure. These sequences were subsequently evaluated for proper folding and structural stability. The study demonstrated that AF2 hallucination could generate diverse cyclic protein oligomers with varied topologies and structures. While initial designs often faced challenges such as overfitting and unrealistic properties, ProteinMPNN addressed these issues, generating viable sequences with higher experimental success rates. Among the 96 tested oligomers, many exhibited high expression levels, correct oligomerization, and stable structures, highlighting the effectiveness of this computational approach in designing functional PNPs.^[43]

In an investigation, Rankovic et al. developed a computational framework combining AF2, RFdiffusion, and ProteinMPNN to design bifaceted PNPs with pseudo-D5 symmetry for biomedical applications, such as colocalizing biological entities. RFdiffusion was used to generate precise protein backbones, ProteinMPNN optimized the amino acid sequences, and AF2 validated the models for stability and accurate assembly. Experimentally, these NPs efficiently colocalized two distinct populations of polystyrene beads, each coated with different proteins, by functionalizing the NP surfaces with de novo protein minibinders. Additionally, RFdiffusion reoriented the subunit termini for genetic fusion while maintaining structural integrity. The modularity of the framework was demonstrated by successfully integrating multiple minibinders without compromising NP assembly quality, highlighting its versatility for biomedical applications.^[44]

4. Conclusion and Future Perspectives

The design of protein oligomers and PNPs via self-assembly represents a significant advancement in biomolecular science, offering immense potential for biotechnology, materials science, and medicine. Recent computational breakthroughs have greatly improved our ability to create symmetric protein assemblies. Tools such as AF2^[45] and RF^[46] enable highly accurate protein structure predictions. Additionally, the advent of AI models like ProteinMPNN and RFdiffusion has facilitated sequence design. ProteinMPNN, leveraging NN architectures, excels in predicting protein sequences that enhance the expression and stability of recombinant proteins.^[47] RFdiffusion has also shown

promise in designing 3D protein crystals with precise lattice structures, expanding the potential for complex biomaterial development.^[48] Given the successful applications of PNPs in drug delivery, multivalent antigens and Abs, and the design of environmentally responsive carriers,^[10,49–52] these AI methodologies have paved the way for the development of advanced smart therapeutics targeting complex diseases such as cancer and infections.

AI applications for PNPs extend beyond design and structure prediction to the evaluation of biological behavior. Researchers have used AF2 to explore cotranslational protein assembly, revealing that protein structure plays a crucial role in translation and cellular function. By combining AF2 predictions with experimental data, they demonstrated that cotranslational assembly is structurally driven and likely conserved across species, providing deeper insights into protein interactions and evolution.^[53] In the context of intravenous therapeutics, AI has shifted the focus from the enhanced permeability and retention effect to active *trans*-endothelial transport. AI-driven analysis of vascular permeability in tumors using genetically modified ferritin nanocages revealed variability in permeability, guiding the design of more effective PNPs for improved therapeutic outcomes.^[54] Additionally, AI has been employed to predict protein corona composition, addressing concerns about the biological effects of nanomaterials and enabling the development of safer and more effective nanocarriers.^[55]

Despite significant advancements, challenges persist in designing and validating symmetric protein oligomers using AI. Achieving structural stability in large assemblies remains a formidable task, and the high computational demands restrict design exploration and scalability. Currently, self-assembling designs have a success rate of only about 10%. To overcome these limitations, future efforts must integrate coarse-grained modeling with atomistic refinement, develop more efficient sampling algorithms, and establish community-wide datasets for rigorous validation of AI-designed PNPs.^[29,56,57] Looking into the future, the full potential of symmetric protein oligomers will be unlocked by blending AI-driven design with high-throughput experimental platforms—such as microfluidic droplet screening and *in vivo* selection assays—to rapidly iterate between *in silico* predictions and empirical performance. Emphasizing modularity, researchers should explore hybrid architectures that combine protein scaffolds with synthetic polymers or small-molecule crosslinkers to fine-tune mechanical properties and trigger responsiveness. Finally, as PNPs move closer to clinical and industrial deployment, it will be essential to address manufacturing scalability, regulatory pathways, and long-term biosafety, ensuring that these smart biomaterials can be translated into real-world solutions for complex diseases, sustainable materials, and beyond.

Acknowledgements

S.A.B. gratefully acknowledges financial support from the National Institutes of Health (NIH, 1R01EB027705) and the *Chaire d'Excellence de Normandie*. M.M.G. acknowledges financial support from the Clinical Research Development Center of Ali-Asghar Children's Hospital for this work (grant no. 34133). [Correction added on 10 July 2025, after first online publication: The Acknowledgment section has been updated in this version.]

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Mohammad Mahmoudi Gomari: conceptualization and writing—original draft. **Mehdi Alidadi:** writing—review and editing. **Neda Rostami:** writing—review and editing. **Sidi A. Bencherif:** supervision, validation, and writing—review and editing.

Keywords

artificial intelligence, biomedical applications, prediction models, protein design, protein-based nanoparticles

Received: January 28, 2025

Revised: April 27, 2025

Published online: June 6, 2025

- [1] R. P. Feynman, *Eng. Sci.* **1960**, 23, 22.
- [2] S. Bayda, M. Adeel, T. Tuccinardi, M. Cordani, F. Rizzolio, *Molecules* **2019**, 25, 112.
- [3] N. Rostami, M. M. Gomari, E. Choupani, S. Abkhiz, M. Fadaie, S. S. Eslami, Z. Mahmoudi, Y. Zhang, M. Puri, F. N. Monfared, E. Demireva, V. N. Uversky, B. R. Smith, S. A. Bencherif, *Small Sci.* **2024**, 4, 2400192.
- [4] S. Hong, D. W. Choi, H. N. Kim, C. G. Park, W. Lee, H. H. Park, *Pharmaceutics* **2020**, 12, 720.
- [5] A. Olshefsky, C. Richardson, S. H. Pun, N. P. King, *Bioconjugate Chem.* **2022**, 33, 2018.
- [6] J. Fu, E. Wu, G. Li, B. Wang, C. Zhan, *Nano Today* **2024**, 55, 102163.
- [7] N. Habibi, A. Mauser, Y. Ko, J. Lahann, *Adv. Sci.* **2022**, 9, 2104012.
- [8] D. Diaz, A. Care, A. Sunna, *Genes* **2018**, 9, 370.
- [9] E. C. Yang, R. Divine, M. C. Miranda, A. J. Borst, W. Sheffler, J. Z. Zhang, J. Decarreau, A. Saragovi, M. Abedi, N. Goldbach, M. Ahlrichs, C. Dobbins, A. Hand, S. Cheng, M. Lamb, P. M. Levine, S. Chan, R. Skotheim, J. Fallas, G. Ueda, J. Lubner, M. Somiya, A. Khmelinskaia, N. P. King, D. Baker, *Nat. Struct. Mol. Biol.* **2024**, 31, 1404.
- [10] J. Marcandalli, B. Fiala, S. Ols, M. Perotti, W. de van der Schueren, J. Snijder, E. Hodge, M. Benhaim, R. Ravichandran, L. Carter, W. Sheffler, L. Brunner, M. Lawrenz, P. Dubois, A. Lanzavecchia, F. Sallusto, K. K. Lee, D. Veisler, C. E. Correnti, L. J. Stewart, D. Baker, K. Loré, L. Perez, N. P. King, *Cell* **2019**, 176, 1420.
- [11] A. Ljubetič, F. Lapenta, H. Gradišar, I. Drobna, J. Aupič, Ž. Strmšek, D. Lainšček, I. Hafner-Bratkovič, A. Majerle, N. Krivec, M. Benčina, T. Pisanski, T. Č. Veličković, A. Round, J. M. Carazo, R. Melero, R. Jerala, *Nat. Biotechnol.* **2017**, 35, 1094.
- [12] M. K. Yazdi, P. Zarrintaj, M. R. Saeb, M. Mozafari, A. B. Bencherif, *Artificial Intelligence in Biomaterials Design & Development*. Woodhead Publishing, **2025**.
- [13] R. Wu, F. Ding, R. Wang, R. Shen, X. Zhang, S. Luo, C. Su, Z. Wu, Q. Xie, B. Berger, J. Ma, J. Peng, *bioRxiv* **2022**, 500999.
- [14] Y. Chen, Y. Xu, D. Liu, Y. Xing, H. Gong, *Nat. Commun.* **2024**, 15, 7400.
- [15] F. Wu, L. Wu, D. Radev, J. Xu, S. Z. Li, *Commun. Biol.* **2023**, 6, 876.
- [16] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, *Science* **2023**, 379, 1123.
- [17] T. Hayes, R. Rao, H. Akin, N. J. Sofroniew, D. Oktay, Z. Lin, R. Verkuil, V. Q. Tran, J. Deaton, M. Wiggert, R. Badkundri, I. Shafkat, J. Gong, A. Derry, R. S. Molina, N. Thomas, Y. A. Khan, C. Mishra, C. Kim, L. J. Bartie, M. Nemeth, P. D. Hsu, T. Sercu, S. Candido, A. Rives, *bioRxiv*, **2025**, 387, 850.
- [18] C. Hsu, R. Verkuil, J. Liu, Z. Lin, B. Hie, T. Sercu, A. Lerer, A. Rives, *bioRxiv* **2022**, 487779.
- [19] Y. Si, C. Yan, *Elife* **2024**, 12, RP92184.
- [20] K. Zheng, S. Long, T. Lu, J. Yang, X. Dai, M. Zhang, Z. Nie, W.-Y. Ma, H. Zhou, *bioRxiv*, **2024**, 583284.
- [21] O. Avraham, T. Tsaban, Z. Ben-Aharon, L. Tsaban, O. Schueler-Furman, *BMC Bioinf.* **2023**, 24, 433.
- [22] M. Baek, I. Anishchenko, I. R. Humphreys, Q. Cong, D. Baker, F. DiMaio, *bioRxiv* **2023**, 542179.
- [23] R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, A. Kang, E. Brackenbrough, A. K. Bera, M. Baek, F. DiMaio, D. Baker, *Science* **2024**, 384, eadl2528.
- [24] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, et al., *Nature* **2024**, 630, 493.
- [25] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *Nat. Methods* **2022**, 19, 679.
- [26] M. Manfredi, C. Savojardo, G. Iadukhin, D. Salomoni, A. Costantini, P. L. Martelli, R. Casadio, *J. Mol. Biol.* **2024**, 436, 168593.
- [27] P. T. Nguyen, B. J. Harris, D. L. Mateos, A. H. González, A. M. Murray, V. Yarov-Yarovoy, *Channels* **2024**, 18, 2325032.
- [28] T. Kosugi, M. Ohue, *Biomedicines* **2022**, 10, 1626.
- [29] J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, *Nature* **2023**, 620, 1089.
- [30] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, et al., *Nature* **2021**, 596, 583.
- [31] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, *Science* **2022**, 378, 49.
- [32] W. Sheffler, E. C. Yang, Q. Dowling, Y. Hsia, C. N. Fries, J. Stanislaw, M. D. Langowski, M. Brandys, Z. Li, R. Skotheim, A. J. Borst, A. Khmelinskaia, N. P. King, D. Baker, *PLOS Comput. Biol.* **2023**, 19, e1010680.
- [33] H. Shen, E. M. Lynch, S. Akkineni, J. L. Watson, J. Decarreau, N. P. Bethel, I. Benna, W. Sheffler, D. Farrell, F. DiMaio, E. Derivery, J. J. De Yoreo, J. Kollman, D. Baker, *Nat. Nanotechnol.* **2024**, 19, 1016.
- [34] G. Ueda, A. Antanasijevic, J. A. Fallas, W. Sheffler, J. Copps, D. Ellis, G. B. Hutchinson, A. Moyer, A. Yasmeen, Y. Tsybovsky, Y.-J. Park, M. J. Bick, B. Sankaran, R. A. Gillespie, P. J. M. Brouwer, P. H. Zwart, D. Veisler, M. Kanekiyo, B. S. Graham, R. W. Sanders,

- J. P. Moore, P. J. Klasse, A. B. Ward, N. P. King, D. Baker, *eLife* **2020**, 9, e57659.
- [35] R. Mout, R. C. Bretherton, J. Decarreau, S. Lee, N. Gregorio, N. I. Edman, M. Ahlrichs, Y. Hsia, D. D. Sahtoe, G. Ueda, A. Sharma, R. Schulman, C. A. DeForest, D. Baker, *Proc. Natl. Acad. Sci.* **2024**, 121, e2309457121.
- [36] A. Khmelinskaia, N. P. Bethel, F. Fatehi, B. B. Mallik, A. Antanasijevic, A. J. Borst, S.-H. Lai, H. Y. Chim, J. Y. J. Wang, M. C. Miranda, A. M. Watkins, C. Ogohara, S. Caldwell, M. Wu, A. J. R. Heck, D. Veessler, A. B. Ward, D. Baker, R. Twarock, N. P. King, *Nat. Struct. Mol. Biol.* **2025**, 188, 764.
- [37] K. Meador, R. Castells-Graells, R. Aguirre, M. R. Sawaya, M. A. Arbing, T. Sherman, C. Senarathne, T. O. Yeates, *Structure* **2024**, 32, 751.
- [38] N. P. Bethel, A. J. Borst, F. Parmeggiani, M. J. Bick, T. J. Brunette, H. Nguyen, A. Kang, A. K. Bera, L. Carter, M. C. Miranda, R. D. Kibler, M. Lamb, X. Li, B. Sankaran, D. Baker, *Nat. Chem.* **2023**, 15, 1664.
- [39] R. J. de Haas, N. Brunette, A. Goodson, J. Dauparas, S. Y. Yi, E. C. Yang, Q. Dowling, H. Nguyen, A. Kang, A. K. Bera, B. Sankaran, R. de Vries, D. Baker, N. P. King, *Proc. Natl. Acad. Sci.* **2024**, 121, e2314646121.
- [40] T. F. Huddy, Y. Hsia, R. D. Kibler, J. Xu, N. Bethel, D. Nagarajan, R. Redler, P. J. Y. Leung, C. Weidle, A. Courbet, E. C. Yang, A. K. Bera, N. Coudray, S. J. Calise, F. A. Davila-Hernandez, H. L. Han, K. D. Carr, Z. Li, R. McHugh, G. Reggiano, A. Kang, B. Sankaran, M. S. Dickinson, B. Coventry, T. J. Brunette, Y. Liu, J. Dauparas, A. J. Borst, D. Ekiert, J. M. Kollman, et al., *Nature* **2024**, 627, 898.
- [41] I. D. Lutz, S. Wang, C. Norn, A. Courbet, A. J. Borst, Y. T. Zhao, A. Dosey, L. Cao, J. Xu, E. M. Leaf, C. Treichel, P. Litvicov, Z. Li, A. D. Goodson, P. Rivera-Sánchez, A.-M. Bratovianu, M. Baek, N. P. King, H. Ruohola-Baker, D. Baker, *Science* **2023**, 380, 266.
- [42] N. Gladkov, E. A. Scott, K. Meador, E. J. Lee, A. D. Laganowsky, T. O. Yeates, R. Castells-Graells, *Protein Sci.* **2024**, 33, e4973.
- [43] B. I. M. Wicky, L. F. Milles, A. Courbet, R. J. Ragotte, J. Dauparas, E. Kinfu, S. Tipps, R. D. Kibler, M. Baek, F. DiMaio, X. Li, L. Carter, A. Kang, H. Nguyen, A. K. Bera, D. Baker, *Science* **2022**, 378, 56.
- [44] S. Rankovic, K. D. Carr, J. Decarreau, R. Skotheim, R. D. Kibler, S. Ols, S. Lee, J. Chun, M. Tooley, J. Dauparas, H. E. Eisenach, M. Glögl, C. Weidle, A. J. Borst, D. Baker, N. P. King, *bioRxiv*, **2024**, 619149.
- [45] L. M. F. Bertoline, A. N. Lima, J. E. Krieger, S. K. Teixeira, *Front. Bioinform.* **2023**, 3, 1120370.
- [46] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhllheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, et al., *Science* **2021**, 373, 871.
- [47] M. Dolorfino, R. Samanta, A. Vorobieva, *bioRxiv* **2024**.
- [48] X. Qiu, H. Li, G. Ver Steeg, A. Godzik, *Biomolecules* **2024**, 14, 339.
- [49] P. J. M. Brouwer, A. Antanasijevic, Z. Berndsen, A. Yasmeeen, B. Fiala, T. P. L. Bijl, I. Bontjer, J. B. Bale, W. Sheffler, J. D. Allen, A. Schorcht, J. A. Burger, M. Camacho, D. Ellis, C. A. Cottrell, A.-J. Behrens, M. Catalano, I. del Moral-Sánchez, T. J. Ketts, C. LaBranche, M. J. van Gils, K. Sliepen, L. J. Stewart, M. Crispin, D. C. Montefiori, D. Baker, J. P. Moore, P. J. Klasse, A. B. Ward, N. P. King, et al., *Nat. Commun.* **2019**, 10, 4272.
- [50] Y.-N. Zhang, J. Paynter, A. Antanasijevic, J. D. Allen, M. Eldad, Y.-Z. Lee, J. Copps, M. L. Newby, L. He, D. Chavez, P. Frost, A. Goodroe, J. Dutton, R. Lanford, C. Chen, I. A. Wilson, M. Crispin, A. B. Ward, J. Zhu, *Nat. Commun.* **2023**, 14, 1985.
- [51] T. Tokatlian, B. J. Read, C. A. Jones, D. W. Kulp, S. Menis, J. Y. H. Chang, J. M. Steichen, S. Kumari, J. D. Allen, E. L. Dane, A. Liguori, M. Sangesland, D. Lingwood, M. Crispin, W. R. Schief, D. J. Irvine, *Science* **2019**, 363, 649.
- [52] J. T. Martin, C. A. Cottrell, A. Antanasijevic, D. G. Carnathan, B. J. Cossette, C. A. Enemu, E. H. Gebru, Y. Choe, F. Viviano, S. Fischinger, T. Tokatlian, K. M. Cirelli, G. Ueda, J. Copps, T. Schiffner, S. Menis, G. Alter, W. R. Schief, S. Crotty, N. P. King, D. Baker, G. Silvestri, A. B. Ward, D. J. Irvine, *npj Vaccines* **2020**, 5, 72.
- [53] S. Mallik, J. Venezian, A. Lobov, M. Heidenreich, H. Garcia-Seisdedos, T. O. Yeates, A. Shiber, E. D. Levy, *Cell* **2025**, 188, 764.
- [54] M. Zhu, J. Zhuang, Z. Li, Q. Liu, R. Zhao, Z. Gao, A. C. Midgley, T. Qi, J. Tian, Z. Zhang, D. Kong, J. Tian, X. Yan, X. Huang, *Nat. Nanotechnol.* **2023**, 18, 657.
- [55] Z. Ban, P. Yuan, F. Yu, T. Peng, Q. Zhou, X. Hu, *Proc. Natl. Acad. Sci.* **2020**, 117, 10492.
- [56] E. C. Yang, R. Divine, C. S. Kang, S. Chan, E. Arenas, Z. Subol, P. Tinker, H. Manninen, A. Feichtenbiner, T. Mustafa, J. Hallowell, I. Orr, H. Haddox, B. Koepnick, J. O'Connor, I. C. Haydon, K.-L. Herpoldt, K. V. Wormer, C. Abell, D. Baker, A. Khmelinskaia, N. P. King, *J. Chem. Educ.* **2022**, 99, 3177.
- [57] B. B. Mallik, J. Stanislaw, T. M. Alawathurage, A. Khmelinskaia, *ChemBioChem* **2023**, 24, e202300117.



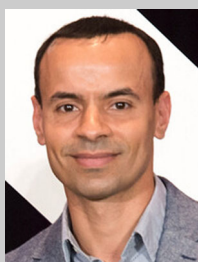
Mohammad Mahmoudi Gomari a researcher specializing in biomolecular engineering and drug discovery, with a primary focus on protein design. He earned his Ph.D. in Medical Biotechnology in 2023 from Iran University of Medical Sciences. Currently, he works as a researcher at Ali-Asghar Hospital, where he studies the behavior of proteins in the context of diseases.



Mehdi Alidadi holds a PhD in Anatomical and Regenerative Medicine. His research is dedicated to understanding degenerative diseases and developing innovative therapeutic and diagnostic strategies to address them.



Neda Rostami earned her Ph.D. in Chemical Engineering from Arak University in 2022. Her primary research focuses on the application of nanotechnology and nanoinformatics in drug delivery systems, with a particular emphasis on polymeric carriers.



Sidi A. Bencherif is a CNRS Senior Researcher at the University of Rouen Normandy and affiliated with Northeastern University. He received his Ph.D. in Chemistry from Carnegie Mellon University. His research primarily focuses on the design and development of advanced biomaterials for regenerative medicine, drug delivery, and cancer immunotherapy.