

The role of bioinformatics algorithms in modern biopharmaceutical design: Progress, challenges, and future perspectives

Mohammad Mostafa Pourseif^{1,2,3,4*}, Seyed Ali Baradaran Hosseini^{1,5}, Seyed Hossein Khoshraftar^{1,6}, Yadollah Omid⁷

¹Research Center for Pharmaceutical Nanotechnology, Biomedicine Institute, Tabriz University of Medical Sciences, Tabriz, Iran

²Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran

³Engineered Biomaterial Research Center (EBRC), Khazar University, Baku, Azerbaijan

⁴Health Science and Technology Park, Tabriz University of Medical Sciences, Tabriz, Iran

⁵Department of Medicinal Chemistry, School of Pharmacy, Urmia University of Medical Sciences, Urmia, Iran

⁶Department of Medical Genetics, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

⁷Department of Pharmaceutical Sciences, Barry and Judy Silverman College of Pharmacy, Nova Southeastern University, Fort Lauderdale, FL, 33328, USA

Article Info



Article Type:
Editorial

Article History:
Received: 28 Oct. 2025
Revised: 11 Nov. 2025
Accepted: 15 Nov. 2025
ePublished: 29 Nov. 2025

Abstract

Bioinformatics algorithms empowered by artificial intelligence (AI), machine learning (ML), and deep learning (DL) are revolutionizing biopharmaceutical design and development. These methods accelerate discovery through rapid in silico prediction of protein structure, function, and immunogenicity, reducing experimental cost and time. Generative and hybrid frameworks, especially those combining AI with physics-informed neural networks (PINNs), enable interpretable, mechanism-aware modeling for enzyme kinetics and protein engineering. Multi-omics integration and graph-based network algorithms support systems-level understanding of therapeutic targets. Despite remarkable progress, challenges persist, including limited data for novel modalities, interpretability gaps, and computational scalability. Recent advances such as AlphaFold 3, OpenFold, and NeuralPlexer, alongside evolving FDA and EMA guidelines for AI-derived therapeutics, are helping bridge innovation and clinical translation. The future of drug discovery will rely on synergistic human–algorithm collaboration to ensure responsible, reproducible, and clinically relevant biopharmaceutical innovation.

Keywords: Predictive algorithms, AI-driven biopharma, Structural bioinformatics, In silico biopharmaceuticals, Machine learning, Drug design, Systems biology

From sequence to structure to function: Algorithmic acceleration

The biological design space is combinatorial by nature; billions of possible variants emerge from even a small protein scaffold. Traditional high-throughput experimentation, while powerful, remains costly, time-consuming, and limited in scope. Algorithms close this gap through (i) sequence-structure prediction e.g., deep-learning (DL) structural models e.g., D-I-TASSER,¹ (ii) molecular docking and interaction scoring,¹ (iii) stability and solubility estimation under pharmaceutical constraints, (iv) immunogenicity and epitope assessments for therapeutic safety.²

The fusion of physics-based simulation with machine learning (ML) models allows us to traverse sequence landscapes with unprecedented accuracy. What once required a year of iterative assays can now occur in days through in silico screening.

AI-driven design: beyond predicting, toward inventing biology

Generative models such as variational autoencoders,³ diffusion models,⁴ and protein language models,⁵ have marked a significant shift in how biological design is approached. We no longer design by manual mutation; instead, we co-design with machines that learn evolutionary constraints and suggest novel therapeutic candidates. As practical examples, it should be mentioned that bioinformatics algorithms are actively driving real-world innovation across diverse classes of biopharmaceuticals. For instance, generative and predictive models are reshaping enzyme engineering by identifying mutations that improve catalytic efficiency in industrial and therapeutic pathways. In antibody development, DL-based affinity maturation can optimize both antigen-binding kinetics and manufacturability profiles early in the discovery pipeline, minimizing



*Corresponding author: Mohammad Mostafa Pourseif, Email: pourseifm@tbzmed.ac.ir



© 2025 The Author(s). This work is published by BioImpacts as an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.

downstream redesign cycles. Vaccine design also benefits significantly from immunoinformatics frameworks capable of predicting conserved,⁶ multi-epitope antigens against rapidly evolving pathogens, thereby accelerating the transition from genomic surveillance to clinical candidate bio-drugs.⁷

Likewise, computational humanization and immunogenicity de-risking strategies assist in tailoring therapeutic proteins to patient-specific contexts, effectively reducing attrition caused by adverse immune responses.⁸ This ongoing transition from predictive analysis to algorithmic creativity represents one of the most notable advances in therapeutic bioinformatics to date. Looking forward, emerging hybrid frameworks that couple generative AI with physics-informed neural networks (PINNs) are poised to further enhance model accuracy and mechanistic interpretability, particularly in enzyme kinetics and catalytic pathway prediction.⁹ By embedding physical conservation laws within generative architectures, these hybrid systems overcome the limitations of purely data-driven models and enable more robust extrapolation to novel biochemical contexts. Recent applications of PINNs in biocatalysis and protein dynamics modeling have demonstrated improved reaction rate estimation and thermodynamic consistency compared to traditional ML approaches,¹⁰ positioning such methods as a promising direction for next-generation bioinformatics algorithms.

Systems-level design and multimodal data integration

Biopharmaceutical function is inherently multilayered. It implies that the linear sequence defines the structure, structure informs function, and ultimately cellular and immunological contexts determine therapeutic outcomes. To address this complexity, modern computational pipelines employ multi-omics integration algorithms such as feature-selection models that fuse transcriptomics, proteomics, and tumor-neoantigen profiles to prioritize more druggable and clinically actionable targets. Additionally, graph-based network inference techniques,¹¹ including protein interaction adaptive graph convolutional network (e.g., PF-AGCN¹²) and immune repertoire network modeling (e.g., NAIR¹³), help to decipher how therapeutic proteins interfere with cell signaling cascades or evade immune activation. Importantly, these algorithmic predictions do not remain static; adaptive ML frameworks, such as reinforcement learning-guided optimization or Bayesian active learning, iteratively update their models throughout in silico design-build-test cycles, as fresh biochemical and biophysical data are incorporated into the evaluation of therapeutic candidates. By converging heterogeneous biological layers into a unified computational interpretation, these methods yield far more reliable forecasts of clinical performance and manufacturability, enabling the early

elimination of weak candidates well before costly pre-clinical development.

Manufacturability and developability: The often-ignored algorithmic frontier

Despite functional breakthroughs, an engineered therapeutic must still be (i) expressible at scale, (ii) stable during purification and storage, (iii) low-viscosity and aggregation-resistant in formulation, (iv) non-immunogenic in population-level diversity. Bioinformatics algorithms, for example, protein stability prediction using AlphaFold-Multimer¹⁴ or sequence optimization with RosettaDesign,¹⁵ increasingly incorporate such bioprocessability metrics early in the design stage. This prevents late-stage failures and aligns innovation with industrial feasibility.

Challenges: where algorithmic optimism meets biological reality

Despite great progress, several pressing limitations remain. First, sparse ground-truth data continues to restrict the reliability of AI-driven models, particularly in emerging therapeutic modalities such as bispecific antibodies,¹⁶ gene-editing enzymes,¹⁷ and de novo protein scaffolds.¹⁸ Unlike natural proteins with decades of accumulated structural, kinetic, and safety data, experimental annotations for these engineered entities are scarce, making models prone to overfitting and reduced real-world validity. This challenge is especially noted in enzyme engineering studies where only a tiny fraction of the mutational landscape has experimentally validated fitness measurements.

Second, interpretability and biophysical rigor remain critical barriers to clinical translation. Black-box neural networks may generate high-performing predictions, yet without mechanistic transparency regarding molecular stability, immunogenicity, or binding pathways, regulatory frameworks cannot confidently endorse algorithm-derived candidates for human use. Recent protein and antibody design pipelines¹⁹ suggest that models incorporating explicit structural or thermodynamic constraints, for example diffusion-based, physics-informed frameworks such as RFdiffusion¹⁸ can outperform purely sequence-based language models such as evolutionary scale modeling (ESM) on several structure, binding, and developability benchmarks.²⁰ Because these physics-informed models expose epitope:paratope interactions and stability features that map onto established biophysical criteria, they are also argued to be more amenable than purely statistical predictors to regulatory review and downstream manufacturability assessment.

Third, generalization across biological contexts is far from guaranteed. A therapeutic protein optimized in silico, for example, for high binding affinity in a specific

host cell line, may exhibit drastically different properties in vivo due to variations in glycosylation, immune response, proteolysis, or microbiome interaction. Studies in mRNA vaccine optimization highlight this gap,^{21–23} where constructs showing strong predicted translation efficiency occasionally underperform in diverse patient populations with distinct innate immune sensitivities.²⁴ Notably, early mRNA vaccine development efforts such as BioNTech's initial trials demonstrated that aggressive codon optimization, while computationally favorable, could unexpectedly reduce protein expression or alter immune activation profiles, underscoring the need for experimentally guided sequence refinement alongside algorithmic design. Recent advances such as the LinearDesign algorithm²⁵ illustrate how jointly optimizing codon usage and structural stability can overcome such pitfalls, achieving markedly improved half-life and translation efficiency compared to traditional codon-optimization benchmarks.

Fourth, computational scalability is still a practical constraint. High-fidelity simulation methods such as long-timescale molecular dynamics or mixed quantum mechanics/molecular mechanics (QM/MM) calculations offer deeper insight into conformational behavior and catalytic mechanisms but remain compute-intensive, limiting their feasibility in early discovery cycles where thousands of variants must be evaluated.²⁶ Nevertheless, recent advances in structure prediction and molecular modeling such as AlphaFold 3,²⁷ OpenFold,²⁸ and NeuralPlexer,²⁹ have begun to alleviate these constraints by achieving near-QM-level accuracy with markedly improved scalability, enabling broader in silico screening and iterative design at reduced computational cost.

Ethical and biosecurity considerations demand parallel innovation in model governance. The same generative design tools capable of discovering novel antitumor cytokines or antibody therapeutics can, in theory, be misappropriated to engineer highly virulent proteins or evade immunological detection. Accordingly, leading organizations are now advocating standardized transparency, access control, and safety guardrails to ensure that algorithmic advancements in biopharmaceuticals remain aligned with global health priorities.

From a translational and regulatory standpoint, agencies such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have begun articulating frameworks for the oversight of AI/ML-derived therapeutics, as highlighted in the FDA 2025 discussion paper on artificial intelligence in drug development. These initiatives emphasize the need for traceability, algorithmic interpretability, and model lifecycle management. Incorporating these governance principles into algorithmic pipelines can not only facilitate regulatory approval but also strengthen clinical

confidence in AI-driven discovery, bridging the current divide between innovation and implementation.

Looking ahead: Human-algorithm collaboration

The future of biopharmaceutical innovation will be shaped by a paradigm in which human expertise and algorithmic intelligence operate as co-architects rather than counterparts. Cross-disciplinary fluency is no longer a desirable skill but an operational necessity: researchers must simultaneously wield advanced computational literacy, encompassing algorithm design, biostatistical robustness, and model reliability and deep molecular and pharmacological insight to contextualize predictions within biological and clinical reality. The role of computation is not to replace biological experimentation, but to transform it into an iterative, data-driven feedback system.

In this emerging framework, experimental measurements continuously recalibrate ML models; those models, in turn, generate novel design hypotheses and prioritize candidates with the highest prospects for therapeutic success; subsequent targeted validation enriches the collective knowledge base, further improving algorithmic accuracy. Such a closed-loop design–build–test–learn ecosystem, powered by scalable simulation, adaptive learning, and mechanistic interpretation, will catalyze more rapid and precise development of next-generation therapeutics. Ultimately, the most transformative breakthroughs will come from seamlessly integrating human intuition with computational optimization, where algorithms become intelligent partners in discovering and engineering the medicines of tomorrow.

Conclusion

Bioinformatics algorithms have evolved into strategic engines of therapeutic discovery, compressing timelines, reducing costs, and unlocking biological territories once inaccessible to experimental science alone. Yet, as we stand at this inflection point, we must maintain rigor, transparency, and interdisciplinary alignment to ensure responsible, clinically impactful innovation. The future of biopharmaceutical design will belong to those who not only understand biology and computation independently, but can engineer the interface between them.

Acknowledgements

The authors would like to express their sincere gratitude to the Research Center for Pharmaceutical Nanotechnology at Tabriz University of Medical Sciences for their support.

Authors' Contribution

Conceptualization: Mohammad Mostafa Pourseif, Yadollah Omid.

Funding acquisition: Mohammad Mostafa Pourseif.

Investigation: Mohammad Mostafa Pourseif, Seyed Ali Baradaran Hosseini, Seyed Hossein Khoshraftar.

Project administration: Mohammad Mostafa Pourseif, Yadollah Omid.

Study Highlights

What is the current knowledge?

- AI, ML, and DL algorithms accelerate protein, antibody, enzyme, and vaccine discovery.
- Bioinformatics enables rapid in silico prediction of structure, function, and immunogenicity.
- Multi-omics integration supports target prioritization and systems-level therapeutic design.
- Computational workflows reduce experimental cost and time in drug development.
- Interpretability and biological context remain major barriers to clinical translation.

What is new here?

- Highlights the shift from predictive modeling to generative algorithmicbiodesign.
- Discusses adaptive learning cycles for continuous model refinement during development.
- Emphasizes manufacturability-aware algorithms in early therapeutic design stages.
- Addresses ethical, regulatory, and biosecurity considerations of generative bioinformatics.
- Envisions human–algorithm collaboration as the future paradigm of biopharmaceutical innovation.

Resources: Mohammad Mostafa Pourseif, Seyed Ali Baradaran Hosseini.

Supervision: Mohammad Mostafa Pourseif, Yadollah Omid.

Validation: Mohammad Mostafa Pourseif, Seyed Ali Baradaran Hosseini, Seyed Hossein Khoshraftar, Yadollah Omid.

Writing-original draft: Mohammad Mostafa Pourseif, Seyed Ali Baradaran Hosseini, Seyed Hossein Khoshraftar.

Writing-review & editing: Mohammad Mostafa Pourseif, Seyed Ali Baradaran Hosseini, Seyed Hossein Khoshraftar, Yadollah Omid.

Competing Interests

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

Data Availability Statement

This editorial does not include new experimental data or proprietary datasets. All information discussed is derived from publicly available literature, computational frameworks, and previously published resources cited within the text. No new data were generated or analyzed in support of this article. Therefore, data sharing is not applicable to this work.

Declaration of AI-assisted Tools in the Writing Procedure

During the preparation of this work, the authors used ChatGPT version 5.1 (OpenAI) to assist with paraphrasing portions of the manuscript and to check the grammar of the translations from Persian into English. After using these tools, the authors thoroughly reviewed, edited, and verified the content and take full responsibility for the accuracy and integrity of the publication.

Ethical Approval

We extend our appreciation to the Ethics Committee of Tabriz University of Medical Sciences for granting ethical approval for this study (Ethical Code: IR.TBZMED.REC.1402.769).

Funding

This research was conducted as part of a PharmD thesis and was

financially supported by the Research Center for Pharmaceutical Nanotechnology at Tabriz University of Medical Sciences (Grant #70390).

References

1. Zhou X, Zheng W, Li Y, Pearce R, Zhang C, Bell EW, et al. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat Protoc* **2022**; 17: 2326–53. doi: 10.1038/s41596-022-00728-0
2. Guo J, Jia Z, Yang Y, Wang N, Xue Y, Xiao L, et al. Bioinformatics analysis, immunogenicity, and therapeutic efficacy evaluation of a novel multi-stage, multi-epitope DNA vaccine for tuberculosis. *Int Immunopharmacol* **2025**; 152: 114415. doi: 10.1016/j.intimp.2025.114415
3. Ochiai T, Inukai T, Akiyama M, Furui K, Ohue M, Matsumori N, et al. Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity. *Commun Chem* **2023**; 6: 249. doi: 10.1038/s42004-023-01054-6
4. Alakhdar A, Poczos B, Washburn N. Diffusion models in de novo drug design. *J Chem Inf Model* **2024**; 64: 7238–56. doi: 10.1021/acs.jcim.4c01107
5. Lam HY, Guan JS, Ong XE, Pincket R, Mu Y. Protein language models are performant in structure-free virtual screening. *Brief Bioinform* **2024**; 25: bbae480. doi: 10.1093/bib/bbae480
6. Salemi A, Pourseif MM, Masoudi-Sobhanzadeh Y, Ansari R, Omid Y. Proteome-wide reverse vaccinology to identify potential vaccine candidates against *Staphylococcus aureus*. *Mol Immunol* **2025**; 183: 296–312. doi: 10.1016/j.molimm.2025.05.016
7. Majidani H, Pourseif MM, Kordi B, Sadeghi MR, Najafi A. TgVax452, an epitope-based candidate vaccine targeting *Toxoplasma gondii* tachyzoite-specific SAG1-related sequence (SRS) proteins: immunoinformatics, structural simulations and experimental evidence-based approaches. *BMC Infect Dis* **2024**; 24: 886. doi: 10.1186/s12879-024-09807-x
8. Ebrahimi SB, Samanta D. Engineering protein-based therapeutics through structural and chemical design. *Nat Commun* **2023**; 14: 2411. doi: 10.1038/s41467-023-38039-x
9. Wang Z, Xie D, Wu D, Luo X, Wang S, Li Y, et al. Robust enzyme discovery and engineering with deep learning using CataPro. *Nat Commun* **2025**; 16: 2736. doi: 10.1038/s41467-025-58038-4
10. Omar SI, Keasar C, Ben-Sasson AJ, Haber E. Protein design using physics informed neural networks. *Biomolecules* **2023**; 13: 457. doi: 10.3390/biom13030457
11. Que J, Xue G, Wang T, Jin X, Wang Z, Cai Y, et al. Identifying T cell antigen at the atomic level with graph convolutional network. *Nat Commun* **2025**; 16: 5171. doi: 10.1038/s41467-025-60461-6
12. Yang S, Su Y, Lin Y, Lin Q, Chen Z. PF-AGCN: an adaptive graph convolutional network for protein-protein interaction-based function prediction. *Bioinformatics* **2025**; 41: btaf473. doi: 10.1093/bioinformatics/btaf473
13. Yang H, Cham J, Neal BP, Fan Z, He T, Zhang L. NAIR: network analysis of immune repertoire. *Front Immunol* **2023**; 14: 1181825. doi: 10.3389/fimmu.2023.1181825
14. Kim AR, Hu Y, Comjean A, Ridiger J, Mohr SE, Perrimon N. Enhanced protein-protein interaction discovery via AlphaFold-Multimer. *bioRxiv* [Preprint]. February 21, 2024. Available from: <https://www.biorxiv.org/content/10.1101/2024.02.19.580970v1>.
15. Liu Y, Kuhlman B. RosettaDesign server for protein design. *Nucleic Acids Res* **2006**; 34: W235–8. doi: 10.1093/nar/gkl163
16. Shadman M, Gopal AK. Bispecific antibodies in action: the reality of engagement. *Blood* **2025**; 146: 2148–50. doi: 10.1182/blood.2025030376
17. Yin Y, Wang Q, Xiao L, Wang F, Song Z, Zhou C, et al. Advances in the engineering of the gene editing enzymes and the genomes: understanding and handling the off-target effects of CRISPR/Cas9. *J Biomed Nanotechnol* **2018**; 14: 456–76. doi: 10.1166/jbn.2018.2537
18. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. De novo design of protein structure and function with

- RFdiffusion. *Nature* **2023**; 620: 1089-100. doi: 10.1038/s41586-023-06415-8
19. Bennett NR, Watson JL, Ragotte RJ, Borst AJ, See DL, Weidle C, et al. Atomically accurate de novo design of antibodies with RFdiffusion. *Nature* **2025**. doi: 10.1038/s41586-025-09721-5
 20. Zhang Q, Chen W, Qin M, Wang Y, Pu Z, Ding K, et al. Integrating protein language models and automatic biofoundry for enhanced protein evolution. *Nat Commun* **2025**; 16: 1553. doi: 10.1038/s41467-025-56751-8
 21. Pourseif MM, Parvizpour S, Jafari B, Dehghani J, Naghili B, Omid Y. A domain-based vaccine construct against SARS-CoV-2, the causative agent of COVID-19 pandemic: development of self-amplifying mRNA and peptide vaccines. *Bioimpacts* **2021**; 11: 65-84. doi: 10.34172/bi.2021.11
 22. Pourseif MM, Masoudi-Sobhanzadeh Y, Azari E, Parvizpour S, Barar J, Ansari R, et al. Self-amplifying mRNA vaccines: mode of action, design, development and optimization. *Drug Discov Today* **2022**; 27: 103341. doi: 10.1016/j.drudis.2022.103341
 23. Omid Y, Pourseif MM, Ansari RA, Barar J. Design and development of mRNA and self-amplifying mRNA vaccine nanoformulations. *Nanomedicine (Lond)* **2024**; 19: 2699-725. doi: 10.1080/17435889.2024.2419815
 24. Panigaj M, Basu Roy T, Skelly E, Chandler MR, Wang J, Ekambaram S, et al. Autonomous nucleic acid and protein nanocomputing agents engineered to operate in living cells. *ACS Nano* **2025**; 19: 1865-83. doi: 10.1021/acsnano.4c13663
 25. Zhang H, Zhang L, Lin A, Xu C, Li Z, Liu K, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* **2023**; 621: 396-403. doi: 10.1038/s41586-023-06127-z
 26. Brunk E, Rothlisberger U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem Rev* **2015**; 115: 6217-63. doi: 10.1021/cr500628b
 27. Fang Z, Ran H, Zhang Y, Chen C, Lin P, Zhang X, et al. AlphaFold 3: an unprecedented opportunity for fundamental research and drug development. *Precis Clin Med* **2025**; 8: pbaf015. doi: 10.1093/pcmedi/pbaf015
 28. Ahdriz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat Methods* **2024**; 21: 1514-24. doi: 10.1038/s41592-024-02272-z
 29. Qiao Z, Nie W, Vahdat A, Miller TF, Anandkumar A. State-specific protein-ligand complex structure prediction with a multiscale deep generative model. *Nat Mach Intell* **2024**; 6: 195-208. doi: 10.1038/s42256-024-00792-z