



OPEN ACCESS

EDITED AND REVIEWED BY

Junyi Li,
Harbin Institute of Technology, China

*CORRESPONDENCE

Yaan J. Jang,
✉ yaan.jang@gmail.com

RECEIVED 14 September 2025

ACCEPTED 24 October 2025

PUBLISHED 31 October 2025

CITATION

Jang YJ (2025) Editorial: Computational protein function prediction based on sequence and/or structural data. *Front. Bioinform.* 5:1705252. doi: 10.3389/fbinf.2025.1705252

COPYRIGHT

© 2025 Jang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Computational protein function prediction based on sequence and/or structural data

Yaan J. Jang*

Department of Biochemistry, University of Oxford, Oxford, United Kingdom

KEYWORDS

protein function prediction, machine learning, deep learning, protein sequence, protein structure

Editorial on the Research Topic

Computational protein function prediction based on sequence and/or structural data

In recent years, the integration of computational and machine learning techniques has dramatically advanced our understanding of protein function and engineering. As proteins underpin virtually every biological process, accurate predictions of their functions—whether based on sequence or structural data—are critical to unraveling their roles in health, disease, and therapeutics. This editorial explores how cutting-edge computational strategies are shaping the future of protein sciences, as illustrated by the contributions in this Research Topic. By embracing both evolutionary data and state-of-the-art technologies like deep learning and molecular dynamics simulations, these studies illuminate new pathways for advancing drug development, vaccine design, and pathogen-interaction research.

At the heart of protein function prediction is the challenging task of deciphering the link between protein sequences, their three-dimensional structures, and the biological functions they support. While protein sequences provide a genetic blueprint, their true functional capabilities are often embedded in the spatial arrangement of amino acids and the resulting folding patterns. Recent breakthroughs in computational biology have allowed researchers to use sequence data and structural features to predict protein functions with unprecedented accuracy (Gligorijević et al., 2021; Jang et al., 2024).

Machine learning and deep learning approaches, such as DeepPredict (Alanazi et al.) and DCMA (Zhang et al.), are enhancing the landscape of protein function prediction. These algorithms leverage large datasets to identify subtle patterns in sequence-structure-function relationships, enabling the prediction of secondary structures, solvent accessibility, and backbone dihedral angles with improved precision. Importantly, these methods do not only enhance prediction accuracy; they also reduce computational demands, enabling large-scale analysis that was previously unfeasible. Several studies in this Research Topic highlight the utility of such AI-based tools for practical biomedical applications, from antigen design to drug discovery.

One of the most compelling uses of these predictive methods is in the realm of vaccine development. For instance, AI-guided approaches have revolutionized

the design of broadly protective antigens by predicting how viral proteins interact with immune receptors (Motiwala et al.; Chesterman et al.). These insights are not just theoretical—they are directly applicable to real-world challenges like the development of vaccines against evolving pathogens. This Research Topic includes contributions that exemplify how computational tools can accelerate vaccine design by identifying critical epitopes and optimizing immune response profiles. In particular, the studies focusing on viral infections, including the investigation of SARS-CoV-2 cytokine interactions and the engineering of antiviral peptides, underscore the role of computational methods in addressing urgent global health threats. One such example is the design of cross-reactive antigens using machine learning, where computational models were employed to predict fHbp properties and design mutants that could potentially be used in next-generation vaccine platforms (Chesterman et al.).

In addition to antigen design, another area where protein function prediction plays a pivotal role is in drug discovery. The ability to predict protein-ligand and protein-protein binding sites has far-reaching implications for designing therapeutic interventions. Many contributors to this Research Topic explore novel techniques for predicting these binding sites, with a particular emphasis on the challenges of pathogen-host interactions. By studying protein-protein interactions in pathogens such as *Enterobacter cloacae* (Motiwala et al.), or host-virus interactions like those between the human immune system and SARS-CoV-2, these studies are advancing our understanding of the molecular underpinnings of disease. By accurately predicting how proteins interact, researchers can identify new drug targets and design more effective treatments. Notably, large-scale molecular docking and PhIP-Seq (Huang et al.) approaches are being employed to uncover complex host-virus interactions, aiding in the discovery of novel therapeutic strategies.

Moreover, this Research Topic highlights the growing importance of high-throughput techniques and large-scale molecular simulations in structural biology. Tools such as PhIP-Seq (Huang et al.), which enables comprehensive antibody profiling, and large-scale molecular docking, which uncovers complex host-virus interactions, are helping to unravel the intricate web of protein interactions that govern biological processes. These approaches are essential for mapping the full spectrum of pathogen-host dynamics and for designing strategies that can mitigate the impact of infectious diseases. Additionally, a computational framework for virtual screening of cytokine and coronavirus nucleocapsid protein interactivity has accelerated the discovery of effective interventions against viral threats, showcasing the complementary nature of deep learning-based models and semi-physicochemical methods (Tomezsko et al.).

Underlying all these efforts is an appreciation for the evolutionary data that drives protein function. Understanding how proteins evolve—and how evolutionary pressures shape their structure and function—is a key aspect of protein engineering. In this context, evolutionary algorithms and molecular dynamics simulations have emerged as invaluable tools. By simulating how proteins have adapted over time and predicting how they might evolve in the future, these methods provide critical insights into how we can design proteins with desired functions. Whether it's the development of enzymes with novel catalytic properties or the engineering of proteins with enhanced binding affinities, evolutionary data are integral to rational protein design. The

study of Nipah virus (NiV) and its interaction with host receptors highlights the importance of understanding viral evolution and using computational approaches to develop novel antiviral peptides for therapeutic applications (Sajjad et al.).

The diversity of studies presented here emphasizes the breadth of protein function prediction and engineering, ranging from the design of therapeutic antibodies to the prediction of protein-protein interactions and the identification of antiviral peptides. Collectively, these contributions highlight the transformative power of computational methods in modern biomedical research. They also reinforce the growing recognition that understanding protein function requires an integrated approach, combining sequence data, structural insights, and evolutionary analysis.

Looking ahead, the convergence of deep learning, molecular dynamics, and high-throughput techniques promises to further enhance our ability to predict and engineer protein functions with ever greater accuracy and efficiency. As these computational strategies continue to evolve, they will likely open new frontiers in protein science, providing invaluable tools for tackling complex challenges in drug discovery, vaccine development, and beyond.

In conclusion, this Research Topic of articles provides a timely and comprehensive overview of the current state of computational protein function prediction, showcasing how innovative technologies are driving progress in both fundamental and applied protein research. The future of protein engineering lies in our ability to leverage these advances to design proteins with specific functions, offering exciting possibilities for the development of novel therapeutics and biotechnologies.

Author contributions

YJ: Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Glgorijević, V., Renfrew, P. D., Kosciolk, T., Leman, J. K., Berenberg, D., Vatanen, T., et al. (2021). Structure-based protein function prediction using graph convolutional networks. *Nat. Commun.* 12, 3168. doi:10.1038/s41467-021-23303-9

Jang, Y. J., Qin, Q. Q., Huang, S. Y., Peter, A. T. J., Ding, X. M., and Kornmann, B. (2024). Accurate prediction of protein function using statistics-informed graph networks. *Nat. Commun.* 15, 6601. doi:10.1038/s41467-024-50955-0