



OPEN ACCESS

EDITED AND REVIEWED BY
Debbie C. Crans,
Colorado State University, United States

*CORRESPONDENCE
Andrea Ilari,
✉ andrea.ilari@cnr.it

RECEIVED 26 May 2025
ACCEPTED 04 June 2025
PUBLISHED 18 June 2025

CITATION
Exertier C and Ilari A (2025) Specialty grand
challenges in theoretical modeling, structure
prediction and design.
Front. Chem. Biol. 4:1635423.
doi: 10.3389/fchbi.2025.1635423

COPYRIGHT
© 2025 Exertier and Ilari. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
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.

Specialty grand challenges in theoretical modeling, structure prediction and design

Cécile Exertier and Andrea Ilari*

Institute of Molecular Biology and Pathology of the National Research Council of Italy, Rome, Italy

KEYWORDS

structure prediction, MD simulation, NMR, X-ray crystallography, cryo-electron microscopy structure-based-drug-design

Introduction

The rapid development of experimental structural biology

Knowledge of the three-dimensional (3D) structure of biological macromolecules (proteins, nucleic acids, and complex macromolecular assemblies) is essential to understand their function and therefore the metabolic process to which they belong. Moreover, the structural information of protein targets involved in physio-pathological processes can be used for structure-based drug design.

Since the first protein structure was determined (the structure of Myoglobin in 1958) (Kendrew et al., 1958) to date more than 237,317 structures of biological macromolecules have been deposited in the Protein Data Bank (<https://www.rcsb.org>) solved with different techniques (194,898 by X-ray crystallography; 14,438 by NMR; 27,021 by Cryo-Electron Microscopy). Determining the structure of Myoglobin by X-ray crystallography took more than 20 years and the methodology of the isomorphous replacement method was used to solve the phase problem. While in the past solving a protein structure involved years of research, today, once a diffracting protein crystal is obtained, determining the 3D structure has become an almost automatic process. It can take just a few hours, including data collection, data indexing, structure solution, and refinement. The rapid expansion of structural biology during the 1990s was driven by several key technological advances in X-ray crystallography, which paved the way for numerous biomolecular structure determinations: the introduction of cryopreservation techniques, which prevented radiation damage of protein crystals, allowing the collection of complete diffraction data sets from single crystals; next-generation X-ray sources provided unprecedentedly bright X-ray beams; modern X-ray detectors replacing the laborious and time-consuming use of photographic film for recording diffraction patterns; and phasing methods were significantly improved through the incorporation of seleno-methionine into proteins, enabling efficient use of anomalous scattering. Finally, the advent of faster computers accelerated data processing, while advancements in computer graphics and software streamlined the construction and refinement of atomic models. Such technological advancements allowed notably the determination of large protein assemblies or the structural investigations of more complex but pharmacologically relevant membrane proteins, adding to the long lists of Nobel prizes, those awarded in Chemistry to Venki Ramakrishnan, Thomas Arthur Steitz and Ada Yonat (Schluenzen et al., 2000; Ban et al., 2000; Wimberly et al., 2000); and Robert Lefkowitz and Brian Kobilka (Rasmussen et al., 2011) respectively for the structure of the ribosome in 2009 and of G protein-coupled receptors in 2012.

Meanwhile, the use of NMR to determine the protein structure allowed the determination of the protein dynamics in solution (Wüthrich, 2003), although its application is still limited by protein size, typically to proteins below 20 kDa. Nevertheless, NMR allows structural analysis under conditions more similar to the cellular environment. It is worth noting, however, that the protein crystals used for X-ray crystallography also contain between 30% and 80% water.

Starting in 1990 thanks to the so called “resolution revolution” performed by the technical innovation introduced by Jacques Dubochet, Joachim Frank and Richard Henderson, who won the Nobel Prize in Chemistry 2017, electron microscopy become a technique on par with X-ray crystallography to solve the structure of protein with a molecular mass higher than 100 kDa (Egelman, 2016). As of now 27,021 structures determined by Cryo-EM are available in the protein data bank.

Recently, structural biology has increasingly focused on membrane proteins, which are crucial for understanding complex metabolic processes. Despite their importance, their structural characterization has been limited by challenges in stabilization and crystallization. While crystallization techniques like Lipidic Cubic Phase (LCP) have helped overcome some barriers (Landau and Rosenbusch, 1996), it is primarily cryo-electron microscopy (Cryo-EM) that has enabled the recent surge in solved membrane protein structures in the Protein Data Bank.

Conventional structural and chemical biology approaches are applied to macromolecules extrapolated from their native context. When this is done, important structural and functional features of macromolecules, which depend on their native network of interactions within the cell, may be lost. To overcome this limitation, in-cell nuclear magnetic resonance (in-cell NMR) has been used since the early 2000s to analyze macromolecules in living cells at atomic resolution (Luchinat and Banci, 2016).

The AlphaFold revolution

Our era is particularly stimulating for the field of structural study of biological macromolecules and their interactions with molecular partners. The competition between various structural prediction methods is shown in the 13th and 14th editions of Critical Assessment of Structure Prediction (CASP 13 and 14) (Kryshtafovych et al., 2019; Alexander et al., 2021), which witnessed the extraordinary success of the artificial intelligence programs AlphaFold and AlphaFold2 (Jumper et al., 2021) into the protein structure prediction field, and was rewarded by a Nobel Prize in Chemistry in 2024 to David Baker, Demis Hassabis and John Jumper. This program, based on machine learning, can predict with a high degree of accuracy the 3D structure of a biological macromolecule. Thanks to this program, more than 214 million protein structures have been predicted (Varadi et al., 2024) giving to the biologists the unique opportunity to investigate at the molecular level the metabolic pathways of different organisms.

However, it must never be forgotten that AlphaFold still provides a prediction, even if an accurate one, of the protein structure, the model may contain regions predicted with low confidence or poor accuracy, and it has been proposed that by implicitly including experimental information on the protein

target, such as a density map, the protein model would additionally improve (Terwilliger et al., 2022). Moreover, the currently available experimentally determined structures of complexes among biological macromolecules or between protein targets and their ligands, are not numerous enough to allow even the most advanced artificial intelligence methods to accurately predict the interactions between macromolecule and their interactors.

New challenges of structural biology

The integrating structural biology to disclose cell function

Looking forward, the synergy between experimental and computational approaches is expected to greatly expand the structural knowledge of biological systems and their interactions. There's a growing awareness that the most meaningful understanding of molecular machines emerges when they are examined within the context of intact cells, where their functions can be observed in a native, complex environment. Reaching this goal will depend on an integrative structural biology strategy, one that not only refines and diversifies structural datasets to support more accurate predictions, but also brings together information from complementary methods.

The further development of cryo-electron tomography will make it possible to obtain the structure of proteins within the cell.

Another challenge lies in understanding the function of intrinsically disordered regions of proteins, which make up 30%-40% of the proteome in eukaryotes and are often critically important for cellular metabolism (such as cell division, DNA transcription, translation, and signaling) (Holehouse and Kragelund, 2024). The fact that these regions can adopt different structures upon binding to different partners, and can undergo rapid conformational changes, makes it necessary to introduce the variable of time in order to study the structure–function relationship in these proteins. Integrative structural biology, time-resolved experimental techniques, molecular dynamics and deep learning-based approaches, such as AlphaFold2, will be of fundamental importance in this field.

Structure-based drug design

In the early 1990s, the first drugs designed using structure-based drug design (SBDD) entered the market. One example of these successes is indinavir, an HIV protease inhibitor (Dorsey et al., 1994). Today, structural biology is widely used in pharmaceutical chemistry to design new lead compounds. High-throughput screening on protein targets, as well as virtual screening, make it possible to identify hit compounds that can then be optimized through the resolution of the protein–inhibitor complex structure. The growing number of protein–ligand/inhibitor complex structures in the Protein Data Bank will enable the application of artificial intelligence and machine learning to better predict interactions between small molecules and biological macromolecules.

Moreover, the development of new and more sophisticated techniques as the fragment screening at the synchrotron light radiation facilities allowed the identification of sub-pocket in the catalytic cavity of the protein target that can be exploited to synthesize more efficient, more specific and less toxic lead compounds.

Conclusion and perspectives

In recent years, structural biology has demonstrated an increasingly direct impact on society, not only through the development of innovative drugs and more targeted therapeutic approaches, but also within the broader context of the United Nations Sustainable Development Goals (SDGs). The ability to predict and model protein structures, enhanced by the use of artificial intelligence algorithms such as AlphaFold, has enabled more effective responses to global health challenges. These advances contribute tangibly to achieving Goal 3 (Good Health and Well-being), by facilitating the discovery of new therapies for complex and rare diseases, and Goal 9 (Industry, Innovation and Infrastructure), by fostering collaborations between research centers, universities, and the pharmaceutical industry. Moreover, the integration of omics approaches, such as proteomics and metabolomics and chemical biology enables a deeper understanding of biomolecular networks, paving the way for personalized medicine based on the identification of specific targets and the selection of optimized treatments for each patient. This multidisciplinary and integrated approach, in addition to strengthening the link between basic research and clinical applications, contributes to promoting a more equitable and sustainable global health.

Together, these advances highlight a future where structural biology, empowered by technological innovation and multidisciplinary collaboration, will continue to bridge the gap between fundamental molecular understanding and real-world applications, contributing not only to scientific progress, but also to a healthier, more sustainable, and more equitable world.

In this dynamic and rapidly evolving landscape, the Theoretical Modelling, Structure, Prediction & Design section of *Frontiers in Chemical Biology* aims to serve as a multidisciplinary platform for advancing our understanding of macromolecular structure–function relationships. The section will welcome original research articles and comprehensive reviews that contribute to both foundational and applied aspects of structural and computational biology, with a particular focus on the chemical dimensions of these investigations.

The scope includes, but is not limited to, experimental and theoretical studies on molecular structures, folding mechanisms, protein and nucleic acid design, evolutionary pathways, and biomolecular interactions. A central emphasis will be placed on structure-based drug design, encompassing both small-molecule discovery and the development of innovative strategies for modulating biological targets.

The integration of structural data with computational modeling, machine learning, and time-resolved experimental techniques offers unprecedented opportunities to explore biomolecular complexity at multiple scales. As structural databases expand and methodologies

become more sophisticated, we foresee a future where predictive modeling will not only complement experimental approaches but will also guide hypothesis generation, drug development, and functional annotation of unknown proteins.

By fostering interdisciplinary collaboration among chemists, biologists, physicists, and computational scientists, this section aspires to contribute to the next generation of discoveries in chemical biology, where structure truly meets function, and where molecular insight translates into therapeutic innovation.

Author contributions

CE: Writing – review and editing. AI: Writing – review and editing, Writing – original draft, Conceptualization.

Funding

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

Acknowledgments

The authors acknowledge the project: “Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC”, acronym “ITACA.SB” (Project No. IR0000009, CUP B53C22001790006), funded by the European Union’s NextGenerationEU under the MUR call 3264/2021 PNRR M4/C2/L3.1.1.

Conflict of interest

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

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

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

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

- Alexander, L. T., Lepore, R., Kryshchak, A., Adamopoulos, A., Alahuhta, M., Arvin, A. M., et al. (2021). Target highlights in CASP14: analysis of models by structure providers. *Proteins Struct. Funct. Bioinforma.* 89 (12), 1647–1672. doi:10.1002/prot.26247
- Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science*, 289(5481), pp. 905–920. doi:10.1126/science.289.5481.905
- Dorsey, B. D., Levin, R. B., McDaniel, S. L., Vacca, J. P., Guare, J. P., Darke, P. L., et al. (1994). L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor. *J. Med. Chem.* 37 (21), 3443–3451. doi:10.1021/jm00047a001
- Egelman, E. H. (2016). The current revolution in cryo-EM. *Biophys. J.* 110 (5), 1008–1012. doi:10.1016/j.bpj.2016.02.001
- Holehouse, A. S., and Kragelund, B. B. (2024). The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell Biol.* 25 (3), 187–211. doi:10.1038/s41580-023-00673-0
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C., (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181 (4610), 662–666. doi:10.1038/181662a0
- Kryshchak, A., Schwede, T., Topf, M., Fidelis, K., and Moulton, J. (2019). Critical assessment of methods of protein structure prediction (CASP)—round XIII. *Proteins Struct. Funct. Bioinforma. Funct. Bioinforma.* 87 (12), 1011–1020. doi:10.1002/prot.25823
- Landau, E. M., and Rosenbusch, J. P. (1996). Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci.* 93 (25), 14532–14535. doi:10.1073/pnas.93.25.14532
- Luchinat, E., and Banci, L. (2016). A unique tool for cellular structural biology: in-cell NMR. *J. Biol. Chem.* 291 (8), 3776–3784. doi:10.1074/jbc.R115.643247
- Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., et al. (2011). Crystal structure of the β_2 adrenergic receptor–Gs protein complex. *Nature* 477 (7366), 549–555. doi:10.1038/nature10361
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., et al. (2000). Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*, 102(5), pp. 615–623. doi:10.1016/S0092-8674(00)00084-2
- Terwilliger, T. C., Poon, B. K., Afonine, P. V., Schlicksup, C. J., Croll, T. I., Millán, C., et al. (2022). Improved AlphaFold modeling with implicit experimental information. *Nat. Methods* 19 (11), 1376–1382. doi:10.1038/s41592-022-01645-6
- Varadi, M., Bertoni, D., Magana, P., Paramval, U., Piduchna, I., Radhakrishnan, M., et al. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res. Nucleic Acids Res.* 52 (D1), D368–D375. doi:10.1093/nar/gkad1011
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M. Jr, Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., et al. (2000). Structure of the 30S ribosomal subunit. *Nature* 407 (6802), 327–339. doi:10.1038/35030006
- Wüthrich, K. (2003). NMR studies of structure and function of biological macromolecules Nobel Lecture. *J. Biomol. NMR Biomol. NMR* 27 (1), 13–39. doi:10.1023/A:1024733922459