Check for updates

OPEN LETTER

When artificial intelligence meets protein research

[version 1; peer review: 2 approved, 1 approved with reservations]

Sonia Longhi[1,2], Salvador Ventura[3-6], Sandra Macedo-Ribeiro[7], Leandro G. Radusky[8], Jovana Kovačević[9,10], R. Gonzalo Parra[11], Miguel A. Andrade-Navarro[12], Andrey V. Kajava[1,13], Zuzana Bednáriková[14], Alexander Monzon[15], Rita Vilaça[16]

[1]Aix Marseille University, Marseille, Provence-Alpes-Côte d'Azur, France
[2]Centre National de la Recherche Scientifique (CNRS), Lab. Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille, Provence-Alpes-Côte d'Azur, France
[3]Autonomous University of Barcelona Institute of Biotechnology and Biomedicine, Bellaterra, Catalonia, Spain
[4]Universitat Autonoma de Barcelona Departament de Bioquimica i Biologia Molecular, Bellaterra, Catalonia, Spain
[5]Institut d'Investigacio i Innovacio Parc Tauli, Sabadell (I3PT-CERCA), Catalonia, Spain
[6]Hospital Universitari Parc Taulí, Sabadell, Spain
[7]Universidade do Porto Instituto de Investigacao e Inovacao em Saude, Porto, Porto District, Portugal
[8]Mimark Diagnostics S.L., Parc Cientific Barcelona, Barcelona, Catalonia, Spain
[9]University of Belgrade Faculty of Mathematics, Belgrade, Serbia
[10]The Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad, Serbia
[11]Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Barcelona, Spain
[12]Johannes Gutenberg University Mainz Institute of Organismic and Molecular Evolution, Mainz, Rhineland-Palatinate, Germany
[13]Centre de Recherche en Biologie cellulaire de Montpellier, Université de Montpellier, CNRS, Montpellier, Montpellier, France
[14]Institute of Experimental Physics Slovak Academy of Sciences, Košice, Košice Region, Slovakia
[15]University of Padua Department of Biomedical Sciences, Padua, Veneto, Italy
[16]REQUIMTE LAQV Porto, Porto, Porto District, Portugal

v1 **First published:** 15 Jul 2025, **5**:185
https://doi.org/10.12688/openreseurope.20628.1
Latest published: 15 Jul 2025, **5**:185
https://doi.org/10.12688/openreseurope.20628.1

Abstract

The 2024 Nobel Prizes in Chemistry and Physics mark a watershed moment in the convergence of artificial intelligence (AI) and molecular biology. This article explores how AI, particularly deep learning and neural networks, has revolutionized protein science through breakthroughs in structure prediction and computational design. It highlights the contributions of 2024 Nobel laureates John Hopfield, Geoffrey Hinton, David Baker, Demis Hassabis, and John Jumper, whose foundational work laid the groundwork for AI tools such as AlphaFold. These tools are transforming our understanding of protein folding, and the dynamics of non-globular proteins, including intrinsically disordered proteins. While AI-driven methods have made predicting protein structures faster and more accessible, they also underscore ongoing scientific challenges, including the dynamics of protein folding and amyloid aggregation. European initiatives, such as

Open Peer Review

Approval Status ✓ ? ✓

	1	2	3
version 1 15 Jul 2025	✓ view	? view	✓ view

1. **Nicola Bordin**, University College London, London, UK
2. **Judith Bernett**, Technical University of Munich, Freising, Germany
Konstantin Pelz, Technical University of Munich, Freising, Germany
3. **Marcelo C. R. Melo**, Colorado State University College of Natural Sciences, Fort

the COST Actions NGP-net (BM1405) and ML4NGP (CA21160), are spearheading efforts to bridge these gaps by integrating AI and experimental data in the study of non-globular proteins. Together, these developments signal a transformative shift in biology, paving the way for novel discoveries in medicine, biotechnology, and materials science.

Keywords
artificial intelligence, machine learning, non-globular proteins, deep learning, structural biology, bioinformatics, protein folding

Collins, USA

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the COST Actions gateway.

This article is included in the Machine Learning collection.

This article is included in the Deep Learning collection.

Corresponding authors: Alexander Monzon (alexander.monzon@unipd.it), Rita Vilaça (ritavilacawork@gmail.com)

Author roles: Longhi S: Writing – Original Draft Preparation, Writing – Review & Editing; **Ventura S**: Writing – Original Draft Preparation, Writing – Review & Editing; **Macedo-Ribeiro S**: Writing – Original Draft Preparation, Writing – Review & Editing; **Radusky LG**: Writing – Original Draft Preparation, Writing – Review & Editing; **Kovačević J**: Writing – Original Draft Preparation, Writing – Review & Editing; **Parra RG**: Writing – Original Draft Preparation, Writing – Review & Editing; **Andrade-Navarro MA**: Writing – Original Draft Preparation, Writing – Review & Editing; **Kajava AV**: Writing – Original Draft Preparation, Writing – Review & Editing; **Bednáriková Z**: Writing – Original Draft Preparation, Writing – Review & Editing; **Monzon A**: Conceptualization, Funding Acquisition, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Vilaça R**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Framework Programme for Research & Innovation as part of the COST Action [CA21160, ML4NGP], as supported by the COST Association (European Cooperation in Science and Technology).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Longhi S *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License , which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Longhi S, Ventura S, Macedo-Ribeiro S *et al.* **When artificial intelligence meets protein research [version 1; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2025, **5**:185
https://doi.org/10.12688/openreseurope.20628.1

First published: 15 Jul 2025, **5**:185 https://doi.org/10.12688/openreseurope.20628.1

Disclaimer

The views expressed in this article are those of the authors. Publication in Open Research Europe does not imply endorsement of the European Commission.

Artificial intelligence (AI) is revolutionizing biology, unlocking mysteries that defied previous understanding. The 2024 Nobel Prizes in Physics and Chemistry recognize this transformation, honoring researchers whose work has reshaped our understanding of biomolecules, including proteins, the molecular machines of life. From neural networks that decode complex biological data to AI-driven models capable of predicting protein structures with near-experimental accuracy, these breakthroughs are accelerating discoveries in medicine, biotechnology, and beyond. As we stand at the intersection of AI and molecular science, it is clear that we are witnessing the dawn of a new era in biology.

Nobel laureates pioneering New Frontiers in protein science

The 2024 Nobel Prizes in Physics and Chemistry reflect the growing impact of AI in revolutionizing biological sciences, particularly in the field of protein structure prediction and *de novo* computational protein design. John Hopfield and Geoffrey Hinton were awarded the 2024 Nobel Prize in Physics for their foundational contributions to artificial neural networks, while the Nobel Prize in Chemistry was awarded to David Baker for his innovations in computational protein design and to Demis Hassabis and John M. Jumper for the development of AlphaFold, an AI system that predicts protein structures with unprecedented accuracy.

The fundamental scientific contributions of these laureates have catalysed advances in machine learning, bioinformatics, and structural biology, opening new avenues for understanding the complex behaviour of proteins, including the fascinating class of non-globular proteins. Together, these achievements offer extraordinary promise for building better working hypotheses to design experiments that advance our broader understanding of biology at the molecular level with implications in medical research, biotechnology and material sciences.

From neural networks to protein science: the role of AI in biology

Hopfield and Hinton's pioneering work on neural networks laid the foundation for today's machine learning models, which have become integral tools for processing and interpreting vast amounts of experimental data. These models can now detect intricate patterns and generate predictions from complex datasets, an advancement crucial for fields like genomics, transcriptomics and proteomics. Their innovations in AI, particularly Hinton's work on deep learning techniques and Hopfield's neural network based on physics principles, have transformed how researchers analyse large-scale data, enabling them to uncover hidden patterns within complex biological data.

These combined contributions enhance our ability to analyse complex datasets with machine learning technologies but also reduce the need for manual intervention, making data interpretation more rapid, efficient and precise. "The 2024 Nobel Prize for Physics underscores how fundamental discoveries can have profound impacts across various scientific domains, including bioinformatics. This has revolutionised the study of biological processes, particularly in areas such as protein structure prediction" says Prof. Jovana Kovacevic.

AlphaFold and the breakthrough in protein structure prediction

A major milestone in applying AI to biology was reached in 2021 when AlphaFold was developed by Demis Hassabis and John M. Jumper demonstrating an unprecedented ability to predict the **three-dimensional structure of proteins** from their amino acid sequence. This breakthrough, rooted in the principles of machine learning laid down by Hopfield and Hinton, has reshaped structural biology, giving access to structural data even in the absence of experimental data obtained through costly and time-consuming techniques. One must not forget that the impactful advancements in AI-based prediction of protein structures stand on the shoulders of experimental protein structure determination. The pioneering development of the Protein Data Bank (PDB), established in 1971, one of the first open-access databases for biological data, played a crucial role in enabling and expediting these breakthroughs in AI-based methodologies for structure prediction. The PDB set a precedent by creating a centralized repository of high-quality, expertly curated, three-dimensional atomic coordinates of proteins and nucleic acids, leveraging AI-driven predictors like AlphaFold to train models. "In the end, AI-based prediction models are improving and expediting experimental determinations of protein structures, being routinely used in the experimental structure determination workflows." says Sandra Macedo-Ribeiro.

While AlphaFold has made significant strides in predicting globular protein structures, non-globular proteins (NGPs) pose unique challenges due to their irregular shapes and the disordered nature of regions thereof. NGPs, which are involved in numerous biological processes, have long defied conventional predictive methods due to their unique properties, such as high flexibility, lack of a predominant equilibrium conformation, and high conformational heterogeneity.

Deep learning models are making strides in predicting NGP functions, interactions, and roles in cellular processes. Neural networks are particularly promising in enhancing functional prediction accuracy and discovering complex protein-protein interactions. Just as physics expands our understanding of the universe, machine learning is helping decode the intricate world of proteins, especially those that defy conventional categorization.

"The availability of AlphaFold models and easy access to ColabFold and AlphaFold2 Colab servers for custom

predictions, have contributed considerably to growing awareness of intrinsically disordered regions (IDRs) in proteins. Long "loopy" regions devoid of regular secondary structure, which are often predicted with low confidence in AlphaFold models, visually catch the attention of scientists. "says Sonia Longhi.

Nobel-winning chemistry: computational protein design and AlphaFold

The Nobel Prize in Chemistry for David Baker highlights another key area where AI is pushing boundaries: *de novo* computational protein design. Baker's research allows scientists to design synthetic proteins for applications in medicine, biotechnology, and materials science. This work complements the achievements of Hassabis and Jumper, who have demonstrated the predictive power of AI with AlphaFold, which is already making an impact in fields like drug discovery and enzyme engineering.

For more than half a century, the "holy grail" of bioinformatics has been about how to decipher the information that is contained within a polypeptide to predict the 3D structure that it would acquire once in solvent, e.g. water.

Sequencing technologies allowed us to obtain the sequences of millions of proteins (approximately 200 million are contained in the UniProt database). However, experimentally solving their corresponding 3D structures would be an unimaginably expensive process, both in time and money. After developing some of the first algorithms aiming to solve the structure prediction problem, in 2003, David Baker produced the first *de novo* designed protein, which showed to be foldable after being expressed in the wet lab (Basanta *et al.*, 2016). However, after some initial success, predicting the structures of complex proteins remained elusive. It was in 2021, two years after its first version, that the group led by John Jumper and Demis Hassabis published AlphaFold2. "This software was able to predict in minutes the structures of proteins that experimental labs have failed to solve for almost a decade. This was the beginning of an unprecedented revolution in molecular biology where the sequence/structure gap was closed and suddenly we started having access to good-quality predictions of structures of a majority of known proteins in the planet." says Dr Gonzalo Parra.

The ability to predict the structure of proteins from their amino acid sequence has profound implications for biomedicine, biochemistry, and biotechnology. As AlphaFold continues to evolve, researchers across the globe are now applying its principles to tackle more complex and irregular protein structures, such as non-globular proteins and protein complexes. From drug design to building synthetic materials, the consequences of these developments are yet to be assessed. One thing is clear: the spotlight right now is on Computational Biology, and it will be so for a long time.

The impact of AI on structural biology: a researcher's perspective

Dr. Leandro Radusky, a structural biologist, recalls the first time he encountered AlphaFold:

"The first time I seriously paid attention to AlphaFold was when the results of the 14th edition of the CASP protein structure prediction competition were announced in 2020. Like most people working in this field, my initial reaction was disbelief. The ability to predict structures from a protein's sequence had made a massive leap, using a completely different approach from what had been employed for decades." After AlphaFold's release, the impact was immediate: "Shortly after the release of the tool's code, the structure predictions of all human proteins and many other organisms of interest became publicly available. Overnight, the number of proteins with reasonably accurate structures available went from a few hundred thousand to millions." Despite this, Radusky raises an important question about the balance between AI-driven methods and traditional scientific approaches:

"Beyond the technological leap, AlphaFold has highlighted a critical choice young researchers are facing today: how much should they rely on traditional interpretive science versus AI-driven approaches? With AI, we are often trading understanding for the ability to solve highly complex problems. As we move forward, the question seems not to be whether to use AI, but how to integrate it thoughtfully into the pursuit of understanding."

The future of AI in protein science

Without any doubt, AI-based tools like AlphaFold represent a breakthrough in structural biology, with extensive applications in protein structure prediction which deserved the Nobel Prize recognition. Yet, this does not signal the end of discoveries in this field, as many challenges in structural biology remain unresolved.

One major challenge is understanding the molecular mechanisms and atomic interactions that guide a polypeptide chain to fold into its specific three-dimensional structure. This has been a longstanding question in structural biology, originally attracting many talented physicists to the field. Despite significant advances, a comprehensive theory of protein folding based on physical principles that would offer a deep understanding of these processes is still missing. "AI-based methods, with their ability to generate reliable structural models, are poised to greatly advance our understanding of protein folding mechanisms. These remaining challenges offer exciting opportunities for groundbreaking advances in protein structure research, with the potential for future Nobel Prizes built on the achievements of AI-driven methods" says Prof. Andrey V. Kajava.

While researchers in the field of protein disorder are certainly excited by the progress brought by AlphaFold and the more recently developed AlphaFold Multimer, which extends predictions to protein complexes, one must remain aware of the current limits of these systems: they produce predictions

that do not take into account the dynamic properties of protein structures and depend on experimentally solved structures, which might be affected by different experimental conditions and molecular partners. Concerning IDRs, AlphaFold appears to have a tendency to overpredict helical structures. Fortunately, AI-based methods are continuously being improved, and the community is advancing in developing new or improving existing AI-driven tools to specifically study intrinsically disordered proteins (IDPs). For instance, machine learning methods are already being used to learn force fields specific for molecular dynamics simulations of IDRs.

Another question arises when approaching complex biological systems, such as amyloid structures. Although biologically relevant as they are associated with debilitating neurodegenerative and other conformational diseases, these structures cannot yet be solved by AI-based approaches owing to their high complexity. Deciphering the amyloid structures formed during protein aggregation is challenging, as protein aggregation is not guided by functional constraints or evolutionary pressure, which typically help maintaining specific interactions. As a result, most of the aggregated species lack the residue coevolution patterns that AlphaFold and similar predictors depend on, creating a significant limitation for structure prediction in this context. Furthermore, the aggregation process generates a variety of structures—intermediates, oligomers, and amyloid fibrils—leading to a highly complex protein landscape. "It's becoming clear that the ability of a single polypeptide sequence to form multiple stable amyloid structures is not rare. Instead, it underscores that the aggregation process is controlled by kinetics rather than thermodynamics. This means that a single protein sequence can produce various stable forms, challenging Anfinsen's principle, which traditionally underpins AI-sequence-based predictions" says Prof. Salvador Ventura.

Nevertheless, these AI-based tools like AlphaFold are highly dependent on experimental data. The recent advances in the resolution of protein structures by electron cryo-EM, which has reached atomic resolution similar to X-ray crystallography in recent years, have revitalised the pace at which novel structures are deposited in databases, addressing the resolution of very large complexes that until recently were beyond the reach of experimental resolution. AI-driven methodologies are also becoming increasingly integrated into electron cryo-EM for resolving large protein complexes at atomic resolution. AI's reliance on detecting protein contacts through correlated mutations has been significantly enhanced by the sequencing of complete genomes, making it possible to derive statistically significant structural information, considering that for many proteins, we have thousands of corresponding homologs computationally translated from the fully sequenced genomes as potential transcripts. "This alone justifies not only efforts to sequence as many organisms as possible but also conservationist initiatives across the planet, as each species' genome contributes to our knowledge of protein variations, and human-made climate change challenges the existence of entire ecosystems." says Prof. Miguel Andrade.

AI as a transformative force in biology

The Nobel Prizes for 2024 mark a turning point for AI's role in advancing complex dogmas in Biology like the protein sequence-structure-function prediction. The integration of Baker's computational design with the foundational principles of AI laid down by Hopfield and Hinton, as well as Hassabis and Jumper's AlphaFold, underscores the transformative potential of machine learning in solving complex biological problems. These innovations will not only deepen our knowledge of biological macromolecules, such as proteins, but will also enhance drug development and the treatment of diseases driven by protein misfolding and pathological aggregation, including neurodegenerative diseases and cancer. As these tools continue to evolve, the scientific community stands on the brink of uncovering new insights into non-globular proteins that are key to novel therapeutic approaches and a deeper understanding of life at the molecular level.

This impressive wealth of three-dimensional protein models significantly enhances awareness of the relevance of protein structure in biological research. While experimental validation remains necessary, they undoubtedly provide outstanding starting points not only for structural biologists aiming to elucidate experimental structures of proteins but also for other researchers studying the physiological and pathological implications of target proteins.

As new developments in AI methods continue to unfold, the scientific community is looking forward to witnessing advancements in predicting protein structural dynamics and conformational biases in IDPs, which represent the next challenge in the quest to decode protein function. In this regard, developing open-access databases with experimental data on protein conformational dynamics and ensembles will provide the foundations for robust high-quality AI-driven predictions.

The Nobel Prizes awarded last year highlight the extraordinary potential of integrating AI into biological research. By applying these AI-driven techniques to non-globular proteins, we now have the tools to solve some of the most challenging questions in biology.

COST actions paving the way to integrate AI and disordered proteins

As we celebrate these groundbreaking achievements, it becomes increasingly clear that the integration of machine learning in protein science is not a passing trend but a transformative force that will shape the future of science for decades to come. While AlphaFold and similar AI models have revolutionized the prediction of globular protein structures, the complexity of non-globular proteins (NGPs) continues to pose unique challenges. Recognizing this gap, European scientific collaboration through COST Actions has played a pioneering role in shaping the research agenda around NGPs.

The COST Action BM1405: Non-globular proteins – from sequence to structure, function and application in molecular physiopathology (NGP-net) was instrumental in establishing

a pan-European research community dedicated to exploring the complexity of NGPs. This Action united experimentalists, computational biologists, and medical researchers, creating a fertile ground for knowledge exchange and cross-disciplinary innovation. The action fostered new standards for disorder annotations, facilitated the sharing of experimental protocols, and catalyzed collaborations that led to key advances in our understanding of protein disorder and its role in disease mechanisms.

Building on these foundations, the COST Action CA21160: Non-globular proteins in the era of Machine Learning (ML4NGP), started in 2022, is driving a new paradigm that integrates machine learning and deep learning technologies to tackle the enduring challenges of NGPs. ML4NGP aligns directly with the scientific momentum recognized by the 2024 Nobel Prizes by emphasizing how AI can be purposefully adapted to understand disordered and dynamic protein systems. This Action is currently developing community benchmarks, curating protein intrinsic disorder datasets, and promoting the development of interpretable AI models tailored for flexible protein regions. By organizing training schools, workshops, and open-access initiatives, ML4NGP ensures that the next generation of researchers has the skills and knowledge to maximize the benefits of these powerful new tools.

Together, NGP-net and ML4NGP form a clear path from scientific understanding to real-world technological applications. Their work highlights that while AI breakthroughs like AlphaFold are impressive, we still need to push further to truly capture the complexity and flexibility of non-globular proteins. In doing so, they are not only contributing to the scientific advancements spotlighted by the Nobel Committee but also ensuring that non-globular proteins, long in the shadow of structured proteins, are finally receiving the attention and methodological innovation they deserve. The Nobel Committee's recognition of AI's potential in advancing protein science is a timely reminder that the biggest challenges in biology may finally be within reach.

Ethics and consent

Ethical approval and consent were not required for this study.

Data availability

No new data were created or analysed in this study. Data sharing is not applicable to this open letter as it only presents opinions and perspectives from the authors.

Acknowledgements

Not applicable.

References

Basanta B, Chan KK, Barth P, *et al.*: **Introduction of a polar core into the *de novo* designed protein Top7.** *Protein Sci.* 2016; **25**(7): 1299–307.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ? ✔

Version 1

Reviewer Report 05 September 2025

https://doi.org/10.21956/openreseurope.22318.r58288

© **2025 Melo M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Marcelo C. R. Melo**
Biochemistry and Molecular Biology, Colorado State University College of Natural Sciences, Fort Collins, Colorado, USA

In this open letter, the authors accurately describe the transformative and long-lasting effects AI models will have (and already have had) on biological research, especially the “extraordinary promise” they offer to medicine, biotechnology, and beyond. However, I believe more evidence is needed to claim that such models have “reshaped our understanding of biomolecules.” As discussed in the letter (and in other recent publications), the protein structure prediction and the protein folding process are different problems. AI models have reliably outperformed previous structure prediction methods, but such models have yet to reveal a new fundamental principle of protein folding. The authors highlight that “Despite significant advances, a comprehensive theory of protein folding based on physical principles that would offer a deep understanding of these processes is still missing.” I would suggest a more careful tone as to avoid conflating the promise and hopes we have for AI methods, and the achievements they have actually already reached. As mentioned in the letter, over reliance on AI can to lead to “trading understanding for the ability to solve highly complex problems.”

With that said, the authors highlight the impact AI models can have in a variety of research endeavors. Moreover, they also describe how their use has drawn attention to complicated, difficult to determine structures, and even intrinsically disordered regions of proteins. Importantly, the authors highlight the fundamental work done by the Protein Data Bank and the decades of structural biology work, without which the AI models discussed here would not exist. Besides the structural databases, I would add that the development of the first AlphaFold models also relied on sequence databases mentioned elsewhere in this open letter.

While the overall tone of the letter is very positive (and I believe it should be given the potential applications and possibility of innovation AI brings to protein research), it is important to remember the reader that such models are far from perfect. At times, structures are still predicted with low accuracy and detail, and generative models still can fail to produce biologically relevant structures altogether. While their overall impact is obviously positive, results from AI models still

require validation and curation, which are new problems that require research and methods development.

The letter has a single citation, but makes claims regarding current research that could be supported by a few more citations. For instance, authors mention that "... machine learning methods are already being used to learn force fields specific for molecular dynamics simulations of IDRs." This is a very important topic, both in the context of IDR dynamics and AI-based structure prediction, so one or more citations here would help support the argument made by the authors. Similar instances are mentions of AI integration into experimental structure determination, and drug development.

Since the field of deep learning, and specifically the applications to biomolecules, moves at impressively fast pace, it is inevitable that any manuscript would age quickly. However, when authors mention AlphaFold Multimer, they could also raise more recent example of AlphaFold 3, which made a very sharp change from sequence-based contact matrix prediction to fully generative structure prediction, and improved how it predicts protein complexes that involve ligands and nucleic acids. The very recent publication of BioEmu also shows how AI-based structure prediction is trying to tackle the dynamical nature of proteins, and even with its own caveats, can provide a broader distribution of predicted structures.

The focus on non globular proteins, their complex dynamics and functions, and how integration between experimental and computational research will drive this new field of research, are well supported in the letter. Authors also propose a path for workforce development and training, as well as for the development of new research directions integrating AI to protein science.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)

Yes

Does the article adequately reference differing views and opinions?

Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Partly

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational biophysics; Protein dynamics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 05 September 2025

https://doi.org/10.21956/openreseurope.22318.r57519

© 2025 Bernett J et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Judith Bernett
TUM School of Life Sciences, Technical University of Munich, Freising, Bavaria, Germany
Konstantin Pelz
TUM School of Life Sciences, Technical University of Munich, Freising, Bavaria, Germany

Summary of the article
In this open letter, the authors present how computational biology receives more public spotlight, as in 2024, the physics Nobel prize was awarded for the invention of artificial neural networks, and the chemistry Nobel prize was awarded for computational protein design and protein structure prediction. They highlight current limitations of AI-driven protein research, particularly in predicting intrinsically disordered proteins, non-globular proteins, and amyloid structures. They finally present their COST actions on non-globular proteins.
While the general topic of the article is timely and interesting to the general research community, we think it would benefit from some revisions. The article is repetitive and too long, given its relatively low information density. The text may be difficult for non-computational biologists to follow, as key concepts are not explained clearly. Moreover, the transitions between sections are weak, and the overall narrative flow requires improvement. Finally, we find the format of the letter with the direct quotes by the authors somewhat confusing.

(1) Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained.)
Yes.
Justification: The purpose of the letter is explained - AI is increasingly important in protein research and can successfully be applied, as demonstrated by the Nobel-awarded AlphaFold.

(2) Does the article adequately reference differing views and opinions?
No.
Justification: The opinions stated are the opinions of the authors, sometimes with direct quotes, but the opinions do not differ much. Other opinion pieces are not mentioned or quoted.
Recommendations: It might make sense to include some more critical voices about the excessive

AI usage, lack of interpretability, excessive use of energy, and untrustworthiness. Moreover, developing breakthrough AI models such as AlphaFold, Enformer, AlphaGenome, or ESM-2 lies beyond the scope of most research labs due to their limited computational and human resources.

(3) Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Partly.
Justification: There is only one cited source, and the article contains bold, overexaggerated, and generalizing statements. Examples of the latter are

- *"[AI is] unlocking mysteries that defied previous understanding",*
- *"Just as physics expands our understanding of the universe, machine learning is helping decode the intricate world of proteins.",*
- *"One this is clear: the spotlight right now is on Computational Biology and it will be so for a long time",*
- *"AI-based methods [...] are poised to greatly advance our understanding of protein folding mechanisms",*
- *"These innovations will not only deepen our knowledge on biological macromolecules [...] but will also enhance drug development and the treatment of diseases",*
- *"[...] the biggest challenges in biology may finally be within reach".*

Statements that could use sources include:

- NGPs are challenging to predict (for AlphaFold)
- NGPs are involved in numerous biological processes
- *"[...] machine learning methods are already being used to learn force fields specific for molecular dynamics simulations of IDRs"*
- *"The action [...] led to key advances in our understanding of protein disorder and its role [...]"*
- Generally, the PDB, AlphaFold, ColabFold, and AlphaFold Multimer should be cited, as well as sources for the definition of NGPs, IDPs/IDRs, and amyloid structures.

Recommendations: Please insert missing sources and tone down these statements. The recent AI models are impressive, but they are still far from understanding the underlying biology, and we have no certainty about what the future will hold.

(4) Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts, and abbreviations are explained)
Partly.
Justification: The language is accessible and the article is easy to read, but key subject-specific terms and concepts are not explained.
Recommendations: It might be useful to add a glossary explaining key concepts like non-globular proteins, intrinsically disordered regions/proteins, amyloid structures, and Anfinsen's principle, making the article accessible to people outside of the fields of bioinformatics and biology.

(5) Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Partly.
Justification: The authors explain how research on non-globular proteins is still needed, which they address(ed) with their COST actions. However, this is not the only field of protein research that could still profit from AI models. Further, as previously pointed out, the article is mostly one-

sided about applying AI to biology; it does not suggest usage guidelines or spark a debate.
Recommendations: It might be useful to add some caveats and some more critical points to stimulate constructive debate within the research community.

(6) Other comments and recommendations

1. The format of the letter is confusing to us. Eight of the eleven authors are directly quoted - does this mean the remaining three conducted interviews with them, making the piece resemble a feature article?
   Recommendation: If this is indeed the case, it would be helpful to state it explicitly. If not, we would suggest removing the direct quotations and instead integrating these statements into the main text, since any content not otherwise attributed can reasonably be understood as representing the authors' own views.

2. The article contains repetitive parts, while not going into enough detail for more complicated concepts. The amazing powers of AI are repeatedly but vaguely stated without examples (*AI catalysed advances in machine learning, bioinformatics, and structural biology; AI uncovers hidden patterns; AI is decoding the world of proteins; AlphaFold has profound implications for biomedicine, biochemistry, and biotechnology; AI advances complex problems in biology; with AI, we can solve some of the most challenging questions in biology; AI is a transformative source*).
   Recommendation: It would be helpful to provide some examples for the transformative powers of AI that are understandable to people outside the field of bioinformatics. What exactly can or was done now that was not possible before? Which concrete success stories were there? Then, these repeated sections could be condensed into one paragraph, freeing up some space to go into more details for, e.g., the glossary.

3. The transition from the Nobel prizes of 2024 to non-globular proteins, intrinsically disordered proteins, and the COST actions is not made very smoothly. Non-globular and intrinsically disordered proteins and amyloid aggregation are mentioned at different places throughout the text, but not followed up upon directly.
   Recommendation: Instead of just repeatedly mentioning NGPs, it could be beneficial to dedicate a paragraph or info box to them: what is this class of proteins, why are they hard to determine experimentally and, hence, harder to predict, and why are they important to study? Then, the transition to the COST actions would become smoother, too. This information could also be worked into an explicit paragraph about limitations of AlphaFold where the concept of intrinsically disordered regions and amyloid structures could be elaborated, too.

4. There are some minor semantical/grammatical/formal issues:
   1. *"In the end, AI-based prediction models are improving [... |, being routinely used in [...] workflows"*: This sentence is not grammatically correct.
   2. *"[...] and the disordered nature of regions thereof"* → some of their regions?
   3. *"**the three-dimensional structure of proteins**"* is the only bold part of the text - please remove this.
   4. *"The impressive wealth of three-dimensional protein models significantly enhances awareness of the relevance of protein structure in biological research."*: This statement does not make sense.
   5. *"says Prof. Miguel Andrade"* - in the author list, the surname is stated as Andrade-

Navarro.
6. Make sure that the abbreviations are introduced the first time the long form is mentioned. This is, e.g., not the case for NGPs.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
No

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Partly

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Partly

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, computational biology, machine learning, implications of data leakage, protein-protein interaction prediction, alternative splicing, drug response prediction, cell-type deconvolution

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 27 August 2025

https://doi.org/10.21956/openreseurope.22318.r57516

© 2025 Bordin N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Nicola Bordin

University College London, London, UK

In their open letter, Longhi and colleagues outline the dramatic impact artificial intelligence had on the fields of protein structure and computational biology, leading to long-lasting implications on research in the larger field of molecular biology.

I agree with all opinions on both merits and limitations of the tools highlighted, particularly the need for further research on non-globular proteins, amyloids and the process of folding.

If the authors are not constrained by word limits, this letter would benefit by mentioning another key factor that changed the field of computational biology concurrently with AlphaFold and protein design: protein language models. pLMs, in conjunction with these novel structure-based predictors are powering the research behind the next ten years of discoveries and should be mentioned as a key aspect of the "ML" part of "ML4NGP".

I have no further comments.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Protein Bioinformatics, AI, Globular Proteins, algorithm design for protein structure, sequence and function.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.