

Perspective

# Structural Biology in the AlphaFold Era: How Far Is Artificial Intelligence from Deciphering the Protein Folding Code?

Nicole Balasco <sup>1,†</sup>, Luciana Esposito <sup>2,†</sup>  and Luigi Vitagliano <sup>2,\*</sup> 

<sup>1</sup> Institute of Molecular Biology and Pathology, National Research Council (CNR), c/o Department Chemistry, Sapienza University of Rome, 00185 Rome, Italy; nicole.balasco@cnr.it

<sup>2</sup> Institute of Biostructure and Bioimaging, Department of Biomedical Sciences, National Research Council (CNR), 80131 Naples, Italy; luciana.esposito@cnr.it

\* Correspondence: luigi.vitagliano@cnr.it

† These authors contributed equally to this work.

**Abstract:** Proteins are biomolecules characterized by uncommon chemical and physicochemical complexities coupled with extreme responsiveness to even minor chemical modifications or environmental variations. Since the shape that proteins assume is fundamental for their function, understanding the chemical and structural bases that drive their three-dimensional structures represents the central problem for an atomic-level interpretation of biology. Not surprisingly, this question has progressively become the Holy Grail of structural biology (the folding problem). From this perspective, we initially describe and discuss the different formulations of the folding problem. In the present manuscript, the folding problem is framed from a historical perspective, effectively highlighting the progress made in the last lustrum. We chronologically summarize the major contributions that traditional methodologies provide in approaching this multifaceted problem. We then describe the recent advent and evolution of predictive approaches based on machine learning techniques that are revolutionizing the field by pointing out the potentialities and limitations of this approach. In the final part of the perspective, we illustrate the contribution that computational approaches will make in current structural biology to overcome the limitations of the reductionist approach of studying individual molecules to afford the atomic-level characterization of entire cellular compartments.

**Keywords:** protein structure predictions; sequence–structure paradigm; sequence–stability relationships; CASP



Academic Editors: Philippe Urban and Adrián Velázquez Campoy

Received: 31 March 2025

Revised: 24 April 2025

Accepted: 2 May 2025

Published: 6 May 2025

**Citation:** Balasco, N.; Esposito, L.; Vitagliano, L. Structural Biology in the AlphaFold Era: How Far Is Artificial Intelligence from Deciphering the Protein Folding Code? *Biomolecules* **2025**, *15*, 674. <https://doi.org/10.3390/biom15050674>

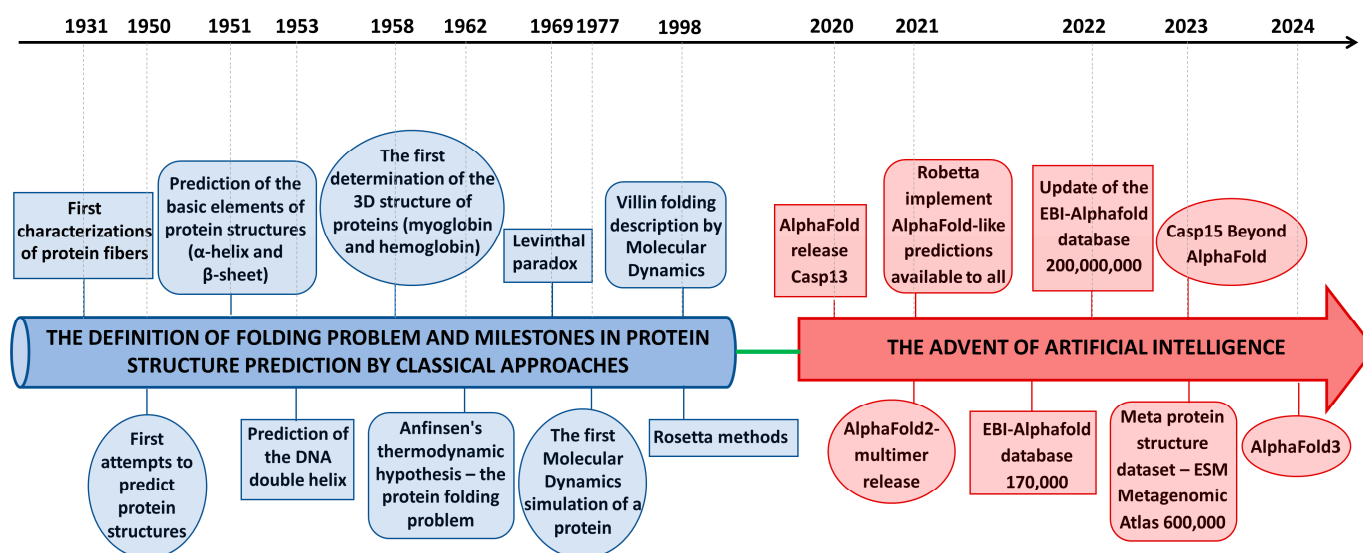
**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. The Central Role of Proteins in Life and Their Extraordinary Chemical and Structural Complexity

Life, so far detected uniquely on Earth, is an extraordinarily intricate phenomenon based on molecules endowed with uncommon chemical and physicochemical complexities. Interestingly, the atoms constituting these biologically active (macro)molecules present intricate three-dimensional organizations susceptible to even minor chemical modifications or environmental variations. The responsiveness of biomolecules to external stimuli represents a landmark of biology. In this context, proteins represent prototypical examples. Indeed, these macromolecules, which play fundamental roles in all biological processes, are typically formed by thousands of atoms, and even replacing a few of them may have devastating effects on their functionality. Frequently, the activities of these “molecular giants with feet of clay” may be undermined by minimal modifications to the local environments. Fragility is only the apparent weakness of proteins, as it represents a key factor for their

ability to play crucial roles in the labyrinth of molecular processes that characterize life. Proteins exploit their structural versatility to react appropriately to external stimuli, establish reversible physiological partnerships, or bind/release substrates and products of reactions. Since the shape that proteins assume is fundamental for their function, understanding the basis that drives their three-dimensional structures represents the central problem for an atomic-level interpretation of biology, in such a way that this question has become the Holy Grail of structural biology [1]. The solution to this problem is complicated by many factors, including the extreme chemical complexity of these molecules, the high variability of the structures they may assume, and the dynamic behavior of their three-dimensional organizations. A simple visual inspection of the repertoire of protein shapes reported in the Protein Data Bank (PDB), an open repository for experimentally resolved structures [2], provides an immediate picture of proteins' global structural variability. The idea of the intrinsic complexity of this subject can be clarified through the illustration of the so-called Levinthal paradox [3,4]. By making simple assumptions, C. Levinthal showed that even for a relatively small protein of 100 amino acid residues, the astronomical number of possible conformations (at least  $3^{200}$ ) cannot be explored in the experimental protein folding process even if under physiological conditions each state is sampled at ultrafast speed (e.g., in the order of  $10^{-15}$  s). An obvious consequence of this observation is that the possible conformational states of proteins cannot be exhaustively evaluated using the concepts and methodologies commonly used in structural and computational chemistry.

However, the possibility, at least theoretically, of tackling the protein folding problem has been suggested by the seminal experiments performed by C. Anfinsen and collaborators. They demonstrated that all the information proteins require to assume well-defined structural organizations resides in their chemistry, i.e., the sequence of amino acid residues constituting their primary structure [5–7]. Indeed, they proved that proteins could be refolded in the test tube after removing the chemical species that induced their unfolding, thus showing, at least for the tested cases, that no other cellular component was essential for gaining a native three-dimensional organization. Although several exceptions to this sequence–structure paradigm have been pointed out over the years, it represents a valid working hypothesis widely accepted by the structural biology community. Not surprisingly, the folding problem has become a central structural biology problem [8]. As detailed in the following paragraphs, the definition of this sequence–structure paradigm has also provided conceptual support to the attempts carried out over many decades to predict protein structures using only sequence information. In the present perspective, to better appreciate the importance of recent advancements in the field, the story of the folding problem is illustrated from a historical viewpoint (Figure 1). After briefly describing the so-called folding problem and a rapid excursus of the traditional methods employed to unravel the basis of the sequence–structure relationship, we delineate the impact that machine-based approaches, such as AlphaFold, have on modern structural biology. We highlight these approaches' strengths and weaknesses by discussing how some structure-oriented biases have paradoxically favored their success. Finally, we show how implementing this approach may significantly contribute to overcoming the limitations of reductionist approaches by allowing the simultaneous study of entire protein families. The fruitful integration of the potential of this approach with innovative structural biology techniques will stimulate future atomic-level studies of the functioning of entire cellular compartments.



**Figure 1.** The chronological unfolding of the protein folding story.

## 2. The Protein Folding Code and Its Formulation(s)

The discovery that the amino acid sequence dictates the three-dimensional structure of proteins has generated intense research on the chemical and structural bases of this relationship. Over the years, different views on the mechanism of protein folding have been proposed, based either on pre-defined pathways characterized by well-defined intermediates or on the shape of the energy driving random folding in a reasonable time frame [9]. Deciphering the structural code linking protein sequence and structure is only part of the folding problem, which aims for a comprehensive understanding of the thermodynamic and kinetic factors that drive proteins to assume well-defined three-dimensional organizations. In its most general formulation, elegantly proposed by Dill and coworkers [8], which fully embodies the chemical and structural complexity of proteins, the folding problem deals with three distinct yet related issues. The first is the physical folding code, which focuses on thermodynamic factors, such as the balance of interatomic forces that favor the three-dimensional native structure(s) of a protein. The second issue is unraveling the folding mechanism, i.e., understanding the kinetic aspects of the pathways that allow proteins to fold quickly. The third issue is predicting the protein structure, focusing on computational issues related to determining protein structures using only sequence information. Despite significant advances observed over the years, elucidating these three aspects of the folding problem is largely incomplete.

## 3. From Ab Initio to Empirical Approaches

The first attempts to predict the spatial organization of proteins using (or sometimes misusing) solely chemical information preceded the determination of the first crystallographic structures, as well as the definition of Anfinsen's sequence–structure paradigm. In this scenario, it may not be surprising that the proposed solutions were far from the correct ones [10], even when they were suggested by the founder of X-ray and protein crystallography [11]. The determination of the first protein structures [12,13] unveiled their non-regular arrangements, which were in contrast with the expected symmetrical organizations suggested by their known propensity to form ordered crystals [14].

Initial attempts to determine the three-dimensional structure of proteins using only sequence information were based on investigating their energy landscape without a priori assuming that they could adopt multiple possible structural states. However, since these approaches strongly rely on key and difficult-to-achieve prerequisites, such as the availability

of accurate force fields, the possibility of performing exhaustive conformational samplings, and accurate description of entropic and solvation effects, their success has been extremely limited. On the other hand, integrating limited experimental information with theoretical modeling has been far more effective. The latter approach has been successfully employed, even in the pioneering years of structural biology, by effectively exploiting the limited information collected with X-ray fiber diffraction studies to generate atomic-level accurate models of the basic secondary structure elements of proteins ( $\alpha$ -helices and  $\beta$ -sheets) [15,16], fibrous proteins (e.g., collagen) [17,18], and DNA [19]. Over the years, the progressive release of experimental structures and their collection in the PDB [2] represented a valuable treasure of information needed for designing innovative empirical strategies in which experimental data were used as the basis of computational approaches. Although the complexity of the sequence–structure problem hindered general solutions, most successful approaches were based on empirical modeling. This fact is evident from inspecting the results from the Critical Assessment of Structure Prediction (CASP) (<https://predictioncenter.org/>, accessed on 1 March 2025) [20], which is a biennial protein-folding challenge that employs scrupulous and objective protocols to evaluate proposed structure prediction solutions. In recent decades, the laboratory led by D. Baker has achieved the most significant accomplishments in the field by developing the Rosetta/Robetta package [21]. In this approach, the preferred structures adopted by fragments of the studied protein were initially derived through the mining of the PDB. Then, the fragments were assembled, and the global fold was evaluated based on energetic considerations. This approach proved to be an effective tool that could provide reliable predictions of protein structures in several cases. More importantly, Rosetta has been successfully applied for the de novo design of proteins endowed with many different folding(s) and functions [22]. Compared to the difficulties in predicting the three-dimensional structure of native proteins, the relative ease of designing proteins with desired structures illustrates the tortuous path that the former faced during evolution, being continuously modified and adapted to operate in different biological contexts. Further information on the state of the art of protein structure prediction methods in the pre-AlphaFold era can be found in detailed literature reviews [23–25].

#### 4. The Advent of Machine Learning Techniques: The AlphaFold Revolution

Over the years, computational studies have exploited the increasingly available experimentally generated three-dimensional protein structures carefully collected and curated in the PDB. They were used to extract the information needed for developing algorithms or as a benchmark for validating the proposed approaches. Applying machine learning to problems in modeling biological systems is over three decades old [26,27]. Considering the commonly used hierarchical description of protein structures, initial studies dealt with predicting protein secondary structures, extensively using deep neural network-based architectures in combination with evolutionary information.

The first successful model using deep learning dates to 1996 when the PHD (profile fed neural network systems from Heidelberg) protein secondary structure prediction method was made available [28]. Since then, the most successful secondary structure predictions have been based on machine learning techniques [29]. The expansion of the available protein structures in the PDB, owing to impressive methodological and technological advances in the appropriate experimental techniques, has enabled the development of ambitious predictive approaches based on machine learning.

DeepMind developed a deep-learning approach based on a specific convolutional neural network known as AlphaFold (AF) by exploiting a vast collection of protein structural data compiled over six decades [30]. Training of AF encompassed combined data

from experimental PDB structures and features derived from multiple sequence alignments (MSAs) to predict distances between pairs of atoms. The first version of AF outperformed the other prediction methods by a significant margin during its first CASP event (CASP13, held in 2018) [31,32]. Indeed, AF achieved a summed z-score of 58.2 in predicting the free modeling structures, i.e., those in which no homologous structure was available, whereas the second-best prediction algorithm reached only 36.6. The impressive performance of AF was announced by the most prestigious scientific journal [33,34] and by the general audience media as an achievement that revolutionized structural biology. However, the real impact of AF on the everyday life of structural biologists at this stage was limited by the lack of user-friendly tools that would permit the replication of these astonishing results with other proteins. The revolution was completed in 2021 with the simultaneous release of AlphaFold2 (AF2) [35] and a novel machine learning-based program developed by D. Baker's group [36]. The latter intentionally replicated the AF protocol to make the approach freely available and implemented it in easy-to-use packages (RoseTTAFold in <https://robetta.bakerlab.org/>, accessed on 1 March 2025).

In the CASP14 competition [37], AF2 was, again, very successful when it blew away its competitors. This approach correctly predicted most of the target protein structures based only on their sequences. The AF2 best-predicted structures presented a median backbone accuracy of 0.96 Å in terms of root mean square deviation (RMSD) on C $\alpha$  atoms compared with the corresponding experimental ones. Notably, the second-best algorithm reached an accuracy of only 2.8 Å in terms of RMSD. Besides the high accuracy, the AF2 model also provided excellent self-evaluation criteria to estimate the reliability of the generated models, thus enabling the confident use of the predictions. This initiative was followed by the release of the AlphaFold Protein Structure Database [38], jointly developed by Google DeepMind and EMBL-EBI (<https://alphafold.ebi.ac.uk/>, accessed on 1 March 2025), at present reporting the three-dimensional structures of the individual polypeptide chains of approximately 200 million proteins cataloged in the UniProt Knowledgebase (<https://www.uniprot.org/>, accessed on 1 March 2025). AF2 no longer employs convolutional neural networks, but a new deep learning architecture, the Transformers, which has been extensively used in the natural language processing field. Moreover, throughout the whole network, this new AF version reinforces the notion of iterative refinement, repeatedly applied to the architecture modules, a feature related to computer vision approaches that remarkably contributed to the accuracy and reduction of training time. Important extensions of AF2 encompass the possibility of generating accurate three-dimensional models of protein homo- and hetero-complexes [39]. Importantly, user-friendly protocols freely available to the community were implemented in Colab [40].

After the appearance of AF, protein structure prediction has experienced tremendous growth with several new methods and applications. At the end of 2022, Meta AI's team, using a different language model architecture (ESMFold) [41] without employing MSAs, generated a structural database for more than 600 million metagenomic proteins. ESMFold is considerably faster than other predictors. The database known as Evolutionary Scale Modelling (ESM) Metagenomic Atlas (<https://esmatlas.com>, accessed on 1 March 2025) includes more than 225 million protein structures predicted with high confidence. The release of AF2 has stimulated the development of several variants by different research groups, often designed to solve specific problems. Interestingly, in the CASP15 competition held in 2022, though the most successful groups developed their own ad hoc tools, all of these were based on AF2, at least partly [42,43].

A further step ahead in this ongoing process is the latest release of AlphaFold3 (AF3), in 2024, with its powerful implementations [44]. AF3 allows, for the first time, not only a prediction of polypeptide chains but also complexes of proteins with other proteins,



nucleic acids, small molecules, ions, and modified residues, which can all be predicted with remarkable accuracy. It should be mentioned, however, that predictions of RNA structures by artificial intelligence (AI)-based approaches could not repeat the success obtained with protein structures [45].

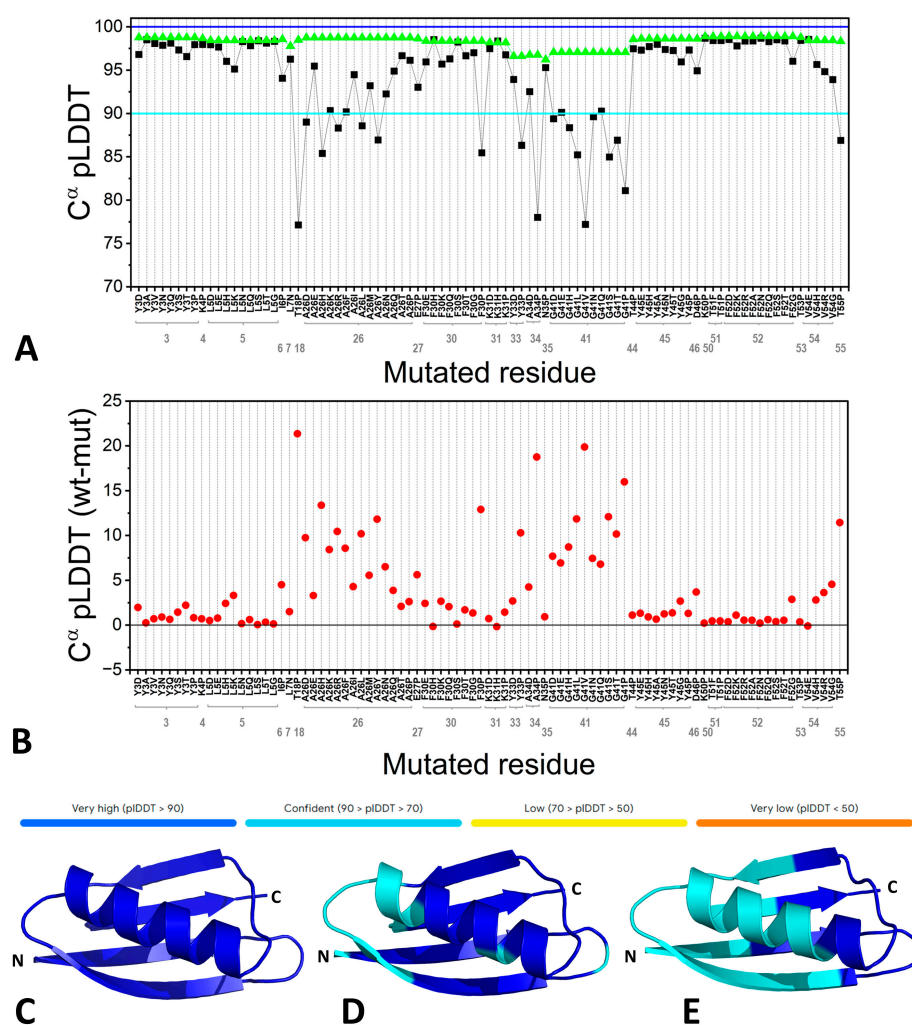
## 5. AlphaFold and the Folding Problem

It is virtually impossible to summarize the impressive impact of AF and its successors in the life sciences [30,35,46–53]. Several tens of thousands of articles have cited original papers reporting this approach. According to its developers, AF has been used by 2 million researchers in over 190 countries (<https://deepmind.google/technologies/alphafold/impact-stories/>, accessed on 1 March 2025). Despite many limitations of the approach [54], interrogating the AF and related servers has become an obligatory step in almost every structural biology study. However, a legitimate question is: How far are we from a deep understanding of protein structural properties?

The molecular complexity of proteins and the necessity to fine-tune their function make the definition of a single or a few structural state(s) an important, yet not definitive, step forward. Indeed, complex dynamic behavior, which is highly diverse in the protein realm, is a fundamental aspect about which AF provides little information. Suppose the same concerns are put forward from the perspective of the folding problem, which, in its most complete formulation (see also Section 2), encompasses not only an understanding of the thermodynamics and mechanism of protein folding, but also the prediction of the protein's three-dimensional structure. The current data suggest that AF and related approaches are close to achieving only the latter goal. While the molecular mechanism underlying protein folding is out of the scope of AF, the accurate prediction of the folded state structure by AF occurs without accurately addressing the related thermodynamic aspects. Indeed, it is well known that AF cannot reproduce the destabilizing effects that single-point mutations induce in the global structure of proteins.

To further illustrate how far AF stretches in replicating these effects, we consider the protein Gbeta1, a model system for studying protein structure–stability relationships. Previous studies elaborated on the effects of single-point mutations on protein stability and did so for every amino acid residue in the protein [55]. Highly destabilizing mutations preventing the expression and characterization of the protein were identified. According to our predictions, AF provided native-like structures for models simultaneously incorporating up to eight mutations (Figure 2C–E and Table 1), where each could completely destabilize the protein structure in the experiments.

The bias toward folded states of AF is well illustrated by the case of human hemoglobin, whose well-known tetrameric structure strongly relies on the heteromeric association of different  $\alpha$  and  $\beta$  chains and the presence of the iron-coordinated heme group. As shown in Figure 3, the structures of the individual chains predicted by AF are identical to those detected in tetrameric native human hemoglobin (HbA). Although these results are generally known to the community of AF users, it is somehow overlooked that the inability of AF to catch the real thermodynamics of the folding process has contributed to its success. Indeed, many, if not most, of the protein structures in the AlphaFold DataBase, generated without taking into account the oligomeric state of proteins or the presence of a prosthetic group (such as the heme) or other stabilizing agents, are thermodynamically unstable despite their similarity to experimental native states.



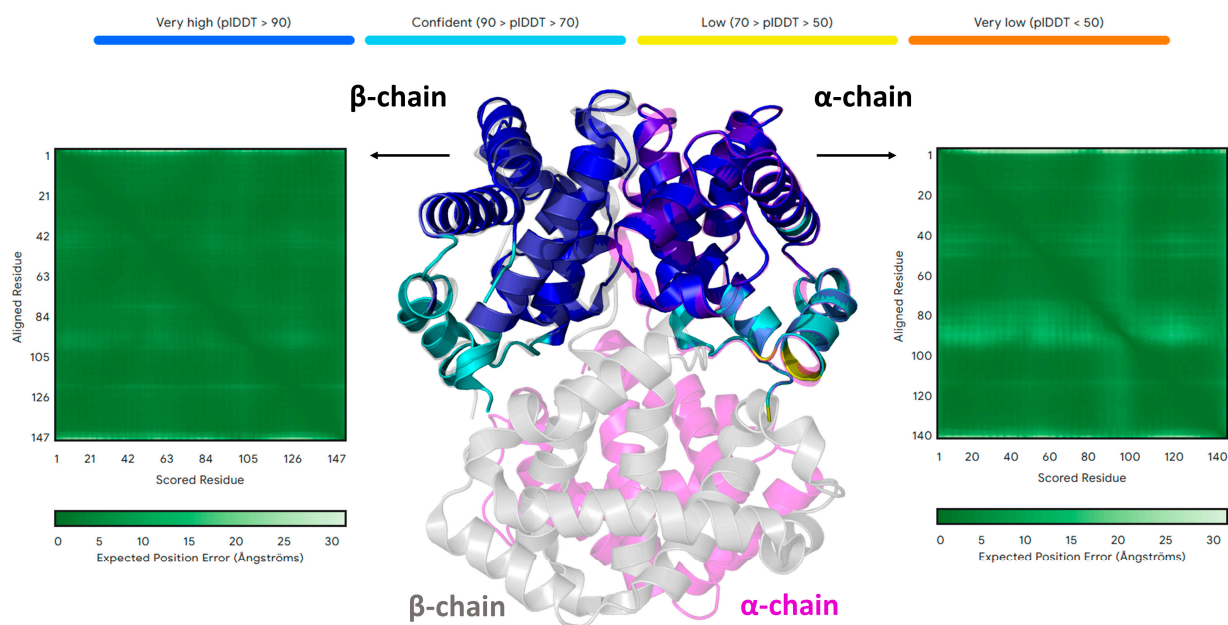
**Figure 2.** The inability of AF to predict the effects of highly destabilizing mutations in the Gbeta1 protein. Predictions were evaluated using the same parameter employed by AF for self-assessment, namely the per-residue local distance difference test (pLDDT) [56], whose values are reported as horizontal colored bars. (A) pLDDT scores for the C $\alpha$  atoms of specific residues of the wild-type protein (wt, green triangles) and single-point mutants (mut, black squares). (B) Difference in pLDDT values between wt and single-point mutants. Cartoon representation of the AF3-predicted models of (C) wt Gbeta1, (D) a mutant protein carrying six destabilizing mutations (Y3A, L5D, A26P, F30G, Y45G, and F52G), and (E) a mutant protein carrying eight destabilizing mutations (Y3A, L5D, A26P, F30G, G41P, Y45G, F52G, and V54G). The models are colored according to the AF pLDDT metric.

**Table 1.** Per-residue local distance difference test (pLDDT) [56] (see Figure 2 for the metric) values for the C $\alpha$  atoms of specific residues in the wild-type Gbeta1 protein, in the variants bearing the single-point mutations, and in the mutants incorporating either six (hexa-mutant) or eight (octa-mutant) of these highly destabilizing mutations.

Residue	pLDDT	Mutation	pLDDT in the Single-Point Mutants	pLDDT in the Hexa-Mutant	pLDDT in the Octa-Mutant
Y3	98.76	Y3A	98.51	90.01	89.11
L5	98.43	L5D	97.93	92.94	90.24
A26	98.75	A26P	96.13	89.03	76.83
F30	98.35	F30G	97.00	90.71	86.31
G41	97.06	G41P	81.09	-	93.96

Table 1. Cont.

Residue	pLDDT	Mutation	pLDDT in the Single-Point Mutants	pLDDT in the Hexa-Mutant	pLDDT in the Octa-Mutant
Y45	98.62	Y45G	95.94	92.23	86.23
F52	98.89	F52G	96.02	93.10	88.60
V54	98.44	V54G	93.90	-	92.43



**Figure 3.** The peculiar ability of AF to predict the structure of individual protein chains without key structural elements, such as prosthetic groups and the oligomeric environment: the case of human hemoglobin (HbA). The AF3-predicted models of the individual HbA chains (central panel), colored according to the AF pLDDT values (upper bar), are superimposed on a couple of the  $\alpha$  (magenta) and  $\beta$  (gray) chains of the HbA tetramer from the high-resolution crystallographic structure (PDB entry 2dn2). The associated predicted aligned error (PAE) matrices for the AF3 models are also shown (left and right panels).

## 6. Conclusions and Perspectives for Structural Biology and Beyond

In the last seven decades, structural biology has experienced several breakthroughs. In its initial stage, structural studies were limited to naturally highly abundant proteins amenable to crystallization. The possibility of producing recombinant proteins has expanded structural biology studies to proteins of extreme biological relevance, which are barely expressed *in vivo*. Despite spectacular technical and methodological advancements that have significantly increased the pace of solving protein structures in the last few decades, the experimental determination of protein structures is still a bottleneck, i.e., a slow and laborious process with no guaranteed success. The development of effective predictive approaches has completely changed the scenario. While a constantly growing community of several thousand experimental structural biologists took nearly sixty years to determine ~200,000 structures of proteins, individual research groups were able to release hundreds of millions of structures in just a few years. Although these structures need validation, they present a reasonable degree of confidence and accuracy globally. This has opened avenues for other applications aimed at clarifying the role of pathogenic mutations [57] or for the structural interpretation of large-scale interactomes [58]. It is worth noting that the rapid determination of protein structures, starting from their sequences,



facilitated by these approaches, allows the simultaneous elucidation of structure–function relationships in entire families, as we have recently shown for KCTD proteins, for which atomic-level structural features were reported for all 25 members of the family [59–61].

Clearly, the ability to predict protein structures from their sequences does not diminish the importance of experimental techniques, especially since the latter can yield results that predictive methods cannot achieve [62,63]. In fact, experimental approaches continue to play a crucial role in defining biological processes at the atomic level. Computational methods, on the other hand, cannot (a) achieve the high accuracy of some crystallographic structures [64], (b) address highly complex systems like those investigated by cryo-electron microscopy [65,66], or (c) provide dynamic properties as NMR does [67].

In general, the availability of these predictive methods is an additional opportunity to eliminate the generally used reductionist approaches (which are focused on the characterization of individual biomolecules) and pursue an atomic-level view of life. In the future, integrating experimental and computational methodologies, including some in their early developmental stages, such as tomography, will likely produce atomic-level visualization of entire cellular compartments and their mutual interactions.

In conclusion, the release of AF represents a further step in pursuing an atomic-level view of life. As also indicated by the Nobel Prize in Chemistry awarded in 2024 (<https://www.nobelprize.org/prizes/chemistry/>, accessed on 1 March 2025), AF currently represents the most successful application of AI to a scientific problem (see also <https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever/>, accessed on 1 March 2025). The impact that this approach is having as a solution to the long-standing problem of determining protein three-dimensional structures from their sequence is revolutionary. However, it should be noted that due to the intrinsic nature of AI, its success is not due to conceptual advancement and has not hitherto provided new intellectual interpretive models for the scientific community. If these considerations are placed in Kuhn’s framework of scientific revolution [68], AF release is a revolution without any paradigm change. Instead of “providing model problems and solutions for a community of practitioners” [68], it is a rather effective tool for solving a fundamental scientific problem. The ability of AF to predict protein structures from their sequences reinforces the old folding paradigm and represents a tool revolution that, however, cannot be considered a mere methodological advancement. Hopefully, the treasure of information generated by this methodology will generate new conceptual paradigms that will help humans understand nature. The irresistible advent of AI in science suggests that other revolutions of this type will soon appear.

**Author Contributions:** N.B.: formal analysis, investigation, writing—review and editing; L.E.: formal analysis, investigation, writing—review and editing; L.V.: conceptualization, validation, writing—original draft. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to thank Maurizio Amendola, Luca De Luca, Massimiliano Mazzocchi, and Giorgio Varriale for their technical support. The CINECA award under the ISCRA initiative (ISCRA C project AF-Koli ID HP10C52U80) is also acknowledged for the availability of high-performance computing resources and support.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Pearce, R.; Zhang, Y. Toward the Solution of the Protein Structure Prediction Problem. *J. Biol. Chem.* **2021**, *297*, 100870. [[CrossRef](#)] [[PubMed](#)]
- Berman, H.M.; Burley, S.K. Protein Data Bank (PDB): Fifty-Three Years Young and Having a Transformative Impact on Science and Society. *Quart. Rev. Biophys.* **2025**, *58*, e9. [[CrossRef](#)] [[PubMed](#)]
- Levinthal, C. Are There Pathways for Protein Folding? *J. Chim. Phys.* **1968**, *65*, 44–45. [[CrossRef](#)]
- Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's Paradox. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 20–22. [[CrossRef](#)]
- Anfinsen, C.B.; Haber, E.; Sela, M.; White, F.H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1309–1314. [[CrossRef](#)]
- Anfinsen, C.B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230. [[CrossRef](#)]
- Anfinsen, C.B.; Scheraga, H.A. Experimental and Theoretical Aspects of Protein Folding. In *Advances in Protein Chemistry*; Elsevier: Amsterdam, The Netherlands, 1975; Volume 29, pp. 205–300. ISBN 978-0-12-034229-7.
- Dill, K.A.; Ozkan, S.B.; Shell, M.S.; Weikl, T.R. The Protein Folding Problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316. [[CrossRef](#)]
- Rumbley, J.; Hoang, L.; Mayne, L.; Englander, S.W. An Amino Acid Code for Protein Folding. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 105–112. [[CrossRef](#)]
- Wrinch, D.M. The Cyclol Hypothesis and the “Globular” Proteins. *Proc. R. Soc. Lond. A* **1937**, *161*, 505–524. [[CrossRef](#)]
- Bragg, W.L.; Kendrew, J.C.; Perutz, M.F. Polypeptide Chain Configurations in Crystalline Proteins. *Proc. R. Soc. Lond. A* **1950**, *203*, 321–357. [[CrossRef](#)]
- Kendrew, J.C.; Bodo, G.; Dintzis, H.M.; Parrish, R.G.; Wyckoff, H.; Phillips, D.C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662–666. [[CrossRef](#)] [[PubMed](#)]
- Perutz, M.F.; Rossmann, M.G.; Cullis, A.F.; Muirhead, H.; Will, G.; North, A.C.T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185*, 416–422. [[CrossRef](#)] [[PubMed](#)]
- Giegé, R. A Historical Perspective on Protein Crystallization from 1840 to the Present Day. *FEBS J.* **2013**, *280*, 6456–6497. [[CrossRef](#)]
- Pauling, L.; Corey, R.B.; Branson, H.R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205–211. [[CrossRef](#)] [[PubMed](#)]
- Pauling, L.; Corey, R.B. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 729–740. [[CrossRef](#)]
- Ramachandran, G.N.; Kartha, G. Structure of Collagen. *Nature* **1955**, *176*, 593–595. [[CrossRef](#)]
- Rich, A.; Crick, F.H.C. The Structure of Collagen. *Nature* **1955**, *176*, 915–916. [[CrossRef](#)]
- Watson, J.D.; Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)]
- Moult, J.; Pedersen, J.T.; Judson, R.; Fidelis, K. A Large-scale Experiment to Assess Protein Structure Prediction Methods. *Proteins* **1995**, *23*, ii–iv. [[CrossRef](#)]
- Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363–382. [[CrossRef](#)]
- Leman, J.K.; Weitzner, B.D.; Lewis, S.M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R.F.; Aprahamian, M.; Baker, D.; Barlow, K.A.; Barth, P.; et al. Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nat. Methods* **2020**, *17*, 665–680. [[CrossRef](#)] [[PubMed](#)]
- Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 681–697. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Y. Progress and Challenges in Protein Structure Prediction. *Curr. Opin. Struct. Biol.* **2008**, *18*, 342–348. [[CrossRef](#)]
- Ginalski, K. Comparative Modeling for Protein Structure Prediction. *Curr. Opin. Struct. Biol.* **2006**, *16*, 172–177. [[CrossRef](#)]
- Rawlings, C.J.; Fox, J.P. Artificial Intelligence in Molecular Biology: A Review and Assessment. *Phil. Trans. R. Soc. Lond. B* **1994**, *344*, 353–363. [[CrossRef](#)]
- Sternberg, M.J.; King, R.D.; Lewis, R.A.; Muggleton, S. Application of Machine Learning to Structural Molecular Biology. *Phil. Trans. R. Soc. Lond. B* **1994**, *344*, 365–371. [[CrossRef](#)]
- Rost, B. PHD: Predicting One-Dimensional Protein Structure by Profile-Based Neural Networks. In *Methods in Enzymology*; Elsevier: Amsterdam, The Netherlands, 1996; Volume 266, pp. 525–539. ISBN 978-0-12-182167-8.
- Pirovano, W.; Heringa, J. Protein Secondary Structure Prediction. In *Data Mining Techniques for the Life Sciences*; Carugo, O., Eisenhaber, F., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2010; Volume 609, pp. 327–348. ISBN 978-1-60327-240-7.
- Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577*, 706–710. [[CrossRef](#)] [[PubMed](#)]
- AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35*, 4862–4865. [[CrossRef](#)]

32. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Protein Structure Prediction Using Multiple Deep Neural Networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins* **2019**, *87*, 1141–1148. [\[CrossRef\]](#)
33. Callaway, E. 'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures. *Nature* **2020**, *588*, 203–204. [\[CrossRef\]](#)
34. Service, R.F. 'The Game Has Changed.' AI Triumphs at Protein Folding. *Science* **2020**, *370*, 1144–1145. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [\[CrossRef\]](#)
36. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Applying and Improving ALPHAFOLD at CASP14. *Proteins* **2021**, *89*, 1711–1721. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [\[CrossRef\]](#)
39. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein Complex Prediction with AlphaFold-Multimer 2021. *bioRxiv* **2021**. [\[CrossRef\]](#)
40. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19*, 679–682. [\[CrossRef\]](#)
41. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379*, 1123–1130. [\[CrossRef\]](#)
42. Genc, A.G.; McGuffin, L.J. Beyond AlphaFold2: The Impact of AI for the Further Improvement of Protein Structure Prediction. In *Prediction of Protein Secondary Structure*; Kloczkowski, A., Kurgan, L., Faraggi, E., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2025; Volume 2867, pp. 121–139. ISBN 978-1-07-164195-8.
43. Liu, J.; Guo, Z.; Wu, T.; Roy, R.S.; Chen, C.; Cheng, J. Improving AlphaFold2-Based Protein Tertiary Structure Prediction with MULTICOM in CASP15. *Commun. Chem.* **2023**, *6*, 188. [\[CrossRef\]](#)
44. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [\[CrossRef\]](#)
45. Kwon, D. RNA Function Follows Form—Why Is It so Hard to Predict? *Nature* **2025**, *639*, 1106–1108. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature* **2021**, *596*, 590–596. [\[CrossRef\]](#)
47. Jumper, J.; Hassabis, D. Protein Structure Predictions to Atomic Accuracy with AlphaFold. *Nat. Methods* **2022**, *19*, 11–12. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Baek, M.; Baker, D. Deep Learning and Protein Structure Modeling. *Nat. Methods* **2022**, *19*, 13–14. [\[CrossRef\]](#)
49. Barrio-Hernandez, I.; Yeo, J.; Jänes, J.; Mirdita, M.; Gilchrist, C.L.M.; Wein, T.; Varadi, M.; Velankar, S.; Beltrao, P.; Steinegger, M. Clustering Predicted Structures at the Scale of the Known Protein Universe. *Nature* **2023**, *622*, 637–645. [\[CrossRef\]](#)
50. Akdel, M.; Pires, D.E.V.; Pardo, E.P.; Jänes, J.; Zalevsky, A.O.; Mészáros, B.; Bryant, P.; Good, L.L.; Laskowski, R.A.; Pozzati, G.; et al. A Structural Biology Community Assessment of AlphaFold2 Applications. *Nat. Struct. Mol. Biol.* **2022**, *29*, 1056–1067. [\[CrossRef\]](#)
51. Burke, D.F.; Bryant, P.; Barrio-Hernandez, I.; Memon, D.; Pozzati, G.; Shenoy, A.; Zhu, W.; Dunham, A.S.; Albanese, P.; Keller, A.; et al. Towards a Structurally Resolved Human Protein Interaction Network. *Nat. Struct. Mol. Biol.* **2023**, *30*, 216–225. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Al-Janabi, A. Has DeepMind's AlphaFold Solved the Protein Folding Problem? *BioTechniques* **2022**, *72*, 73–76. [\[CrossRef\]](#)
53. Ahdriz, G.; Bouatta, N.; Floristean, C.; Kadyan, S.; Xia, Q.; Gerecke, W.; O'Donnell, T.J.; Berenberg, D.; Fisk, I.; Zanichelli, N.; et al. OpenFold: Retraining AlphaFold2 Yields New Insights into Its Learning Mechanisms and Capacity for Generalization. *Nat. Methods* **2024**, *21*, 1514–1524. [\[CrossRef\]](#)
54. Wang, L.; Wen, Z.; Liu, S.-W.; Zhang, L.; Finley, C.; Lee, H.-J.; Fan, H.-J.S. Overview of AlphaFold2 and Breakthroughs in Overcoming Its Limitations. *Comput. Biol. Med.* **2024**, *176*, 108620. [\[CrossRef\]](#)
55. Nisthal, A.; Wang, C.Y.; Ary, M.L.; Mayo, S.L. Protein Stability Engineering Insights Revealed by Domain-Wide Comprehensive Mutagenesis. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 16367–16377. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. IDDT: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests. *Bioinformatics* **2013**, *29*, 2722–2728. [\[CrossRef\]](#) [\[PubMed\]](#)

57. Cheng, J.; Novati, G.; Pan, J.; Bycroft, C.; Žemgulytė, A.; Applebaum, T.; Pritzel, A.; Wong, L.H.; Zielinski, M.; Sargeant, T.; et al. Accurate Proteome-Wide Missense Variant Effect Prediction with AlphaMissense. *Science* **2023**, *381*, eadg7492. [[CrossRef](#)] [[PubMed](#)]
58. Schmid, E.W.; Walter, J.C. Predictomes, a Classifier-Curated Database of AlphaFold-Modeled Protein-Protein Interactions. *Mol. Cell* **2025**, *85*, 1216–1232.e5. [[CrossRef](#)]
59. Esposito, L.; Balasco, N.; Smaldone, G.; Berisio, R.; Ruggiero, A.; Vitagliano, L. AlphaFold-Predicted Structures of KCTD Proteins Unravel Previously Undetected Relationships among the Members of the Family. *Biomolecules* **2021**, *11*, 1862. [[CrossRef](#)]
60. Esposito, L.; Balasco, N.; Vitagliano, L. AlphaFold Predictions Provide Insights into the Structural Features of the Functional Oligomers of All Members of the KCTD Family. *Int. J. Mol. Sci.* **2022**, *23*, 13346. [[CrossRef](#)]
61. Balasco, N.; Esposito, L.; Smaldone, G.; Salvatore, M.; Vitagliano, L. A Comprehensive Analysis of the Structural Recognition between KCTD Proteins and Cullin 3. *Int. J. Mol. Sci.* **2024**, *25*, 1881. [[CrossRef](#)]
62. Terwilliger, T.C.; Liebschner, D.; Croll, T.I.; Williams, C.J.; McCoy, A.J.; Poon, B.K.; Afonine, P.V.; Oeffner, R.D.; Richardson, J.S.; Read, R.J.; et al. AlphaFold Predictions Are Valuable Hypotheses and Accelerate but Do Not Replace Experimental Structure Determination. *Nat. Methods* **2024**, *21*, 110–116. [[CrossRef](#)]
63. Read, R.J.; Baker, E.N.; Bond, C.S.; Garman, E.F.; Van Raaij, M.J. AlphaFold and the Future of Structural Biology. *Acta Crystallogr. D Struct. Biol.* **2023**, *79*, 556–558. [[CrossRef](#)]
64. Bijak, V.; Szczygiel, M.; Lenkiewicz, J.; Gucwa, M.; Cooper, D.R.; Murzyn, K.; Minor, W. The Current Role and Evolution of X-Ray Crystallography in Drug Discovery and Development. *Expert. Opin. Drug Discov.* **2023**, *18*, 1221–1230. [[CrossRef](#)]
65. Bai, X.; McMullan, G.; Scheres, S.H.W. How Cryo-EM Is Revolutionizing Structural Biology. *Trends Biochem. Sci.* **2015**, *40*, 49–57. [[CrossRef](#)] [[PubMed](#)]
66. Callaway, E. The Revolution Will Not Be Crystallized: A New Method Sweeps through Structural Biology. *Nature* **2015**, *525*, 172–174. [[CrossRef](#)] [[PubMed](#)]
67. Markwick, P.R.L.; Malliavin, T.; Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput. Biol.* **2008**, *4*, e1000168. [[CrossRef](#)] [[PubMed](#)]
68. Kuhn, T.S.; Hacking, I. *The Structure of Scientific Revolutions*, 4th ed.; University of Chicago Press: Chicago, IL, USA, 2012; ISBN 978-0-226-45811-3.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.