



## OPEN ACCESS

## EDITED BY

Andrei Surguchov,  
University of Kansas Medical Center,  
United States

## REVIEWED BY

Irina G. Sourgoutcheva,  
University of Kansas Medical Center,  
United States  
Jie Hou,  
Saint Louis University, United States

## \*CORRESPONDENCE

Panagiotis Vlamos  
✉ vlamos@ionio.gr

RECEIVED 17 October 2023

ACCEPTED 19 December 2023

PUBLISHED 05 January 2024

## CITATION

Krokidis MG, Dimitrakopoulos GN,  
Vrahatis AG, Exarchos TP and Vlamos P (2024)  
Challenges and limitations in computational  
prediction of protein misfolding in  
neurodegenerative diseases.  
*Front. Comput. Neurosci.* 17:1323182.  
doi: 10.3389/fncom.2023.1323182

## COPYRIGHT

© 2024 Krokidis, Dimitrakopoulos, Vrahatis,  
Exarchos and Vlamos. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC  
BY\)](#). The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Challenges and limitations in computational prediction of protein misfolding in neurodegenerative diseases

Marios G. Krokidis, Georgios N. Dimitrakopoulos,  
Aristidis G. Vrahatis, Themis P. Exarchos and Panagiotis Vlamos\*

Bioinformatics and Human Electrophysiology Laboratory, Department of Informatics, Ionian University, Corfu, Greece

## KEYWORDS

protein misfolding, structure prediction, neurodegeneration, fold recognition, deep learning algorithms

## Introduction

Proteins serve as the primary functional agents within biological systems and play an integral role in almost every aspect of cellular processes. Proteins are constructed as polymers, comprising monomers or smaller constituent units, known as amino acids. Life employs a repertoire of 20 distinct amino acids as the fundamental building blocks for the synthesis of proteins. The peptide chain possesses all the necessary covalent bonds meticulously formed within its structure (Anfinsen, 1973; Alberts et al., 2002). However, in order to fulfill its vital biological role, the peptide chain must assume a precise and distinctive conformation known as the protein's native structure. It is exclusively within this native structure that a protein can execute its designated biological functions. Proteins that deviate from their intended conformation not only fail to perform their proposed functions but may also precipitate severe consequences.

Neurodegenerative diseases frequently entail the misfolding and aggregation of particular proteins, leading to the formation of abnormal and harmful structures (Park et al., 2017). Addressing this misfolding as a therapeutic strategy presents distinct challenges in drug discovery and development (Scannevin, 2018). These challenges stem from various factors, including the dynamic nature of the protein species involved and the ambiguity surrounding which forms of a particular disease protein (such as monomers, oligomers, or insoluble aggregates) primarily contribute to cellular toxicity. Over the long term, neurodegenerative disease proteins invariably lead to synaptic dysfunction, loss, and eventual neuronal cell death (Surguchov and Surguchev, 2022). The precise mechanisms by which diverse misfolded disease proteins trigger neurotoxicity remain uncertain, and these mechanisms appear to vary depending on the particular protein species involved (Wilson et al., 2023). Misfolded disease proteins are believed to primarily exert their effects through toxic gain-of-function and/or dominant-negative effects, although instances of loss-of-function effects have also been documented. For gain-of-function mechanism a few examples could be included neurotoxic signaling, synaptic deficits, impairment of proteasomal or lysosomal degradation and axonal transport while increased vulnerability to stress, mitochondrial dysfunction and impairment of synaptic dynamic have been linked with to loss-of-function mechanism (Winklhofer et al., 2008). Proteostasis disruption is a common feature which characterizes misfolding disease proteins (Cuanalo-Contreras et al., 2013). Loss of protein functionality arises from early degradation, mislocalization or aggregation such as aggregation-prone proteins amyloid-beta, prion protein, a-synuclein

and tau. A $\beta$  oligomers are able to modulate synaptic transmission (Zhang et al., 2022), mutations in tau have been associated with microtubule destabilization and impaired axonal transport of substances (Mietelska-Porowska et al., 2014) and alpha-synuclein inhibits mitochondrial protein import machinery (Lurette et al., 2023). In Parkinson's disease the  $\beta$ -strand segments ( $\beta$ 1 and  $\beta$ 2) of  $\alpha$ -synuclein which involved in interactions within amyloid fibrils were detected using AlphaFold2 and all-atom MD simulation (Rani et al., 2023). Focusing on human Alsin, a protein implicated in a rare neurological disorder called infantile-onset ascending hereditary spastic paralysis (IAHSP), a computational strategy based on AI-based protein structure databases was performed including structural information derived from AlphaFoldDB (Sebastiano et al., 2022). MOVA, a new in silico method for predicting variant pathogenicity using AlphaFold2 was developed and positional information for structural variants were extracted analyzing 12 ALS-causative genes (Hatano et al., 2023). Structural states of functional oligomers of all members of the KCTD family which are proteins containing a (K) Potassium Channel Tetramerization Domain with highly involvement in neurological and neurodevelopmental processes was performed (Esposito et al., 2022).

## The thermodynamic hypothesis of protein folding

A hallmark feature of a living system is the capacity for even its most intricate molecular components to self-assemble accurately and reliably. Protein folding can be broadly dissected into three interrelated facets: Folding process, mechanical aspect of folding and predicting native structures (Bryngelson et al., 1995; Dill et al., 2008). The first one encompasses inquiries into the kinetics of peptide chain folding, including the rate at which it occurs and the intermediate structural configurations that manifest between the initial conformation and the ultimate native structure. The second central to this aspect is the exploration of the underlying forces that drive and stabilize the folding process and the last one of particular relevance to biological research is the capacity to predict a protein's native structure from its constituent peptide chain.

Over the past six decades, the prevailing paradigm within the realm of protein folding has revolved around the concept that proteins undergo folding processes that result in a reduction in Gibbs free energy (expressed as a negative  $\Delta G$ ) (Sorokina et al., 2022). The thermodynamic hypothesis of protein folding, especially the notion that the native state represents the most stable configuration, effectively constituting the global minimum of Gibbs free energy ( $G$ ), possesses an inherent allure and intuitive appeal. Moreover, this perspective significantly streamlines the development of theories and models, as it obviates the necessity to unravel the mechanisms by which a protein attains its unique native conformation. Indeed, by definition, the global minimum is a singular state. However, when the supposition is made that the native conformation occupies a local minimum rather than the global one within the Gibbs free energy landscape, it introduces a considerable degree of complexity (Lazaridis and Karplus, 2002). This complexity stems from the need to elucidate the processes

by which this specific local minimum is selected from among numerous other local minima.

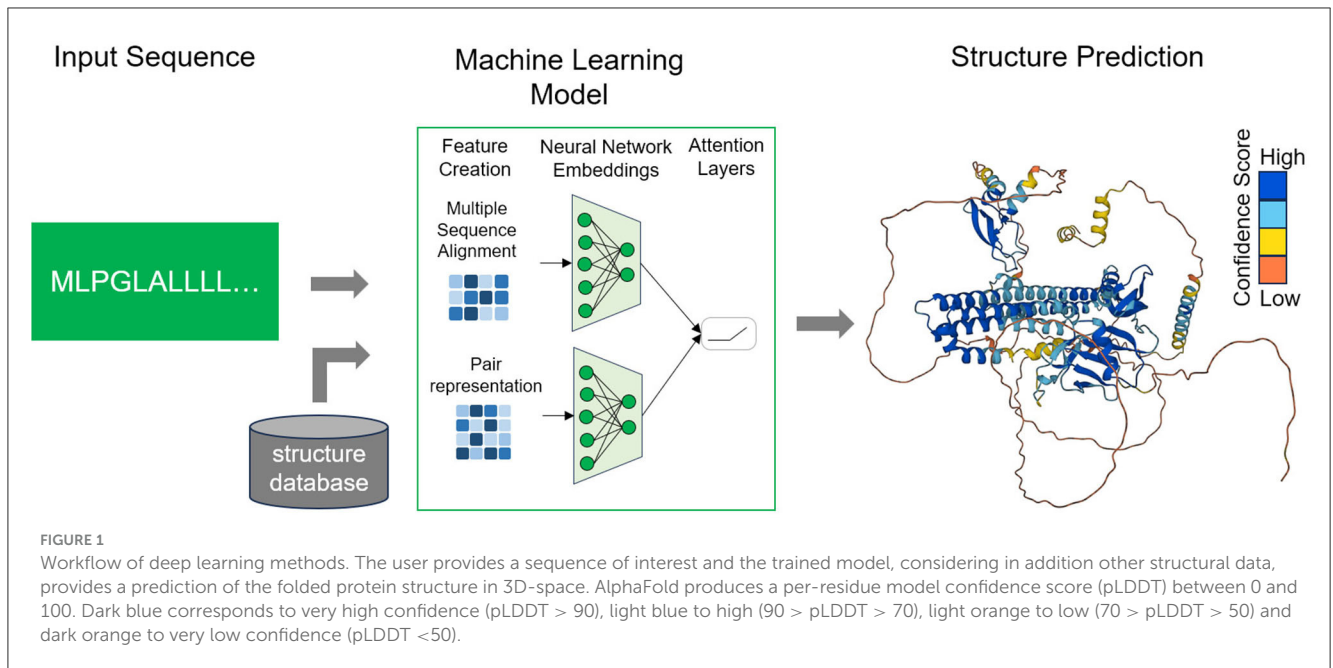
## Computational approaches to protein structure prediction

Computational methods for protein structure modeling are divided into distinct categories depending on the type of information they use: homology modeling and threading rely on structural data from similar proteins, whereas the ab initio method operates independently of such templates (Xu et al., 2008; Hameduh et al., 2020). Template-free approaches, followed by ab initio methods, exhibited superior performance in numerous instances. Numerous computational studies have explored solutions for protein conformation prediction, including evolutionary algorithms, Monte Carlo simulations and HP models (Istrail and Lam, 2009; Li et al., 2011; Tsay and Su, 2013). However, these approaches often struggle to efficiently search the vast conformational space of proteins.

MODELER stands out as a significant milestone among pioneering computational tools. Its primary application lies in the realm of homology or comparative modeling of protein three-dimensional (3D) structures (Webb and Sali, 2016). To harness the capabilities of MODELER, a fundamental requirement is the provision of a sequence alignment, aligning the target sequence with sequences from established, closely related protein structures referred to as templates. In return, it generates a comprehensive model that encompasses all non-hydrogen atoms within the protein structure (Kuntal et al., 2010). In addition to its fundamental function, this software offers a versatile array of supplementary features. These include the ability to perform de novo modeling of protein structure loops, assist in fold assignments, facilitate the alignment of two or more protein sequences or structures, and proficiently cluster protein structure.

Deep learning (DL) techniques have achieved notable progress in developing end-to-end differentiable models and directly forecasting dihedral angles or 3D protein structures (Torrisi et al., 2020; Pakhrin et al., 2021) (Figure 1). However, it is important to note that these approaches, as currently reported, still heavily rely on domain-specific input features in addition to the primary amino acid sequence. This reliance on extra information can pose challenges when attempting to generalize predictions for novel protein sequences lacking any prior knowledge of these specific input features. For instance, these methods frequently require the use of templates, which are explicit sequence-to-structure mappings generated through computational procedures (AlQuraishi, 2019; Qin et al., 2020). Templates facilitate the transformation of raw amino acid sequences into three-dimensional structures by referencing existing protein structures with similar sequences. Another common reliance is on co-evolutionary data, which is built upon the assumption that pairs of residues displaying highly correlated mutations are likely to be positioned closely in the protein's three-dimensional structure (Noé et al., 2020; Xu et al., 2021).

Addressing the significant challenge posed by the enigmatic protein folding problem, meticulous organization characterizes



the Critical Assessment of Protein Structure Prediction (CASP) competition. DeepMind's AlphaFold, the CASP13 winner, follows a methodology that embraces two firmly established concepts deeply ingrained in scientific literature (Senior et al., 2020). Firstly, it employs co-evolutionary analysis to map the co-variation of residues within protein sequences to physical interactions in protein structures. Secondly, it leverages deep neural networks to adeptly discern patterns in protein sequences and co-evolutionary couplings, subsequently translating them into contact maps.

In CASP14, a highly evolved iteration, denoted as AlphaFold2, emerged as the victorious solution, developed by Jumper et al. (2021). This model employs a DL network architecture that seamlessly integrates both physical and biological knowledge within a dual-track framework. To delve into the intricacies, the output of AlphaFold2 consists of the three-dimensional coordinates of all heavy molecules, complemented by a confidence score. This score is derived from an amalgamation of multiple sequence alignments (MSAs) and pairwise residue features. ColabFold was also used as an implementation of the AlphaFold framework that uses the MMseqs2 algorithm to promptly compile MSAs. Both can provide structural in specific proteins in involved in Alzheimer's disease (Santuz et al., 2022; Efraimidis et al., 2023). RoseTTAFold presents itself as an additional deep learning model that builds upon concepts initially developed by AlphaFold2. This model employs a three-track neural network, utilizing inputs at three distinct levels: the 1D sequence level, the 2D distance map level, and the 3D coordinate level, all in pursuit of protein structure prediction (Baek et al., 2021).

In another approach, the recurrent geometric network (RGN) employs a sequence of amino acids and position-specific scoring matrices (PSSMs) as its input. This method culminates in the prediction of a 3D protein structure (Chowdhury et al., 2022). Notably, the RGN model primarily relies on mathematical equations that pertain to chemical properties and employs a

recurrent neural network to encode the protein sequence. Lastly, a new deep learning architecture, DeepPotential, was developed for the prediction of protein structural geometry, integrating multiple unary and pairwise features as inputs into a hierarchical deep residual neural network (Li et al., 2022).

The availability of precise protein structure prediction has driven progress across diverse fields with AlphaFold and RosettaFold exhibiting competence in the prediction of intricate protein assemblies. Similarly, several other approaches have honed their focus on the realm of protein-protein interactions (PPIs). One such method, Struct2Graph, is founded on a 3D-structure-based graph attention network, specifically tailored for PPI prediction (Baranwal et al., 2022; Soleymani et al., 2022). In essence, this technique entails the acquisition of low-dimensional feature embeddings from the graph structures of individual proteins. Another field closely related to protein folding is the effort to design novel proteins designed to possess specific desired functionalities with ProteinMPNN standing out as a noteworthy example, having demonstrated its utility in the design of monomers, cyclic oligomers, protein nanoparticles, and protein-protein interfaces (Dauparas et al., 2022).

In a parallel vein, ProteinSolver is used, underpinned by graph convolutional neural networks (CNNs), which adeptly addresses the challenge of constraint satisfaction within the realm of protein topologies (Strokach et al., 2020). This model, trained on protein sequences, acquires the capacity to discern the constraints inherent in protein structures. Subsequently, it leverages this knowledge to generate new protein sequences capable of folding into user-defined shapes.

Integrating computational predictions, such as those generated by AlphaFold or RoseTTAFold with cryoelectron microscopy (cryo-EM) data can provide a synergistic approach to achieving more accurate and reliable protein structure determination such as helicase-primase D5 (Hutin et al., 2022) and interleukin-27

signal complex (Jin et al., 2022). Emphasis should be given on inclusion of experimental information from density maps however a mechanism for assimilating this information in a form compatible with modeling should be required. In several examined cases, enhancing the deep learning prediction involves utilizing templates derived from the initial prediction and employing automatic rebuilding with the density map. This entire process can serve as an initial step in structure determination, generating a docked algorithm model that may surpass the accuracy of one predicted without the density map (Terwilliger et al., 2022).

## Limitations and improvements

AlphaFold2 gained prominence in drug discovery not just for its remarkable accuracy, but also for its contribution of predicted protein structures to a publicly accessible database (Borkakoti and Thornton, 2023). This resource has significantly simplified the drug discovery process, particularly in the context of small molecule-based research, by reducing the necessity for recurrent structure predictions. However, AlphaFold2 such as every recent DL method, lacks the capability to predict critical elements of protein structures as the algorithm is unable to directly predict three-dimensional structures solely from a raw sequence (Xu, 2019; Jumper et al., 2021). Moreover, it cannot efficiently predict intrinsically disordered proteins/regions and loops as well it showed weak performance on reverse docking (Wong et al., 2022). Although lacking stable structures, intrinsically disordered proteins (IDPs) play a vital role in numerous biological processes linked to neurodegeneration. Identification and annotation of intrinsically disordered regions (IDRs) could be precisely predicted through deep learning-based disorder predictors. Recently, DeepDPR has been proposed which can achieve satisfactory results compared with other methods in predicting IDRs consisting of four distinct steps such as Masking layer, a TimeDistributed module, a Bi-LSTM network and 4) a fully connected network (Yang et al., 2023). Current designs of deep learning models could explore a broad space of inputs and network topologies, demonstrating diversity in their approach to predicting IDRs (Zhao and Kurgan, 2022). A challenge is the combination of physicochemical attributes, considering factors such as hydrogen bonding, contact potential energy, hydrophobicity and molecular size with folding prediction through AI approaches. AlphaFold2 uses as an input amino acid sequence to generate a MSA based on several databases of protein sequences to identify which parts of the sequence are mutation prone, determining correlation between them. In this combination of bioinformatics and physical approaches, the use of physical and geometric approaches to generate features that learn from PDB data is demonstrated, leading to a network that effectively learns from the limited data available while efficiently handling the complexity and diversity inherent in structural data (Jumper et al., 2021; Bertoline et al., 2023). Similarly, RGNs require both Position Specific Scoring Matrices (PSSMs) and the raw sequence to make predictions and are currently unable to harness information pertaining to secondary protein structure (AlQuraishi, 2021). Several extensions of the AlphaFold algorithm have been

created to tackle these challenges. For example, AlphaFold-Multimer trained on data with known stoichiometry, tailored for predicting multimeric interfaces while preserving a high degree of accuracy within individual protein chains while the web-based utility AlphaKnot is designed to assess entanglement within protein models resolved using AlphaFold, leveraging pLDDT confidence values (Niemyska et al., 2022; Zhu et al., 2023). Novel modeling methods derived from natural language processing for protein structure prediction have gain attention, such as ESMfold and EMBER2 to determine evolutionary, structural and functional patterns from massive protein sequence databases (Weißer et al., 2022; Lin et al., 2023). Concurrently improving the accuracy of a predictive model could simplify the detection of required adjustments for obtaining a better match with the density map. These prospects suggest that employing an iterative approach to incorporate information from the density map into structure prediction has the potential to increase the overall modeling accuracy (Terwilliger et al., 2022). Identifying protein interactions and integrating experimental data should be taken into account to refine accuracy in predicting the structures of proteins associated with neurodegenerative diseases, such as Alzheimer's and Parkinson's disease, and better understand their underlying molecular mechanisms and their functional implications. In this direction, improvements should include enhancing the ability to predict interactions between specific proteins involved in neurodegeneration, handling intrinsically disordered proteins and proteins embedded in cell membranes, and integrating experimental data from studies like X-ray microscopy, nuclear magnetic resonance or cryo-electron microscopy. In a recent study, the ReFOLD4 refinement approach represents a significant advancement in improving AlphaFold2 models, exhibiting a commendable ability to maintain high accuracy in localized regions (Adiyaman et al., 2023).

## Conclusions

Recent progress in AI models has brought substantial changes to the study of protein folding. Predicting protein structures from sequences and exploring protein energy landscapes through simulations now heavily involve machine learning techniques. Additionally, machine learning is instrumental in designing force-fields and extracting essential information from extensive simulation datasets, enhancing the understanding of rare events like folding/unfolding transitions. While certain challenges persist, these methods are expected to play a pivotal role in advancing the field of protein folding and dynamics in the near term. Continued progress is being made in addressing more intricate unresolved questions and help exploring numerous disorders as it is widely acknowledged that protein misfolding represents a prominent hallmark in the majority of neurodegenerative diseases. Our existing knowledge base largely derives from decades of research focused on purified proteins studied in vitro, such as  $\beta$ -amyloid in Alzheimer's disease or  $\alpha$ -synuclein in Parkinson's disease, and the extent of its applicability to protein folding within the intricate microenvironment of a living cell continues to be a persistent concern. Anticipating the forthcoming developments in tools and databases within the coming years, each incremental enhancement

is expected to effectively address current limitations and broaden the usability and applicability of predicted protein structures.

## Author contributions

MK: Conceptualization, Writing—original draft, Writing—review & editing. GD: Conceptualization, Visualization, Writing—original draft, Writing—review & editing. AV: Visualization, Writing—original draft, Writing—review & editing. TE: Conceptualization, Supervision, Writing—original draft, Writing—review & editing. PV: Funding acquisition, Supervision, Writing—original draft, Writing—review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research has been co-financed by the European Union and Greek national funds through the Operational Program

Competitiveness, Entrepreneurship and Innovation, under the call Regional Excellence (Research Activity in the Ionian University, for the study of protein folding in neurodegenerative diseases) (FOLDIT) MIS 5047144.

## 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.

## 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

- Adiyaman, R., Edmunds, N. S., Genc, A. G., Alharbi, S. M., and McGuffin, L. J. (2023). Improvement of protein tertiary and quaternary structure predictions using the ReFOLD refinement method and the AlphaFold2 recycling process. *Bioinf. Adv.* 12, vbad078 doi: 10.1093/bioadv/vbad078
- Alberts, B., Johnson, A., Lewis, J., Raff, M., and Roberts, K. (2002). *The Shape and Structure of Proteins. Molecular Biology of the Cell, 4th Edn.* New York, NY: Garland Science.
- AlQuraishi, M. (2019). End-to-end differentiable learning of protein structure. *Cell Syst.* 8, 292–301. doi: 10.1016/j.cels.2019.03.006
- AlQuraishi, M. (2021). Machine learning in protein structure prediction. *Curr. Opin. Chem. Biol.* 65, 1–8. doi: 10.1016/j.cbpa.2021.04.005
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223–230. doi: 10.1126/science.181.4096.223
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754
- Baranwal, M., Magner, A., Saldinger, J., Turali-Emre, E. S., Elvati, Kozarekar, S., et al. (2022). Struct2Graph: a graph attention network for structure based predictions of protein–protein interactions. *BMC Bioinf.* 23, 370. doi: 10.1186/s12859-022-04910-9
- Bertoline, L. M., Lima, A. N., Krieger, J. E., and Teixeira, S. K. (2023). Before and after AlphaFold2: an overview of protein structure prediction. *Front. Bioinf.* 3, 1120370. doi: 10.3389/fbinf.2023.1120370
- Borkakoti, N., and Thornton, J. M. (2023). AlphaFold2 protein structure prediction: implications for drug discovery. *Curr. Opin. Struct. Biol.* 78, 102526. doi: 10.1016/j.sbi.2022.102526
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins Struct. Funct. Bioinf.* 21, 167–195. doi: 10.1002/prot.340210302
- Chowdhury, R., Bouatta, N., Biswas, S., Floristean, C., Kharkar, A., Roy, K., et al. (2022). Single-sequence protein structure prediction using a language model and deep learning. *Nat. Biotechnol.* 40, 1617–1623. doi: 10.1038/s41587-022-01432-w
- Cuanalo-Contreras, K., Mukherjee, A., and Soto, C. (2013). Role of protein misfolding and proteostasis deficiency in protein misfolding diseases and aging. *Int. J. Cell Biol.* 2013, 1–15. doi: 10.1155/2013/638083
- Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378, 49–56. doi: 10.1126/science.ade21187
- Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008). The protein folding problem. *Annu. Rev. Biophys.* 37, 289–316. doi: 10.1146/annurev.biophys.37.092707.153558
- Efraimidis, E., Krokidis, M. G., Exarchos, T. P., and Lazar, T. (2023). *In silico* structural analysis exploring conformational folding of protein variants in Alzheimer's disease. *Int. J. Mol. Sci.* 24, 13543. doi: 10.3390/ijms241713543
- Esposito, L., Balasco, N., and Vitagliano, L. (2022). Alphafold predictions provide insights into the structural features of the functional oligomers of all members of the KCTD family. *Int. J. Mol. Sci.* 23, 13346. doi: 10.3390/ijms232113346
- Hameduh, T., Haddad, Y., Adam, V., and Heger, Z. (2020). Homology modeling in the time of collective and artificial intelligence. *Comp. Struct. Biotechnol. J.* 18, 3494–3506. doi: 10.1016/j.csbj.2020.11.007
- Hatano, Y., Ishihara, T., and Onodera, O. (2023). Accuracy of a machine learning method based on structural and locational information from AlphaFold2 for predicting the pathogenicity of TARDBP and FUS gene variants in ALS. *BMC Bioinf.* 24, 206. doi: 10.1186/s12859-023-05338-5
- Hutin, S., Ling, W. L., Tarbouriech, N., Schoehn, G., Grimm, C., Fischer, U., et al. (2022). The Vaccinia virus DNA helicase structure from combined single-particle cryo-electron microscopy and AlphaFold2 prediction. *Viruses* 14, 2206. doi: 10.3390/v14102206
- Istrail, S., and Lam, F. (2009). Combinatorial algorithms for protein folding in lattice models: a survey of mathematical results. *Commun. Inf. Syst.* 9, 303–346. doi: 10.4310/CIS.2009.v9.n4.a2
- Jin, Y., Fyfe, K., Gardner, S., Wilmes, S., Bubeck, D., Moraga, I., et al. (2022). Structural insights into the assembly and activation of the IL-27 signaling complex. *EMBO Rep.* 23, e55450. doi: 10.15252/embr.202255450
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kuntal, B. K., Aparoy, P., and Reddanna, P. (2010). EasyModeller: a graphical interface to MODELLER. *BMC Res. Notes* 3, 1–5. doi: 10.1186/1756-0500-3-226
- Lazaridis, T., and Karplus, M. (2002). Thermodynamics of protein folding: a microscopic view. *Biophys. Chem.* 100, 367–395. doi: 10.1016/S0301-4622(02)00293-4
- Li, Y., Zhang, C., Yu, D. J., and Zhang, Y. (2022). Deep learning geometrical potential for high-accuracy ab initio protein structure prediction. *IScience* 25, 104425. doi: 10.1016/j.isci.2022.104425
- Li, Y. W., Wüst, T., and Landau, D. P. (2011). Monte Carlo simulations of the HP model (the “Ising model” of protein folding). *Comput. Phys. Commun.* 182, 1896–1899. doi: 10.1016/j.cpc.2010.12.049
- Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379, 1123–1130. doi: 10.1126/science.ade2574

- Lurette, O., Martín-Jiménez, R., Khan, M., Sheta, R., Jean, S., Schofield, M., et al. (2023). Aggregation of alpha-synuclein disrupts mitochondrial metabolism and induce mitophagy via cardiolipin externalization. *Cell Death Disease* 14, 729. doi: 10.1038/s41419-023-06251-8
- Mietelska-Porowska, A., Wasik, U., Goras, M., Filipek, A., and Niewiadomska, G. (2014). Tau protein modifications and interactions: their role in function and dysfunction. *Int. J. Mol. Sci.* 15, 4671–4713. doi: 10.3390/ijms15034671
- Niemyska, W., Gren, B. A., Nguyen, M. L., Garstka, W., da Bruno Silva, F., Rawdon, E. J., et al. (2022). AlphaKnot: server to analyze entanglement in structures predicted by AlphaFold methods. *Nucleic Acids Res.* 50, W44–W50. doi: 10.1093/nar/gkac388
- Noé, F., De Fabritiis, D., and Clementi, G. (2020). Machine learning for protein folding and dynamics. *Curr. Opin. Struct. Biol.* 60, 77–84. doi: 10.1016/j.sbi.2019.12.005
- Pakhrin, S. C., Shrestha, B., Adhikari, B., and Kc, D. B. (2021). Deep learning-based advances in protein structure prediction. *Int. J. Mol. Sci.* 22, 5553. doi: 10.3390/ijms22115553
- Park, H., Baumann, M., Dunlop, J., Frydman, J., Kopito, R., McCampbell, A., et al. (2017). Protein misfolding in neurodegenerative diseases: implications and strategies. *Transl. Neurodegen.* 6, 1–13. doi: 10.1186/s40035-017-0077-5
- Qin, Z., Wu, L., Sun, H., Huo, S., Ma, T., Lim, E., et al. (2020). Artificial intelligence method to design and fold alpha-helical structural proteins from the primary amino acid sequence. *Extreme Mech. Lett.* 36, 100652. doi: 10.1016/j.eml.2020.100652
- Rani, K., Pal, A., Gurnani, B., Agarwala, S., Sasmal, D. K., Jain, N., et al. (2023). An innate host defense protein  $\beta$ 2-microglobulin keeps a check on  $\alpha$ -synuclein amyloid assembly: implications in Parkinson's disease. *J. Mol. Biol.* 435, 168285. doi: 10.1016/j.jmb.2023.168285
- Santuz, H., Nguyen, H., and Sterpone, F. (2022). Small oligomers of A $\beta$ 42 protein in the bulk solution with AlphaFold2. *ACS Chem. Neurosci.* 13, 711–713. doi: 10.1021/acscchemneuro.2c00122
- Scannevin, R. H. (2018). Therapeutic strategies for targeting neurodegenerative protein misfolding disorders. *Curr. Opin. Chem. Biol.* 44, 66–74. doi: 10.1016/j.cbpa.2018.05.018
- Sebastiano, M. R., Ermondi, G., Hadano, S., and Caron, G. (2022). AI-based protein structure databases have the potential to accelerate rare diseases research: AlphaFoldDB and the case of IAHSPL/Alsin. *Drug Discovery Today* 27, 1652–1660. doi: 10.1016/j.drudis.2021.12.018
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710. doi: 10.1038/s41586-019-1923-7
- Soleymani, F., Paquet, E., Viktor, H., Michalowski, W., and Spinello, D. (2022). Protein–protein interaction prediction with deep learning: a comprehensive review. *Comput. Struct. Biotechnol. J.* 20, 5316–5341. doi: 10.1016/j.csbj.2022.08.070
- Sorokina, I., Mushegian, A. R., and Koonin, E. V. (2022). Is protein folding a thermodynamically unfavorable, active, energy-dependent process? *Int. J. Mol. Sci.* 23, 521. doi: 10.3390/ijms23010521
- Strokach, A., Becerra, D., Corbi-Verge, C., Perez-Riba, A., and Kim, M. (2020). Fast and flexible protein design using deep graph neural networks. *Cell Syst.* 11, 402–411. doi: 10.1016/j.cels.2020.08.016
- Surguchov, A., and Surguchev, A. (2022). Synucleins: new data on misfolding, aggregation and role in diseases. *Biomedicines* 10, 3241. doi: 10.3390/biomedicines10123241
- Terwilliger, T. C., Poon, B. K., Afonine, V., Schlicksup, C. J., Croll, T. I., Millán, C., et al. (2022). Improved AlphaFold modeling with implicit experimental information. *Nat. Methods* 19, 1376–1382. doi: 10.1038/s41592-022-01645-6
- Torrisi, M., Pollastri, G., and Le, Q. (2020). Deep learning methods in protein structure prediction. *Comput. Struct. Biotechnol. J.* 18, 1301–1310. doi: 10.1016/j.csbj.2019.12.011
- Tsay, J. J., and Su, S. C. (2013). An effective evolutionary algorithm for protein folding on 3D FCC HP model by lattice rotation and generalized move sets. *Proteome Sci.* 11, 1–14. doi: 10.1186/1477-5956-11-S1-S19
- Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinf.* 54, 5–6. doi: 10.1002/cpbi.3
- Weißnow, K., Heinzinger, M., and Rost, B. (2022). Protein language-model embeddings for fast, accurate, and alignment-free protein structure prediction. *Structure* 30, 1169–1177. doi: 10.1016/j.str.2022.05.001
- Wilson, D. M., Cookson, M. R., Van Den Bosch, L., Zetterberg, H., Holtzman, D. M., Dewachter, I., et al. (2023). Hallmarks of neurodegenerative diseases. *Cell* 186, 693–714. doi: 10.1016/j.cell.2022.12.032
- Winklhofer, K. F., Tatzelt, J., and Haass, C. (2008). The two faces of protein misfolding: gain-and loss-of-function in neurodegenerative diseases. *The EMBO J.* 27, 336–349. doi: 10.1038/sj.emboj.7601930
- Wong, F., Krishnan, A., Zheng, E., Stark, H., Manson, A., Earl, A. M., et al. (2022). Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery. *Mol. Syst. Biol.* 18, e11081. doi: 10.15252/msb.202211081
- Xu, J. (2019). Distance-based protein folding powered by deep learning. *Proc. Nat. Acad. Sci.* 116, 16856–16865. doi: 10.1073/pnas.1821309116
- Xu, J., Jiao, F., and Yu, L. (2008). Protein structure prediction using threading. *Protein Struct. Prediction* 12, 91–121. doi: 10.1007/978-1-59745-574-9\_4
- Xu, J., Mcpartlon, M., and Li, J. (2021). Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat. Mach. Int.* 3, 601–609. doi: 10.1038/s42256-021-00348-5
- Yang, Z., Wang, Y., Ni, X., and Yang, S. (2023). DeepDRP: Prediction of intrinsically disordered regions based on integrated view deep learning architecture from transformer-enhanced and protein information. *Int. J. Biol. Macromol.* 253, 127390. doi: 10.1016/j.ijbiomac.2023.127390
- Zhang, H., Jiang, X., Ma, L., Wei, W., Li, Z., Chang, S., et al. (2022). Role of A $\beta$  in Alzheimer's-related synaptic dysfunction. *Front. Cell Dev. Biol.* 10, 964075. doi: 10.3389/fcell.2022.964075
- Zhao, B., and Kurgan, L. (2022). Deep learning in prediction of intrinsic disorder in proteins. *Comput. Struct. Biotechnol. J.* 20, 1286–1294. doi: 10.1016/j.csbj.2022.03.003
- Zhu, W., Shenoy, A., and Elofsson, A. (2023). Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics.* 39, btad424. doi: 10.1093/bioinformatics/btad424