LEVERAGING THE STRENGTH OF ARTIFICIAL INTELLIGENCE IN SOLVING PROTEIN STRUCTURES BY ALPHAFOLD-2- A MODERN APPROACH TO UNDERSTAND PROTEIN DYNAMICS

Abstract

Proteins are unique macromolecules made up of a long chain of amino acids and are classified based on their function, structure, shape, chemical composition and solubility in different solvents. A wide variety of proteins are prone to misfold and create intracellular or extracellular aggregates that cause severe cellular malfunction. The importance of the protein folding problem was recognized and put forward 50 years back by distinguished scientists. Understanding the dynamics of protein folding is crucial and this can help us predict the ultimate configuration of functional protein. Many of the life-threatening diseases are caused by the misfolding of proteins. The reason for the misfolding can be point mutations since the three-dimensional structure of proteins depends on the primary sequence of its amino acid. Despite fifty years of research, we still need to fill the knowledge gap and accelerate our understanding, particularly in computational biology for the accurate prediction of protein structure. Homology modeling is utilized to predict protein structure in absence of experimental structure. Artificial intelligence, machine learning, and deep learning are being used extensively by researchers to computationally estimate a protein's structure based only on its amino acid sequence. AlphaFold which is in a second iteration tool has changed the perception about protein folding by solving the unsolved structures.

Keywords: Homology modeling, Artificial intelligence, Machine learning, Deep Learning, AlphaFold

Authors

Rik Ganguly
Department of Biotechnology and Bioinformatics, North-Eastern Hill University
Shillong, Meghalaya, India

Shashi Kumar Yadav
Department of Biotechnology and Bioinformatics, North-Eastern Hill University
Shillong, Meghalaya, India

Prosperwell Ingty
Department of Biotechnology and Bioinformatics, North-Eastern Hill University
Shillong, Meghalaya, India

Angneh Ngoruh
Department of Biotechnology and Bioinformatics, North-Eastern Hill Universit
Shillong, Meghalaya, India

Atanu Bhattacharjee
Department of Biotechnology and Bioinformatics, North-Eastern Hill University
Shillong, Meghalaya, India
atanubioinfo@gmail.com

I. INTRODUCTION

Proteins are the building block of macromolecules and are found in all living systems including prokaryotes and eukaryotes. Proteins are unique amongst the macromolecules in underpinning every reaction occurring in a biological system (1). Proteins are made up of amino acids, which are attached to long-chain fatty acids called polypeptides. The polypeptide chains fold into their final three-dimensional structure to constitute a functional protein. Twenty different types of amino acids can be combined to make a protein. Amino acids are coded by a combination of three nucleotides called codon which is determined by the sequence of genes (2).

Proteins are classified based on their function, structure, shape, chemical composition and solubility in different solvents (3). Folding and unfolding of proteins are crucial ways of regulating biological activity and targeting proteins to different cellular locations. Aggregation of misfolded proteins that escape the cellular quality-control mechanisms is a common feature of a wide range of highly debilitating and increasingly prevalent diseases (4). The folding of proteins is primarily driven by peptide bonds, hydrogen bonds, di-sulphide bonds and hydrophobic bonds.

Central dogma and protein folding: Proteins undergo reversible structural changes in performing their biological function. There are four protein structures, Primary, Secondary, tertiary and quaternary (5). The primary structure is held together by peptide bonds that are made during the process of protein biosynthesis. Secondary structure refers to highly regular local sub-structures on the actual polypeptide backbone chain and it is defined by patterns of hydrogen bonds between the main-chain peptide groups. Two main types of secondary structure, the α-helix, random coils and the β-strand or β-sheets, were suggested in 1951 by Linus Pauling *et al* (6).

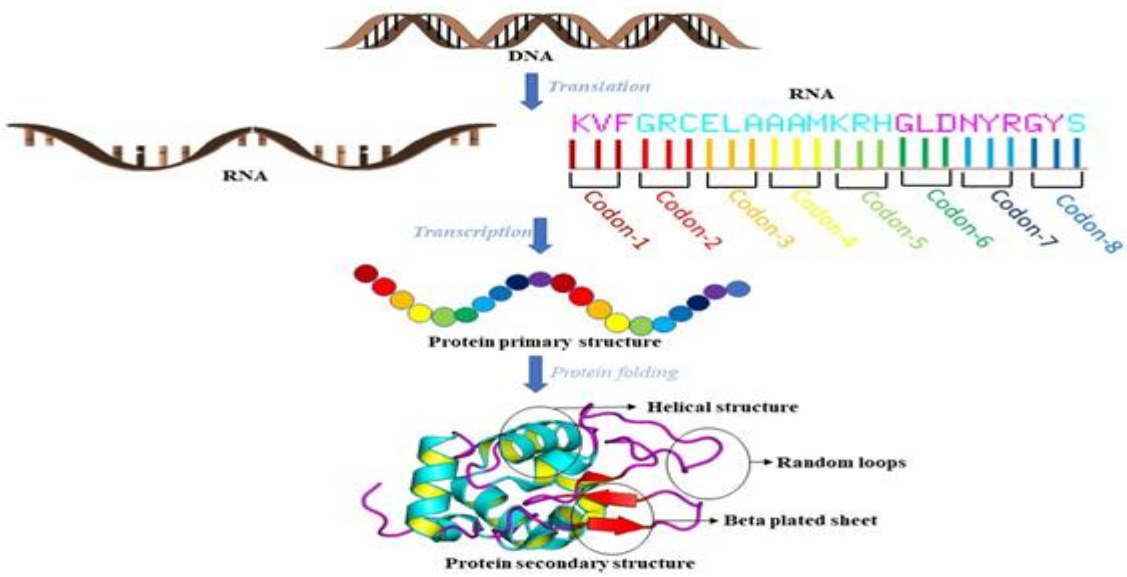

Figure 1: shows the central dogma of life which leads to protein synthesis and finally post-translational modification which gives rise to the secondary structure of a protein.

The Tertiary structure refers to the three-dimensional structure created by a single protein molecule (a single polypeptide chain). The tertiary protein structure folding is driven by the non-specific hydrophobic interactions. The quaternary structure of a protein is the three-dimensional structure consisting of the aggregation of two or more individual polypeptide chains (subunits) that operate as a single functional unit (multimer). There are several types of proteins: antibodies, enzymes, hormonal proteins, structural proteins, storage proteins, receptor proteins and transport proteins (7).

Different proteins can misfold and form extracellular or intracellular aggregates that initiate catastrophic cellular dysfunction. Particularly challenging examples of such disorders occur in the post-mitotic environment of the neuron and include Alzheimer's and Parkinson's diseases. Understanding some of the principles of protein folding has helped to explain how such diseases arise and will help in the field of medical science (8).

II. PROTEIN FOLDING PROBLEM

Some of the fundamental questions about protein folding arose during the 1950s such as how mRNA codons dictate the amino acid sequence (9). The information about the native three-dimensional structure of the proteins is present in the amino acid sequence but how biologically proteins fold so fast (10). Christian Anfinsen was one of the pioneer scientists in the biochemistry field who attempted in the year 1959 to merge the recently developing area of protein chemistry with classical genetics in his manuscript entitled “The molecular basis of evolution. He set the platform for the expansion of molecular biology based on the determination of nucleic acids and protein sequence and shared the Noble Prize with Moore and Stein in chemistry in the year 1972 for his work on the link between the amino acid sequence and the biologically active protein conformation (11).

1. **Energy landscape theory of protein:** Thermodynamically protein tends to fold from open to compact in the lowest free energy possible which is considered the most stable state (12). The stability of the protein increases as it starts folding into the local structure and finally to the global structure. There are many research gaps in understanding the mechanisms of protein folding, the kinetics of partially structured folding intermediate, and measuring the interatomic interaction in nano to microsecond time scale (13). We still do not have a distinct real-time evaluation technique which could take snapshots in a nano-second time scale while the primary sequence of amino acids is folding to its native structure (14). The free energy landscape of protein is a statistical analysis of protein folding, it is a well-founded model that describes how protein folded into its native structure. Very few compact and low free energy conformations of the folded protein fit in the narrow bottom of the funnel-shaped energy landscape (Shown in figure 2 A & B) (15). A landscape which appears favourable on a global scale can be unfavourable on the local scale. We still could not figure out how the different amino acid residues of the same protein follow different folding routes for the common native structure (16). Force fields are used in computer simulations to study molecular dynamics (17). It gives the mathematical expression of energy exchange in the system on coordinates of its components, it describes how the interatomic forces are acting upon one another. Empirical force field-based simulations have some limitations such as they cannot give

information about the distinct structure of electronic arrangements, its excitations, making and breaking of bonds, and charge transfer (18).

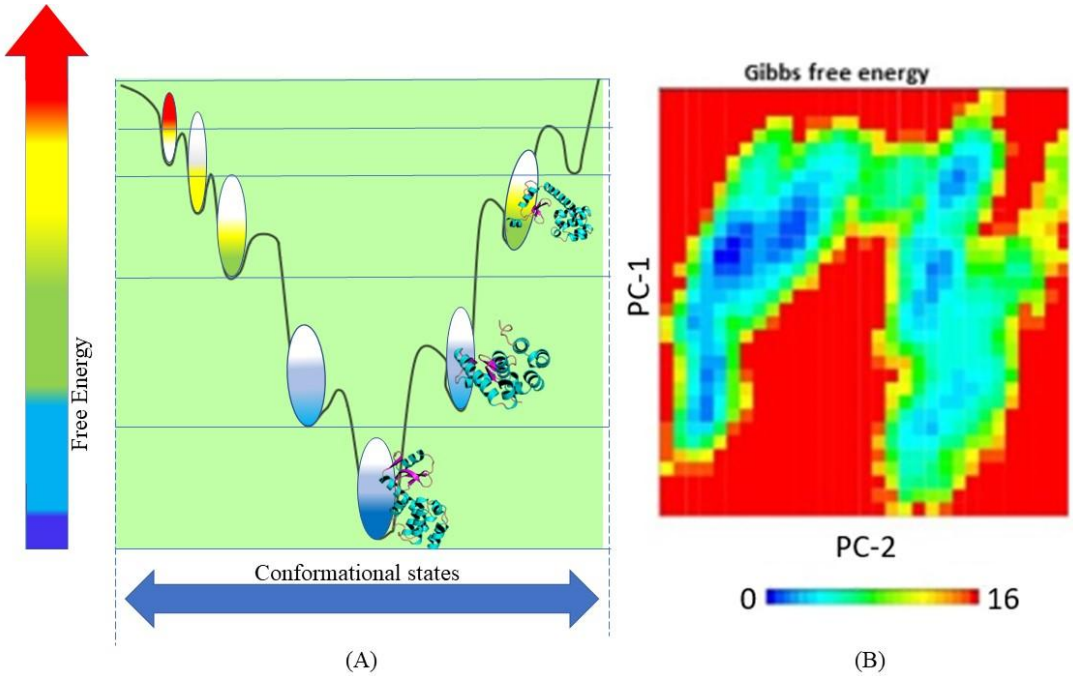

Figure 2: Showing (A) the varying free energy funnel with respect to different conformational states of a protein. (B) Gibbs free energy landscape for the major two principal components in two-dimensional representation (Ganguly *et al*., 2022**).**

2. **Consequences of protein-misfolding:** More than half a century of research revealed that various forces contribute to the protein structure like hydrogen bonding, electrostatic interactions, Van der wall interactions, hydrophobic interactions, backbone angle preferences and chain entropy (19). Many diseases are also caused by misfolding of proteins like Creutzfeldt-Jakob disease, type 2 diabetes, Alzheimer etc. study of these diseases provide us with more clue about the significance of misfolding of protein and its related consequences. The last 50 years was a period of enormous advancement in the protein folding problem. With techniques like energy landscapes, single-molecule methods, fast temperature jump methods along with bioinformatics tools we can address the current problem of protein folding with more accuracy (20).

III. ARTIFICIAL INTELLIGENCE IN PROTEIN STRUCTURE PREDICTION

A method for predicting protein structure known as comparative or homology modelling is based on the basic insight that proteins with related sequences have related structures (21). To gain an understanding of the structure and function of these proteins in the absence of experimental structures, computational approaches are employed to predict 3D protein models. Protein models created using a variety of automated techniques can be found in repositories such as the SWISS-MODEL (http://swissmodel.expasy.org/SWISS-MODEL.html), Protein Model Portal (http://proteinmodelportal.org) (22), and Modbase (http://modbase.compbio.ucsf.edu) (23). For the benefit of biologists and experimentalists

working on structural genomics and biomedical research projects, these servers offer models that can be used as starting points. But without human intervention, errors brought on by incorrect sequence alignment and the inability to recognize and accurately model domains, such as loop and ligand-binding regions, are amplified, leading to low-accuracy generated models, which restricts their applicability to drug discovery projects (24,25). Research is currently being done on the creation and advancement of homology modeling refinement tools for drug discovery (26).

Researchers are increasingly using Artificial Intelligence (AI) methods to computationally estimate a protein's structure based solely on its amino acid sequence (27, 28, 29). AlphaFold is a DeepMind AI system that creates cutting-edge predictions of protein structures from their amino acid sequences (27). Artificial intelligence software created by DeepMind is well renowned. Its primary goal is to push the limits of artificial intelligence by creating computers that can figure out how to tackle any hard problem on their own, without any training or prerequisite knowledge. A powerful general-purpose learning algorithm set is created, and an AI is subsequently made by combining them (30). Deep learning enables computational models, which are made up of several processing layers, to learn representations of data at various levels of abstraction. The state-of-the-art has been significantly enhanced by these techniques in many other fields, including drug discovery and genomics, as well as speech recognition, visual object recognition, object detection, and many more (31). In a competition in March 2016, Google's artificial intelligence (AI) computer program AlphaGo defeated Lee Se-dol, the top Go player in the world, winning 4 of 5 games (32). AlphaGo Zero can self-train at computer speeds without human assistance since it has enough data to create, play, and evaluate any legal game that might exist in its world (33,34).

1. **Journey from CASP to AlphaFold-2:** AlphaFold is now in its second iteration which is developed from the Critical Assessment of Structure Prediction (CASP14) algorithm, the protein structure predicting software using Artificial intelligence. In between the AlphaFold1 came into existence in 2019 (35). If the accuracy in predicting the structural similarity between the experimentally deduced structures and the predicted one comes to 90 and above it means there are very less discrepancies in prediction. Before AlphaFold came into existence this accuracy lied in between 30-40 % in accuracy for different versions of CASP (36). For the first time, AlphaFold has achieved an accuracy of above 80 % which has become the game changer in the field of structural prediction. And the second iteration has even enhanced this accuracy to near 90 % (37).

Table 1. Shows the number of available experimentally derived structures present in the Protein data bank (*"PDB Current Holdings Breakdown"*. *RCSB*.).

Experimental Method	Proteins	Nucleic Acids	Protein/Nucleic Acid complexes	Other	Total
X-ray diffraction	135170	2097	6945	4	144216
NMR	11337	1325	264	8	12934
Electron microscopy	3475	35	1136	0	4646
Hybrid	155	5	3	1	164
Other	286	4	6	13	309
Total:	150423	3466	8354	26	**162269**

Proteins have a high level of individuality and diversity even among the same family, each amino acid sequence gives rise to an intricate arrangement of the amino acid which leads to 3-dimensional spatial arrangement and interactions which gives rise to protein folding. There are about 8 million sequences present in a non-redundant protein sequence database and compare to that the experimentally derived structures present in the Protein data bank (RCSB-PDB) are very less (Shown in Table 1) (38). As there is a huge gap between the available sequences and the derived structure for a protein highly accurate algorithm with a high confidence rate can help fill the gap by modelling and predicting the accurate protein structures. AlphaFold not only helped in predicting the protein structures to a high level of confidence but also predicted the low electron density region to the nearest probable loop structure (39). On a multimer challenge, the AlphaFold 2 predicted a very highly accurate protein-peptide complex formation which was found to be more fitting as compared to other commercial software (40). AlphaFold protein structure database (https://AlphaFold.ebi.ac.uk) which contains several protein structures which are biologically important and are not present in a publicly available protein data bank. AlphaFold structures were tested with crystal structures found to be very similar as compared to other homology-modelled structures (41). AlphaFold has mitigated several experimental limitations to study protein dynamics such as membrane-bound proteins as well as large proteins such as nuclear pore complex which are difficult to isolate and crystalize. There are an enormous number of proteins which are important from fundamental research point of view which we don't have much knowledge about with respect to their structural and functional aspects (42).

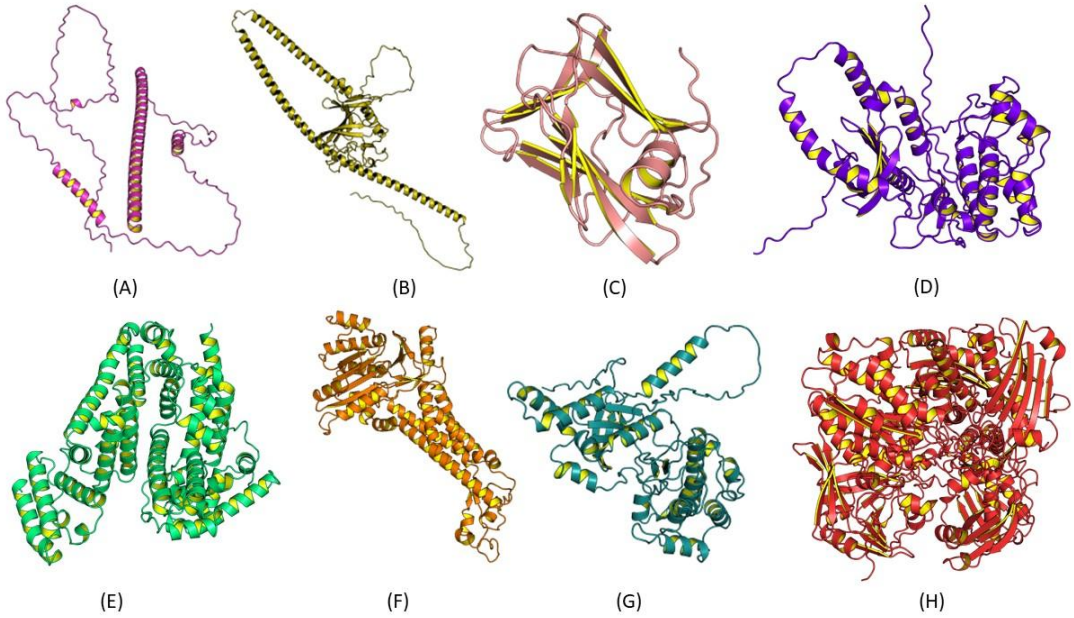

Figure 3: Showing some biologically important protein structures which are not derived experimentally but present in alfa fold protein repository such as (A) Leucine zipper motif present in humans, (B) Angiopoietin-like 3, (C) Transcription factor IIB in *homo sapiens*, (D) Serine/threonine protein kinase (E) Human albumin protein (F) NADPH oxidase-1, (G) human F-box protein (H) Xanthine oxidase.

2. **Solving the yet to be solved protein galaxy:** Proteins such as Leucine zipper (LZ) structural motifs which are also known as leucine scissors found in several different transcription factors with 20-40 amino acids of an alpha helix which is rich in leucine in every seventh position (43). LZ binds to the major groove of the DNA in the "ACGT" rich repeat within the promoter or enhancer region with the N-terminal domain and helps in the process of transcription. The structure of the leucine zipper motif in a human was not available in PDB and is derived by AlphaFold with a high level of accuracy (shown in figure 3. A). Angiopoietin-like 3 (ANGPLT-3) protein inhibits lipoprotein lipase which is an enzyme that degrades lipoproteins. Several recent studies have predicted that ANGPLT-3 can be an excellent therapeutic target against several cardiovascular conditions (44). AlphaFold derived complete protein structure for ANGPLT-3 (shown in figure 3.B) was not available in any publicly available protein database. Factor in transcription Growth hormone-secreting pituitary adenomas' production of the aryl hydrocarbon receptor-interacting protein (AIP) protein is controlled by GTF2B also known as Transcription factor IIB, which also influences tumour behaviours. This can be a probable marker for cancer biology and due to the lack of experimental structures, there was less scope in the field of structure-based drug discovery (45). AlphaFold has helped in deriving this important biological macromolecule very precisely (Shown in figure 3.C). AlphaFold derived protein structure of Serine/Threonine Kinase was not present previously and was derived with a high confidence level (shown in figure 3.D) (46), It is an effective enzyme target in the treatment of ovarian cancer patients. Figure 2. E is

showing the most abundant protein in blood plasma, human serum albumin protein and the AlphaFold derived structure was found highly similar to the experimentally derived structure (47). NADPH oxidase-1 (nox-1) is a major player in Reactive oxygen species (ROS) production and plays an important role in clinical conditions such as pulmonary ischemia which can cause a catastrophic outcome (48), AlphaFold structural database for the first time made this structure publicly available (Shown in figure 3.F) for Insilco experimentations. Human F-box protein which helps in the protein-protein interaction in the process of ubiquitination plays a very significant physiological role, AlphaFold has reported the complete accurate structure of F-box protein (Shown in figure 3.G) (49). The production of uric acid depends on the enzyme xanthine oxidoreductase (XOR). inhibition of excess uric acid production improves patients who have reduced cardiovascular function. For the first time, the structure of XOR has been derived by AlphaFold 2 (shown in figure 3.H) (50)**.**

3. **AlphaFold architecture:** The AlphaFold 2 architecture consists of three important parts. The first step includes "**Embedding**" which involves a multiple sequence alignment of the query sequence with others and finding template structures. No coordinates are associated with any of this data, which is not embedded in 3D space (51). The second part is called "**The Trunk**" which involves a pairwise alignment of the residue-residue graph edges and the sequence-residue graph edges. Similar to relative distances in 3D space and relative angles, the residue-residue edges convey pairwise information between all residues. The edges of the sequence residues can include information on sequence evolution. Pairwise distances may be predicted using this information, but more significantly, it is given to the structure module so that it can create the structure's 3-dimensional coordinates. Strings of amino acids are used by the "Embedding" and the "Trunk" to generate matrix descriptions of possible interactions and relationships (52). The third and final module includes "**The Structure module**" which refines backbone coordinates and predicts side chains using a 3D equivariant transformer architecture. This network's job is to predict new red triangle placements and orientations as well as the confidence score (53).

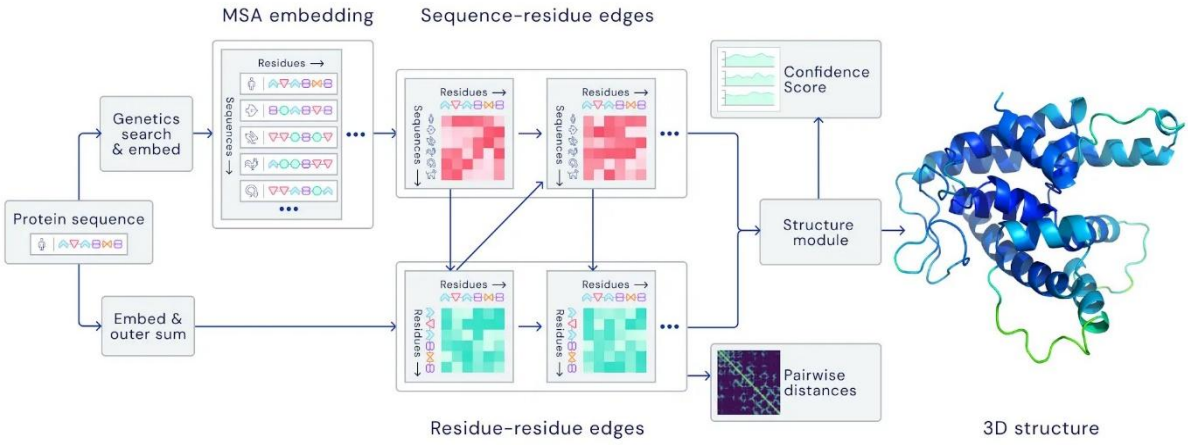

Figure 3. Showing the overall architecture of AlphaFold-2 (DeepMind blog, November 2020).

The backbone is represented by oriented red triangles. It becomes complicated in the structural module. Unexpectedly, 3D coordinates appear. Normal neural networks are unable to recognise coordinates. Coordinates are just integers by default. Even though the energy doesn't change when the backbone is rotated, these values do. Although dealing with global rotations accurately imposes a learning challenge that is far from straightforward, a neural network might presumably learn this. The structural module of AlphaFold 2 is not required to learn this. Its design combines it. To do this, it combines the ideas of equivariance and self-attention (also known as the transformer mechanism) (54). Transformer and self-attention are two terms that are frequently used interchangeably to refer to a neural network mechanism that acts on a collection of objects (an object may, for example, be an atom or an amino acid) and enables the querying of particular data. A self-attention layer updates the nodes' characteristics as it maps from set to set, or in our instance, graph to graph. It examines one thing at a time, says a carbon atom, and makes enquiries about nearby objects based on the data or attributes associated with that carbon atom (55). For instance, it could be very helpful to look for nitrogen atoms nearby given what we currently know about the position of the carbon atom. The second machine learning concept that is used is called equivariance. The most effective approach to demonstrate it is to use CNN's (convolutional neural networks). Convolutional layers shift their output by three pixels to the right for every three pixels that the input image is moved to the right because they are translation equivariant. If the input is shifted a few pixels to the right, the problem isn't entirely new. Utilizing this symmetry and treating the two inputs equally is essential since doing so reduces overfitting, saves parameters, and speeds up learning (56).

With the development of high-performance computing and cloud computing research on artificial intelligence has enhanced and emerging with a lot of solutions to several problems that were present in the society for ages. AI is being used in different aspect of life from share market stock prediction to navigate a car, from translating any language using your phone camera to defeating the world famous go player multiple times. The advancement in the field of medicine will not only make AlphaFold as a game changer but it will definitely solve several diseases related to protein dysfunction due to improper folding, protein overexpression, mutations. AlphaFold has helped the bioinformatics community by giving a solution to the unsolved structures, thereby leading to development of structure-based inhibitor design and also understanding protein dynamics in details. It is expected that in coming days the tool will further enhance its capabilities with advancement of Artificial intelligence and deep mining.

REFERENCES

[1] Whitford, D. (2013). Proteins: structure and function. *John Wiley & Sons*.
[2] Branden, C. I., & Tooze, J. (2012). Introduction to protein structure. *Garland Science*.
[3] Murzin, A. G. (1996). Structural classification of proteins: new superfamilies. *Current Opinion in Structural Biology*, 6(3), 386-394.
[4] Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, *426*(6968), 884-890.
[5] Damodaran, S. (2008). Amino acids, peptides and proteins. *Fennema's Food Chemistry*, 4, 217-329.
[6] Schulz, G. E., & Schirmer, R. H. (2013). Principles of protein structure. *Springer Science & Business Media*.

[7] Prediction of tight turns and their types in proteins. *Analytical Biochemistry*, 286(1), 1-16.
[8] Selkoe, D. J. (2003). Folding proteins in fatal ways. *Nature*, *426*(6968), 900-904. Chou, K. C. (2000).
[9] Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, *426*(6968), 884-890.
[10] Fersht, A. R. (2008). From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nature Reviews Molecular Cell Biology*, *9*(8), 650-654.
[11] Young, M. (1995). Christian B. Anfinsen (1916-1995) Remembering His Life and His Science. *Protein Science: A Publication of the Protein Society*, *4*(10), 2237.
[12] Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences*, *89*(18), 8721-8725.
[13] Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature structural biology*, *4*(1), 10-19.
[14] Voelz, V. A., Bowman, G. R., Beauchamp, K., & Pande, V. S. (2010). Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1− 39). *Journal of the American Chemical Society*, *132*(5), 1526-1528.
[15] Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry*, *24*(6), 1501-1509.
[16] Englander, S. W., Mayne, L., & Krishna, M. M. (2007). Protein folding and misfolding: mechanism and principles. *Quarterly Reviews of Biophysics*, *40*(4), 1-41.
[17] Kadau, K., Germann, T. C., & Lomdahl, P. S. (2006). Molecular dynamics comes of age: 320 billion atom simulation on BlueGene/L. *International Journal of Modern Physics C*, *17*(12), 1755-1761.
[18] Schulz, R., Lindner, B., Petridis, L., & Smith, J. C. (2009). Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer. *Journal of Chemical Theory and Computation*, *5*(10), 2798-2808.
[19] Ponder, J. W., & Case, D. A. (2003). Force fields for protein simulations. *Advances in Protein Chemistry*, *66*, 27-85.
[20] Uversky, V. N., & Dunker, A. K. (2010). Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1804*(6), 1231-1264.
[21] Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, *29*, 291-325.
[22] Arnold, K., Kiefer, F., Kopp, J., Battey, J. N., Podvinec, M., Westbrook, J. D., Berman, H. M., Bordoli, L., & Schwede, T. (2009). The Protein Model Portal. *Journal of Structural and Functional Genomics*, *10*(1), 1-8.
[23] Pieper, U., Eswar, N., Davis, F. P., Braberg, H., Madhusudhan, M. S., Rossi, A., Marti-Renom, M., Karchin, R., Webb, B. M., Eramian, D., Shen, M. Y., Kelly, L., Melo, F., & Sali, A. (2006). MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research*, *34*(Database issue), D291-D295.
[24] Dalton, J. A., & Jackson, R. M. (2007). An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics (Oxford, England)*, *23*(15), 1901-1908.
[25] Fischer, D (2006). Servers for protein structure prediction. *Current Opinion in Structural Biology*, *16*(2), 178-182.
[26] Kairys, V., Gilson, M. K., & Fernandes, M. X. (2006). Using protein homology models for structure-based studies: approaches to model refinement. *The Scientific World Journal*, *6*, 1542-1554.
[27] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., … Hassabis, D.

(2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589.

[28] Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., … Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science (New York, N.Y.), 373*(6557), 871-876.

[29] Ramanathan, A., Ma, H., Parvatikar, A., & Chennubhotla, S. C. (2021). Artificial intelligence techniques for integrative structural biology of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 66, 216-224.

[30] Holcomb, Sean D.; Porter, William K.; Ault, Shaun V.; Mao, Guifen; Wang, Jin (2018). [ACM Press the 2018 International Conference - Honolulu, HI, USA (2018.03.09-2018.03.11)] Proceedings of the 2018 International Conference on Big Data and Education - ICBDE '18 - Overview on DeepMind and Its AlphaGo Zero AI., (), 67–71.

[31] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

[32] Granter, S. R., Beck, A. H., & Papke, D. J., Jr (2017). AlphaGo, Deep Learning, and the Future of the Human Microscopist. *Archives of Pathology & Laboratory Medicine*, *141*(5), 619-621.

[33] Bollinger, T. (2021). Why AlphaFold is Not Like AlphaGo. *Academia Letters*, Article 728.

[34] Moult, J. (2005). A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Current Opinion in Structural Biology*, *15*(3), 285-289.

[35] Akdel, M., Pires, D. E., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., ... & Beltrão, P. (2021). A structural biology community assessment of AlphaFold 2 applications. *BioRxiv*.

[36] Anon, H. (2014). Hard data: it has been no small feat for the protein data bank to stay relevant for, vol. 100, 000 structures. *Nature*, *509*(7500), 260-268.

[37] Si, D., Moritz, S. A., Pfab, J., Hou, J., Cao, R., Wang, L., ... & Cheng, J. (2020). Deep learning to predict protein backbone structure from high-resolution cryo-EM density maps. *Scientific reports*, *10*(1), 1-22.

[38] Gao, M., Nakajima An, D., Parks, J. M., & Skolnick, J. (2022). AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature Communications*, *13*(1), 1-13.

[39] Hu, L., Salmen, W., Sankaran, B., Lasanajak, Y., Smith, D. F., Crawford, S. E., ... & Prasad, B. V. (2022). Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography. *Communications Biology*, *5*(1), 1-8.

[40] Barbarin-Bocahu, I., & Graille, M. (2022). The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: a case-study report. *Acta Crystallographica Section D: Structural Biology*, *78*(4).

[41] Bornberg-Bauer, E., Rivals, E., & Vingron, M. (1998). Computational approaches to identify leucine zippers. *Nucleic Acids Research*, *26*(11), 2740-2746.

[42] Lupo, M. G., & Ferri, N. (2018). Angiopoietin-like 3 (ANGPTL3) and atherosclerosis: lipid and non-lipid related effects. *Journal of Cardiovascular Development and Disease*, *5*(3), 39.

[43] Ha, I. L. H. O., Roberts, S., Maldonado, E., Sun, X. I. A. O. Q. I. N. G., Kim, L. U., Green, M., & Reinberg, D. (1993). Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes & Development*, *7*(6), 1021-1032.

[44] Kandel, E. S., & Hay, N. (1999). The regulation and activities of the multifunctional serine/threonine kinase Akt/PKB. *Experimental Cell Research*, *253*(1), 210-229.

[45] Sheehan, A. L., Carrell, S., Johnson, B., Stanic, B., Banfi, B., & Miller Jr, F. J. (2011). Role for Nox1 NADPH oxidase in atherosclerosis. *Atherosclerosis*, *216*(2), 321-326.

[46] Edelman, A. M., Blumenthal, D. K., & Krebs, E. G. (1987). Protein serine/threonine kinases. *Annual Review of Biochemistry*, *56*(1), 567-613.

[47] Curry, S. (2002). Beyond expansion: structural studies on the transport roles of human serum albumin. *Vox Sanguinis*, *83*, 315-319.
[48] Sheehan, A. L., Carrell, S., Johnson, B., Stanic, B., Banfi, B., & Miller Jr, F. J. (2011). Role for Nox1 NADPH oxidase in atherosclerosis. *Atherosclerosis*, *216*(2), 321-326.
[49] Wang, Z., Liu, P., Inuzuka, H., & Wei, W. (2014). Roles of F-box proteins in cancer. *Nature Reviews Cancer*, *14*(4), 233-247.
[50] Harrison, R. (2002). Structure and function of xanthine oxidoreductase: where are we now?. *Free Radical Biology and Medicine*, *33*(6), 774-797.
[51] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., & Hassabis, D. (2021). Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics*, *89*(12), 1711-1721.
[52] Thompson, M. C., Yeates, T. O., & Rodriguez, J. A. (2020). Advances in methods for atomic resolution macromolecular structure determination. *F1000Research*, *9*.
[53] Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, *33*(7), 2302-2309.
[54] Skolnick, J., Gao, M., Zhou, H., & Singh, S. (2021). AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *Journal of Chemical Information and Modeling*, *61*(10), 4827-4831.
[55] Cao, S. (2021). Choose a transformer: Fourier or galerkin. *Advances in Neural Information Processing Systems*, *34*, 24924-24940.
[56] Billings, W. M., Hedelius, B., Millecam, T., Wingate, D., & Della Corte, D. (2019). ProSPr: democratized implementation of AlphaFold protein distance prediction network. *BioRxiv*, 830273.
[57] Ganguly, R., Mylliemngap, B. J., & Bhattacharjee, A. (2022). Discovery of a novel inhibitor against urokinase-type plasminogen activator, a potential enzyme with a role in atherosclerotic plaque instability. Journal of biomolecular structure & dynamics, 1–11. Advance online publication.

