18th National and 3rd International Conference of Iranian Biophysical chemistry

هجدهمین همایش ملی و سومین همایش بین المللی بیوشیمی فیزیک ایران

25-26 Des, 2024, University of Hormozgan

۶-۵ دی ماه ۱۴۰۳، دانشگاه هرمزگان

A review of AlphaFold, a method for protein structure prediction

Abazar Abouzarkhani Fard[1]*, Zohreh Zahraei[2]

1. Department of Cell and Molecular Biology, Faculty of Chemistry, University of Kashan, Kashan, Iran, E-mail address: abouzarkhanifardamirreza@gmail.com
2. Department of Cell and Molecular Biology, Faculty of Chemistry, University of Kashan, Kashan E-mail address: zahraei@kashanu.ac.ir

Abstract

Introduction: Proteins are regarded as one of the most important biological macromolecules. Identifying their structure will help researchers understand protein's function and practical applications. In the past, experimental and classical methods such as X-ray crystallography or NMR were used, with advantages and disadvantages. Today, new methods based on artificial intelligence have made a big change in predicting the structure of proteins, an example of which is the method called AlphaFold. **Methodology**: Articles related to the present topic were reviewed from databases such as Google Scholar and Pubmed from 2019 to 2024. **Results**: AlphaFold, a method for predicting the structure of proteins in three versions has been introduced to the world. This method, based on deep learning based on convolutional neural networks, processes input data that are often collected from PDB and accurately and quickly predicts the structure of proteins. This method can be widely used in medical, biological, education, industrial and production cases. **Discussion and conclusion**: Different methods are used in protein structure prediction. New methods that are based on artificial intelligence have attracted the attention of researchers because they have high accuracy and speed. However in order to gain confidence, they need to be measured.

Key words: Protein, Protein structure prediction, Artificial intelligence, Deep neural network, AlphaFold

1. Introduction

Proteins are considered one of the most important biological macromolecules in the body of living organisms, and they have important functions that are enzymatic, structural, hormonal, and transportation. From the structural point of view, proteins are classified into four levels. The primary structure includes the sequence of amino acids that are put together by peptide bonds. Secondary structures are created by connecting amino acid sequences by hydrogen bonds, such as alpha helix and beta sheets. The third structure of proteins, three-dimensional (3D), results from interactions between the amino acids' side chains (R), often formed by hydrophobic, disulfide and ionic bonds. The fourth structure is created by juxtaposition of several polypeptide chains (subunits). Predicting the structure of proteins has long been an important and challenging issue because there is this view that the structure and function of proteins are affected by each other and are very important in the dynamics of proteins. First, classical and experimental methods were used, such as X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR), each with its advantages and disadvantages. For example, you can get detailed information about the structure using the X-ray crystallography method. However, it requires crystallizing the protein and spending a lot of time to perform the test process. In the NMR method, you can examine the proteins in natural conditions, but this method is for the detection of small proteins and requires special equipment [1]. Today, with the advancement of science and technology and the invention of methods based on artificial intelligence, huge developments have taken place in the field of detecting and predicting the structure of proteins, and the subject of this article is one of these practical and accurate methods called AlphaFold. Before dealing with this method, we intend to make an introduction about artificial intelligence, its levels and its objective applications.

Artificial intelligence (AI), as one of the leading sciences in the field of computer science, has made a big change in the world. Artificial intelligence is a whole that can be divided into smaller parts, such as machine learning and deep learning. In the AlphaFold method, we focus on deep learning, which uses neural networks to classify, predict and process patterns. The neural network works similarly to the human brain's neural network and examines data in processing units called neurons. The neural network comprises three parts: The first is the input layer, through which information is entered. The second part is called hidden layers, where the input data is processed. Finally, in the third part of the output layer, the processed data is presented to us as results. If the neural network has many hidden layers, it is called Deep Neural Networks (DNNs). Deep neural networks, also called recurrent neural network (RNN), are used in data processing, such as text. Another type is the Convolutional Neural Network (CNN), recognizes and processes images. The AlphaFold method often combines a

convolutional neural network (CNN) with biological methods to predict the structure of proteins [2, 3].

In this article, we will provide an overview of the AlphaFold method and examine its mechanism of action. We will also carefully analyze its advantages and disadvantages and study its applications in different sciences.

2 – Methodology

First, databases such as Google Scholar and PubMed were reviewed. After screening, qualified articles that were highly related to the present topic from 2019 to 2024 were selected and reviewed.

3 - Results

3-1 Approach, History

As an up-to-date and powerful method for predicting the structure of proteins, the AlphaFold has created a huge change in biology. AlphaFold was first developed by Deepmind company, a subsidiary of Alphabet (the parent company of Google). So far, three versions of it have been published. AlphaFold 1 was released in December 2018, AlphaFold 2 in July 2021 and AlphaFold 3 in May 2024. AlphaFold 1 first participated in the CASP13 (Critical Assessment of Techniques for Protein Structure Prediction) competition, which is a biennial competition for protein structure prediction, and surprised everyone [1, 4]. In CASP14, AlphaFold 2 was introduced and evaluated, which had some improvements compared to the previous one. In this competition, the methods of predicting the structure of proteins are based on two principles. First, evolutionary analysis is used to map the remaining changes in the protein sequence with physical contact in the protein structure. They used deep neural networks to accurately identify patterns in protein sequences [1].

AlphaFold 1 could predict the single chain structure of proteins, but AlphaFold 2 could also predict proteins with high complexity. AlphaFold 2 codes are available to the public and can be obtained for free on GITHUB. AlphaFold 3 is not yet commercially available to the public, but bright horizons can be imagined for it [1].

3-2 Advantages and disadvantages

There is much discussion about the advantages and disadvantages of this method. In this section, we are trying to examine these issues. AlphaFold 1 can identify single-chain protein structures, but AlphaFold 2, while maintaining this ability, tried to predict and identify more

complex and larger proteins by improving its patterns and building blocks. Due to the fact that AlphaFold uses a deep neural network to process information, the speed and volume of processing are also higher. Along with the mentioned benefits, the AlphaFold method also has some disadvantages, which we will mention. One of the disadvantages of the AlphaFold method is the need for a lot of data (ACCURATE). On the other hand, alpha has not been very successful regarding the allosteric mechanisms of alpha fold. Of course, this issue can be resolved with molecular dynamics (MD) methods, which also involve relatively high costs. Also, this method cannot correctly predict the dimensions of other protein structures, such as metal ions, cofactors and other ligands. It is also important that despite the public interest in using new methods in various sciences, experimental methods and the tools and methods of doctors and biologists are still efficient, and we should use the combination of different methods in our predictions [1, 5-9].

3-3 Objective applications of AlphaFold

Predicting the structure of proteins has many applications in medicine, industry, education, etc., which we will discuss now. So far, more than 180,000 protein structures have been discovered and identified, and they have been made available for researchers in the PDB database. Despite these efforts and the huge amount of information, there are still no structures of pathogenic agents and many other proteins that are essential for human health [7]. Diagnosing and predicting of the three-dimensional structure of proteins can be a tool for researchers in producing small molecule drugs or therapeutic antibodies, ultimately leading to more effective and selective development of drugs for diseases such as cancer, neurological and infectious disorders. Methods based on artificial intelligence, such as AlphaFold, can be used to make and design proteins that have a therapeutic aspect and are used in the treatment of diseases such as Parkinson's, heart failure, oncology, and immunology [1]. Among other uses of AlphaFold, it is used in education to make complex lesson concepts more concrete and understandable for students with the help of computer models [1]. Another application of predicting the structure of proteins is that new proteins can be made. These artificial proteins have desired properties, such as new catalytic activity, which can ultimately lead to the creation of biosensors and new drugs [1].

Structure-based drug discovery (SBDD) is another important field widely used in medicine and biology. It is expected that methods such as AlphaFold, especially its version 2, will play a constructive and important role in this case and be used. In order to improve the accuracy of AlphaFold 2 predictions, it can be combined with methods such as molecular dynamics [10]. AlphaFold can be used to study the molecular interactions of viral infections and design drugs suitable for these infections. So far, AlphaFold has been mentioned in

extensive research related to viruses such as monkeypox (MPXV), hepatitis E (HEV) and other human and animal viruses [11].

4- Discussion and Conclusion

Predicting and determining the structure of proteins is important for the mutual relationship between structure and function. Empirical and classical forecasting techniques, despite their popularity, also have disadvantages. On the other hand, with the emergence of artificial intelligence and its use in various sciences, the luck of human society applying accurate methods based on artificial intelligence in the biological and medical sciences are increasing. One of these techniques is called alpha fold, which accurately and quickly predicts the structure of proteins and has many applications in medicine, education, industry production, etc. It is important to mention that each method has its own advantages and disadvantages, and blindly accepting without measuring the results of a particular technique will not be very reliable and acceptable. Therefore, we invite researchers in the fields of biology, medicine, pharmaceutical sciences and computer science to do more research.

Refrences

[1] Chen, L., Li, Q., Nasif, K. F. A., Xie, Y., Deng, B., Niu, S., ... & Xie, C. Y. (2024). AI-driven deep learning techniques in protein structure prediction. International journal of molecular sciences, 25(15), 8426

[2] Torrisi, M., Pollastri, G., & Le, Q. (2020). Deep learning methods in protein structure prediction. Computational and Structural Biotechnology Journal, 18, 1301-1310

[3] Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. Minimally Invasive Therapy & Allied Technologies, 28(2), 73-81.

[4] AlQuraishi, M. (2019). AlphaFold at CASP13. Bioinformatics, 35(22), 4862-4865.

[5] Gomes, P. S., Gomes, D. E., & Bernardi, R. C. (2022). Protein structure prediction in the era of AI: Challenges and limitations when applying to in silico force spectroscopy. Frontiers in Bioinformatics, 2, 983306.

[6] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., ... & Hassabis, D. (2021). Protein complex prediction with AlphaFold-Multimer. biorxiv, 2021-10.

[7] Nussinov, R., Zhang, M., Liu, Y., & Jang, H. (2022). AlphaFold, artificial intelligence (AI), and allostery. The Journal of Physical Chemistry B, 126(34), 6372-6383.

[8] Perrakis, A., & Sixma, T. K. (2021). AI revolutions in biology: The joys and perils of AlphaFold. EMBO reports, 22(11), e54046.

[9] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. nature, 596(7873), 583-589.

[10] Paiardini, A. (2023). Protein Structure Prediction in Drug Discovery. Biomolecules, 13(8), 1258.

[11] Gutnik, D., Evseev, P., Miroshnikov, K., & Shneider, M. (2023). Using AlphaFold predictions in viral research. Current Issues in Molecular Biology, 45(4), 3705-3732