Title

Research of predicting protein three-dimensional structure by Alphafold2

Author information

Yantong Ma,College of Life Science and Technology,Gansu agricultural university,730070,Lanzhou, China, 3193127256@qq.com

Xu Lu, College of Horticuture,Gansu agricultural universuty,730070 Lanzhou, China, 2457180577@qq.com

Lei Ma,College of Agromomy,Gansu agricultural university,730070,Lanzhou,China, 2467972188@qq.com

Jinyu Bao,College of Horticuture, Gansu agricultural universuty,730070,Lanzhou,China, 1031817055@qq.com

Jie Yang,College of Life Science and Technology ,Gansu agricultural university, 730070,Lanzhou, China, 546760375@qq.com

Kang Li, College of Life Science and Technology ,Gansu agricultural university, 730070,Lanzhou, China, 552185789@qq.com

Sheng Li, College of Life Science and Technology,Gansu agricultural university, 730070， Lanzhou, China, lish@gsau.edu.cn

Shaoying Ma, Basic Experimental Teaching Center,Gansu agricultural university, Lanzhou, China, mashy@gsau.edu.cn

Joint first author: Yantong Ma and Xu Lu

Corresponding author: Sheng Li and Shaoying Ma.

Abstract: In this paper, firstly, Alphafold2 was used to calculate the three-dimensional structure (predicting structure) of three proteins 2LY9, 6Y4F and 6YJ1, and compared with the protein structure (experimental structure) published by NCBI. Then, the quality of predicting structure was evaluated. The results showed that the RMSD values of the predicting structure were less than 3 Å, which was indicated that the difference distance between the predicted structure and the experimental structure was less than 0.3nm. TM score was higher than 0.5, which showed that the predicting structure by Alphafold2 had similar folding structure with the experimental structure. The ramachandran plot showed that the residues in most favoured regions [A,B,L], in additional allowed regions [a,b,l,p] and in generously allowed regions [~a,~b,~l,~p] was accounted for more than 90% of the whole number of non-glycine and non-proline residues. The above results showed that the function of predicting three-dimensional structure of protein by Alphafold2 was very highly accurate. Subsequently, Alphafold2 was used to predict the three-dimensional structure of protein related to auxin synthesis, and the stereo chemical quality of protein structure was checked through residue geometry and overall geometry. The ramachandran plot showed that the residues in most favoured regions [A,B,L], in additional allowed regions [a,b,l,p] and in generously allowed regions [~a,~b,~l,~p] was accounted for more than 90% of the whole number of non-glycine and non-proline residues. Therefore, the predicted structure by Alphafold2 could highly be as the experimental structure of protein.

Keywords: Alphafold2; Three-dimensional structure; Protein; Quality evaluation

Abbreviations: TM-score The template modeling score; RMSD Root mean square deviation

Statements and Declarations

There is no conflict of interest

Acknowledgments

This work was supported by the Industrial support of Gansu Provincial Department of Education (2021CYZC-55), Major special projects of Gansu Provincial (20ZD7NA007-01), National Natural Science Foundation of China (31860067), Key research and development projects of Gansu Provincial (21YF5NA090), Natural Science Foundation of Gansu Provincial (21JR7RA822), and the Innovation Fund of Gansu Provincial Higher Education (2021B-136).

Research of predicting protein three-dimensional structure by Alphafold2

Abstract: In this paper, firstly, Alphafold2 was used to calculate the three-dimensional structure (predicting structure) of three proteins 2LY9, 6Y4F and 6YJ1, and compared with the protein structure (experimental structure) published by NCBI. Then, the quality of predicting structure was evaluated. The results showed that the RMSD values of the predicting structure were less than 3 Å, which was indicated that the difference distance between the predicted structure and the experimental structure was less than 0.3nm. TM score was higher than 0.5, which showed that the predicting structure by Alphafold2 had similar folding structure with the experimental structure. The ramachandran plot showed that the residues in most favoured regions [A,B,L], in additional allowed regions [a,b,l,p] and in generously allowed regions [~a,~b,~l,~p] was accounted for more than 90% of the whole number of non-glycine and non-proline residues. The above results showed that the function of predicting three-dimensional structure of protein by Alphafold2 was very highly accurate. Subsequently, Alphafold2 was used to predict the three-dimensional structure of protein related to auxin synthesis, and the stereo chemical quality of protein structure was checked through residue geometry and overall geometry. The ramachandran plot showed that the residues in most favoured regions [A,B,L], in additional allowed regions [a,b,l,p] and in generously allowed regions [~a,~b,~l,~p] was accounted for more than 90% of the whole number of non-glycine and non-proline residues. Therefore, the predicted structure by Alphafold2 could highly be as the experimental structure of protein.

Keywords: Alphafold2; Three-dimensional structure; Protein; Quality evaluation

1 Introduction

Protein is essential for life. Predicting protein structure is of great significance for studying protein function [1-4]. Analyzing protein structure is an important part of proteome planning, which helps to understand the role of proteins, how proteins exercise biological functions, and the interaction between proteins (or other molecules), which is very important for biology, medicine and pharmacy [5]. For unknown or newly discovered proteins, functional annotation can be carried out through sequence analysis and three-dimensional structure prediction to confirm functional units or domains, so as to provide a reliable basis for designing new proteins or transforming existing proteins [6].

It is reported that more than 100,000 protein structure have been determined [7], but this is only a small part of the billions of known proteins [8-9]. “Predicting the 3D structure of protein given an amino acid sequence”, a seemingly difficult problem, will be solved with the emergence of Alphafold2 [10]. At present, the methods with high accuracy for predicting protein three-dimensional structure are mainly divided into two categories: Rosetta with three track nerve as the basic prediction logic and Alphafold2 [11]. By using the neural network based method Alphafold2, the three-dimensional structure of almost the whole human proteome has been predicted and available（https://www. Alphafold2. EBI. Ac.uk) [12]. At 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP14), Alphafold2 won the champion because the predicted protein three-dimensional structure by Alphafold2 are extremely close to experimental structure of protein (https://predictioncenter.org/casp14/index.cgi)[13]. In this study, firstly, the experimental structure of three proteins 2LY9, 6Y4F and 6YJ1 with the three-dimensional structure predicted by Alphafold2 was compared to confirm the reliability of Alphafold2 function of predicting protein structure. Then, prediction structure and protein quality evaluation of 10 homologous proteins related to auxin synthesis was carried out, which will contribute to prediction structure and function research of protein.

2 Experimental Design

2.1 Confirming the reliability of Alphafold2 function of predicting protein structure

Firstly, calculating the three-dimensional structure of the protein 2LY9 [14], 6Y4F [15] and 6YJ1 [14] by Alphafold2 and plotted by Pymol. Then, comparing the three-dimensional structure of the protein 2LY9, 6Y4F and 6YJ1 by Alphafold2 with by NCBI to confirm the reliability of Alphafold2 function of predicting protein structure.

2.2 Predicting the three-dimensional structure of protein by Alphafold2

The three-dimensional structure of protein (Mtarf03,10,12,23; PsARF15,17,32,33,36; Atarf17) predicted by Alphafold2 was performed. The quality assessment analysis of three-dimensional structure of protein was performed.

3 Data download and processing

3.1Data download

Protein 2LY9, 6Y4F, 6YJ1, three-dimensional structure data were downloaded in NCBI, and amino acid sequence files of 10 protein homologous families (Mtarf03,10,12,23; PsARF15,17,32,33,36; Atarf17) related to auxin synthesis were downloaded.

3.2 Alphafold2 calculates the spatial structure of proteins

Alphafold2 was used to determine the spatial coordinates of the selected base sequence and calculate the spatial structure of the protein.

3.3 Alphafold2 predicts protein structure analysis

The UCLA-DOE SAVE platform was used to perform PROCHECK analysis, PROVE analysis and ERRAT analysis for the protein three-dimensional structure. TM-score and R.M.S.D were calculated between the prediction structure and the experiment structure on RCSB platform (https://www.rcsb.org/alignment).

4 Results

4.1 Alignment of Protein structure

4.1.1 Protein 2LY9 three-dimensional structure alignment

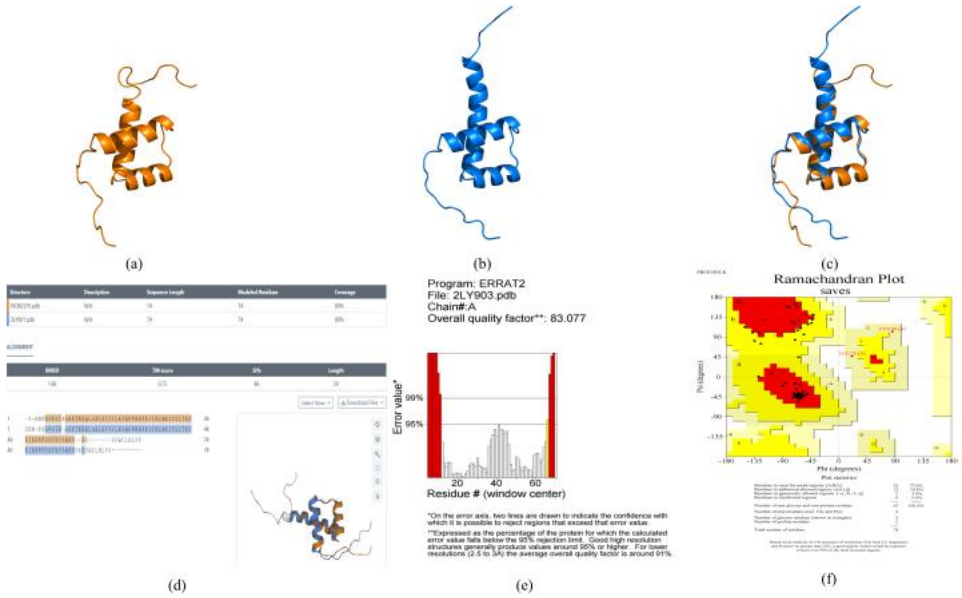

Fig.1 Alignment of the three-dimensional structure of protein 2LY9. a,The three-dimensional protein structure(experimental structure) data downloaded by NCBI(Image by pymol). b,The protein three-dimensional structure(predicting structure) by Alphafold2 (Image by Pymol). c,The comparison diagram about the three-dimensional structure in NCBI and the three-dimensional structure obtained by Alphafold2 (the comparison image by pymol). d,The protein structure comparison TM-score and R.M.S.D. e,The protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error difference, ** represents the percentage of proteins whose calculation error value is lower than the 95% rejection limit.f,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. The same below .

Fig.1 Protein 2LY9 three-dimensional structure alignment

In Fig.1a,b,c, it can be seen that the experimental structure and the predicted structure is exceedingly similar and the contact ratio is high. As shown in Fig.1d, the results showed that the protein alignment TM-score was 0.73, R.M.S.D was 1.85 Å < 3 Å, which indicates that the three-dimensional structure distance of experimental structure and the predicted structure in spatial coordinates is 0.186nm, and they have similar folding position (TM score=0.73). The overall score of the 2LY9 experimental structure is 83.077 (Fig.1e). As shown in Fig.1f, the total number of residues of predicted structure of 2LY9 was 74.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 67, 2, 3 and 2. For prediction structure of protein 2LY9 , the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 52,13,2 and 0, and was accounted for 77.6%, 19.4%, 3.0% and 0.0% of the number of non-glycine and non-proline residues.

4.1.2 Protein 6Y4F three-dimensional structure alignment

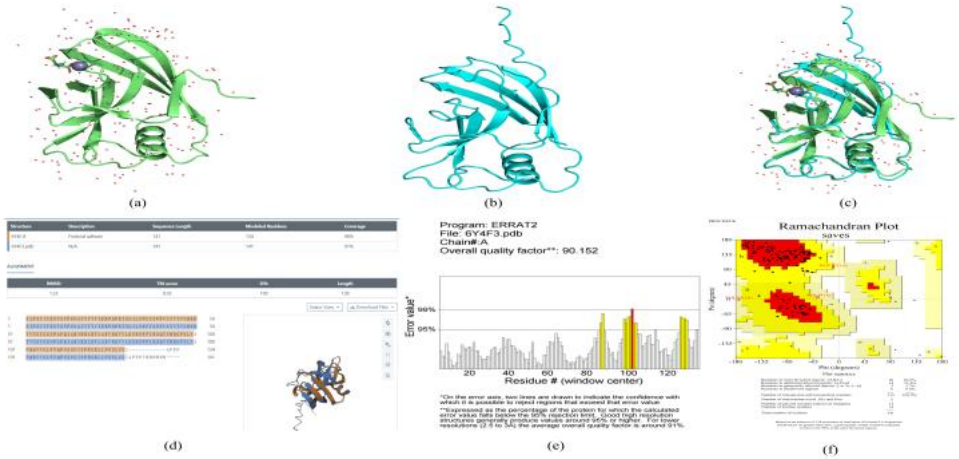

Fig.2 Alignment of the three-dimensional structure of protein 6Y4F. a,The three-dimensional protein structure(experimental structure) data downloaded by NCBI(Image by pymol). b,The protein three-dimensional structure(predicting structure) by Alphafold2 (Image by Pymol). c,The comparison diagram about the three-dimensional structure in NCBI and the three-dimensional structure obtained by Alphafold2 (the comparison image by pymol). d,The protein structure comparison TM-score and R.M.S.D. e,The protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error difference, ** represents the percentage of proteins whose calculation error value is lower than the 95% rejection limit.f,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform.

Fig.2 Protein 6Y4F three-dimensional structure alignment

In Fig.2a,b,c, it can be seen that the experimental structure and the predicted structure is exceedingly similar and the contact ratio is high. In Fig.2d, the results showed that the protein alignment TM-score was 0.93, R.M.S.D was 1.22 Å < 3 Å, which indicates that the three-dimensional structure distance of experimental structure and the predicted structure in spatial coordinates is 0.122nm,and they have similar folding position (TM score=0.93). The overall score of the 6Y4F experimental structure is 90.152 (Fig.2e). As shown in Fig.2f, the total number of residues of predicted structure of 6Y4F was 141.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 113, 2, 13 and 13. For prediction structure of protein 6Y4F, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 96, 14, 3 and 0, and was accounted for 85.0%, 12.4%, 2.7% and 0.0% of the number of non-glycine and non-proline residues.

4.1.3 Protein 6YJ1 three-dimensional structure alignment

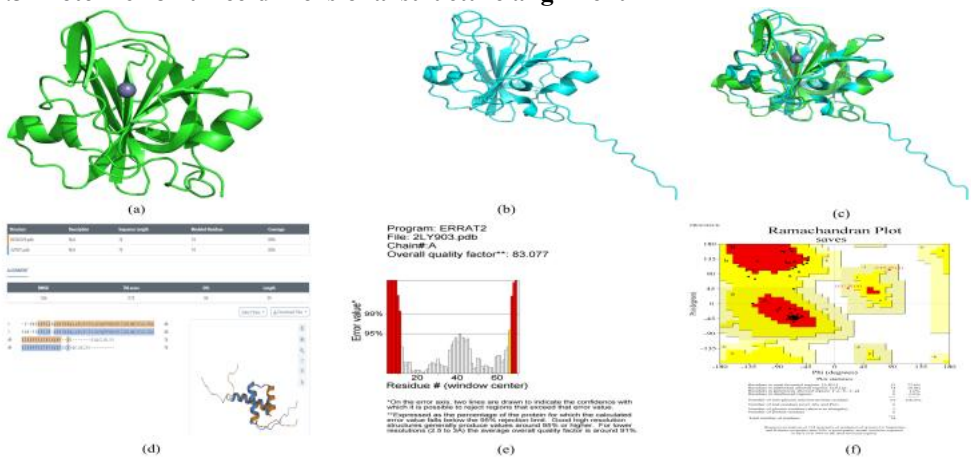

Fig.3 Alignment of the three-dimensional structure of protein 6YJ1. a,The three-dimensional protein structure(experimental structure) data downloaded by NCBI(Image by pymol). b,The protein three-dimensional structure(predicting structure) by Alphafold2 (Image by Pymol). c,The comparison diagram about the three-dimensional structure in NCBI and the three-dimensional structure obtained by Alphafold2 (the comparison image by pymol). d,The protein structure comparison TM-score and R.M.S.D. e,The protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error difference, ** represents the percentage of proteins whose calculation error value is lower than the 95% rejection limit.f,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform.

Fig.3 Protein 6YJ1 three-dimensional structure alignment

In Fig.3a,b,c, it can be seen that the experimental structure and the predicted structure is exceedingly similar and the contact ratio is high. In Fig.3d, the results showed that the protein alignment TM-score was 0.96,

R.M.S.D was 1.1 Å < 3 Å, which indicates that the three-dimensional structure distance of experimental structure and the predicted structure in spatial coordinates is 0.111nm,and they have similar folding position (TM score = 0.96). The overall score of the 6YJ1 experimental structure is 83.077(Fig.3e). As shown in Fig.3f, the total number of residues of predicted structure of 6Y4F was 74.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 67, 2, 3and 2. For prediction structure of protein6YJ1, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 52, 13, 2 and 0, and was accounted for 77.6%, 19.4%, 3.0% and 0.0% of the number of non-glycine and non-proline residues.

The above results indicated that the overall score of protein three-dimensional structure obtained by Alphafold2 conforms to the conformation of protein model and the rules of stereochemistry. In other words, it was proved that the predicted structure by Alphafold2 is credible. Then, prdicting structure of proteins related with auxin synthesis was performed by Alphafold2.

4.2 Predicting the three-dimensional structure of the protein by Alphafold2

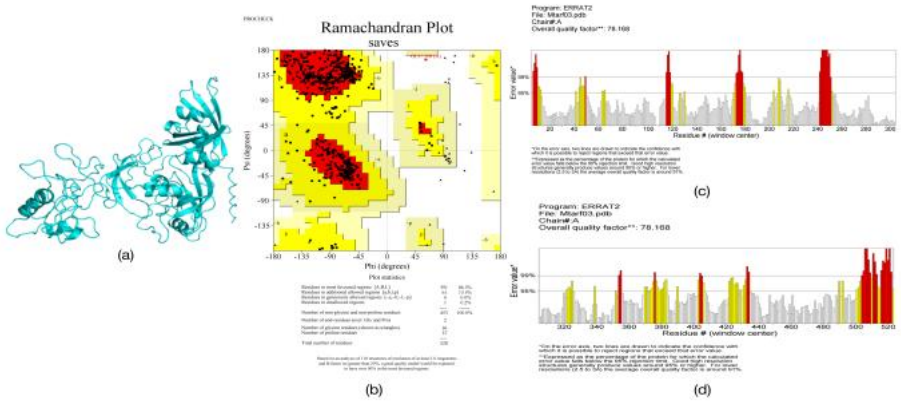

Fig.4, the three-dimensional structure of Mtarf03 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.4 Protein Mtarf03 three-dimensional structure and analysis

In Fig.4b, the total number of residues was 528.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 453, 2, 36 and 37. For prediction structure of protein Mtarf03, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 391,61,0 and 1, and was accounted for 86.3%, 13.5%, 0.0% and 0.2% of the number of non-glycineand non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Mtarf03 was 78.168 (Fig.4c).

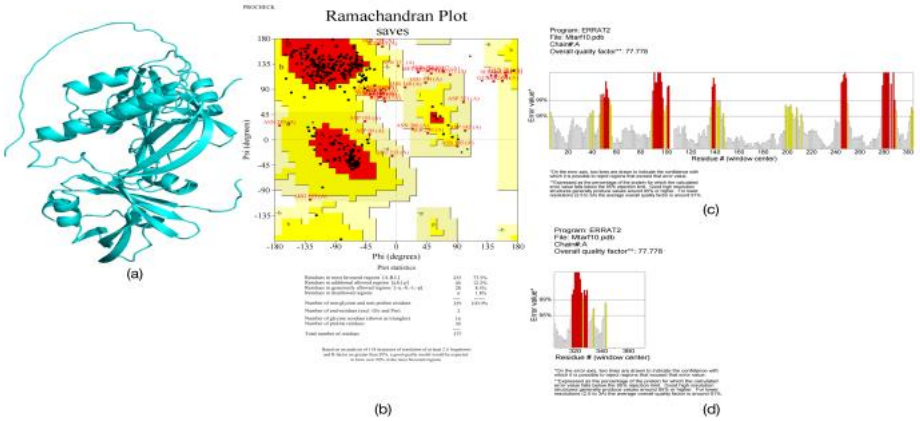

Fig.5, the three-dimensional structure of Mtarf10 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.5 Protein Mtarf10 three-dimensional structure and analysis

In Fig.5b, the total number of residues was 377.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 329, 2, 16 and 30. For prediction structure of protein Mtarf10, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 255, 40, 28 and 6, and was accounted for 77.5%, 12.2%, 8.5% and 1.8% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Mtarf10 was 77.778 (Fig.5c).

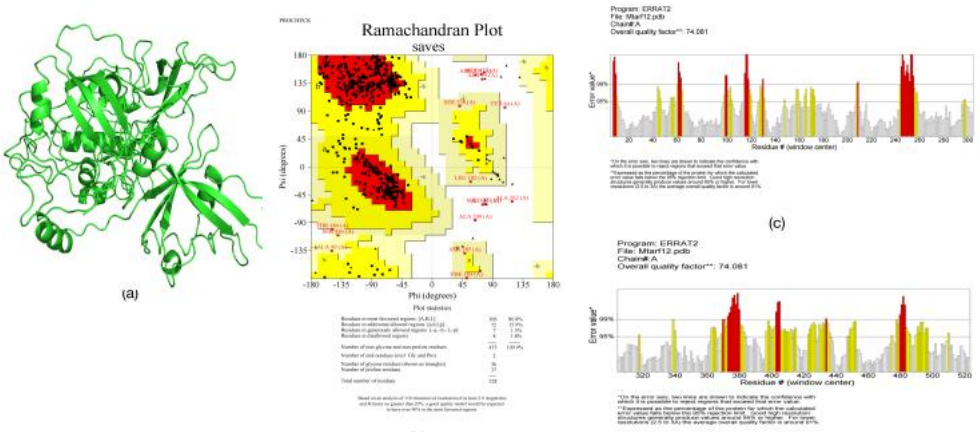

Fig.6, the three-dimensional structure of Mtarf12 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c, Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.6 Protein Mtarf12 three-dimensional structure and analysis

In Fig.6b, the total number of residues was 528.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 453, 2, 36 and 37. For prediction structure of protein Mtarf12, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 366, 72, 7 and 8, and was accounted for 80.8%, 15.9%, 1.5% and 1.8% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Mtarf12 was 74.081 (Fig.6c).

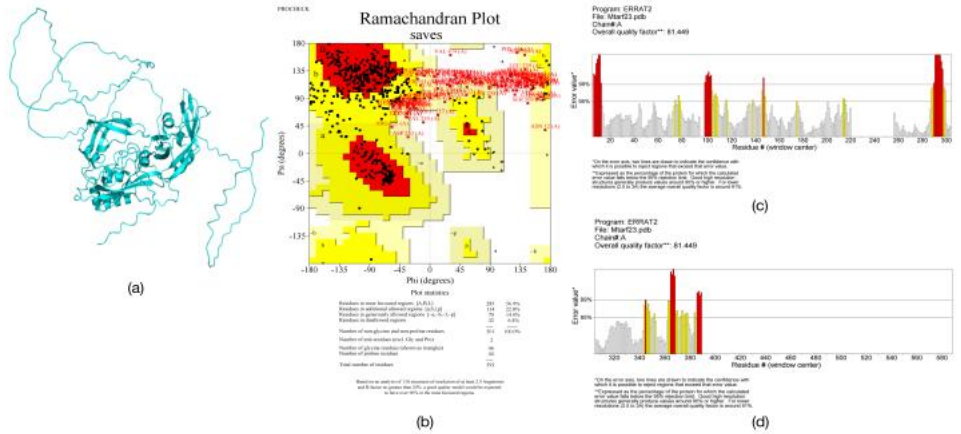

Fig.7, the three-dimensional structure of Mtarf23 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c, Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.7 Protein Mtarf23 three-dimensional structure and analysis

In Fig.7b, the total number of residues was 593.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 501, 2, 46 and 44. For prediction structure of protein Mtarf23, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in

disallowed regions, was 285, 114, 70 and 32, and was accounted for 56.9%, 22.8%, 14.0% and 6.4% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Mtarf23 was 81.449 (Fig.7c).

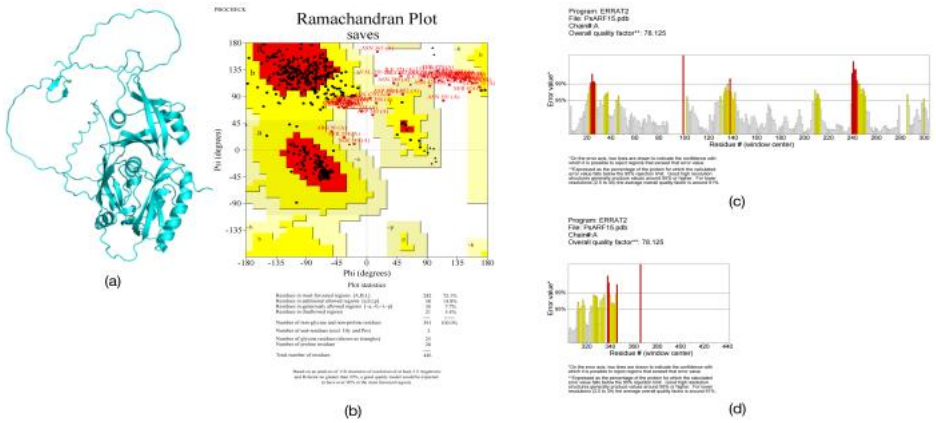

Fig.8, the three-dimensional structure of Psarf15 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c, Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.8 Protein Psarf15 three-dimensional structure and analysis

In Fig.8b, the total number of residues was 446.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 391, 2, 25 and 28. For prediction structure of protein PsARF15, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions[~a,~b,~l,~p],in disallowed regions, was 282, 58, 30 and 21, and was accounted for 72.1%, 14.8%, 7.7% and5.4% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Psarf15 was 78.125(Fig.8c).

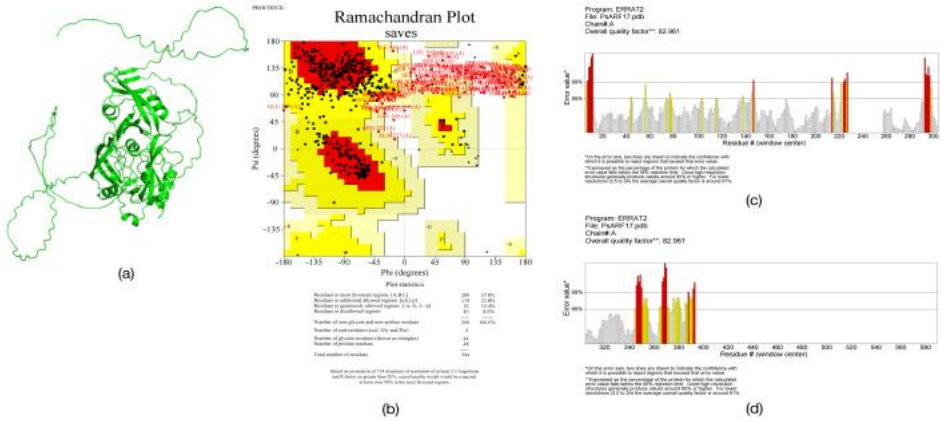

Fig.9, the three-dimensional structure of Psarf17 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.9 Protein Psarf17 three-dimensional structure and analysis

In Fig.9b, the total number of residues was 594.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 500, 2, 44 and 48. For prediction structure of protein PsARF17, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 289, 119, 52 and 40, and was accounted for 57.8%, 23.8%, 10.4% and 8.0% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Psarf17 was 82.961 (Fig.9c).

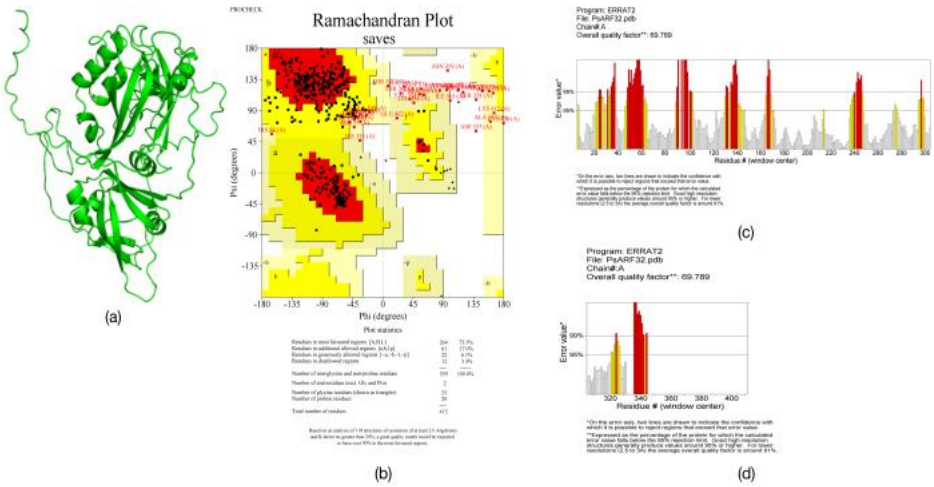

Fig.10, the three-dimensional structure of Psarf32 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.10 Protein Psarf32 three-dimensional structure and analysis

In Fig.10b, the total number of residueswas 415.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 359, 2, 25 and 29. For prediction structure of protein PsARF32, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 264, 61, 22 and 12, and was accounted for 73.5%, 17.0%, 6.1% and 3.3% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Psarf32 was 69.789 (Fig.10c).

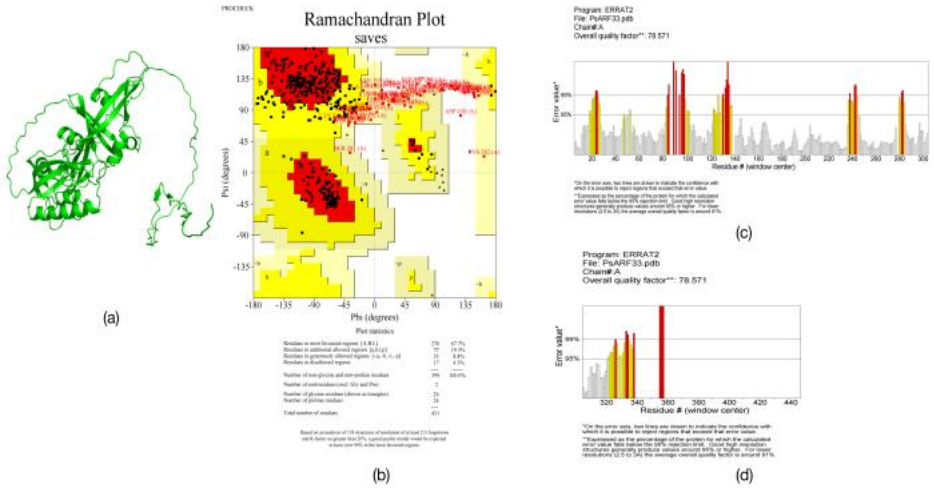

Fig.11, the three-dimensional structure of Psarf33 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform.c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.11 Protein Psarf33 three-dimensional structure and analysis

In Fig11b, the total number of residues was 451.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 399, 2, 26 and 24. For prediction structure of protein PsARF33, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 270, 77, 35 and 17, and was accounted for 67.7%, 19.3%, 8.8% and 4.3% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Psarf33 was 78.571 (Fig.11c).

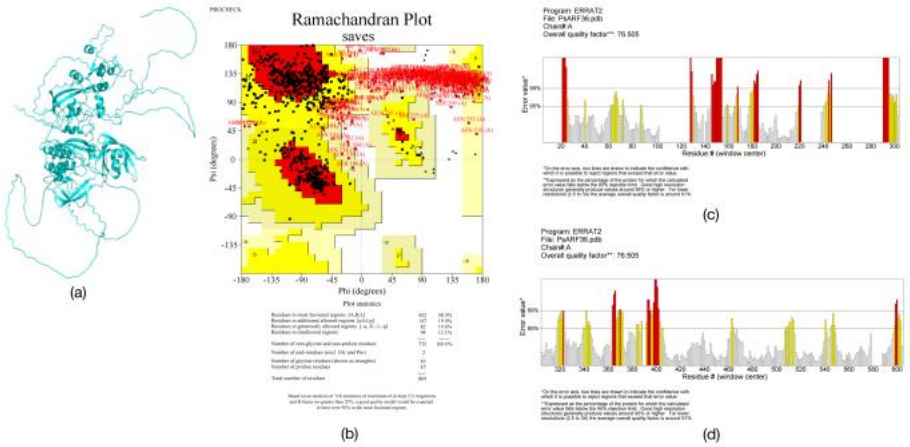

Fig.12, the three-dimensional structure of Psarf36 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.12 Protein Psarf36 three-dimensional structure and analysis

In Fig.12b, the total number of residues was 869.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 775, 2, 45 and 47. For prediction structure of protein PsARF36, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 452, 147, 82 and94, and was accounted for 58.3%, 19.0%, 10.6% and12.1% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Psarf36 was 76.505 (Fig.12c).

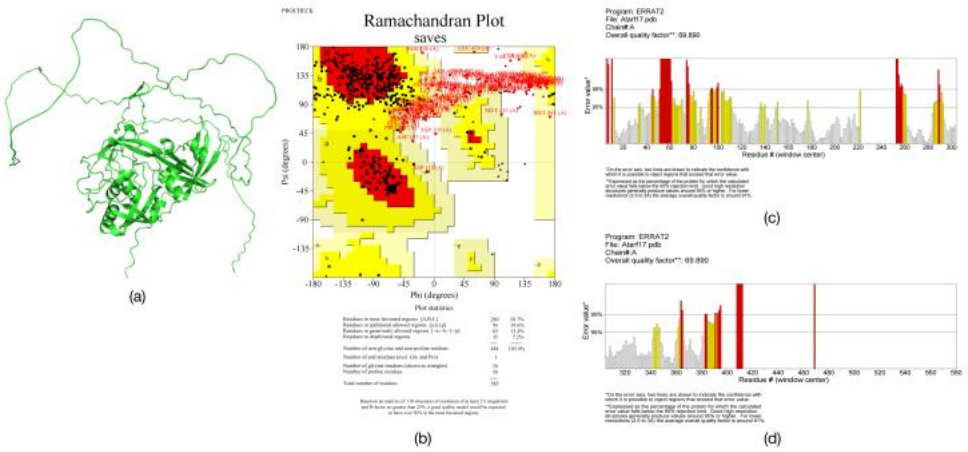

Fig.13, the three-dimensional structure of Atarf17 protein. a.The three-dimensional structure of the protein calculated by alphfold2 (Image by pymol). b,The ramachandran plot by PROCHECK analysis of protein three-dimensional structure on UCLA-DOE SAVE platform. c,Protein error evaluation analysis. The two lines on the error axis represent the confidence of rejecting the area exceeding the error value. ** represents the percentage of protein whose calculated error value is lower than the 95% rejection limit.

Fig.13 Protein Atarf17 three-dimensional structure and analysis

In Fig.13b, the total number of residues was 585.The number of non-glycine and non-proline residues, the end-residue (excluding glycine and proline), the glycine residues (shown in triangle) and the proline residue was 484, 1, 50 and 50. For prediction structure of protein Atarf17, the number of residues in most favorable regions [A,B,L], in additional allowed regions [a,b,l,p], in generously allowed regions [~a,~b,~l,~p], in disallowed regions, was 289, 95, 65 and 35, and was accounted for 59.7%, 19.6%, 13.4% and 7.2% of the number of non-glycine and non-proline residues. Protein error assessment analysis showed that the overall score of AlphFold2 predicted the three-dimensional structure of Atarf17 was 69.890 (Fig.13c).

In conclusion, the protein three-dimensional structure predicted by Alphafold2 conforms to the rules of conformational stereochemistry of the selected protein model.

5 Disscussion

So far, experts and scholars are working hard to solve the problem of "Predicting the 3D structure of protein given an amino acid sequence"[16], but the accuracy score of predicting the 3D model can only reach about 40 (full score was 100). However, the accuracy score of predicting the 3D model can reach about 92.4 due to the use of Alphafold2.There was only one atomic width difference between the predicted protein structure by Alphafold2 and the real structure of the protein, which really solved the problem of protein folding [17]. In 2020, Alphafold2 was selected as one of the top ten breakthroughs of science, known as the "revolutionary" breakthrough in structural biology and a milestone in the field of protein research [18].

In this study, the prediction structure of proteins 2LY9, 6Y4F, 6YJ1by Alphafold2 are compared with the experimental structure. It is found that the RMSD is less than 3 Å. The overlap gap of protein spatial structure is less than 0.3nm between experimental structure and prediction structure, and the m-score is higher than 0.5, which indicates that the three-dimensional structure of the protein by Alphafold2 has a similar folding with the known structure [13]. In summary, the prediction function of protein three-dimensional structure of Alphafold2 is credible and the prediction accuracy can reach the experimental level.

After the three-dimensional structure of 10 homologous proteins was accurately predicted by Alphafold2, the three-dimensional structure of these proteins was pre checked, and the three-dimensional chemical quality of protein structure was checked by analyzing the geometric structure of residues and the overall geometric structure. The ramachandran plot showed that the residues in most favoured regions [A,B,L], in additional allowed regions [a,b,l,p] and in generously allowed regions [~a,~b,~l,~p] was accounted for more than 90% of the whole number of non-glycine and non-proline residues. The above conclusion shows that the three-dimensional structure conformation of 10 auxin synthesis related homologous proteins predicted by Alphafold2 conforms to the rules of stereochemistry [19-21].

Conflict of interest

The authors declare that they have no conflict of interest.

Reference

[1] Cramer Patrick. AlphaFold2 and the future of structural biology. [J]. Nature structural & molecular biology, 2021, 28(9):704-705. https://doi.org/10.1038/s41594-021-00650-1

[2] Michael C. Thompson, Todd O. Yeates, Jose A. Rodriguez. Advances in methods for atomic resolution macromolecular structure determination. [J]. F1000 Research, 2020, 9:667. https://doi.org/10.12688/f1000research.25097.1

[3] Xiao-chen Bai, Greg McMullan, Sjors H.W Scheres. How cryo-EM is revolutionizing structural biology. [J]. Trends in Biochemical Sciences. 2015, 40(1):49-57. https://doi.org/10.1016/j.tibs.2014.10.005

[4] Mariusz Jaskolski, Zbigniew Dauter, Alexander Wlodawer. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. [J]. The FEBS journal, 2014, 281(18):3985-4009. https://doi.org/10.1111/febs.12796

[5] Kurt Wüthrich. The way to NMR structure of proteins. [J]. Nature Structural Biology, 2001, 8(11):923-5. https://doi.org/10.1038/nsb1101-923

[6] Junna Hayashi, Jennifer Ton, Sparsh Negi, Daniel E. K. M. Stephens, Dean L. Pountney, Thomas Preiss, John A. Carver. The Effect of Oxidized Dopamine on the Structure and Molecular Chaperone Function of the Small Heat-Shock Proteins, αB-Crystallin and Hsp27. [J]. International Journal of Molecular Sciences, 2021, 22(7):3700. https://doi.org/10.3390/ijms22073700

[7] Susanne E. Ulbrich, Thomas Frohlich, Katy Schulke, Eva Englberger, Nadine Waldschmitt, Georg J. Arnold, Horst-Dieter Reichenbach, Myriam Reichenbach, Eckhard Wolf, Heinrich H.D. Meyer, Stefan Bauersachs. Evidence for Estrogen-Dependent Uterine Serpin (SERPINA14) Expression During Estrus in the Bovine Endometrial Glandular Epithelium and Lumen1. [J]. Biology of Reproduction, 2009, 81(4):795-805. https://doi.org/10.1095/biolreprod.108.075184

[8] wwPDB, consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. [J]. Nucleic Acids Research, 2019, 8; 47(D1):D520-D528. https://doi.org/10.1093/nar/gky949

[9] Alex L. Mitchell, Alexandre Almeida, Martin Beracochea, Miguel Boland, Josephine Burgin, Guy Cochrane, Michael R. Crusoe, Varsha Kale, Simon C. Potter, Lorna J. Richardson, Ekaterina Sakharova, Maxim Scheremetjew, Anton Korobeynikov, Alex Shlemov, Olga Kunyavskaya, Alla Lapidus, Robert D.

Finn. MGnify: the microbiome analysis resource in 2020. [J]. Nucleic Acids Research. 2020, 8; 48(D1):D570-D578. https://doi.org/10.1093/nar/gkz1035

[10] Martin Steinegger, Milot Mirdita, Johannes Söding. Protein-level assembly increases protein sequence recovery from metagenomic samples manifold. [J]. Nature Methods, 2019, 16(7):603-606. https://doi.org/10.1038/s41592-019-0437-4

[11] Zweckstetter Markus. NMR hawk-eyed view of AlphaFold2 structure. [J]. Protein science: a publication of the Protein Society, 2021, 30(11): 2333-2337. https://doi.org/10.1002/pro.4175

[12] Jad Abbass, Jean-Christophe Nebel. Reduced fragment diversity for alpha and alpha-beta protein structure prediction using rosetta. [J]. Protein and peptide letters, 2017, 24(3):215-222. https://doi.org/10.2174/0929866523666161216124019

[13] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. [J]. Nature, 2021: 583-589. https://doi.org/10.1038/s41586-021-03819-2

[14] Madej Thomas, Lanczycki Christopher J, Zhang Dachuan, Thiessen Paul A, Geer Renata C, Marchler-Bauer Aron, Bryant Stephen H. MMDB and VAST+: tracking structural similarities between macromolecular complexes. [J]. Nucleic acids research, 2014, 42(Database issue): D297-303. https://doi.org/10.1093/nar/gkt1208

[15] Fischer Daniel, Elofsson Arne, Rice Danny, Eisenberg David. Assessing the performance of fold recognition methods by means of a comprehensive benchmark. Pac Symp Biocomput. 1996:300-318.

[16] Andriy Kryshtafovych, Krzysztof Fidelis, John Moult. Progress from CASP6 to CASP7. [J]. Proteins. 2007, 69 (8):194-207. https://doi.org/10.1002/prot.21769

[17] Cramer Patrick. AlphaFold2 and the future of structural biology. [J]. Nature structural & molecular biology, 2021, 28(9): 704-705. https://doi.org/10.1038/s41594-021-00650-1

[18] ZHAO Yun-bo. Can AI Prediction Replace Scientific Experiments?—Focusing on AlphaFold Cracking the Protein Folding Puzzle. [J]. Medicine and Philosophy, 2021, 42(06):17-21.

[19] Rob W.W.Hoo, Chris Sande, Gerrit Vrien. Objectively judging the quality of a protein structure from a ramachandran plot. [J]. Computer Applications in the Biosciences Cabios, 1997, 13(4):425-30. https://doi.org/10.1093/bioinformatics/13.4.425

[20] Laskowski R A, Macarthur M W, Moss D S. PROCHECK: a program to check the stereochemical quality of protein structure. [J]. 1993. https://doi.org/10.1107/S0021889892009944

[21] Giacomo J, Alessandro P. PyMod 3: a complete suite for structural bioinformatics in PyMOL. [J]. Bioinformatics, 2020. https://doi.org/10.1093/bioinformatics/btaa849