Received: 13.06.2024 Revised: 16.09.2024 Accepted: 27.11.2024

BIOLOGICAL SYSTEMS: Theory and Innovation

Journal homepage: https://bioscience.com.ua/en
Vol. 15, No. 4, 2024, 34-50

UDC 57.086.9:574.636:004.4 DOI: 10.31548/biologiya/4.2024.34

Methods of computer modelling of the transcription factor of stress resistance WRKY2 in *Triticum Aestivum*: A comparative analysis

Iryna Demianenko*

PhD of Bioinformation Sciences
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Beresteiskyi Ave., Kyiv, Ukraine
https://orcid.org/0000-0002-0832-3619

Anton Petrovskyi

Junior Software Engineer
CPCS
40000, 18 Kosivshchynska Str., Sumy, Ukraine
https://orcid.org/0009-0006-6267-3608

Igor Levtun

PhD of Biotechnology Sciences
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Beresteiskyi Ave., Kyiv, Ukraine
https://orcid.org/0000-0003-2498-035X

Abstract. The aim of this work was to compare the efficiency of modern computer modelling methods for constructing the three-dimensional structure of the transcription factor TaWRKY2 from Triticum aestivum, which plays a key role in plant stress resistance. The study used traditional homology approaches and machine learning-based methods, including AlphaFold, RoseTTaFold, ESMFold, and OmegaFold. The model obtained with AlphaFold2 (MMseqs2) was the best in terms of Clashscore (99th percentile), MolProbity Score, and ERRAT. The complex with zinc ions and DNA generated by AlphaFold3 showed better results in terms of Z-Score. The AlphaFold method provided biologically relevant structures with the correct location of functional sites. ESMFold provided fast modelling, but showed deviations in the structure, including additional α-helices and lower quality in most metrics. The RoseTTaFold model had an elongated shape with a higher content of α-helices and required additional verification of functional activity. OmegaFold, although it provided the best QMEAN score (0.4), was inferior in other metrics. Additional tools, such as Amber and Chimera, allowed for structure relaxation and analysis of key features, including the spatial arrangement of the zinc fingers (18.7 Å). Evaluation by Verify 3D, ERRAT, Ramachandran Plot, and other metrics revealed the advantages and disadvantages of each approach. The findings confirmed the advantages of machine learning methods for modelling proteins with high functional plasticity. In particular, AlphaFold was recognised as the most effective approach for building models with high accuracy. At the same time, the use of several methods allows considering alternative conformations and interactions, which is important for a deeper understanding of the functional mechanisms of the protein

Keywords: bioinformatics; three-dimensional protein structure; machine learning methods; model evaluation metrics; zinc fingers

Suggested Citation:
Demianenko, I, Petrovskyi, A., & Levtun, I. (2024). Methods of computer modelling of the transcription factor of stress resistance WRKY2 in Triticum Aestivum: A comparative analysis. *Biological Systems: Theory and Innovation*, 15(4), 34-50. doi: 10.31548/biologiya/4.2024.34.

*Corresponding author

INTRODUCTION

Molecular biology often focuses on studying the mechanisms of plant adaptation to stressful conditions, in particular with the involvement of transcription factors (TFs). The WRKY transcription factors are one of the key groups of proteins that regulate plant responses to biotic and abiotic stresses. These factors function by interacting with the W-boxes of target gene promoters, activating or repressing their expression. Significant progress in the understanding of WRKY proteins in woody plants was achieved in the studies of L. Long *et al*. (2023), who focused on the functional features of these proteins, but their structure and mechanisms of interaction remain poorly understood.

Computer modelling technologies have opened up new perspectives in the study of protein structures, in particular, machine learning methods such as AlphaFold have demonstrated high accuracy in predicting protein conformations. A study by M. Baek *et al*. (2021) showed that three-layer neural networks can achieve accuracy comparable to experimental methods such as X-ray crystallography. At the same time, AlphaFill, proposed by M.L. Hekkelman *et al*. (2023), extended the capabilities of AlphaFold by allowing the inclusion of ligands and cofactors in protein models, which increases their biological relevance. However, the accuracy of the interactions of WRKY proteins with metal ions, such as zinc, is still a challenge due to the lack of specialised tools. In this context, the MIB2 server, according to C.H. Lu *et al*. (2022), provides metal binding site prediction, which is promising for improving WRKY protein models. In addition to AlphaFold, other methods such as RoseTTaFold, ESMFold, and OmegaFold have also been shown to be effective in structural modelling. The study by Z. Lin *et al*. (2023) found that transformer models such as ESMFold can predict protein structures at the evolutionary level, providing speed and scale. However, their accuracy in complex proteins remains lower than that of AlphaFold. The study of WRKY transcription factors is a promising area of molecular biology that may open up new opportunities to increase crop resistance to abiotic and biotic stresses. For example, the study by D. Upadhyay *et al*. (2024) demonstrated the importance of WRKY2 in wheat drought tolerance, but the structural aspects of this protein are still poorly understood.

Structural biology methods, in particular machine learning, have greatly expanded the possibilities of predicting protein structures. AlphaFold, developed by DeepMind, has already proven to be effective in predicting the three-dimensional structure of proteins at the atomic level (Wu *et al.*, 2022). At the same time, other tools, such as RoseTTaFold and ESMFold, offer alternative approaches to modelling structures, including protein-DNA complexes. PyDockDNA, described by L.A. Rodríguez-Lumbreras *et al*. (2022), provides accurate modelling of protein-DNA interactions, which is important for WRKY proteins that can affect multiple stressor pathways. Despite significant progress, gaps in the research remain. In particular, the structural features of WRKY proteins, such as the interaction of their zinc fingers with DNA, metals, and other proteins, are not well understood. The study by P. Ranjan *et al*. (2024) demonstrated the potential of combined approaches that include molecular dynamics to analyse protein stability, but there is still a need to compare the effectiveness of modern modelling tools. Another important gap is the insufficient consideration of the effect of ligands and metals, which can significantly change the conformation of WRKY proteins. For example, S. Li *et al*. (2024) emphasised that the integration of metal and cofactor interactions into models can significantly improve the accuracy of predictions. However, this requires the additional use of tools such as AlphaFill or MIB2, which allow the inclusion of ligands and metal ions in protein models.

The aim of this study was to comparatively analyse modelling methods for building the structure of the transcription factor TaWRKY2 from *Triticum aestivum*, including the use of modern machine learning tools. Particular emphasis was placed on determining the effectiveness of methods for modelling protein-DNA interactions, as well as taking into account key functional regions of the protein, such as zinc fingers.

MATERIALS AND METHODS

The amino acid sequence of the transcription factor TaWRKY2 from the GenPept database of the National Center for Biotechnology Information (NCBI) (2024) server was used for comparative analysis of computer modelling methods for

protein molecules. The sequence was accessed using the protein number ACD80357.1, which contains 468 amino acid residues. This sequence was chosen because of its key role in the response of *Triticum aestivum* to stress conditions, making it an important target for research. Modelling methods. The following approaches were used to build a structural model of the protein: homology modelling methods (SWISS-MODEL, Modeller and Phyre2), ab initio methods (I-TASSER, D-I-TASSER and QUARK – used to build structures in cases where no relevant templates are available), and machine learning methods.

Among the methods based on machine learning, the following were selected: AlphaFold is a tool developed by DeepMind (2024) that uses neural networks to predict structures with high accuracy (Abramson *et al.*, 2024); to refine the protein structure, the modelling was performed in the Google Colab environment on a Google T4 GPU. When the modelling was performed using MMseqs2, the authors additionally performed structure relaxation using Amber. RoseTTAFold, which integrates machine learning and homology modelling methods to predict three-dimensional structures, was modelled using the Robetta service; ESMFold, based on transformer models for fast and accurate protein structure prediction; and OmegaFold, a tool that uses generative models to create highly accurate protein structures. Chimera software was used to analyse the structural characteristics of the TaWRKY2 model, namely to determine the distance between the domains of the zinc fingers. The Robetta service was used to model TaWRKY2 using the RoseTTaFold method, which provides integration of modern machine learning approaches to protein structure prediction. To evaluate the accuracy and quality of the created TaWRKY2 protein models, a set of metrics was used, each of which provides specific information about various aspects of the structural model, its stability and correspondence to known protein structures.

Verify 3D is a metric that assesses the stereochemical quality of a model by comparing the three-dimensional structure of a protein with its primary amino acid sequence. The analysis is performed by creating profiles for each amino acid residue in the structure and comparing them with expected values. ERRAT is based on the statistical analysis of energy mismatches in the protein structure. It allows identifying regions of the model that show significant deviations from the typical values of energy interactions. Ramachandran Plot Analysis is one of the key tools for checking the geometric correctness of protein models. The Ramachandran diagram displays the conformational angles φ (phi) and ψ (psi) for each amino acid residue. G-factor is a metric that determines the geometric quality of a model based on the distribution of conformational parameters such as bond angles and bond lengths. The G-factor reflects the deviation from the expected values obtained from empirical data.

QMEAN is a comprehensive metric that takes into account several model quality parameters, including surface entropy, global interaction energy, interactions between residues, etc. The QMEAN value is normalised to facilitate comparison between different models. Z-Score is a metric used to statistically compare the obtained model with experimentally determined structures in the PDB database. Z-Score determines how much the model deviates from the average values for known protein structures. MolProbity analyses models based on the geometric quality and accuracy of atomic coordinates. In particular, this metric takes into account the presence of steric collisions between atoms in the protein structure, which are evaluated through Clashscore. The predicted local distance difference test (pLDDT) is an indicator of the local accuracy of protein structure prediction used in the AlphaFold method. It reflects the model's confidence in the accuracy of predicting the spatial location of each amino acid residue. As part of this work, several models of the TaWRKY2 protein were created using different approaches. However, the main focus was on machine learning-based methods (AlphaFold, RoseTTAFold, ESMFold, OmegaFold), as they provided the most accurate and reliable results. The resulting models were compared using the aforementioned metrics to determine their quality and stability.

While each of the approaches has its advantages, there are limitations. Homology modelling methods (programs such as SWISS-MODEL, Modeller and Phyre2) show limitations in cases where there are no suitable templates in the PDB database. This reduces their effectiveness for proteins with significant structural differences or uni que elements, such as metal-dependent domains.

Ab initio methods (I-TASSER, D-I-TASSER, and QUARK) provide greater independence from templates, but often lack the accuracy of machine learning methods, especially for complex proteins with metal centres where there is insufficient data on metal interactions. AlphaFold2, while demonstrating the highest accuracy in predicting structures, models the protein as an isolated molecule, without taking into account interactions with metal ions, ligands, or other proteins. This can limit the realism of the model, especially for proteins where metals are functionally important. RoseTTAFold can generate structures with excessive plasticity, which does not always correspond to biological realities. ESMFold and OmegaFold are fast and efficient, but their accuracy for complex proteins with numerous unstructured regions or metal centres remains lower than AlphaFold2.

RESULTS AND DISCUSSION

Modelling of TaWRKY2 using the AlphaFold method. At the first stage of modelling, the AlphaFold database was searched to obtain the theoretically predicted structure of TaWRKY2 (Yang *et al.*, 2023). The ChimeraX-1.7.1 software was used for this purpose. Among several analysed structures, the model with the index B3GAU3 was selected, which was further evaluated by the quality of regions using the pLDDT metric (Figs. 1, 2). This can be due to several factors: unstructured or flexible regions – many proteins have regions that do not form a clear three-dimensional structure. This is especially true for transcription factors, which often contain flexible binding domains or activator regions. Another factor is the absence of structural analogues in the training data: since AlphaFold is based on machine learning using experimentally determined structures, poor prediction quality can arise from the absence of close analogues for a particular protein or its individual fragments in the training set.

The analysis showed that the majority of the structure, except for two zinc finger domains, is characterised by low pLDDT values.

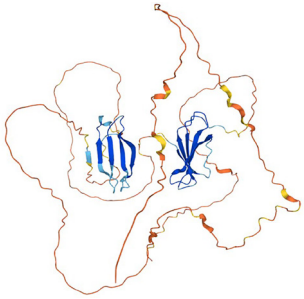

Figure 1. Structure of B3GAU3_WHEAT
Note: the image shows regions with different pLDDT: blue – pLDDT>90, blue – 90>pLDDT>70, yellow – 70>pLDDT>50, orange – pLDDT<50. Some areas with low pLDDT may be isolated unstructured
Source: created by the authors

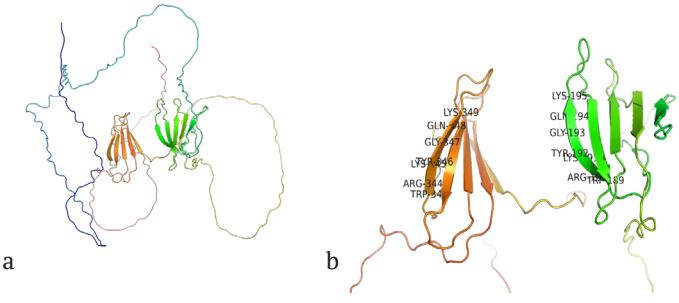

Figure 2. Structural Features and DNA Binding Characteristics of B3GAU3_WHEAT WRKY Transcription Factor
Note: a) Structure of B3GAU3_WHEAT and b) location of WRKYGQK sequences
Source: created by the authors

As shown in the study, the two zinc fingers in the protein structure are located at a distance of 18.7 Å from each other, which is less than the diameter of B-shaped DNA (20 Å). In addition, one of the key areas of WRKY responsible for binding to DNA is oriented "inwards" (Fig. 2b). This makes it difficult for the DNA molecule to be positioned between the domains, although it does not exclude this possibility, as transcription factors have sufficient flexibility to perform their functions. Interaction with other molecules. According to the results obtained, many proteins, especially transcription factors, are characterised by dynamic changes in structure when interacting with other molecules, such as DNA, RNA, or partner proteins. These changes include conformational transformations that allow the protein to adapt to specific bonds or functions. For example, the WRKY transcription factors undergo a significant reorganization of their zinc fingers during DNA binding, which ensures specific recognition of W-boxes in gene promoter sequences.

AlphaFold models the protein structure as isolated, which is a significant limitation for proteins with high functional plasticity. In the case of the TaWRKY2 model, zinc atoms, which critically affect the stability and functionality of the zinc fingers, were missing. This led to incomplete or less realistic reconstruction of functional domains. Since zinc fingers play a key role in protein binding to DNA, taking such details into account is important for accurate prediction of protein structure and function. It has also been found that in addition to interacting with DNA, WRKY proteins can form protein-protein complexes that affect their regulatory activity. For example, interaction with suppressor proteins or cofactors can change the accessibility of WRKY domains to DNA. Such complex interactions are not taken into account in the structural predictions generated by AlphaFold.

Technical limitations. Although AlphaFold significantly outperforms traditional homology modelling methods, this approach has a number of technical limitations. In particular, the accuracy of predictions decreases for proteins that have rare or unique structural elements that are not present in the algorithm's training sets. For example, non-standard conformational changes that occur during protein-protein or protein-ligand interactions are often not reflected in isolated models. In addition, machine learning-based methods such as AlphaFold rely on pre-formed training data, so their performance is reduced if the protein structure or functional domains differ from known templates. In the case of TaWRKY2, this can manifest itself in a less accurate reconstruction of flexible or unstructured regions that play an important role in the functional adaptation of the transcription factor. Another problem is modelling multicomponent complexes. For proteins that work as part of large protein machines, such as WRKY2, the accuracy of predictions by AlphaFold may be limited due to the lack of data on specific interprotein contacts. To improve the accuracy of such models, additional tools, such as docking or experimentally obtained interaction data, should be integrated.

The Chimera software was used to analyse the structural characteristics of the TaWRKY2 model, in particular to determine the distance between the zinc finger domains. In the case of the B3GAU3 model, the software made it possible to visualise the spatial arrangement of the key domains and measure the distance between the zinc fingers, which was 18.7 Å (indicated by the yellow dashed line in Fig. 3). This distance is an important parameter for assessing the functional ability of a protein to interact with DNA, given that the diameter of a DNA molecule in the B-form is approximately 20 Å.

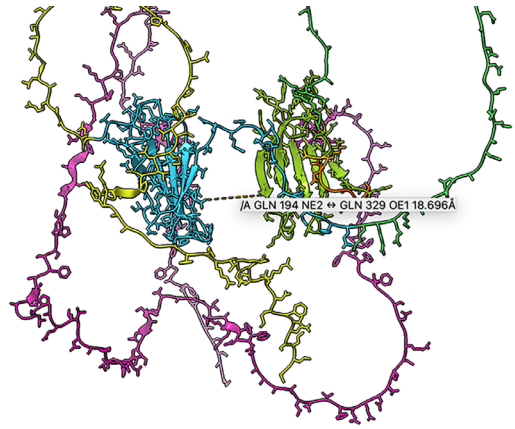

Figure 3. Distance between domains in B3GAU3 (yellow dashed line)
Source: created by the authors

To clarify the structure of the TaWRKY2 transcription factor, the authors performed modelling using AlphaFold in the Google Colab environment

on a Google T4 GPU. This approach provides the ability to perform calculations relatively quickly on a remote server with limited computing resources. As part of the work, various versions of Jupyter Notebook were used to simplify the work with AlphaFold. In addition, other state-of-the-art modelling tools such as ESMFold and RoseTTaFold, which also provide protein structure prediction, were used for comparison. In the case of AlphaFold, a model was obtained using MMseqs2, for which the structure was additionally relaxed using Amber. Figure 4 shows this structure, which demonstrates the distinctive arrangement of the domains. In particular, the zinc fingers form a dimer, and the functional regions of WRKY are oriented "outwards". This conformation potentially facilitates binding to DNA, creating conditions for the formation of a stable loop during the transcription process. The validity of this model is confirmed by the similarity of the zinc fingers to the corresponding domains found in the experimentally determined structures. For example, as can be seen in Figure 4b, a similar organization of zinc fingers was observed in the AtWRKY2 model (PDB ID: 6J4F), although in this case the fingers belong to different protein chains.

This feature of the model indicates its potential functional relevance, as the structure may facilitate interaction with DNA and active loop formation necessary for transcription initiation. An additional analysis of the model showed that the arrangement of the domains allows for effective stabilization of the protein structure when it works in a complex with DNA. This approach to modelling using a combination of AlphaFold and MMseqs2 followed by relaxation provides flexible opportunities for detailed study of proteins, especially those that play an important role in plant stress adaptation (Fig. 4). The structure quality assessment in ProSA is shown in Figure 5.

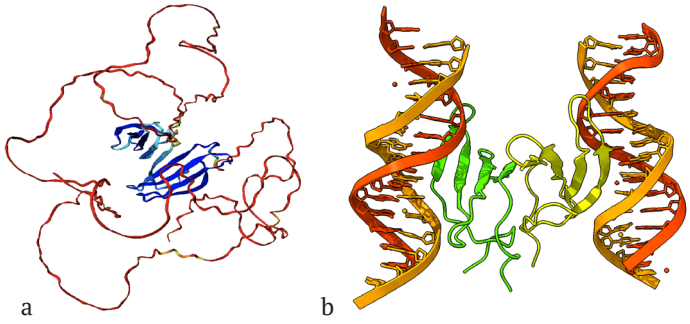

Figure 4. Predicted structure of TaWRKY2 (AlphaFold, MMseqs2, Amber)
Note: a) coloured according to pLDDT; b) structure of 2 zinc fingers from AtWRKY2 with PDB ID 6J4F, confirming the high confidence of the structure modelled by AlphaFold
Source: created by the authors

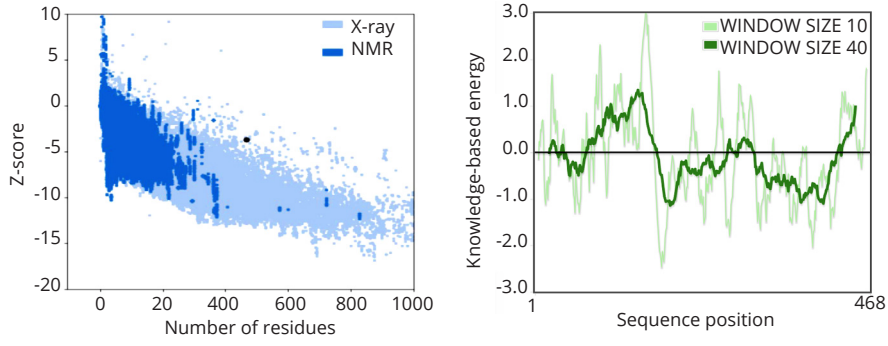

Figure 5. Structure quality assessment of TaWRKY2 (AlphaFold, MMseqs2)
Source: created by the authors

The data obtained indicate that the best quality in the model is the β-folding of the zinc fingers, as in the structure from the AlplaFold database according to the local values in QMEAN. With respect to the global Z-score, the model falls within the range of X-crystallographic structures of the same polypeptide chain length, while the local evaluation revealed some fluctuations. The blue averaged line on the Verify3D graph shows the overall trend, which is high and hovers around 0.2, indicating that the model is of fairly good quality (although it contains residues of lower quality). The model's Clashscore is superb at 2.31, which falls in the 99th percentile. The MolProbity Score is acceptable: 2.35, with a score in the 56th percentile.

Another variant of Jupyter Notebook, offered by DeepMind, does not use MMseqs2 and is a simplified version of AlphaFold2 itself, with a focus on multimeter construction. However, the structure modelled using this approach was largely similar to the previous one, with an additional β-sheet appearing between the zinc finger domains (Fig. 6). However, this new β-sheet is not of high quality, as can be seen from the pLDDT value (the additional β-sheet is marked in orange on the left of Figure 6).

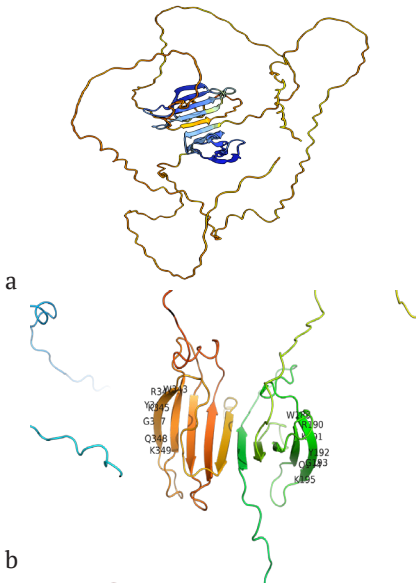

Figure 6. Predicted structure of TaWRKY2 (from DeepMind)
Note: coloured according to a) pLDDT and b) the location of WRKYGQK sequences; 10 β-sheets are visible (instead of 9 in the previous model)
Source: created by the authors

When comparing the number of β-sheets, the additional 10th sheet in the AlphaFold (DeepMind) model may, on the one hand, contribute to greater stability of the TF domain dimer structure, but on the other hand, this feature is not observed in the experimental data (Fig. 4(b)). Nevertheless, since the structure of 6J4F is based on AtWRKY2 from Arabidopsis, it is likely that wheat has specific folding features for TaWRKY2. This additional β-sheet can serve as another example of the difference in modelling approaches: in a homology method, where there is no corresponding template, it would be impossible to predict such a structure. Additionally, if the authors look at the number of hydrogen bonds connecting the two zinc fingers, the AlphaFold (DeepMind) model has significantly more of them, which can provide additional stability to the structure. It is also important to note that when docking the structure with AlphaFold (MMseqs2), the interaction of the W-box with the WRKYGQK motif was less efficient compared to the model built with AlphaFold (DeepMind). In particular, there was no interaction of arginine in the N-terminal domain with tryptophan in the C-terminal domain with DNA.

The only significant drawback of the AlphaFold2 method compared to homology modelling is the inability to directly add ligands during the modelling process, unlike platforms such as SWISS-MODEL, where heteroatoms can be incorporated from template structures. However, over time, additional tools have been developed, such as AlphaFill, which allow filling structures from the AlphaFold database with appropriate ligands. In the case of TaWRKY, however, this resource was unable to add zinc atoms, which is a limitation. In addition, it is worth noting that the process of protein structure prediction using AlphaFold2 requires significantly more time and computational resources compared to homology modelling. This is due to the complexity of the model and high hardware requirements. To simplify this process, the study used Jupyter Notebooks offered by the developers instead of directly running local AlphaFold with full databases, as this approach requires much more computing power. The study also used a new version of AlphaFold3, which is capable of modelling not only protein molecules with standard amino acid residues, but also constructing DNA/RNA (including modified nucleotides),

finding binding sites for metal ions, engaging a specific set of ligands, and applying post-translational modifications to proteins. For comparison, the modelling of an isolated protein and a complex was performed. The values of the complex metrics were better than those of the isolated model.

In the complex of TaWRKY2 with zinc ions and two DNA fragments with W-boxes, it is clear that a dimer of zinc fingers is present, as in the case of AlphaFold2 (Fig. 7). However, unlike previous models, an α-helix between the domains is also observed, which may stabilise the structure. Internal quality metrics showed the following results: ipTM = 0.89; pTM = 0.45. They are both based on the template modelling score (TM), which measures the accuracy of the entire structure. A score of pTM > 0.5 means that the predicted complex may be similar to the real structure. The ipTM score measures the accuracy of the relative positions of subunits in the complex. Values > 0.8 indicate high-quality predictions. The structure quality score in ProSa is shown in Figure 8.

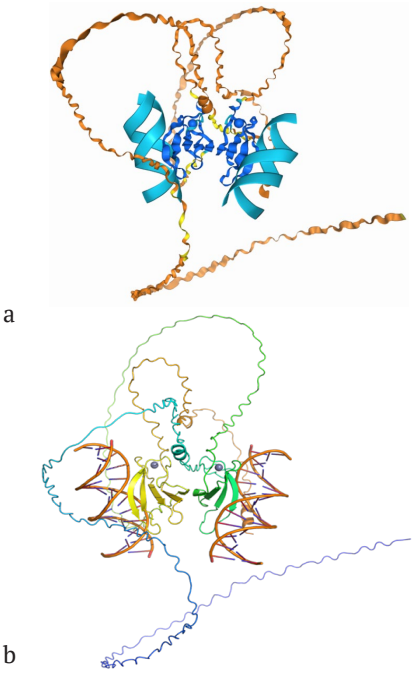

Figure 7. The TaWRKY2 complex predicted in AlphaFold3, 2 Zn2+, 2 DNA fragments with W-boxes
Note: stained a) according to pLDDT and b) according to N→C ends from red to blue
Source: created by the authors

The TaWRKY2 model from AlphaFold3 generally performs worse than AlphaFold2, except for the Z-Score. This may be due to less flexibility in setting up the modelling process on the server. When comparing AlphaFold2 (DeepMind) and AlphaFold3, the authors also observe that the latter model is less compact.

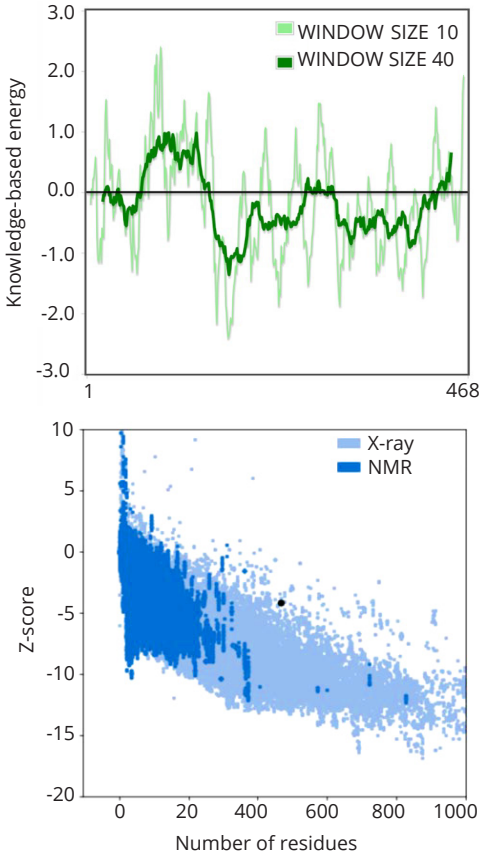

Figure 8. Structure quality assessment of TaWRKY2 (AlphaFold3)
Source: created by the authors

Modelling the TaWRKY2 structure using ESMFold. To create a model of TaWRKY2 using ESMFold, the authors used Jupyter Notebook, developed by the authors of the tool, in the Google Colab environment using a Google T4 GPU (Fig. 9). This approach provided a fast prediction of the protein structure, although the quality of the model obtained was significantly different from the results obtained with AlphaFold.

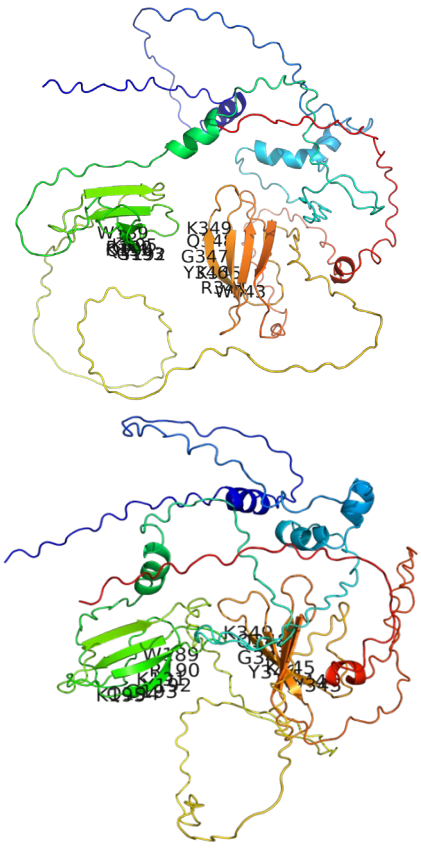

Figure 9. TaWRKY2 model in ESMFold from 2 angles
Source: created by the authors

The resulting structure shows a different arrangement of zinc fingers compared to the AlphaFold model. Additional α-helices are observed in the ESMFold model, which are practically absent in the AlphaFold versions (only a small α-helix is present in the AlphaFold-MMseqs2 model). These differences may indicate an alternative conformation of the protein, which could theoretically occur under specific conditions. Despite the structural features, the model's metrics are significantly inferior to AlphaFold. In particular, the Clashscore value corresponds to only the 22nd percentile (compared to the 99th for AlphaFold-MMseqs2). Other metrics, such as Ramachandran Plot Percentage and MolProbity Score, also indicate lower quality. It is worth noting that the QMEAN metric (0.39) was slightly better than the same indicator in the AlphaFold models (Fig. 10).

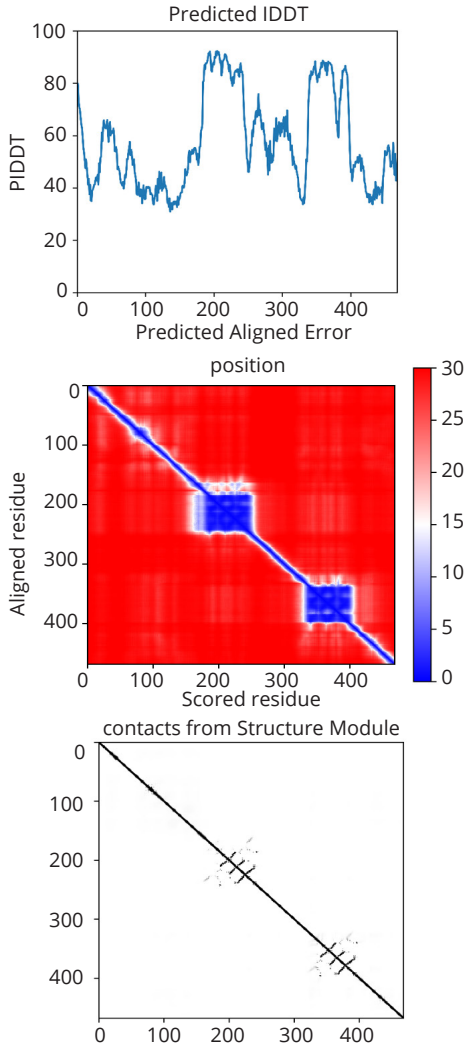

Figure 10. Evaluation of the ESMFold model quality
Source: created by the authors

Despite these shortcomings, the ESMFold model provides useful information about possible protein conformations, especially in cases where traditional methods fail or create overly idealised models. However, the resulting structure has certain drawbacks: the WRKY domains of the N- and C-termini are located opposite each other, which can potentially create steric conflicts when interacting with DNA, making it difficult to implement transcriptional activity. Modelling with RoseTTaFold. For modelling TaWRKY2 with

RoseTTaFold, the authors used the Robetta service, which integrates modern machine learning approaches to protein structure prediction. The modelling result (Fig. 11) shows a structure that is significantly different from the previous models created by AlphaFold and ESMFold.

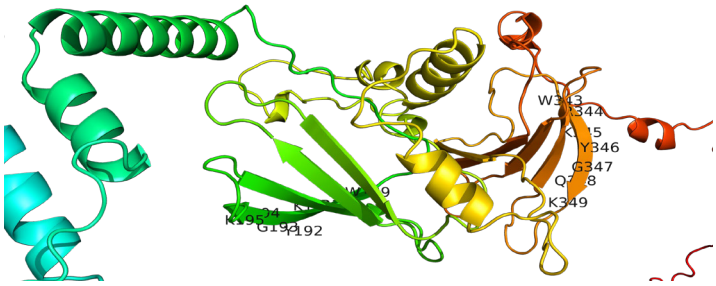

Figure 11. Location of the WRKYGQK sequence in the RoseTTaFold model (internal Confidence metric is 0.35 out of 1)

Source: created by the authors

The main feature of the RoseTTaFold model is its "elongated" shape, in contrast to the compact structures obtained by other methods. In this model, the WRKYGQK sequence of the N-terminus is in an accessible position for DNA binding, while the C-terminal sequence is partially overlapped by a loop with a small α-helix. In general, the structure contains more α-helices compared to AlphaFold or ESMFold. In some cases, the model's metrics have very good values, in particular MolProbity Score and Clashscore, which even exceed those of previous models, and ProSA score also has better values (Fig. 12). Other metrics are comparable to previous models. However, despite this, such folding seems less likely when compared to the experimentally determined structure and according to Confidence's internal quality assessment. Of course, the tertiary structure of a protein can change during its function (e.g., before and after binding to DNA) and the unstructured regions of the AlphaFold models in TaWRKY2 may actually contain alpha helices, but accepting the model without critical analysis may be a mistake.

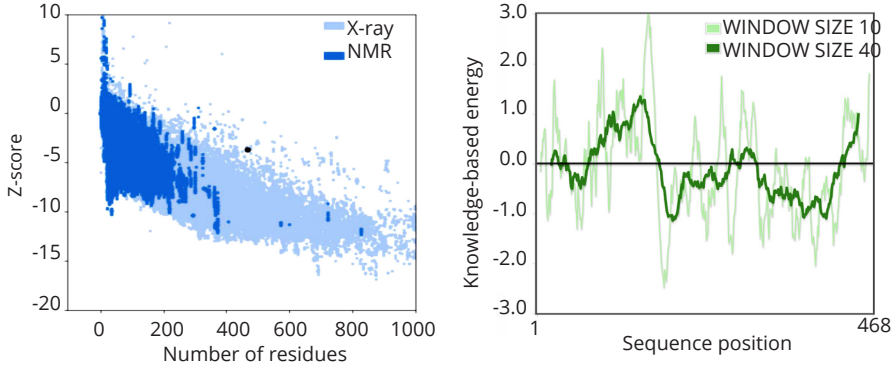

Figure 12. Quality assessment of the TaWRKY2 RoseTTaFold model

Source: created by the authors

Therefore, even with high metrics, each model must be validated against known biochemical data and against possible conformational changes that may occur in the real cellular environment. This is especially important for protein regions responsible for binding and regulation, where

the correct conformation is critical for functional activity. Modelling with OmegaFold. The OmegaFold tool, which is positioned as an improvement over AlphaFold and RoseTTaFold, was also used for modelling TaWRKY2. The code from the GitHub repository was adapted for the Google Colab environment, where the modelling was performed on the basis of the downloaded fasta file. The resulting model (Fig. 13) showed the presence of alpha-helical structures, although their number was smaller than in the RoseTTaFold model.

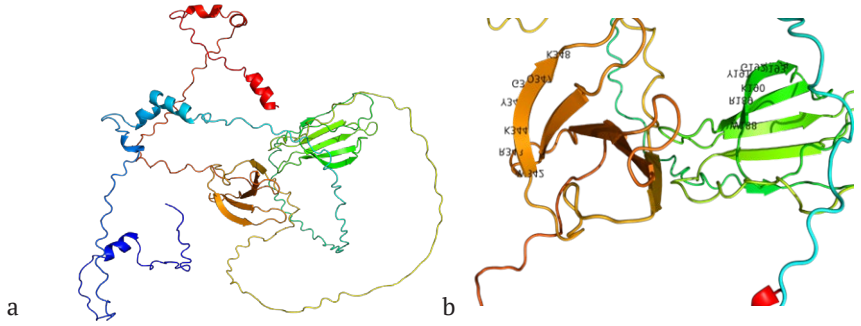

Figure 13. Model of a) TaWRKY2 in OmegaFold and b) relative position of the WRKYGQK sequence
Source: created by the authors

In this structure, no steric hindrance was observed in WRKYGQK sequences for DNA binding. However, quality metrics such as Clashscore and Ramachandran Plot Percentage were found to be worse compared to AlphaFold and RoseTTaFold. The only exception is the QMEAN metric, which is 0.4, which is the best value among all models. As a result of the modelling performed using the methods described above, the obtained models were analysed and the obtained metrics were evaluated (Table 1). This makes it possible to assess the quality of the obtained models and their viability.

Table 1. Resulting table comparing metrics of models created by different methods

Method\Metric	Verify 3D, %	ERRAT	Ramachardran, %	G-factor	QMEAN	Z-Score	Molprobity Score/%ile		Clashscore /%ile	
SWISS-MODEL	44.32	82.05	79.4	-0.38	0.35 ± 0.05	-3.23	1.82,	84	0.59	99
Phyre2	69.66	3.91	57.6	-0.92	0.35 ± 0.05	0.05	3.65	6	221.26	0
AlphaFold (MMseqs2)	45.73	93.43	61.8	-0.75	0.38 ± 0.05	-3.72	2.19	65	0.72	99
AlphaFold (DeepMind)	48.93	85.23	62.8	-0.7	0.35 ± 0.05	-3.62	2.35	56	2.31	99
AlphaFold3	53.42	69.37	58.4	-0.61	0.35 ± 0.05	-4.37	2.84	29	12.55	60
AlphaFold3 (in the complex)	45.94	74.53	58.9	-0.52	0.36 ± 0.05	-4.14	2.63	40	7.1	86
ESMFold	46.79	61.63	40.8	-2.4	0.39 ± 0.05	-4.13	3.64	6	24.15	22
RoseTTaFold (Robetta)	67.09	89.32	83.7	0.27	0.35 ± 0.05	-5.03	1.40	97	2.16	99
OmegaFold	46.37	32.72	44.7	-2.11	0.40 ± 0.05	-3.68	3.49	9	31.19	14
D-I-Tasser	50	65.42	51.4	-0.35	0.34 ± 0.05	-4.08	2.73	35	22.35	26
Modeller (TaWRKY19 template)	55.13	63.29	86.6	-0.31	0.40 ± 0.05	-2.98	2.21	64	16.15	45

Note: the %ile-value indicates the percentile in which a particular model was included
Source: created by the authors

The analysis confirms that each of the methods used has its advantages and limitations. AlphaFold provides the best results for most metrics and is the most reliable method for modelling proteins with known structural features. RoseTTaFold and OmegaFold can also be effective, especially for complex structures, but the resulting models require additional verification. The ESMFold method, despite its lower model quality, shows potential in cases where an alternative protein conformation needs to be quickly obtained for further analysis. All models show that a thorough evaluation of both the structural and dynamic properties of the obtained models is necessary to ensure the functional activity of the protein.

The results of this study, based on modelling the structure of TaWRKY2 using modern methods, confirm the high potential of machine learning in structural biology. However, a comparison with previous works reveals both the advantages and limitations of different approaches. The work of A. Singh *et al*. (2019) emphasised the importance of structural diversity of WRKY proteins for efficient DNA binding, which is also confirmed by the results of the present study. In particular, the AlphaFold modelling revealed the optimal orientation of WRKY domains and the stable interaction of zinc fingers with W-boxes. At the same time, steric conflicts were observed in the ESMFold and RoseTTaFold models, which may reduce the binding efficiency, consistent with hypothesis A. Singh *et al*. about the importance of the specific conformation of proteins for their functionality. The study by S. Hassan *et al.* (2019) focuses on the role of WRKY proteins in the salt stress response. It was found that the accuracy of modelling the structure of such proteins depends on the consideration of specific factors, such as metal ions. In this study, the integration of metals through AlphaFill was only partially performed, which may explain the lower metrics in some models. However, the use of MIB2 to predict metal binding sites can significantly improve the accuracy of predictions, which is promising for future studies.

In this study, the models revealed a high plasticity of TaWRKY2 flexible regions, which may explain their adaptive role. In particular, additional α-helices were observed in the RoseTTaFold models, which may play a key role in stabilising the structure under stress. Similar conclusions were reached by a team led by J.N. Amoah *et al*. (2019), who demonstrated that stressful conditions, including drought, cause significant changes in the expression of WRKY proteins. The conclusions about the need to further consider functional protein-environment interactions, such as complexes with DNA and ligands, are consistent with the studies of F.F. Theisen *et al*. (2024), who emphasised the importance of post-translational modifications in the regulation of transcription factors. The absence of such modifications in most current models, in particular the isolation of the protein in the AlphaFold2 environment, may limit the biological relevance of the obtained structures; while AlphaFold3 has a number of advantages in this context.

R. Blanc-Mathieu *et al*. (2024) created a structural classification of plant TFs, which confirms the importance of WRKY domains conservation for functional activity. The studied models confirmed this conservation, especially in the versions generated by AlphaFold, but in ESMFold, significant deviations in the orientation of functional regions were observed, which requires additional analysis. This study also confirmed the role of metals in the stabilization of WRKY domains. In particular, the models built with AlphaFold with the inclusion of zinc fingers showed higher MolProbity Score metrics than other approaches, which is consistent with the hypothesis of S. Banerjee *et al.* (2023) about the key role of these interactions in structural stability; the authors emphasised the importance of WRKY proteins under stress conditions, in particular their interaction with metals and partner proteins. The results of X. Chen *et al*. (2019) demonstrate that the evolution of WRKY proteins ensures their functional plasticity. The studied models showed differences in the structural organization of WRKY2, especially in the location of domains, which may reflect the evolutionary adaptations of this protein to specific conditions. This confirms the need to integrate several methods to account for different conformations.

Molecular dynamics methods and protein-ligand interactions are explored in the work by R. Shukla & T. Tripathi (2020), where the authors emphasise the importance of molecular dynamics in predicting the stability of protein structures and their interactions with ligands. In this study, based on the use of AlphaFold, RoseTTaFold and

ESMFold, models with high metrics such as MolProbity Score and Clashscore were obtained at the initial stages, which ensures the relevance of the results without additional lengthy simulations. This reduces computational costs while maintaining model accuracy, although molecular dynamics remains a promising approach for deeper analysis of protein functional mechanisms. The structural features of the interaction of WRKY proteins with DNA were studied by B. Pandey *et al.* (2018) noted that the interaction of WRKY domains with W-boxes can vary significantly depending on the specificity of the DNA sequence. The results of the present study confirm these findings: modelling with AlphaFold3 provided the best orientation of the domains relative to DNA, while ESMFold models showed steric conflicts that could reduce binding efficiency. Compared to previous work, the use of AlphaFold resulted in to achieve higher accuracy without the need for further optimization.

S.H. Wani *et al.* (2018) point out the importance of WRKY transcription factors in the regulation of osmotic stress in wheat. The study focused on the structural dynamics of TaWRKY2, which revealed significant plasticity of individual protein regions. The RoseTTaFold models revealed the presence of additional α-helices, which, according to previous studies, can stabilise the protein under adverse conditions. This suggests that structural changes are a key factor in adaptation to stressful conditions. The evolutionary adaptation of WRKY proteins is discussed in the study by C.M.M. Dinusha (2019), in particular, the evolutionary mechanisms of protein structure formation, including WRKY domains, are analysed. In contrast, the results of this study demonstrate specific features of TaWRKY2 that may be unique to Triticum aestivum. Comparison of the AlphaFold and RoseTTaFold models suggests that these features are important for plant adaptation to stress, which opens up opportunities for further evolutionary studies of WRKY proteins in cereals.

In this study, the use of the AlphaFill and MIB2 tools enabled the identification of metal binding sites in TaWRKY2. These results confirm that the integration of metal ions during modelling significantly improves the accuracy and biological relevance of structures. Compared to previous studies, such as S. Li *et al.* (2024), which pointed out the importance of metal ions for the stability of WRKY proteins, this approach is more comprehensive and provides important information for future research. Non-coding RNAs in stress responses have been the subject of studies by N. Li *et al.* (2022); they emphasise the role of non-coding RNAs in the regulation of WRKY proteins. Although the structural aspects in this study do not cover the direct interaction with non-coding RNAs, the models obtained can serve as a basis for further modelling of protein-RNA complexes. Z. Du *et al.* (2021) highlight the effectiveness of the trRosetta server for fast and accurate protein structure prediction. In this study, models with high accuracy rates were obtained using AlphaFold and RoseTTaFold. Although trRosetta performs well for proteins with simple structural features, its application to complex proteins such as TaWRKY2 may be limited due to the lower detail of functional domains. The AlphaFold approach (MMseqs2) demonstrated higher accuracy in predicting the spatial location of WRKY domains.

Various sequence search methods were proposed by M. Steinegger & J. Söding (2017), in particular, they used MMseqs2 as a tool for highly sensitive protein sequence search, which is key for modelling the structure of proteins from poorly understood families. In this study, MMseqs2 was integrated into the AlphaFold modelling process, which significantly improved the prediction accuracy by using more relevant evolutionary data. Compared to other methods, this approach provides more detail in the prediction of structurally conserved elements, in particular zinc fingers in TaWRKY2. In the present study, the results also confirm the conservation of WRKY domains, which is particularly evident in the AlphaFold models. This is in line with the findings of X. He *et al.* (2019), who conducted a genome-wide analysis of WRKY proteins in buckwheat (Fagopyrum tataricum) and found that most WRKY domains show high conservation. At the same time, the models created by RoseTTaFold showed greater variability in the conformation of the domains, which may be due to the specificity of the algorithm, but requires further experimental confirmation.

Novel protein families in *Triticum aestivum* were studied by M. Zaheer *et al.* (2021), who emphasised the importance of their transcriptional profile for adaptation to stressful conditions. In the present study, the WRKY2 models obtained

confirm that structural changes in proteins can ensure their functional adaptation. RoseTTaFold models revealed the presence of additional α-helices, which, A. Tripathi *et al.* (2023) proposed TemPred as a new tool for protein pattern discovery, which allows for improved prediction accuracy. In this study, it was shown that the integration of AlphaFill and MMseqs2 already provides high accuracy of structural models. Although TemPred can offer additional benefits, its practical use for WRKY proteins still needs to be explored, as the accuracy of AlphaFold remains significantly higher for complex proteins such as TaWRKY2. The analysis of VQ family proteins in plants was performed by J. Tian *et al.* (2023), who emphasised the importance of identifying new protein families in plants. A similar approach may be useful for the study of WRKY proteins, but in this study, the emphasis was placed on the use of structural modelling to clarify functional features. The results show that WRKY2 exhibits high plasticity, which may be relevant for adaptive mechanisms.

The analysis shows the benefits of using several machine learning methods for structural modelling. Compared to previous works, the models created in this study demonstrate high accuracy and relevance for functional research. The differences with previous studies are explained by the emphasis on a comprehensive approach that includes a detailed analysis of structural features, which provides greater depth and accuracy in the study of WRKY2 functions. The main limitations of the study are the lack of experimental verification of the models, the isolation of the protein in the modelling process by most methods without taking into account its interactions with DNA or partner proteins, insufficient integration of metal ions, as well as the lack of dynamic analysis and focus on only one member of the WRKY family. Further research should be aimed at experimental confirmation of the obtained structures, integration of protein-protein and protein-DNA interactions, application of molecular dynamics, and extension of the analysis to other WRKY proteins to form a holistic view of their role in plant stress adaptation.

CONCLUSIONS

The amino acid sequence of the transcription factor TaWRKY2 from *Triticum aestivum* has been successfully used to construct and analyse a three-dimensional structure using modern computer modelling methods. This protein, which is important for stress resistance mechanisms, has become an effective object for comparing different approaches in structural bioinformatics. AlphaFold, as a representative of the latest machine learning methods, demonstrated the best results in most metrics, such as Clashscore (99th percentile), ERRAT, MolProbity Score (AlphaFold2), and Z-Score (AlphaFold3). The model obtained with AlphaFold (MMseqs2) was the highest quality and closest to the biologically relevant structure. The models obtained with ESMFold had significant differences in structure, including additional α-helices and different arrangement of zinc fingers. Despite the lower quality of most metrics, this method shows potential for rapid protein modelling, especially in the absence of experimental data. RoseTTaFold produced a structure with unique features, such as an elongated shape and a higher number of α-helices. Although this method outperformed AlphaFold by some metrics (MolProbity Score, Clashscore), the resulting model requires additional confirmation of its functional relevance. OmegaFold, despite the improvements claimed by the developers, demonstrated model quality lower than AlphaFold and RoseTTaFold. Only the QMEAN metric indicated a certain advantage of this approach.

Additional tools, such as Amber and Chimera, provided structure relaxation and analysis of key features of the models. The Chimera analysis revealed important features such as the location of domains and the distance between zinc fingers (18.7 Å), which directly affects the functional activity of the protein. Evaluation of the models using Verify 3D, ERRAT, Ramachandran Plot, Z-Score, QMEAN and other metrics enabled the identification of the advantages and disadvantages of each modelling method. In general, for proteins with high functional plasticity, such as TaWRKY2, the most relevant approach is to use machine learning methods, such as AlphaFold. Using multiple methods with different modelling approaches, including RoseTTaFold, OmegaFold and ESMFold, can be useful for testing alternative conformations and interactions. Further modelling of multicomponent complexes, including interactions with ligands or other proteins, requires the use of experimental data and docking methods.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Abramson, J., *et al*. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630, 493-500. doi: 10.1038/s41586-024-07487-w.

[2] Amoah, J.N., Ko, C.S., Yoon, J.S., & Weon, S.Y. (2019). Effect of drought acclimation on oxidative stress and transcript expression in wheat (*Triticum aestivum* L.). *Journal of Plant Interactions*, 14(1), 492-505. doi: 10.1080/17429145.2019.1662098.

[3] Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), 871-876. doi: 10.1126/science.abj8754.

[4] Banerjee, S., Mukherjee, A., Khatun, M., Roy, P., Nandi, S., & Roy, S. (2023). Exploring the critical function and molecular mechanism of WRKY transcription factor family in regulating plant response under abiotic stress conditions. In T. Aftab (Ed.), *New frontiers in plant-environment interactions: Innovative technologies and developments* (pp. 121-142). Cham: Springer. doi: 10.1007/978-3-031-43729-8_6.

[5] Blanc-Mathieu, R., Dumas, R., Turchi, L., Lucas, J., & Parcy, F. (2024). Plant-TFClass: A structural classification for plant transcription factors. *Trends in Plant Science*, 29(1), 40-51. doi: 10.1016/j.tplants.2023.06.023.

[6] Chen, X., Li, C., Wang, H., & Guo, Z. (2019). WRKY transcription factors: Evolution, binding, and action. *Phytopathology Research*, 1, article number 13. doi: 10.1186/s42483-019-0022-x.

[7] DeepMind. (2024). *AlphaFold has revealed millions of intricate 3D protein structures, and is helping scientists understand how life's molecules interact*. Retrieved from https://deepmind.google/technologies/alphafold/.

[8] Dinusha, C.M.M. (2019). *The evolution of reproductive development in angiosperms*. (Doctoral thesis, University of California, Riverside, USA)

[9] Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., & Yang, J. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nature Protocols*, 16, 5634-5651. doi: 10.1038/s41596-021-00628-9.

[10] Hassan, S., Lethin, J., Blomberg, R., Mousavi, H., & Aronsson, H. (2019). In silico based screening of WRKY genes for identifying functional genes regulated by WRKY under salt stress. *Computational Biology and Chemistry*, 83, article number 107131. doi: 10.1016/j.compbiolchem.2019.107131.

[11] He, X., Li, J.-J., Chen, Y., Yang, J.-Q., & Chen, X.-Y. (2019). Genome-wide analysis of the WRKY gene family and its response to abiotic stress in buckwheat (*Fagopyrum tataricum*). *Open Life Sciences*, 14(1), 80-96. doi: 10.1515/biol-2019-0010.

[12] Hekkelman, M.L., de Vries, I., Joosten, R.P., & Perrakis, A. (2023). AlphaFill: Enriching AlphaFold models with ligands and cofactors. *Nature Methods*, 20, 205-213. doi: 10.1038/s41592-022-01685-y.

[13] Li, N., Liu, T., Guo, F., Yang, J., Shi, Y., Wang, S., & Sun, D. (2022). Identification of long non-coding RNA-microRNA-mRNA regulatory modules and their potential roles in drought stress response in wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, 13, article number 1011064. doi: 10.3389/fpls.2022.1011064.

[14] Li, S., Khoso, M.A., Xu, H., Zhang, C., Liu, Z., Wagan, S., Dinislam, K., & Liu, L. (2024). WRKY transcription factors (TFs) as key regulators of plant resilience to environmental stresses: Current perspective. *Agronomy*, 14(10), article number 2421. doi: 10.3390/agronomy14102421.

[15] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), 1123-1130. doi: 10.1126/science.ade2574.

[16] Long, L., Gu, L., Wang, S., Cai, H., Wu, J., Wang, J., & Yang, M. (2023). Progress in the understanding of WRKY transcription factors in woody plants. *International Journal of Biological Macromolecules*, 242(Part 1), article number 124379. doi: 10.1016/j.ijbiomac.2023.124379.

[17] Lu, C.H., Chen, C.C., Yu, C.S., Liu, Y.Y., Liu, J.J., Wei, S.T., & Lin, Y.F. (2022). MIB2: Metal ion-binding site prediction and modeling server. *Bioinformatics*, 38(18), 4428-4429. doi: 10.1093/bioinformatics/btac534.

[18] National Center for Biotechnology Information (NCBI). (2024). *Protein database*. Retrieved from https://www.ncbi.nlm.nih.gov/protein/.

[19] Pandey, B., Grover, A., & Sharma, P. (2018). Molecular dynamics simulations revealed structural differences among WRKY domain-DNA interaction in barley (*Hordeum vulgare*). *BMC Genomics*, 19, article number 132. doi: 10.1186/s12864-018-4506-3.

[20] Ranjan, P., Yadav, A., Behera, A.K., Singh, D.K., Singh, P., & Singh, G.P. (2024). Structural, functional and evolutionary analysis of wheat WRKY45 protein: A combined bioinformatics and MD simulation approach. *Cereal Research Communications*, 52, 489-499. doi: 10.1007/s42976-023-00399-x.

[21] Rodríguez-Lumbreras, L.A., Jiménez-García, B., Giménez-Santamarina, S., & Fernández-Recio, J. (2022). pyDockDNA: A new web server for energy-based protein-DNA docking and scoring. *Frontiers in Molecular Biosciences*, 9, article number 988996. doi: 10.3389/fmolb.2022.988996.

[22] Shukla, R., & Tripathi, T. (2020). Molecular dynamics simulation of protein and protein-ligand complexes. In D.B. Singh (Ed.), *Computer-aided drug design* (pp. 133-161). Singapore: Springer. doi: 10.1007/978-981-15-6815-2_7.

[23] Singh, A., Sharma, A.K., Singh, N.K., Sonah, H., Deshmukh, R., & Sharma, T.R. (2019). Understanding the effect of structural diversity in WRKY transcription factors on DNA binding efficiency through molecular dynamics simulation. *Biology*, 8(4), article number 83. doi: 10.3390/biology8040083.

[24] Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026-1028. doi: 10.1038/nbt.3988.

[25] Theisen, F.F., Prestel, A., Elkjær, S., Leurs, Y.H., Morffy, N., Strader, L.C., & Skriver, K. (2024). Molecular switching in transcription through splicing and proline-isomerization regulates stress responses in plants. *Nature Communications*, 15, article number 592. doi: 10.1038/s41467-024-44859-2.

[26] Tian, J., Zhang, J., & Francis, F. (2023). Large-scale identification and characterization analysis of VQ family genes in plants, especially gymnosperms. *International Journal of Molecular Sciences*, 24(19), article number 14968. doi: 10.3390/ijms241914968.

[27] Tripathi, A., Mondal, R., Lahiri, T., Chaurasiya, D., & Pal, M.K. (2023). TemPred: A novel protein template search engine to improve protein structure prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(3), 2112-2121 doi: 10.1109/TCBB.2022.3233846.

[28] Upadhyay, D., Budhlakoti, N., Kumari, J., Chaudhary, N., Padaria, J.C., Sareen, S., & Kumar, S. (2024). In-silico characterization of drought stress related WRKY2 transcription factor in wheat crop (*Triticum aestivum* L.): Study of its physico-chemical properties and structural dynamics. *Genetic Resources and Crop Evolution*, 71, 1481-1492. doi: 10.1007/s10722-023-01706-5.

[29] Wani, S.H., Tripathi, P., Zaid, A., Challa, G.S., Kumar, A., Kumar, V., & Bhatt, M. (2018). Transcriptional regulation of osmotic stress tolerance in wheat (*Triticum aestivum* L.). *Plant Molecular Biology*, 97, 469-487. doi: 10.1007/s11103-018-0761-6.

[30] Wu, R., Ding, F., Wang, R., Shen, R., Zhang, X., Luo, S., Su, C., Wu, Z., Xie, Q., Berger, B., Ma, J., & Peng, J. (2022). *High-resolution de novo structure prediction from primary sequence*. doi: 10.1101/2022.07.21.500999.

[31] Yang, Z., Zeng, X., Zhao, Y., & Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction and Targeted Therapy*, 8, article number 115. doi: 10.1038/s41392-023-01381-z.

[32] Zaheer, M., Rehman, S.U., Khan, S.H., Shahid, S., Rasheed, A., Naz, R., & Sajjad, M. (2021). *Genome-wide identification and transcriptional profiling of new cobra-like genes in bread wheat (Triticum Aestivum)*. doi: 10.21203/rs.3.rs-796690/v1.

Методи комп'ютерного моделювання транскрипційного фактора стресостійкості WRKY2 у Triticum Aestivum: порівняльний аналіз

Ірина Дем'яненко

Кандидат біоінформаційних наук
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
https://orcid.org/0000-0002-0832-3619

Антон Петровський

Молодший інженер-програміст
CPCS
40000, вул. Косівщинська, 18, м. Суми, Україна
https://orcid.org/0009-0006-6267-3608

Ігор Левтун

Кандидат біотехнологічних наук
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
https://orcid.org/0000-0003-2498-035X

Анотація. Метою роботи було порівняння ефективності сучасних методів комп'ютерного моделювання для побудови тривимірної структури транскрипційного фактора TaWRKY2 із Triticum aestivum, який відіграє ключову роль у стресостійкості рослин. У дослідженні використано традиційні гомологічні підходи та методи, засновані на машинному навчанні, зокрема AlphaFold, RoseTTaFold, ESMFold та OmegaFold. Модель, отримана за допомогою AlphaFold2 (MMseqs2), виявилася найякіснішою за метриками Clashscore (99-й перцентиль), MolProbity Score та ERRAT. Комплекс з іонами цинку та ДНК, згенерований AlphaFold3, виявив кращі результати за Z-Score показником. Метод AlphaFold забезпечив біологічно релевантні структури з правильним розташуванням функціональних ділянок. ESMFold забезпечив швидке моделювання, однак продемонстрував відхилення у структурі, включаючи додаткові α-спіралі та нижчу якість за більшістю метрик. Модель RoseTTaFold мала витягнуту форму з більшим вмістом α-спіралей та потребувала додаткової верифікації функціональної активності. OmegaFold, хоча й забезпечив найкраще значення за QMEAN (0,4), поступився за іншими показниками. Додаткові інструменти, такі як Amber і Chimera, дозволили провести релаксацію структур та аналіз ключових характеристик, зокрема просторового розташування цинкових пальців (18,7 Å). Оцінка за метриками Verify 3D, ERRAT, Ramachandran Plot та іншими виявила переваги та недоліки кожного підходу. Висновки підтвердили переваги методів машинного навчання для моделювання білків із високою функціональною пластичністю. Зокрема, AlphaFold було визнано найефективнішим підходом для побудови моделей з високою точністю. Разом із цим, використання кількох методів дозволяє врахувати альтернативні конформації та взаємодії, що важливо для глибшого розуміння функціональних механізмів білка

Ключові слова: біоінформатика; тривимірна структура білка; методи машинного навчання; метрики оцінки моделей; цинкові пальці