Expert Systems With Applications 286 (2025) 127853

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Enhancing protein structure predictions: DeepSHAP as a tool for understanding AlphaFold2

Sabbir Ahmed Sibli [a,*], Vlasios Panagiotis Panagiotou [b], Christos Makris [b]

[a] Department of Electrical & Computer Engineering, University of Patras, Rio, Patras, 26504, Greece
[b] Department of Computer Engineering & Informatics, University of Patras, Rio, Patras, 26504, Greece

ARTICLE INFO

Keywords:
AlphaFold2
OpenFold
ColabFold
DeepSHAP
MLP
Proteins
pLDDT
PAE
MSA
ExplainableAI
Interpretability

ABSTRACT

The AlphaFold2 introduces an entirely new era in the field of computational biology by achieving outstanding results in protein structure predictions. Despite its remarkable performance, the model's prediction pipeline is highly complex, and the Deep Learning model itself is a black box, leaving researchers uncertain about how specific predictions are made. Explainable AI (XAI) offers a promising way for improving the transparency and trustworthiness of such models, making their predictions understandable. This research presents how DeepSHAP can be applied to explain AlphaFold2's predictions through the OpenFold and ColabFold frameworks. DeepSHAP is designed to integrate Shapley Additive exPlanations (SHAP) values and Deep Learning Important Features (DeepLIFT) with a complex deep learning model to interpret the contribution of individual input features for a specific prediction. This research illustrates how particular amino acids are responsible for defining different structures of a protein. In particular, by computing SHAP values and combining them with DeepLIFT, it examines a number of different structures of a protein and determines influential amino acids responsible for having the maximum impact on AlphaFold2 prediction. The results demonstrate the promising potential of this approach to enhance interpretability and provide insights for the transparency of AI-based protein structure prediction models.

1. Introduction

Protein structure prediction is a critical research and application area in the field of Bioinformatics. It encompasses the appropriate understanding of protein functions, biological evolution, and protein engineering. Deep learning models like AlphaFold2 have revolutionized the protein structure prediction industry by achieving an accuracy rate equivalent to that of experimental approaches (Jumper et al., 2021). Fig. 1 refers to an interactive comparison between experimental and computationally predicted structures. Like other deep learning models, the predictions of AlphaFold2 also act as a black box, which is very difficult to understand (Senior et al., 2020). This obscurity hinders the formulation of new hypotheses, the validation of current theories, and the ability to design novel proteins based on the predictions. To fully grasp the benefits of a predicted protein structure, it is important to understand how a protein is folded from both in-situ and in-silico perspectives. Research has shown that certain amino acids from a sequence are more crucial in building a specific protein structure than the others present in that sequence (Anfinsen, 1973). Understanding the relationship between those particular amino acids and the resulting protein structure allows researchers to perform protein synthesis with specific structures or modify natural proteins to achieve specific functionality. Such protein engineering is essential for advancing other fields, like drug development for serious diseases (Dokholyan et al., 2003).

Understanding the internal mechanisms of Artificial Intelligence (AI) models is always a point of focus (Doshi-Velez & Kim, 2017). Deep learning models are more complex than other AI or machine learning models because of one or more extensive neural networks. It brings a huge difficulty to interpret the outcome or debug the models for future purposes (Zhang, Lipton, Li, & Smola, 2021). On the other hand, researchers are not satisfied with just predictive performance; they want to understand why their models make certain predictions (Gilpin et al., 2018). Structural biologists are always interested in connecting a primary amino acid sequence to its three-dimensional structure, or tertiary structure (Anfinsen, 1973).

The use of Explainable Artificial intelligence (XAI) can be a solution to interpret deep learning models. XAI models are designed in such a way that they can identify how the input features affect the model's prediction (Adadi & Berrada, 2018). This study interprets AlphaFold2's predictions inferred from OpenFold by using the explainable AI (XAI)

* Corresponding author.
E-mail addresses: sabbirahmedsibli@upatras.gr (S.A. Sibli), panagiotov@ac.upatras.gr (V.P. Panagiotou), makri@ceid.upatras.gr (C. Makris).

https://doi.org/10.1016/j.eswa.2025.127853
Received 28 December 2024; Received in revised form 9 April 2025; Accepted 21 April 2025
Available online 12 May 2025
0957-4174/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

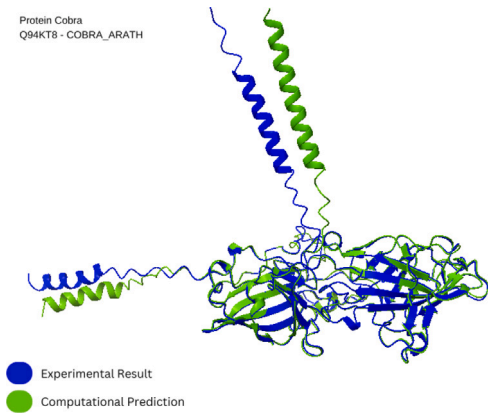


Fig. 1. A comparison between experimental folded structure and predicted structure using AlphaFold2 of COBRA protein (Q94KT8) (UniProt, 0000a).

framework, DeepSHAP. DeepSHAP provides a clear idea of how specific amino acids affect a 3D protein structure by integrating Shapley values with the input features (Lundberg & Lee, 2017). The results are ratified by biological principles, which guarantees the explanations' validity in terms of both protein engineering and biology (Lundberg, Erion, & Lee, 2018). DeepSHAP works by combining Deep Learning Important Features (DeepLIFT) with Shapley Additive exPlanations (SHAP) values to provide insights for the complex neural network models (Shrikumar, Greenside, & Kundaje, 2017). DeepLIFT can identify the contribution of each feature to the model's output and assign importance scores to those features accordingly. On the other hand, SHAP values further refine these scores to ensure that the total contribution of all the features does not differ from the output of the original model (Lundberg & Lee, 2017). This approach is very useful for understanding AlphaFold2. It can provide valuable insights into how specific amino acids from the input sequence influence the final three-dimensional protein structure prediction (Lundberg et al., 2018).

The rest of the paper is organized as follows:

- The **Methodology** section includes a discussion of the dataset, the proposed models, and the details of the algorithms which are being used.
- The **Experiment & Result Analysis** and **Discussion** sections present and discusses the results, respectively.
- Finally, the paper concludes with the **Conclusion** section.

2. Literature review

In recent years, the union of artificial intelligence and Bioinformatics has provided remarkable progress in protein structure prediction. Incorporating XAI methods in this field has become very important due to the necessity of transparency and interpretability in complex models. The main focus of this research is to explain AlphaFold2's predictions using an XAI technique called DeepSHAP. This literature review delves into the current view regarding using XAI for protein structure prediction.

One of the first kinds of AI-based protein structure predictors was Iterative Threading Refinement, in short, I-TASSER, introduced in 2005 by the Zhang Lab at the University of Michigan (Roy, Kucukural, & Zhang, 2010). I-TASSER is an online server which predicts protein structures by threading to find homologous protein structures in the Protein Data Bank (PDB). It builds an initial model using ab initio modeling, which later selects the final model based on energy and structural quality assessments. The iterative nature of I-TASSER allows the model to improve the predicted protein structures through refinement and optimization. Later on, the AlphaFold series by DeepMind pioneered the prediction of three-dimensional protein structures from a given amino acid sequence. After the massive success of AlphaFold (Senior et al., 2020), DeepMind gradually released AlphaFold2 (Jumper et al., 2021) and AlphaFold3 (Abramson et al., 2024). AlphaFold was introduced in 2018 and utilizes a single convolutional neural network (CNN) to predict protein distances and angles based on the evolutionary information from protein databases. AlphaFold2, released in 2020, significantly improved the accuracy by integrating a novel attention-based neural network called transformers and end-to-end differentiable models to predict protein structure directly from a given amino acid sequence. AlphaFold3 showed more powerful results with larger datasets. Combining the principles of I-TASSER and AlphaFold, Meta introduced a language-based model named ESMFold (Lin et al., 2022), which represents another significant advancement in protein structure prediction. ESMFold utilizes evolutionary information and deep learning techniques to predict protein structures. Firstly, it extracts evolutionary information from multiple sequence alignments (MSA) to capture the sequence and structure relationship. After that, it predicts inter-residue distance and secondary structure information using a deep learning model. Finally, ESMFold generates the 3D structure based on predicted distance maps and the secondary structure information. It can refine the prediction using energy minimization techniques.

Despite the success in revolutionizing protein structure prediction, the black-box nature of these models poses a significant challenge. The complexity of the underlying algorithms makes it difficult for researchers to understand the specific reasons behind the predictions. This lack of transparency can hinder trust and acceptance among scientists who require not only accurate predictions but also an understanding of the mechanisms driving these predictions.

To address this challenge, the development of explainable models has become essential. The explainable model aims to make the decision-making processes of the complex nature of protein folding more transparent and interpretable. However, not much work has been done yet; scientists are still working on this issue. This paper discusses both traditional and XAI directions as a literature review.

2.1. Site-directed mutagenesis and protein folding

Site-directional mutagenesis is a fundamental technique in molecular biology which is used to find out the roles of specific amino acids behind a folded protein (Gillam, Jahnke, & Smith, 1978). This method works with the intentional alteration of amino acids in a sequence by systematic deletion or substitution to observe the changes in the protein's function. The expected outcome is finding the key residues that are crucial for the stability of protein folding.

Though site-directional mutagenesis is a powerful approach, it is significantly time-consuming and labor-intensive. For every mutation, the protein must be precisely modified, and then the effects must be determined through an intensive biophysical and structural assessment (Carter, 1986). Overall, this technique is very powerful to get the intuition on how a protein is being folded. For example, a mutation in the hemoglobin gene can drive into sickle-cell anemia (Ingram, 1957).

After the introduction of AI-based models to predict protein structure, mutagenesis became very handy as it can be done computationally (Ribeiro, Singh, & Guestrin, 2016). Besides that, explainable AI models have started taking positions in explaining the underlying mechanisms of how protein structure prediction models work.

2.2. Counterfactual reasoning based XAI

In general, counterfactual reasoning is used to verify a model by systematically altering input data and observing the changes in output. In terms of protein structure prediction, this method works by modifying specific amino acids in the protein sequence and analyzing how it affects the predicted three-dimensional structure. Researchers can gain

insight by identifying which amino acids are crucial for the protein folding process (Pearl et al., 2000).

One such existing framework is ExplainableFold (xFold) (Tan & Zhang, 2023). Though the authors claimed that this model could be used for any AI-based protein structure prediction model, it was introduced to interpret the AlphaFold2 model's prediction. ExplainableFold uses counterfactual reasoning to delete or substitute certain amino acids to understand the dependency of that protein structure on the specific residues. For example, if a single change of an amino acid in the sequence results in a significant change in the predicted structure, it indicates that the amino acid plays the most important role in that protein's conformation. Conversely, if the structure remains mostly unchanged, the amino acid may be less impactful.

This methodology not only enhances the interpretability of AlphaFold2's predictions but also provides valuable biological insights that can be used in experimental designs and further research (Gillam et al., 1978).

However, this research proposes an existing Deep Learning explanation method, DeepSHAP, to the best of current knowledge, which has never been applied for particularly interpreting protein structure prediction. As mentioned earlier in this section, the existing counterfactual method changes certain features of input to see how it affects the prediction, showing which features are important for changing the outcome. In contrast, DeepSHAP assigns importance scores to each feature by looking at how much each one contributes to the prediction, based on a combination of SHAP values and DeepLIFT techniques (detailed in Section 3.5). This way, unlike the counterfactual methods, DeepSHAP provides a clear picture of feature importance without altering the input and keeping it intact, a fact that makes it computationally and biologically superior to existing methods, which alternate the original input.

3. Methodology

3.1. Dataset

AlphaFold2 uses a curated dataset to achieve protein structure predictions. The dataset includes protein sequences from sources like UniProt, BFD and MGnify, along with sequence alignments generated by tools such as JackHMMER, HHblit,s and Uniclust30 (Jumper et al., 2021). Additionally, structural templates from the Protein Data Bank (PDB) and other modeling resources are integrated to improve prediction accuracy. More specifically, the OpenFold (Ahdritz et al., 2022) and ColabFold (Kim et al., 2024; Mirdita et al., 2022) implementations are integrated in the research and load the official pre-trained AlphaFold parameters.

In this research, several protein sequences are used and analyzed, with their entry codes being Q94KT8 (UniProt), 2KIZ (PDB), 5IQ5 (PDB), 7P27 (PDB), 7P2O (PDB), and 1GFL (PDB). After the initial use of the Q94Kt8 sequence, the rest of the sequences of the Protein Data Bank are being used due to their detailed characterization by nuclear magnetic resonance (NMR) spectroscopy. All of these structures are excellent test cases for this particular research, not only because of their biological significance but also because they represent the full spectrum of protein structure determination techniques. Specifically, 'Q94KT8' is a computationally predicted protein using the AlphaFold2 framework. The '1GFL' protein, derived from X-ray crystallography, provides a high-resolution, experimentally validated reference structure, essential for benchmarking accuracy and structural fidelity in the predictions. And the '2KIZ', '5IQ5', '7P27', and '7P2O' proteins are especially valuable as they are resolved using detailed characterization through NMR spectroscopy, offering deep insight into protein dynamics, conformational flexibility, and molecular interactions—areas often missed by static structural models.

3.1.1. Q94KT8 (COBRA) (UniProt, 0000a)

As a first scenario, the UniProt Q94KT8 sequence is being used. It comprises bioinformatics amino acids with annotations of the protein sequence, functional domains, post-translational modifications, and interactions. This particular protein molecule is selected for its significance in the examination of the functional aspects of plant cell wall organization and growth. Arabidopsis thaliana serves as a crucial model organism in plant biology research, offering insights into biological processes that can benefit other plant species (Koornneef & Meinke, 2010).

3.1.2. 2KIZ (Bank, 0000a)

'2KIZ' represents the NMR solution structure of Arkadia's RING-H2 E3 ubiquitin ligase domain, a critical protein involved in amplifying TGF-β signaling by promoting the degradation of its negative regulators. The structural characterization of this RING-H2 domain, along with its interaction properties with the ubiquitin-conjugating enzyme UbcH5b of E2, provides significant information on its conformational dynamics and ubiquitination mechanism, essential for understanding cellular signaling pathways and regulatory mechanisms. Studying this protein elucidates fundamental processes involved in signal transduction and molecular communication within biological systems, potentially contributing to broader biomedical applications.

3.1.3. 5IQ5 (Bank, 0000b)

The '5IQ5' structure represents the NMR solution structure of the Mayaro virus macro domain, a critical viral protein whose detailed structural and dynamic characterization sheds light on its nucleotide and RNA binding selectivity. This protein domain has been shown to preferentially bind adenosine diphosphate ribose (ADP-r) and adenine-rich RNAs, significantly contributing to the understanding of its structure–function relationship and highlighting its potential as a target for antiviral drug design, making it highly relevant for public health research.

3.1.4. 7P27 (Bank, 0000c)

The '7P27' sequence corresponds to the NMR solution structure of the Chikungunya virus macro domain, highlighting its adaptation to selective inhibitor binding, particularly GS-441524, the active metabolite of the antiviral drug remdesivir. This structural characterization provides valuable insights into mechanisms of viral replication and host immune response modulation, underscoring its importance for developing targeted antiviral therapeutics and informing global public health strategies.

3.1.5. 7P20 (Bank, 0000d)

The '7P20' protein is the NMR solution structure of the SUD-C domain of SARS-CoV-2, the virus responsible for the COVID-19 pandemic. This structure provides vital insights necessary for understanding the viral-host interactions critical for infection and replication, thus significantly aiding the development of targeted antiviral treatments and supporting ongoing public health efforts and pandemic preparedness.

3.1.6. 1GFL (GFP) (Bank, 0000e)

Finally, the '1GFL' sequence from the Protein Data Bank is explored due to its precise atomic-level structure obtained by X-ray crystallography. This crystal structure of the recombinant wild-type green fluorescent protein (GFP) reveals a unique cylindrical shape comprising 11 strands of β-sheet with an α-helix inside, known as the β-can motif. The protected fluorophore within this structure is integral to its fluorescence properties and allows investigation into specific mutations that alter GFP's spectral characteristics. This high-resolution, experimentally validated model allows rigorous benchmarking and validation of the predictive methods, ensuring structural accuracy and reliability. Incorporating '1GFL' strengthens the overall methodological framework by offering a robust reference point essential for validating the computational predictions against experimentally confirmed structural data.

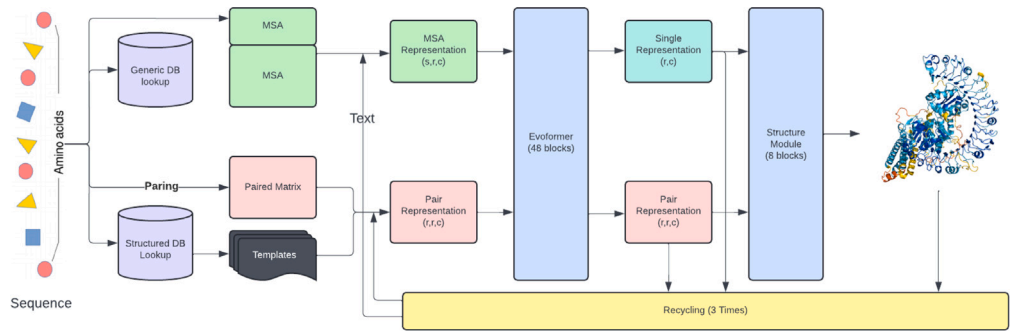


Fig. 2. The pipeline of AlphaFold2 Architecture.

Table 1
Overview of AlphaFold2.

Features	Description
Inputs	Amino acid sequence and similar sequences (MSA).
Outputs	Protein structure: X, Y, Z coordinates of each atom and different angles.
Quality metric	Global distance test, template modeling score, predicted local distance test.
Network description	Evoformer refines pair and MSA representations; the Structure module provides coordinates, recycling outputs as inputs 4x.
Number of parameters	93M, plus 4x recycling.
Evaluation sets	CASP-14 structures, Protein Data Bank after a specific date.
Training sets	Protein Data Bank: supervised (crystallography) and semi-supervised (high-confidence structures).
Training loss	Frame-aligned point error, auxiliary losses.

3.2. The AlphaFold2 model

3.2.1. Architecture

DeepMind's AlphaFold2 represents a revolutionary step forward in the prediction of three-dimensional protein structures. Using massive evolutionary data and machine learning, AlphaFold2 predicts protein structures with astonishing accuracy - a process that has presented considerable hurdles for decades (Jumper et al., 2021). The visualization in Fig. 2 provides an overview of AlphaFold2 architecture from a high-level perspective.

The workflow of AlphaFold2 begins with extracting protein sequences from databases such as UniProt. These sequences are then aligned using different multiple-sequence alignment tools, identifying homologous sequences across different species. This alignment process is crucial for maintaining protein structure and functions (Finn, Clements, & Eddy, 2011). After that, AlphaFold2 searches for structural templates in the Protein Data Bank (PDB). If a homologous structure is found, it simplifies the prediction task by providing a template. However, given that such templates are available for less than 0.1% of proteins AlphaFold2 often operates without them, relying on its robust computational methods to generate accurate predictions (Steinegger et al., 2019).

At the core of AlphaFold2 is a neural network based on the transformer architecture, comprising two main modules (Vaswani et al., 2017):

1. **The Evoformer:** The Evoformer iteratively refines the MSA and pair representation, employing cross-attention mechanisms to integrate information from both rows and columns of the alignment. This iterative refinement helps detect patterns of interaction between amino acids.
2. **The Structure module:** The Structure module then uses these refined embeddings to predict the 3D structure of the protein. This involves a detailed process of iterative updates and structure relaxation using physical simulation methods like OpenMM.

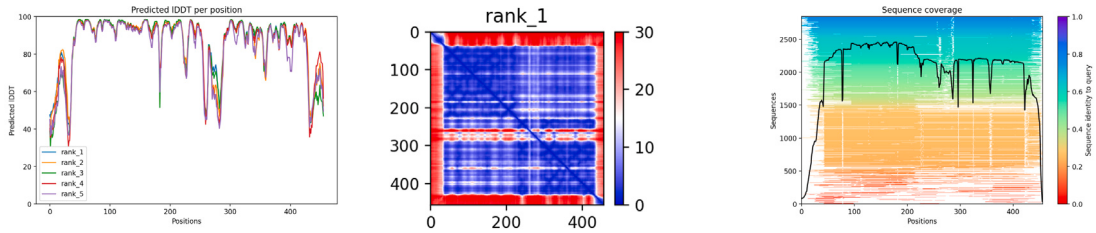
AlphaFold's end-to-end training approach includes innovative architectures designed for optimal information flow and several new loss functions. A notable feature is the recycling neural networks, which iteratively refine predictions to enhance accuracy. Table 1 describes the overview of the entire AlphaFold2 architecture.

Innovative methods like multi-head attention and scaled dot-product attention help the entire process by improving the model's ability to capture complex relationships between residues (Vaswani et al., 2017). The key to AlphaFold2's success is its capacity for prediction refinement through self-distillation and recycling, a process by which the model iteratively changes its predictions to achieve higher stability and accuracy. The final predicted structures are guaranteed to be highly accurate because of this iterative improvement.

3.2.2. Outputs

Besides the 3D structure prediction (in PDB format), the AlphaFold2 model generates some other important outputs depicts on Fig. 3,

- **Prediction Confidence Metrics (pLDDT) plot:** It provides a per-residue confidence score known as pLDDT (predicted Local Distance Difference Test), which indicates the model's confidence in the position of each residue for five ranked models (rank 1 to rank 5). The range is from 0 to 100, and the higher, the better.
- **Predicted Aligned Error (PAE) plot:** These plots show the predicted error in the distance between residue pairs, which helps to evaluate the overall reliability of the predicted structure. In this case, lower values refer to higher confidence. (Axes) Both axes represent the residue positions in the protein sequence. (Color Bar) The color bar ranges from blue to red, indicating the predicted alignment error in angstroms. Blue represents a low error (high confidence), while red represents a high error (low confidence). (Diagonal Line) The blue diagonal line from the top left to the bottom right represents the self-alignment of residues, where the predicted error is naturally low because a residue aligns with itself.
- **Multiple Sequence Alignment (MSA) Coverage plot:** These data are important for understanding the evolutionary relationships within the protein family. The color gradient helps identify which regions of the protein have high sequence identity (blue/purple) and which regions are more variable (red/orange). The black line represents the number of sequences aligned at each position. Dips in the black line indicate regions with fewer aligned sequences



(a) pLDDT (b) PAE (c) MSA Coverage

Fig. 3. AlphaFold2 output visualization.

Table 2
Overview of OpenFold.

Features	Description
Inputs	Amino acid sequence, MSAs, optional structural templates, and experimental constraints.
Outputs	3D structure (atomic coordinates, residue orientations), pLDDT scores, and aligned error metrics.
Quality metric	Global distance test, template modeling score, predicted local distance test.
Core modules	MSA processing, Evoformer for feature refinement, structure prediction module.
Innovations	Open-source, modular framework; designed for XAI integration and domain-specific flexibility.
Compatibility	Supports custom databases, datasets, and alignment tools.
Training loss	Modular loss functions; adaptable to user-defined constraints for specific applications.

3.3. The OpenFold model - A flexible AlphaFold2 alternative

3.3.1. Overview

OpenFold (Ahdritz et al., 2024) is an open-source protein structure prediction framework derived from AlphaFold2, emphasizing flexibility, transparency, and adaptability. While it retains the fundamental architecture of AlphaFold2, OpenFold introduces modularity that enables researchers to customize components for specific applications. This makes OpenFold particularly useful for integrating XAI techniques, such as ExplainableFold (Tan & Zhang, 2023) and DeepSHAP, to thoroughly study and understand specialized protein systems. This research integrates OpenFold in the place of AlphaFold2 while maintaining its accuracy and computational confidence. Table 2 shows a general overview of the OpenFold model.

3.3.2. Architecture and workflow

OpenFold operates on the foundational architecture of AlphaFold2 but introduces significant enhancements in modularity, transparency, and adaptability to support research-focused applications. Its design emphasizes component-level flexibility, enabling researchers to fine-tune or replace individual stages of the prediction pipeline. The system is organized into discrete submodules for multiple sequence alignment (MSA) processing, template retrieval, feature embedding, and structure prediction, each of which can be independently customized or extended. A key strength of OpenFold lies in its compatibility with custom datasets, alternative alignment tools, and domain-specific modeling constraints, making it particularly well-suited for advanced research in Structural Bioinformatics. This modularity facilitates the integration of external interpretability frameworks such as DeepSHAP, allowing fine-grained residue-level attribution and greater insight into the decision-making process of the model. By decoupling components traditionally hardcoded in AlphaFold2, OpenFold empowers experimentation with novel loss functions, structural constraints, and input modalities, positioning it as a versatile and extensible platform for cutting-edge protein modeling.

OpenFold operates similarly to AlphaFold2 but with enhanced modularity to support research-focused adaptations. It features key submodules for MSA processing, template generation, and structure prediction. OpenFold also supports the integration of custom datasets, alignment tools, and interpretability frameworks, offering researchers a versatile platform for protein modeling. The OpenFold outputs are the same as in the classic AlphaFold2 architecture, including a complete protein structure in PDB format, along with pLDDT scores for each residue, indicating the local confidence of the prediction. In addition, Predicted Aligned Error (PAE) matrices are generated to assess the expected positional error between residue pairs.

3.3.3. Sub-modules

1. **MSA Processing:** OpenFold leverages external tools such as JackHMMER, UniRef and HHblits for multiple sequence alignment (MSA). The pipeline allows users to substitute or augment these with alternative alignment methods tailored to specific use cases, such as low-homology sequences.
2. **Template Search:** The template search module identifies structural templates from sources like the Protein Data Bank (PDB). OpenFold provides flexibility to integrate novel databases or custom-defined templates, enhancing its utility in de novo design or comparative studies of underexplored protein families.
3. **Evoformer (Feature Refinement):** The Evoformer processes MSA and pairwise features using a transformer-based architecture. It refines the embeddings iteratively, with customizable parameters to accommodate specific biological or computational constraints. This allows researchers to optimize the model for challenging datasets or engineered sequences.
4. **Structure Prediction Module:** OpenFold's structure module predicts atomic coordinates, residue orientations, and secondary structure elements. Unlike AlphaFold2, OpenFold permits the addition of new loss functions or constraints, facilitating the incorporation of experimental data or novel objectives.
5. **Explainability and Transparency:** OpenFold is inherently designed to support interpretability tools. By integrating XAI methods like DeepSHAP, researchers can quantify the impact of individual residues on structural predictions, bridging the gap between computational predictions and biological insights.

3.3.4. Key features and advantages

1. **Flexibility:** OpenFold's modularity enables the replacement or augmentation of any component, allowing researchers to address specific challenges, such as low-sequence-homology datasets or novel structural motifs.
2. **Transparency**: The open-source design supports advanced XAI techniques, such as DeepSHAP, for interpreting predictions and understanding residue-level contributions to folding.
3. **Scalability**: OpenFold is computationally efficient and supports high-performance computing, making it suitable for large-scale studies and high-throughput scenarios.
4. **Customizability**: The framework allows researchers to integrate domain-specific datasets, experimental constraints, and novel evaluation metrics, broadening its applicability across diverse research areas.

Table 3
Comparison between OpenFold and AlphaFold2.

Feature	OpenFold	AlphaFold2
Modularity	Highly modular, customizable	Monolithic, limited flexibility
Transparency	Fully open-source, well-documented	Partially open-source, some obfuscated components
Experimental integration	Supports targeted modifications	Optimized for general structure prediction
Performance	Almost identical accuracy, customizable	State-of-the-art but less adaptable
Amino acid interaction analysis	Enables SHAP-based residue-level insights	Lacks built-in importance explanations
Protein engineering applications	Supports rational design	Primarily for structure prediction
Computational efficiency	Scalable, adaptable computing	High demand, limited optimization

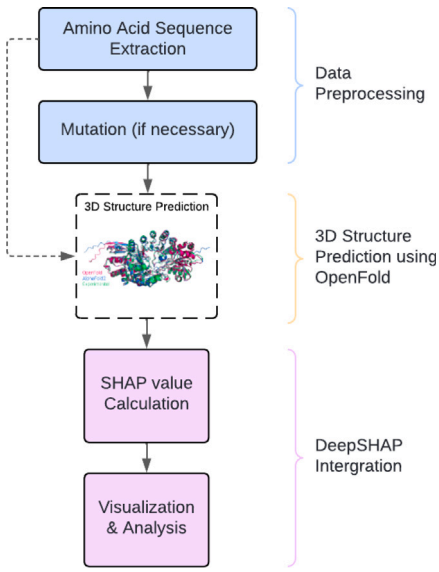


Fig. 4. Workflow of the proposed model for OpenFold.

3.3.5. Comparison to original AlphaFold2 model

While OpenFold leverages the strengths of AlphaFold2, its primary distinction lies in its modular design and transparency. These features make it more adaptable to experimental workflows and enhance its utility in hypothesis-driven research, such as exploring specific amino acid interactions or engineering proteins with unique properties. Table 3 portrays the comparison between these two models based on different aspects.

In conclusion, OpenFold offers a highly flexible and transparent platform for protein structure prediction, bridging the gap between computational modeling and biological understanding. Its adaptability makes it a valuable tool for cutting-edge research in structural biology and protein engineering.

3.3.6. Proposed workflow for OpenFold

Fig. 4 shows the workflow of the proposed model for OpenFold. The process begins with **Data Preprocessing**, where the Q94KT8 protein sequence is extracted from the UniProt database. This sequence undergoes mutation to simulate various scenarios. The modified sequences are then fed into the **OpenFold** (AlphaFold2) model for a 3D Structure Prediction, generating detailed predictions of the protein's three-dimensional conformation. Following this, the **DeepSHAP Integration** phase involves calculating SHAP (SHapley Additive exPlanations) values, which quantify the contribution of each amino acid to the predicted structure. This is achieved through a **SHAP Value Calculation step**. Finally, the model includes an **Interpretation Module** for **Visualization and Analysis**, which helps in understanding the significant amino acids and their roles in protein folding. This structured approach not only provides accurate predictions but also makes the decision-making process of AlphaFold2 transparent and interpretable, thereby fostering trust and facilitating deeper insights into protein structure and function.

3.4. ColabFold - another feasible AlphaFold2 alternative

3.4.1. Architectural overview

ColabFold (Kim et al., 2024; Mirdita et al., 2022) is a high-performance yet lightweight alternative to AlphaFold2 designed to enable fast, accessible protein structure prediction without requiring heavy computational resources. Built as a simplified and accelerated wrapper around AlphaFold2, ColabFold replaces traditional MSA generation methods with MMseqs2-based searches, enabling rapid multiple-sequence alignment in seconds rather than hours. Its portability and cloud compatibility make it particularly appealing for interactive research environments, such as Google Colab.

ColabFold operates as a streamlined interface to AlphaFold2, replacing traditional alignment steps and enabling rapid execution in accessible environments. At its core, ColabFold simplifies the multiple sequence alignment (MSA) generation pipeline by using MMseqs2, an ultra-fast sequence search algorithm that retrieves homologous sequences from protein databases. This MSA generation stage is significantly faster than HHblits or JackHMMER while maintaining the high alignment quality necessary for accurate predictions. Following MSA construction, ColabFold feeds the input data through AlphaFold2's original architecture, comprising the Evoformer and Structure module. The Evoformer refines sequence and pairwise representations using attention mechanisms, while the Structure module predicts 3D atomic coordinates and angles using a learned geometric model. ColabFold supports both template-based and template-free structure predictions, depending on user configuration.

The ColabFold outputs are the same as in the classic AlphaFold2 and OpenFold, including a complete protein structure in PDB format, pLDDT residue scores, Predicted Aligned Error (PAE) matrices and an MSA Sequence Coverage plot.

3.4.2. Sub-modules

ColabFold retains AlphaFold2's core predictive architecture, including Evoformer blocks and the structural module, but modifies its input pipeline for broader accessibility:

1. **MSA Generation via MMseqs2 and Acceleration:** ColabFold replaces AlphaFold2's and OpenFold's traditional HHblits/JackHMMER MSA pipeline with MMseqs2, a highly efficient sequence search tool to generate high-quality MSAs. MMseqs2 enables fast homologous sequence retrieval from large protein databases such as UniRef and BFD, significantly reducing alignment time while maintaining alignment diversity and depth essential for accurate structure prediction. The retrieved MSA is formatted in a manner compatible with AlphaFold2's Evoformer input requirements.
2. **Template Search Flexibility:** ColabFold optionally supports structural template usage via MMseqs2-template searches or remote PDB70/PDB100 alignments. Unlike OpenFold, template integration in ColabFold is streamlined for fast inference and is typically disabled when speed is prioritized. If enabled, template information is used alongside MSA data to guide folding decisions, especially in regions with low evolutionary signal.

Table 4
Comparison between ColabFold and AlphaFold2.

Feature	ColabFold	AlphaFold2
Modularity	Lightly modular, customizable via wrappers and scripting	Monolithic, limited flexibility
Transparency	Fully open-source and well-documented	Partially open-source; core weights and training pipeline remain obfuscated
Experimental integration	Supports targeted adaptation and lightweight custom workflows	Difficult to modify; designed for general-purpose usage
Performance	Near-identical accuracy, highly optimized for speed	State-of-the-art accuracy but slower and more compute-intensive
Amino acid interaction analysis	Enables SHAP-based residue-level importance explanations	Lacks built-in importance explanations
Protein engineering applications	Facilitates rational design through interpretable and rapid predictions	Primarily focused on structure prediction, less adaptable for design tasks
Computational efficiency	Highly scalable; runs efficiently on local GPUs or in cloud environments	Requires HPC-level infrastructure; high memory and time cost
MSA generation	Fast MMseqs2-based alignment	Slow, JackHMMER/HHblits-based alignment
Template search	Optional, fast, or disabled for speed	Enabled by default; HHSearch on PDB70
Deployment flexibility	Cloud-ready, Google Colab compatible	Requires dedicated GPU clusters or TPUs
User accessibility	Beginner-friendly, minimal setup, strong community tools	Complex installation, steep learning curve

3. **Feature Encoding and Model Input Preparation** Once the MSA (and optional template) is generated, ColabFold converts the input into the necessary feature dictionaries. These include sequence identity matrices, positional encodings, residue indices, and evolutionary coupling data. These are packed into the feature format expected by AlphaFold2's Evoformer module, ensuring compatibility with the downstream neural network layers.
4. **Structure Prediction via AlphaFold2 Core Modules:** ColabFold runs the unmodified Evoformer and Structure module from AlphaFold2. The Evoformer applies attention-based transformations to MSA and pairwise representations, while the Structure module infers atomic coordinates and torsion angles using invariant point attention and geometric reasoning. Outputs include the predicted 3D structure (PDB) and per-residue pLDDT scores, representing local confidence.
5. **Cloud-Native Design:** Its minimal setup and GPU acceleration make it ideal for non-HPC environments.
6. **Integrated Confidence Scoring:** Outputs include pLDDT scores and Predicted Aligned Error (PAE) plots to evaluate structural accuracy.

3.4.3. *Comparison to the original AlphaFold2 architecture*

While ColabFold utilizes the same predictive backbone as AlphaFold2, its primary distinction lies in its optimized accessibility and execution speed. Designed for lightweight deployment, ColabFold replaces AlphaFold2's compute-intensive MSA generation pipeline with a significantly faster MMseqs2-based alignment engine, enabling structure prediction in resource-constrained environments without compromising core accuracy. Unlike the monolithic and hardware-demanding AlphaFold2 implementation, ColabFold is optimized for cloud execution, including platforms like Google Colab, making it suitable for rapid prototyping, classroom education, or mobile Bioinformatics workflows. Table 4 summarizes the comparison between Original AlphaFold2 and ColabFold.

3.4.4. *Proposed model for ColabFold with DeepSHAP integration*

Fig. 5 illustrates the proposed model for ColabFold integrating DeepSHAP. An amino acid sequence is fed to ColabFold as input, generating a predicted 3D structure and pLDDT confidence scores. The same sequence is independently encoded and passed to a custom-trained Multi-Layer Perceptron (MLP) that learns to predict pLDDT scores from sequence data. DeepSHAP is applied to the MLP to obtain residue-level attribution of confidence predictions. This layered explanation enables interpretability of both the structural prediction and its associated reliability, providing a dual-view of model behavior from both functional and confidence perspectives.

As mentioned before, in the implementation of the proposed model for ColabFold, a Multi-Layer Perceptron (MLP) that is trained separately on one-hot encoded sequences to predict pLDDT scores is introduced.

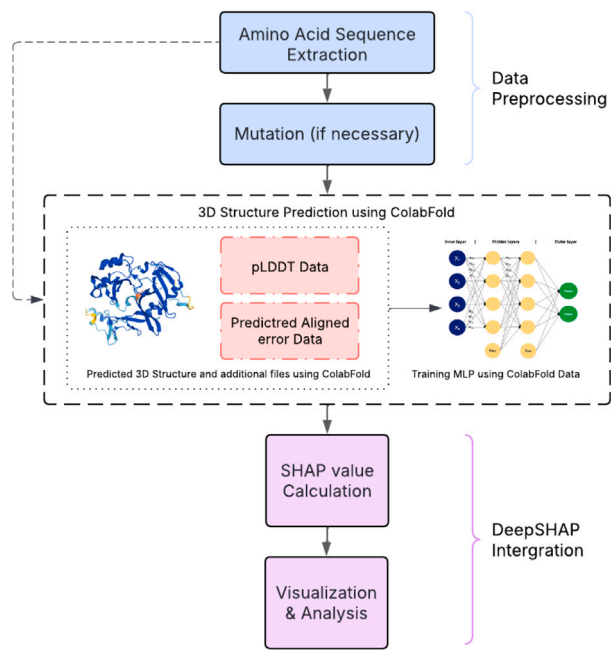


Fig. 5. Workflow of the proposed model for ColabFold.

This Lightweight Confidence Estimation via MLP allows confidence scoring to be explained independently of the structure prediction module. When integrated with DeepSHAP, this model enables residue-level attribution of confidence, bridging the gap between structural prediction and reliability analysis. The procedure and outputs of ColabFold's prediction interpretation via DeepSHAP and MLP involve the following steps:

1. **Execution of ColabFold's Prediction:** Inputting the original amino acid sequence into ColabFold to obtain the predicted 3D protein structure and associated residue-level confidence (pLDDT) scores in JSON format.
2. **Preparation of Training Dataset:** Extract the residue-level pLDDT scores from the ColabFold-generated JSON file. Encode the input protein sequence using a one-hot representation, creating a paired dataset of encoded sequences and corresponding residue-level pLDDT confidence scores.
3. **MLP Model Training and Optimization:** Design and training of a Multi-Layer Perceptron (MLP) neural network to learn the mapping between one-hot encoded sequences and the residue-level pLDDT confidence scores obtained from ColabFold. The MLP architecture consists of:

- *Input Layer:* Sized according to the one-hot encoded amino acid sequence (sequence length × 20).
- *Hidden Layers:* Two fully connected layers with 128 and 64 neurons, respectively, each using ReLU (Rectified Linear Unit) activation.
- *Output Layer:* A linear regression layer that produces a pLDDT confidence score for each residue.
- *Loss Function:* Mean Squared Error (MSE).
- *Optimizer:* Adam optimizer with a learning rate of 0.001.
- *Validation:* 5-fold cross-validation to ensure generalization and reduce overfitting risk.
- *Training Duration:* Model trained until convergence with early stopping based on validation loss.

This optimized architecture ensures that the surrogate MLP accurately approximates ColabFold's residue-wise confidence predictions, enabling subsequent SHAP interpretability.

4. **Model Validation and RMSD Evaluation:** Evaluate the MLP's predictive performance by computing metrics such as RMSD between the original ColabFold-generated pLDDT scores and the MLP-predicted scores to ensure accurate surrogate modeling.
5. **DeepSHAP Application to MLP Model:** Utilize DeepSHAP's DeepExplainer with the trained MLP surrogate model to perform residue-level attribution analysis, quantifying each residue's contribution to the predicted confidence scores.
6. **Interpretation and Biological Validation:** Interpret SHAP values to identify residue-level influence on confidence predictions, analyzing the results in the context of established biological knowledge (e.g., validating that residues in biologically irrelevant regions, such as the COBRA [380–390] cleavage site, exhibit consistently low SHAP attribution). Correlate SHAP attributions with known protein structural and functional information, reinforcing biological plausibility and model reliability.

Through these explicit steps, the developed **ColabFold-MLP-DeepSHAP** interpretative framework enables rigorous explainability analysis, effectively overcoming ColabFold's interpretability limitations and significantly broadening the applicability and robustness of DeepSHAP across diverse protein models and scenarios.

3.5. OpenFold and ColabFold model explanation using DeepSHAP XAI framework

3.5.1. General overview

DeepSHAP works by combining Deep Learning Important Features (DeepLIFT) with Shapley Additive exPlanations (SHAP) values to provide insights for complex neural network models. This integration aims to improve consistency, accuracy, and alignment with human intuition in model explanations.

A significant advantage of DeepSHAP is that it does not require a large number of protein cases to validate experimental results, as it is not a neural network-based approach. Unlike Deep Learning models, which rely on extensive training datasets and suffer from potential biases introduced by dataset composition, DeepSHAP operates independently of training data. By directly computing feature attributions through a principled integration of SHAP values and DeepLIFT, it eliminates the need for large-scale data dependency, ensuring robust and unbiased interpretability. This independence from pre-trained models and dataset limitations makes DeepSHAP a highly accessible, computationally efficient, and scientifically rigorous tool for understanding protein structure predictions. Its ability to provide immediate, biologically meaningful insights without requiring prior knowledge of similar protein structures establishes it as a valuable method for researchers seeking reliable explanations in structural Bioinformatics.

DeepSHAP represents a paradigm shift in the interpretability of AI-driven protein structure predictions, addressing one of the most pressing challenges in computational biology—the need for transparency and robustness without the constraints of data dependency. Unlike neural network-based approaches, which necessitate extensive protein datasets for model training and validation, DeepSHAP circumvents these limitations by providing direct feature attributions through a mathematically grounded integration of SHAP values and DeepLIFT. This allows for high-resolution interpretability without introducing biases from data selection or model overfitting. More importantly, its independence from training data means that DeepSHAP can be applied universally across diverse protein families, making it a powerful tool for researchers working in areas with limited experimental data. By eliminating the need for exhaustive training while ensuring biologically meaningful insights, DeepSHAP not only enhances the transparency of AI-based protein modeling but also democratizes access to explainable AI in Bioinformatics. This methodological advancement establishes DeepSHAP as a gold standard for unbiased, computationally efficient, and biologically interpretable protein structure analysis—paving the way for more trustable AI-driven discoveries in structural biology and drug design.

3.5.2. SHAP values

SHAP values, derived from cooperative game theory, allocate a fair share of the prediction to each feature. The properties of SHAP values make them particularly useful for feature attribution:

1. **Local Accuracy**: The explanation model closely approximates the original model's output for a specific input.
2. **Missingness**: Absent Features do not contribute to the prediction.
3. **Consistency**: If a feature's contribution to the prediction increases or remains the same in a new model, its attributed importance should not decrease.

The SHAP value for a feature i is given by Lundberg and Lee (2017):

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!} \left[f_x(z') - f_x(z' \setminus i) \right] \tag{1}$$

where,

- $\phi_i(f,x)$: The SHAP value for feature i given the model f and input x. It quantifies the contribution of feature i to the prediction.
- $z' \subseteq x'$: All possible subsets of the feature set x'. Here, x' represents the set of all input features.
- $|z'|$: The number of features in the subset z'.
- M: The total number of input features.
- $f_x(z')$: The model's prediction for the input subset z'.
- $f_x(z' \setminus i)$: The model's prediction for the input subset z' without feature i.

The equation calculates the SHAP value by summing the marginal contributions of feature i over all possible subsets of features z', weighted by the number of features in z' and the total number of features M (Lundberg & Lee, 2017). Fig. 6 shows how shape values attribute to all the features based on changing the model prediction while depending on that particular feature.

According to Fig. 6, base value $E[f(z)]$ might be predicted when any features from the current output $f(x)$ are unknown. ϕ_0 indicates to the base prediction rather than stick into 0, which makes sense. On the other hand ϕ_1, ϕ_2 and ϕ_3 indicate the saturation of the prediction from the base prediction according to specific feature values, whereas ϕ_4 indicates somehow any feature impacted the model's prediction returning to the base prediction value.

3.5.3. DeepLIFT

DeepLIFT (Deep Learning Important Features) is a method designed to interpret deep learning models' predictions by quantifying each input feature's contribution to the model's output. It does this by comparing the model's output for a given input x with its output for

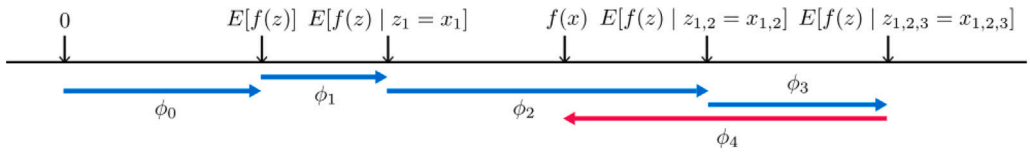


Fig. 6. Changing of SHAP values based on the model's prediction in single ordering (Lundberg & Lee, 2017).

a reference input r, which typically represents a baseline or neutral state (Shrikumar et al., 2017).

The contribution score of DeepLIFT can be calculated with the following equation (Shrikumar et al., 2017),

$$\sum_{i=1}^{n} C_{\Delta x_i \Delta o} = \Delta o \tag{2}$$

where $o = f(x)$ is the model output, $\Delta o = f(x) - f(r)$, $\Delta x_i = x_i - r_i$, and r is the reference input.

The method ensures that the sum of all contributions equals the total output difference, a property known as "summation-to-delta". The contributions are guaranteed to be consistent, and their sum reflects the true change in the model's output due to the input features.

3.5.4. Integrating SHAP and DeepLIFT

DeepSHAP combines SHAP values for smaller network components into SHAP values for the entire network. This is done by recursively passing DeepLIFT's multipliers, redefined in terms of SHAP values, backwards through the network:

1. Compute SHAP values for each network component.
2. Combine these values using a chain rule similar to back-propagation.

The integration is illustrated with the following equations (Sundararajan, Taly, & Yan, 2017):

Multiplier for Feature x_j and Output f_3,

$$m_{x_j f_3} = \frac{\phi_i(f_3, x)}{x_j - E[x_j]} \tag{3}$$

Multiplier for Intermediate Node y_i and Output f_j,

$$m_{y_i f_j} = \frac{\phi_i(f_j, y)}{y_i - E[y_i]} \tag{4}$$

Combining Multipliers (Chain rule),

$$m_{y_i f_3} = \sum_{j=1}^{2} m_{y_i f_j} m_{x_j f_3} \tag{5}$$

Approximate SHAP Value for Output (Linear approximation),

$$\phi_i(f_3, y) \approx m_{y_i f_3}(y_i - E[y_i]) \tag{6}$$

where in all cases, $\phi_i(f_i, x_i/y_i)$ indicates SHAP value for feature x or y with respect to the output f. These equations ensure an efficient combination of SHAP values for individual components to approximate the values for the entire model.

DeepSHAP enhances computational efficiency by reducing the reliance on heuristic linearization and improving sample efficiency. Kernel SHAP, a variant used in DeepSHAP, provides more accurate results with fewer model evaluations compared to other sampling methods. Studies indicate that DeepSHAP's explanations align well with human intuition, making it easier for users to trust and understand model predictions (Lundberg et al., 2018).

3.5.5. Computational complexity analysis of DeepSHAP

The computational setup consisted of a DGX server equipped with NVIDIA A100 GPUs, enabling the efficient execution of DeepSHAP's calculations. The inference time for a single protein sequence of approximately 400 residues is measured to be 8 min, depending on the batch size and sequence length. The computational complexity of DeepSHAP is influenced by three primary factors: the number of input features (protein sequence length N), the number of reference samples (R), and the depth of the neural network (D). The overall computational cost is given by:

$$O(N \times R \times D) \tag{7}$$

which scales linearly with the number of amino acids and reference samples, making it significantly more efficient than counterfactual-based methods like ExplainableFold.

ExplainableFold requires:

$$O(N) \tag{8}$$

full model runs as it iterates over each amino acid independently, making it computationally expensive for long protein sequences. In contrast, DeepSHAP leverages backpropagation-based attributions, reducing the number of model evaluations required while maintaining interpretability.

Additionally, DeepSHAP is more memory-efficient than Kernel SHAP, which typically scales as:

$$O(2^N) \tag{9}$$

for exact Shapley value computation, making it impractical for high-dimensional problems (Lundberg & Lee, 2017). The combination of DeepLIFT and SHAP significantly reduces sampling overhead while maintaining interpretability and local consistency.

By integrating DeepSHAP within the OpenFold framework, this study introduces a scalable and biologically meaningful method for explaining AlphaFold2's predictions. The ability to quantify the relative importance of individual amino acids without direct sequence modifications makes DeepSHAP a powerful and computationally feasible tool for protein structure interpretation.

4. Experiment & result analysis

4.1. Protein structure prediction using OpenFold

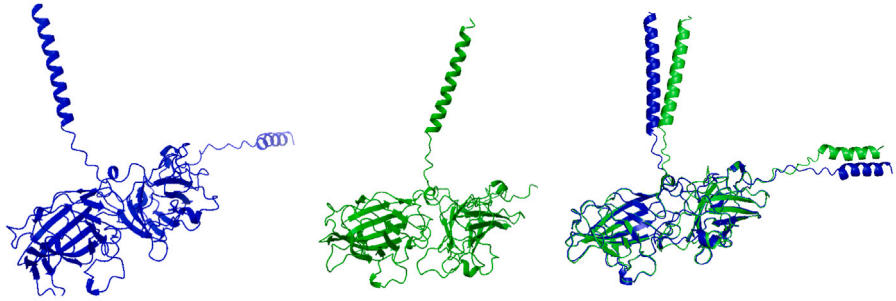
4.1.1. Data preparation

Considering the computational resource constraints and to facilitate a clear demonstration of the proposed approach, the 'Q94KT8' protein sequence as the test case for evaluating the OpenFold framework is chosen. However, one can choose any type of protein sequence. First, the Q94KT8 protein sequence is extracted from the UniProt database. Multiple sequence alignments (MSAs) are generated using JackHMMER, HHblits, and Uniclust30. Structural templates from the Protein Data Bank (PDB) are also incorporated to guide the predictions. The produced output data are presented in Fig. 8.

4.1.2. Application of mutations

After that, various mutations are introduced to the protein sequence to simulate different scenarios. These mutations involved systematic deletion or substitution of specific amino acids to observe the resulting changes in the predicted 3D structure, as analyzed in 2.1. For example, the protein sequence mentioned consists of 456 amino acids where position number 161 to 170 is,

$GAA\mathbf{G}TTN$



(a) OpenFold predictions (b) PyMOL-aligned structures

Fig. 7. Comparison between OpenFold 3D structures of the COBRA protein: (a) Prediction from sequence (b) Alignment using PyMOL (Blue): original, (Green): mutated.

For mutation purposes, these residues have been changed to,

$GAAATTN$

This specific mutation is introduced to examine how altering a single amino acid affects the protein's structure, providing insights into its functional dynamics and potential implications for biological processes.

4.1.3. 3D structure prediction and observation of mutations

The original and mutated sequences are fed into the OpenFold (AlphaFold2) model to generate detailed predictions of the protein's three-dimensional conformation. This advanced neural network architecture, including the Evoformer and structure modules, is utilized for this task. Fig. 7 shows the view of the structure from the COBRA original protein sequence and a mutated sequence using OpenFold.

If these two structures are more closely observed, it is clear that there is a structural difference between them because of this single mutation. This can also be simply verified by an RMSD (Root Mean-Square Deviation) (contributors, 2024b) calculation between them. RMSD is a characteristic geometric metric that is widely used in comparing the similarity of experimental and predicted structures in 3D space (x, y, z axis) and is expressed in Ångstrom (Å). RMSD metric is defined by the following equation (DSpace, 0000):

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(x_{ci}-x_{di})^2+(y_{ci}-y_{di})^2+(z_{ci}-z_{di})^2\right]} \tag{10}$$

To visualize this $RMSD$ calculation between the two structures, a simple structure alignment and $RMSD$ calculation using PyMOL (Schrödinger & DeLano, 2020) is made. The computed $RMSD$ value is 0.283 Å, which indicates the structural change due to the single amino acid mutation. For a single amino acid mutational change, this $RMSD$ value can be considered as relatively high, although the structures remain almost structurally identical ($RMSD < 3$ Å) since they are very closely related models that differ in exactly one amino acid. This indicates the structural (and, potentially, biological) importance of amino acid mutations and the undeniable need for an explainability framework towards this. The effect of a single amino acid mutation can also be further analyzed in a more expensive "biologically-oriented" way by using **Molecular Dynamics (MD) Simulations**. By calculating $RMSD$ and $Radius\ of\ Gyration$ (R_g) (contributors, 2024a) over time both for the initial (wildtype) and mutated structure's MD simulation, it is possible to see the mutational effects on proteins during the simulation time. However, the MD simulations are unfortunately computationally expensive and require pretty high computational power to be executed. Furthermore, MD simulations integrate many other in-silico biological parameters (i.e. Force Field, Pressure, Temperature, Ionization, Solvation etc.) that have a crucial effect on the molecular study depending on the simulation's numerical and computational accuracy and simulation's specific purpose (i.e. nanomaterial simulation, polymer design, molecular binding etc.). Additionally, Fig. 8 shows the side-by-side comparison of other outputs between original and mutated COBRA sequences predicted by OpenFold.

Table 5
Summary of SHAP values for key residues in Q94KT8.

Amino acid position	SHAP value	Contribution to structure stability
45	0.32	High
67	0.29	Moderate
120	−0.15	Low
152	0.27	Moderate
189	−0.35	High

Therefore, it is important to have an algorithmic-oriented and computationally-friendly way to predict and evaluate the amino acid effects on protein structures, focusing only on the static folded PDB output, without the urgent need and high dependence on complex in-silico biological parameters and experimental information like MD simulations have.

4.1.4. OpenFold model interpretation using DeepSHAP

After feeding the protein sequence into the OpenFold (AlphaFold2) model, DeepSHAP has been applied to both the input sequence and the model's output. SHAP (SHapley Additive exPlanations) values are calculated for each prediction to quantify the contribution of each amino acid to the predicted structure. This involved:

1. Applying DeepLIFT to assign importance scores to each input feature by comparing the activation of neurons to a reference activation (Shrikumar et al., 2017).
2. Using SHAP values to refine these contributions, ensuring they are consistent and additive (Lundberg & Lee, 2017).
3. Combining SHAP values for individual network components into SHAP values for the entire network through recursive passing of DeepLIFT's multipliers.

By analyzing the SHAP values, this research identified key amino acids that play a critical role in maintaining the protein's conformation. For example, mutations in specific residues led to significant changes in the predicted structure, indicating their importance in protein folding.

Table 5 summarizes SHAP values and contributions to structural stability for key residues in the COBRA protein (Q94KT8). It lists positions 45, 67, 120, 152, and 189, with SHAP values indicating the impact on prediction accuracy. Positive SHAP values (e.g., 0.32 at position 45 and 0.29 at position 67) suggest these residues improve prediction accuracy, while negative values (e.g., −0.15 at position 120 and −0.35 at position 189) indicate a decrease in accuracy. Positions 45 and 67, contributing highly and moderately to structural stability, respectively, enhance accuracy, whereas position 120, with a low stability contribution, reduces accuracy. Position 152 moderately increases accuracy, and position 189, despite its high stability contribution, decreases prediction accuracy. This highlights the dual role of certain residues in structural stability and prediction accuracy. Fig. 9 shows the view of how calculated SHAP values impact the model's prediction accuracy.

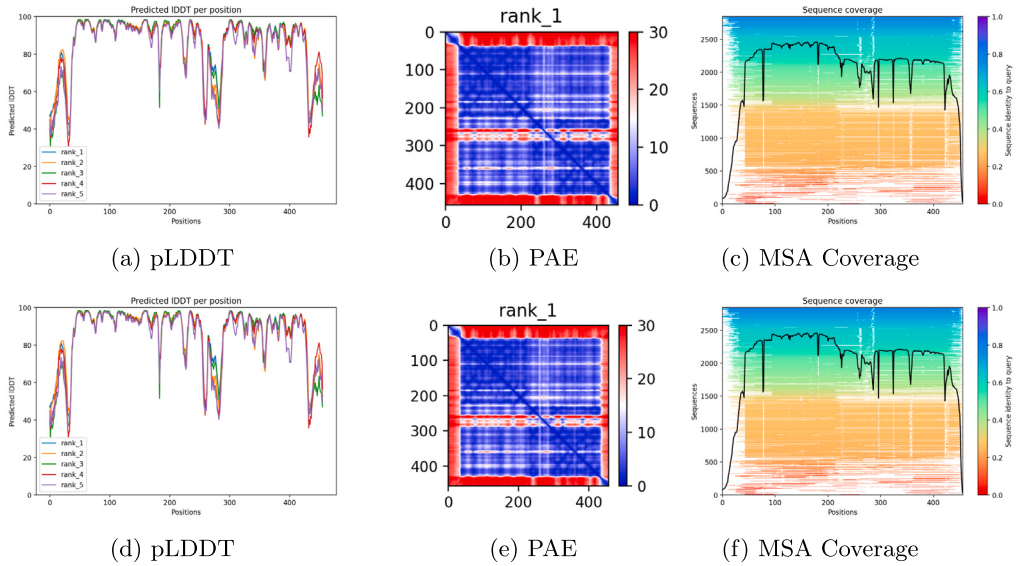


Fig. 8. OpenFold output visualizations for the Q94KT8 sequence (first row: original; second row: mutated).

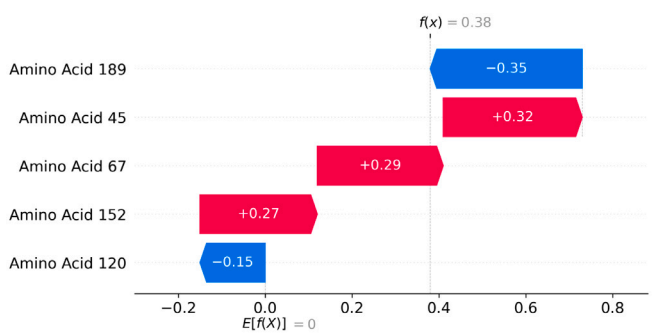


Fig. 9. Waterfall plot showing the contribution of calculated SHAP values on model's prediction accuracy.

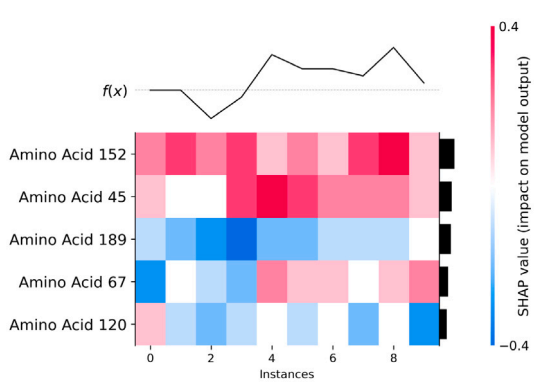


Fig. 10. Heatmap showing impacts of SHAP values for particular amino acid positions.

A heatmap as Fig. 10 is also being illustrated for better intuitions. Each cell in the heatmap corresponds to a specific amino acid position and its SHAP value, with colors ranging from blue to red. Blue shades represent negative SHAP values, indicating a decrease in prediction accuracy, while red shades represent positive SHAP values, indicating an increase in accuracy. The intensity of the color reflects the magnitude of the SHAP value.

Furthermore, the Q94KT8 protein is involved in determining the orientation of cell expansion, probably by playing an important role in cellulose deposition. It may act by recruiting cellulose-synthesizing complexes to discrete positions on the cell surface. A partial protein missing the N-terminal signal peptide was reported to complement a yeast mutant defective in phytochelatin synthesis. It is, therefore, possible that COBRA binds divalent metals (UniProt, 0000b). Also, it is known that the Q94KT8 protein is anchored to the glycosylphosphatidylinositol (GPI) through a predicted cleavage site at the amino acid position ASN-384 (Asparagine in the 384th position), which is ASN-385 in the original Q94KT8 protein sequence PDB file. This site is crucial for its GPI substitution and the subsequent membrane anchoring. Mutations at or around this position (e.g., N384T, G385 W, S387P) or truncation from the cleavage site significantly impair the protein's proper functioning and localization (Roudier et al., 2005).

Specifically, ASN-385 in the COBRA protein is predicted to be part of the site where the glycosylphosphatidylinositol (GPI) anchor attaches. However, it is not the amino acid itself that binds directly to the GPI. Instead, this position is part of the cleavage and recognition site, where the protein undergoes post-translational modification. The GPI anchor is attached to the newly exposed C-terminus after a specific cleavage event near this site. After cleavage at this site, the GPI anchor is covalently attached to the newly formed C-terminus, enabling the protein to be anchored to the membrane. So, while ASN-385 is neither an actual binding residue for GPI nor acts as a post-folding parameter, it is essential for the post-translational modification that leads to GPI anchoring (Schindelman et al., 2001). Therefore, ASN-385 amino acid is "biologically" important (at least in terms of post-translation modification) and therefore a SHAP value would be very insightful and useful to be computed for it too, and for its amino acid "neighborhood" which is from 380th until the 390th amino acid (amino acid range [380, 390]).

By analyzing the SHAP values within the amino acid range [380, 390], it can be seen if these biologically significant residues have a critical role not only in post-translational modifications but also in the COBRA's structural folding too, i.e. in stabilizing the COBRA protein's conformation. This in-depth analysis through DeepSHAP offers valuable insights into their contribution to structural stability, providing an all-around biological evaluation in terms of both post-translational modifications (from literature) and protein folding (from DeepSHAP). Moreover, by elucidating the folding contributions of these residues, this approach uncovers hidden mechanisms that are essential for further understanding their biological significance. These findings could serve as a crucial foundation for future enhanced large-scale wet

Table 6
Summary of SHAP values for amino acid range [380,390].

Amino acid position	SHAP value	Contribution to structure stability
380	0.0	None
381	0.0	None
382	0.0	None
383	0.0	None
384	0.0	None
385	0.0	None
386	0.0	None
387	0.0	None
388	0.0	None
389	0.0	None
390	0.0	None

lab studies about COBRA's post-translational modifications, particularly those that lead to GPI anchoring.

Table 6 summarizes SHAP values and contributions to structural stability for, based on literature, key residues in the COBRA protein (Q94KT8). It lists positions 380 until 390 (range [380, 390]), with SHAP values indicating the impact on prediction accuracy and structural folding.

From literature (Roudier et al., 2005; Schindelman et al., 2001; UniProt, 0000b), it is known and mentioned that these residues are important in terms of post-translational modification and generally not in terms of protein tertiary structure folding and stability. Therefore, it is expected that the SHAP values for the amino acid range [380, 390] would be close or equal to zero.

Indeed, the method produces the expected output values, which verifies and aligns with the established biological findings and the up-to-date experimental biological knowledge derived from literature about COBRA's structure and demonstrates the accuracy and reliability of this approach in also identifying residues that do not contribute to tertiary structure folding or stability. By assigning a SHAP value of zero to these positions, this methodology corroborates the established experimental biological understanding of the COBRA protein. This result validates that this interpretative framework, based on DeepSHAP and OpenFold, can effectively differentiate between residues influencing structural stability and those playing roles mainly in other independent biological processes, such as post-translational modifications.

4.2. Protein structure prediction using ColabFold

4.2.1. Data preparation

In the ColabFold case, due to its cloud-based computational efficiency and feasibility, all of the protein sequences presented in 3.1 are studied.

'Q94KT8' (COBRA). For the 'Q94KT8' case, the applied methodology is the same as in 4.1.1 and 4.1.3 with the OpenFold framework. Fig. 11(a) shows the structure of the original and mutated COBRA sequence using ColabFold, and Fig. 11(b) shows the comparison between them after alignment using PyMOL.

Again, to visualize the $RMSD$ difference between the two different (original and mutated) structures, they are aligned, and their RMSD difference using PyMOL (Schrödinger & DeLano, 2020) is calculated. The computed $RMSD$ value is 0.157 Å, which is lower than the 0.283 Å in the OpenFold framework. This lower RMSD indicates that the ColabFold model, despite employing a simplified and faster alignment pipeline, achieves highly accurate structural predictions comparable to, or even surpassing, those of OpenFold. Also, most residues in both plots show high confidence (scores > 90), while certain regions—particularly near the N- and C-termini and several loop regions exhibit lower confidence, indicated by dips in the score. The consistency across ranks suggests the robustness of the prediction.

Furthermore, since the RMSD value is different, it is obvious that ColabFold "understands" and represents the single mutational effect differently than OpenFold. This is a parameter that should be considered for the DeepSHAP analysis in the following chapters, keeping in mind that different (or pretty similar) SHAP values may be assigned to the Q94KT8 (COBRA) amino acids when applying DeepSHAP in the two frameworks (OpenFold and ColabFold).

The other ColabFold-generated outputs for the original and mutated 'Q94KT8' (COBRA) sequence are depicted by the (Fig. 12) and it is observed that the corresponding outputs are almost identical, which agrees with the low RMSD value calculated before since these two proteins only differ in one amino acid.

'2KIZ'. The '2KIZ' structure is retrieved from the Protein Data Bank, and afterwards, its sequence is inserted as input in ColabFold. The 3D structures of '2KIZ' (original from NMR, with two Zinc ions bounded, and from ColabFold) are visualized in Fig. 13.

The RMSD difference between the aligned structures is 2.262 Å, which clearly shows that ColabFold does make predictions with great confidence and accuracy for this sequence case, as discussed in 3.4.4 regarding the RMSD value cutoffs. Additional outputs generated by ColabFold execution are shown in Fig. 14

'5IQ5'. The '5IQ5' amino acid sequence is retrieved from the Protein Data Bank and, afterwards, inserted as an input in ColabFold. Fig. 15 shows the 3D structures of '5IQ5' (from NMR and ColabFold predicted).

The RMSD difference between the aligned structures is 1.287 Å, which clearly shows that ColabFold does make predictions with great confidence and accuracy for this sequence case, as discussed in 4.1.3 regarding the RMSD value cutoffs. The other generated outputs from the ColabFold execution are visualized in Fig. 16.

'7P27'. The '7P27' structure is retrieved from the Protein Data Bank, and afterwards, the sequence is inserted as input in ColabFold. The 3D structures of '7P27' (from NMR and ColabFold predicted) are visualized in Fig. 17.

The RMSD difference between the aligned structures is 2.039 Å, which clearly shows that ColabFold does make predictions with great confidence and accuracy for this sequence case, as discussed in 3.4.4 regarding the RMSD value cutoffs. The other generated outputs from the ColabFold execution are visualized in Fig. 18.

'7P2O'. The '7P2O' structure is retrieved from the Protein Data Bank and then inserted as input in ColabFold. The 3D structures of '7P2O' (from NMR and ColabFold predicted) are visualized in Fig. 19.

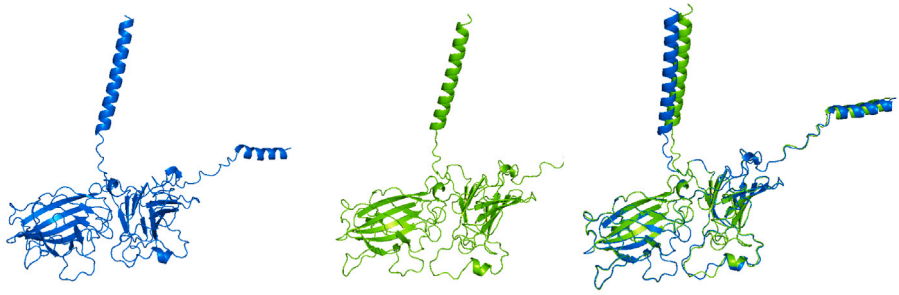
The RMSD difference between the aligned structures is 2.618 Å, which clearly shows that ColabFold does make predictions with great confidence and accuracy for this sequence case, as discussed in 4.1.3 regarding the RMSD value cutoffs between the predicted and native structures. The other generated outputs from the ColabFold execution are shown in Fig. 20.

'1GFL' (Green Fluorescent Protein). The '1GFL' structure is retrieved from the Protein Data Bank and afterwards used as input in ColabFold. The 3D structures of '1GFL' (original from X-ray crystallography and predicted from ColabFold) are visualized in Fig. 21.

The RMSD difference between the aligned structures is 0.310 Å, which clearly shows that ColabFold does make predictions with exceptional confidence and accuracy for this sequence case, as discussed in 3.4.4 regarding the RMSD value cutoffs between predicted and native structures. Additional generated outputs from the ColabFold execution are depicted in Fig. 22

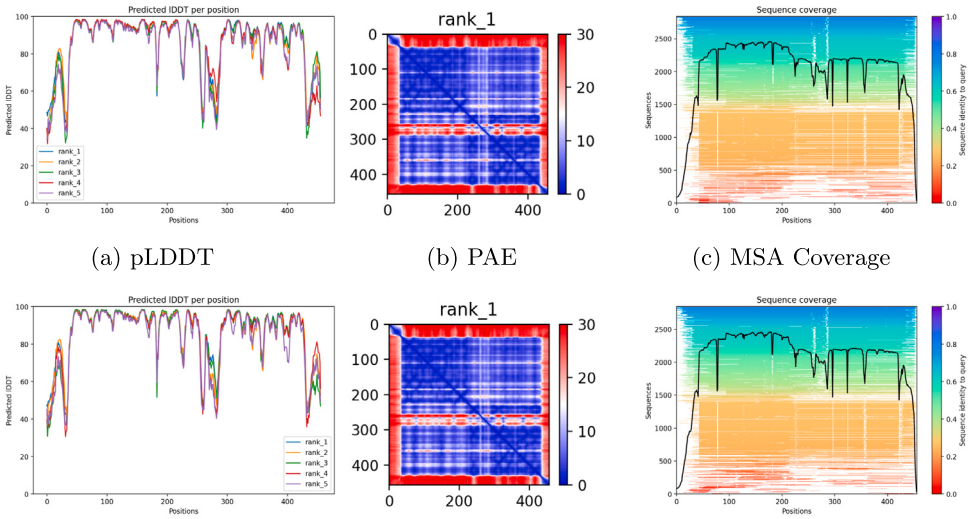
4.2.2. ColabFold model interpretation using DeepSHAP

Unlike OpenFold, ColabFold's architecture is optimized primarily for lightweight, cloud-based execution. Therefore, it does not readily support direct DeepSHAP application due to its online-based execution pipeline. Specifically, as ColabFold is primarily accessible as a remote service, DeepSHAP attribution cannot be directly applied to its



(a) ColabFold Predicted Structures (b) PyMOL-aligned Visualization

Fig. 11. Comparison between ColabFold 3D structures of the COBRA protein: (a) Prediction from sequence. (b) Alignment of original and mutated structures using PyMOL. (Blue): original; (Green): mutated.



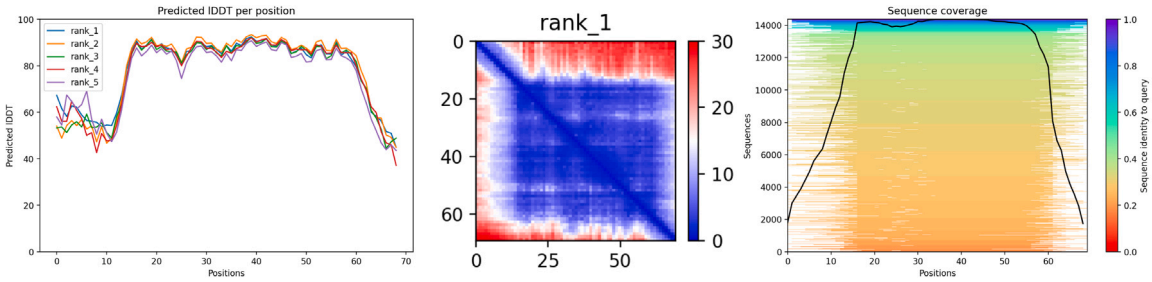
(a) pLDDT (b) PAE (c) MSA Coverage

(d) pLDDT (e) PAE (f) MSA Coverage

Fig. 12. ColabFold output visualizations for the Q94KT8 sequence (first row: original; second row: mutated).

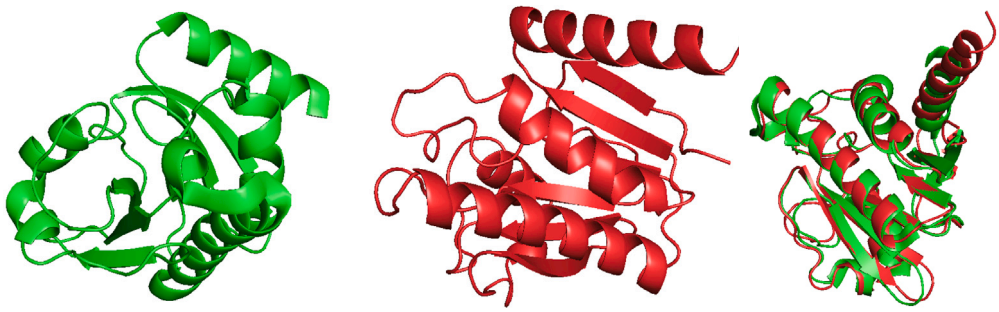
(a) NMR (b) ColabFold (c) Aligned

Fig. 13. Comparison of '2KIZ' 3D structures: (a) NMR, (b) ColabFold Predicted, (c) Both aligned.



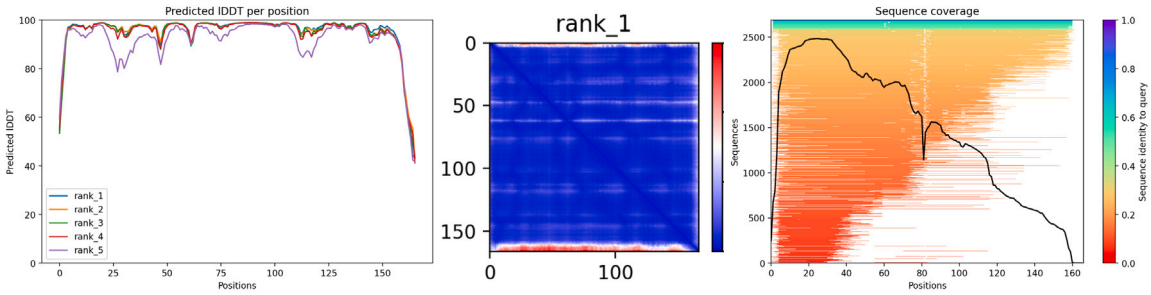
(a) pLDDT (b) PAE (c) MSA Coverage

Fig. 14. ColabFold output visualizations for structure '2KIZ': (a) pLDDT confidence, (b) Predicted Alignment Error (PAE), (c) MSA Coverage.



(a) NMR (b) ColabFold (c) Aligned

Fig. 15. Comparison of '5IQ5' 3D structures: (a) NMR, (b) ColabFold predicted, (c) Both aligned.

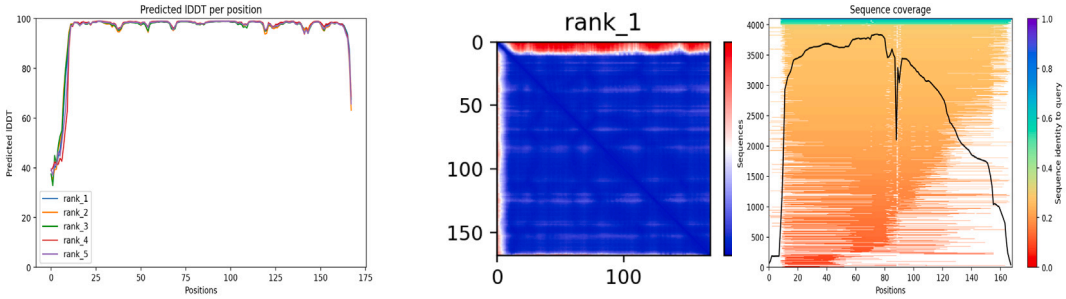


(a) pLDDT (b) PAE (c) MSA Coverage

Fig. 16. ColabFold output visualizations for structure '5IQ5': (a) pLDDT confidence, (b) Predicted Alignment Error (PAE), (c) MSA Coverage.

(a) NMR (b) ColabFold (c) Aligned

Fig. 17. Comparison of '7P27' 3D structures: (a) NMR (b) ColabFold predicted (c) Both Aligned.



(a) pLDDT (b) PAE (c) MSA Coverage

Fig. 18. ColabFold output visualizations for structure '7P27': (a) pLDDT confidence, (b) Predicted Alignment Error (PAE), (c) MSA coverage.

internal neural modules. To overcome this limitation, ColabFold's accessible prediction outputs are leveraged, notably the confidence scores (pLDDT), which encapsulate underlying decision-making information from the model.

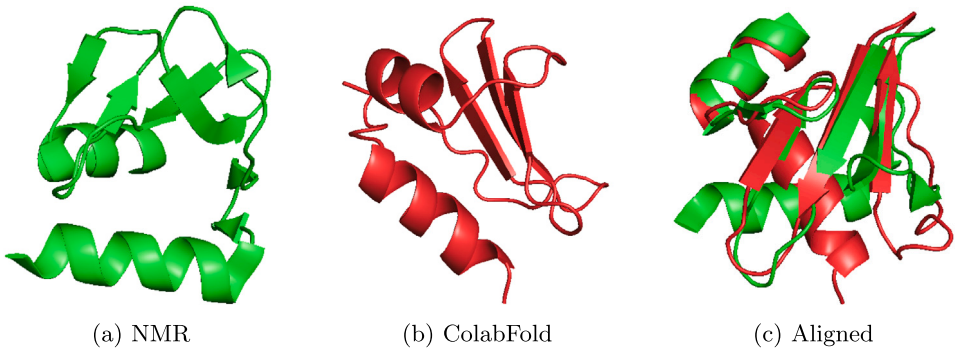


Fig. 19. Comparison of '7P2O' 3D structures: (a) NMR (b) ColabFold predicted (c) Both aligned.

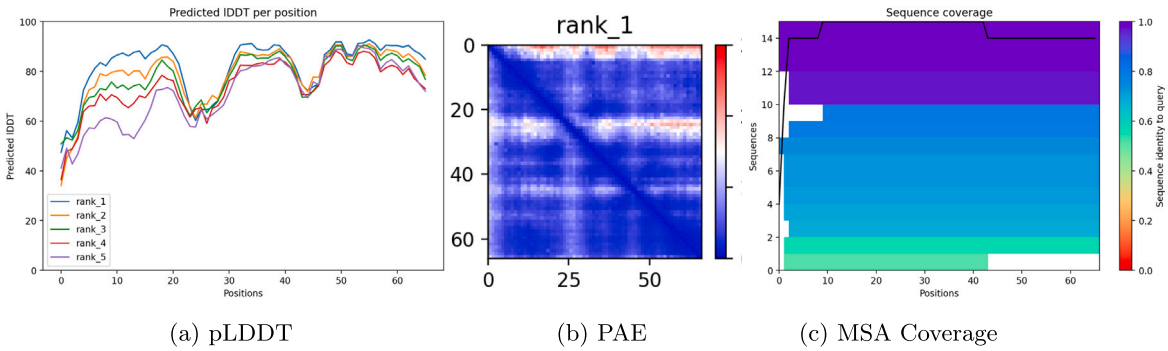


Fig. 20. ColabFold output visualizations for structure '7P2O': (a) pLDDT confidence, (b) Predicted Alignment Error (PAE), (c) MSA coverage.

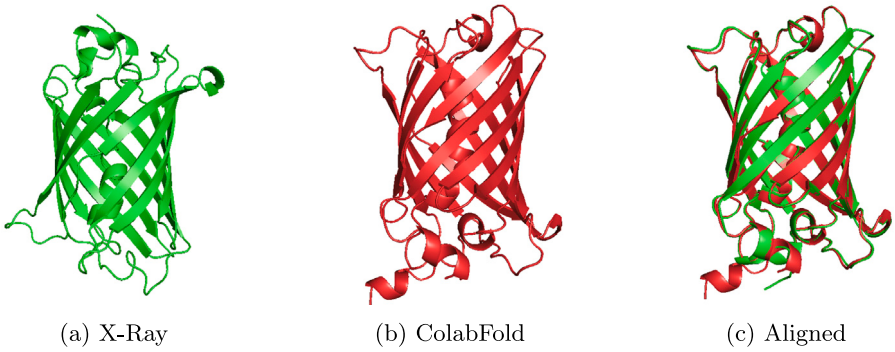


Fig. 21. Comparison of '1GFL' 3D structures: (a) X-ray Crystallography (b) ColabFold Predicted (c) Both aligned.

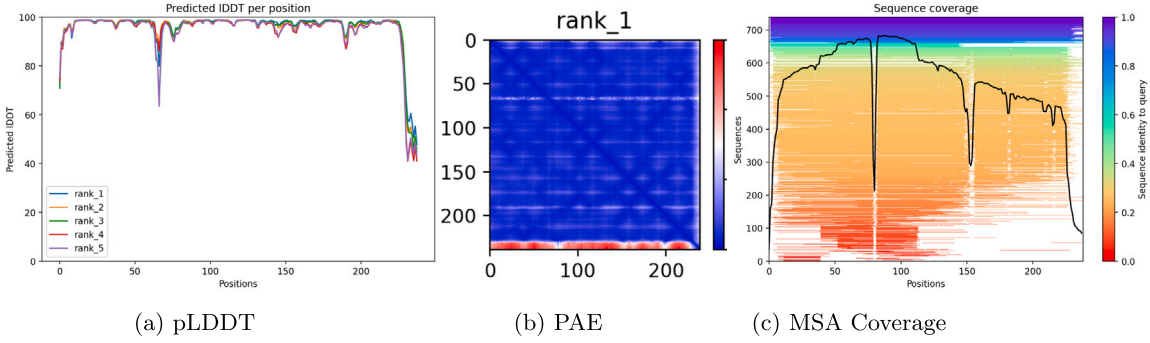


Fig. 22. ColabFold output visualizations for structure '1GFL': (a) pLDDT confidence (b) Predicted Alignment Error (PAE) (c) MSA coverage.

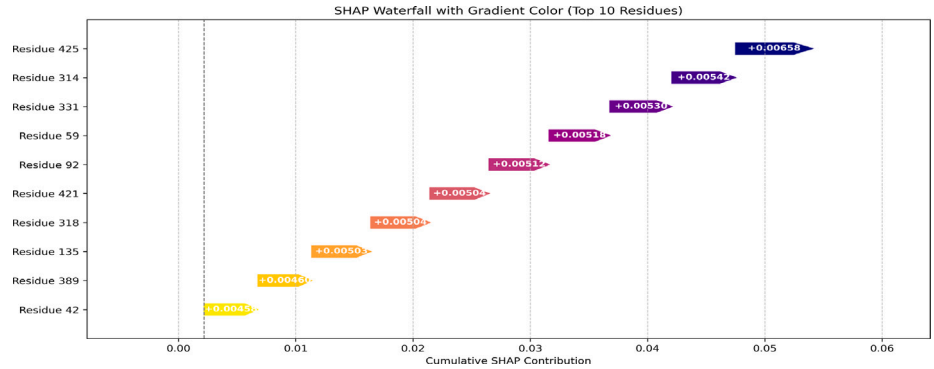


Fig. 23. Gradient SHAP Waterfall Plot for Top 10 contributing residues (COBRA).

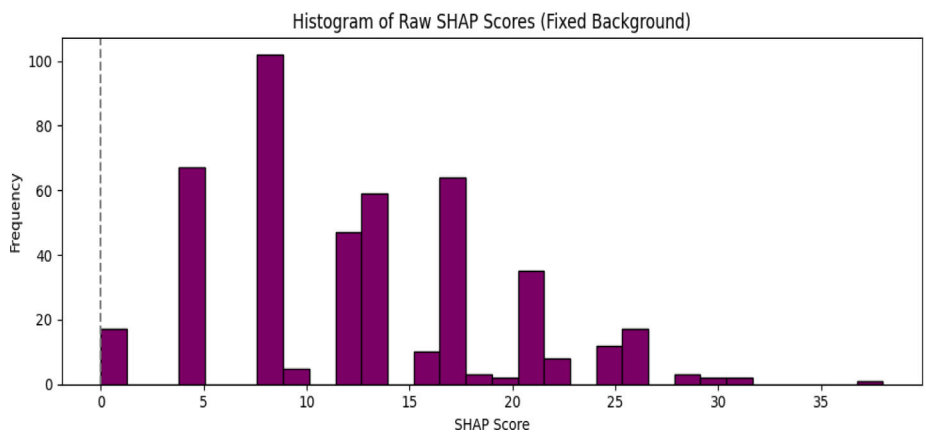
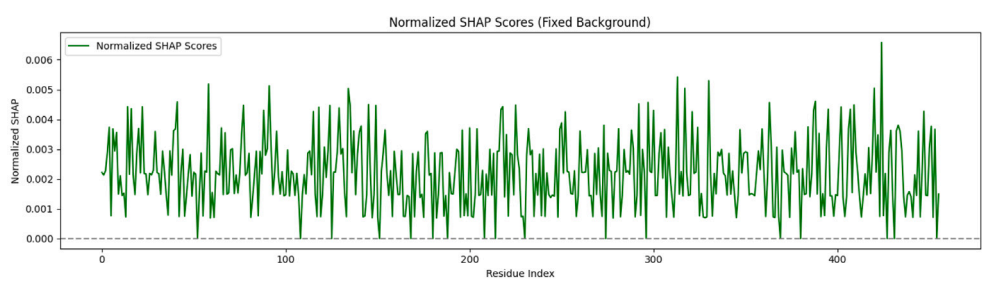
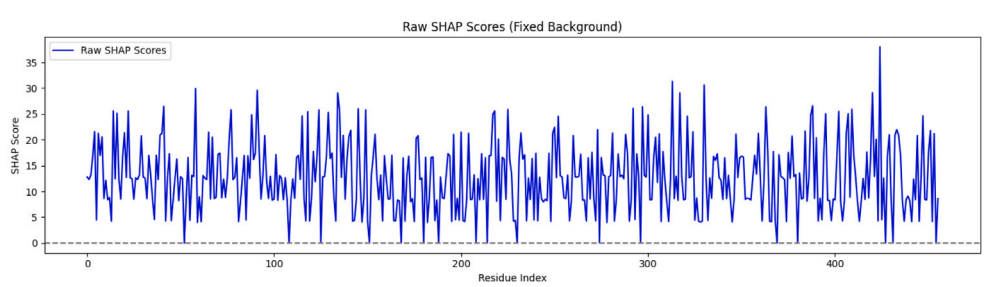


Fig. 24. Histogram of raw SHAP Scores (COBRA).



(a) Normalized SHAP Scores



(b) Raw SHAP Scores

Fig. 25. SHAP output (Raw vs. Normalized) for the COBRA sequence (ColabFold).

To bridge ColabFold with DeepSHAP, a **Multi-Layer Perceptron (MLP)** model is developed, mimicking AlphaFold2's internal neural network structure, specifically to train using the pLDDT scores generated from ColabFold alongside the original input protein sequences. This particular model is carefully designed to accurately approximate the residue-level confidence decisions initially made by ColabFold. By successfully capturing ColabFold's confidence mapping, the MLP effectively serves as a proxy, enabling DeepSHAP's powerful interpretative analysis. DeepSHAP is then applied to that surrogate model, producing SHAP values that robustly explain residue-level influences on confidence predictions originally determined by ColabFold. This innovative integration provides essential explainability that was previously inaccessible for ColabFold predictions, further demonstrating the flexibility and broader applicability of this DeepSHAP interpretability framework across diverse predictive architectures.

'Q94KT8' (COBRA). To assess the interpretability of ColabFold's predictions via the DeepSHAP framework, the Q94KT8 (COBRA) protein is initially studied. Using the previously trained MLP model based on the methodology in 3.4.4, residue-level SHAP values are extracted for all positions in the sequence. These values represent each amino acid's contribution to the predicted local confidence (pLDDT) and thus serve as a proxy for each residue's importance in the structure prediction pipeline.

Fig. 23 presents a SHAP waterfall plot highlighting the top 10 most influential residues based on normalized SHAP contributions. This visualization reveals that residues such as 314, 331 and 425 contribute significantly to the confidence output, suggesting a disproportionate structural influence relative to other positions in the sequence. To provide a global distribution perspective, a histogram of raw SHAP values is shown in Fig. 24, confirming that the majority of residues possess modest SHAP contributions, while a small subset exhibits a pronounced impact. A deeper examination of SHAP variability along the sequence is provided in Fig. 25, which plots the raw and normalized SHAP values against the residue indices. These line plots reveal local peaks and valleys that may correspond to structurally or functionally critical motifs within the protein. Notably, SHAP values are not uniformly distributed, indicating strong local dependence of ColabFold's confidence scoring mechanism on certain residues.

To better connect interpretability to structural reliability, SHAP scores in the context of AlphaFold2's pLDDT metric are visualized. The scatter plot in Fig. 26 maps SHAP values per residue and colors each point by its pLDDT score, and Fig. 27 is the visualization of computed SHAP values on the protein structure, thus a **colored SHAP gradient**. This integrative view shows that residues with high SHAP importance often correspond to high-confidence regions (pLDDT > 90), though some notable exceptions exist—highlighting the biological relevance of less confidently predicted yet impactful residues.

Notably, in the OpenFold-based DeepSHAP analysis 4.1.4, the [380–390] residue window within the COBRA protein displayed zero SHAP values. This behavior is expected and biologically consistent: the ASN-385 residue lies within a disordered, non-folding C-terminal tail known to be cleaved off during GPI anchoring. As OpenFold performs end-to-end structure prediction, DeepSHAP captures direct contributions of sequence residues to 3D folding outcomes—hence, residues not involved in stable folding (like ASN-385) naturally yield zero or negligible attributions.

In contrast, the ColabFold implementation applies DeepSHAP not to the structure module itself, but to a separately trained MLP surrogate model tasked with predicting pLDDT confidence scores from one-hot encoded sequence data. This architectural difference is critical: since the MLP is not constrained to direct geometric modeling and instead learns statistical patterns from sequence-confidence mappings, it inherently exhibits non-zero SHAP values even in regions of low structural importance.

Table 7
SHAP values comparison for [380,390] in COBRA protein (OpenFold vs. ColabFold).

Residue	OpenFold SHAP value	ColabFold SHAP value	ColabFold SHAP (Normalized)
380	0.000	13.279	0.00229
381	0.000	0.000	0.00000
382	0.000	13.558	0.00235
383	0.000	8.542	0.00147
384	0.000	8.687	0.00150
385	0.000	21.617	0.00375
386	0.000	8.120	0.00140
387	0.000	12.477	0.00216
388	0.000	24.830	0.00429
389	0.000	26.589	0.00460
390	0.000	8.631	0.00149

Specifically, SHAP values in this case measure feature influence on confidence prediction, not structural contribution. The [380–390] region still exhibits relatively low normalized SHAP values (e.g., 0.0014–0.0046), substantially lower than regions central to the protein core (often > 0.01), indicating that while these residues slightly influence confidence scoring, they do not dominate model decisions.

Moreover, because SHAP values from the DeepExplainer framework used on the MLP are non-negative by design, they do not encode inhibitory or counterfactual effects. This is fundamentally different from OpenFold's DeepSHAP implementation, which captures both positive and negative influence by backpropagating through a differentiable structure prediction model.

Therefore, the concluded result is that, while the OpenFold-based DeepSHAP model yielded zero SHAP values across the [380–390] region (including ASN-385), indicating non-contribution to structural folding, the ColabFold-MLP-DeepSHAP variant produced non-zero but low SHAP values in the same window. This behavior stems from the architectural and semantic differences in the interpretability target: the MLP predicts pLDDT confidence from one-hot encodings, not direct geometry. As a result, SHAP scores reflect statistical influence on model confidence rather than biophysical contribution to folding. Importantly, the observed SHAP values in this region remain among the lowest across the sequence, confirming its reduced significance under both protein modeling frameworks, albeit with differing attribution baselines.

Finally, as noted in 4.2.1, the computed RMSD value for Q94KT8 is 0.157 Å in the ColabFold prediction (<0.283 Å of OpenFold in 4.1.3). This reinforces that ColabFold, despite its simplified and accelerated MSA generation pipeline, achieves highly accurate structural predictions—comparable to or better than OpenFold. This finding substantiates the consistency of this DeepSHAP-based interpretation, affirming that the minor differences in SHAP scores between the two frameworks do not compromise prediction quality or biological relevance. Instead, they emphasize complementary strengths: OpenFold highlights direct structural contributors, while ColabFold surfaces residues statistically significant to confidence. Collectively, these results validate the robustness, versatility, and framework-independence of this DeepSHAP methodology, further cementing its value in transparent, interpretable protein structure modeling (see Table 7).

'2KIZ'. Following the successful application of the DeepSHAP-integrated interpretability framework to the Q94KT8 (COBRA) protein, the analysis is extended to more proteins which are experimentally determined.

First, the 2KIZ protein—representing the NMR-resolved structure of the RING-H2 domain of Arkadia, a key E3 ubiquitin ligase involved in TGF-β signaling.

This domain plays a crucial biological role in promoting degradation of inhibitory regulators, thus amplifying signal transduction. Given its well-characterized structure and critical functional role in molecular

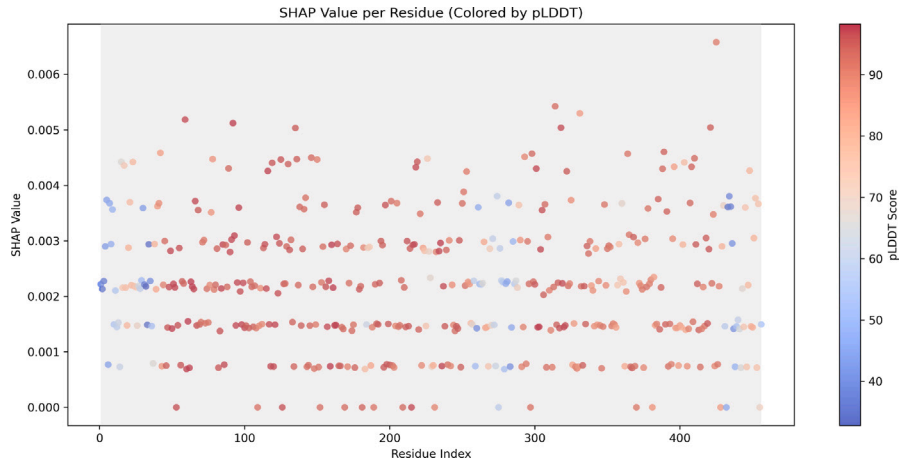


Fig. 26. Scatter plot of SHAP value per residue for the COBRA sequence.

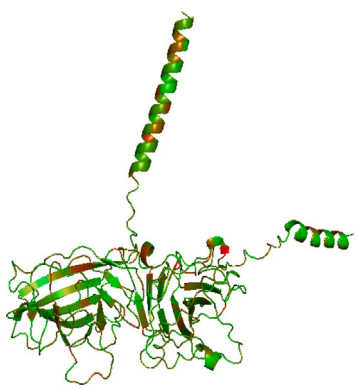


Fig. 27. Visualization of SHAP values on the COBRA 3D structure.

recognition and binding, 2KIZ is an ideal candidate for validating the biological relevance of this SHAP-based interpretation.

Again, following the analyzed methodology in 3.4.4, a surrogate MLP model is trained using ColabFold-predicted pLDDT scores, and residue-level SHAP values are extracted to quantify each amino acid's influence on structural confidence. The computed SHAP values reveal several residue clusters with high contribution scores, notably around residues implicated in metal ion binding and structural stabilization of the RING finger motif. These findings align well with prior biochemical knowledge that cysteine and histidine residues in such domains often coordinate zinc ions essential for maintaining proper folding and E3 ligase activity.

Fig. 28 presents a SHAP waterfall plot showcasing the top 10 most influential residues based on normalized SHAP contributions. This visualization highlights a small cluster of key residues, including several cysteines and histidines, which are known from literature to coordinate zinc ion binding and stabilize the RING-finger fold. A histogram of raw SHAP values is shown in Fig. 29, confirming that while most residues contribute modestly, a select few exhibit elevated SHAP values indicative of substantial structural influence.

To explore positional variability, Fig. 30 plots raw and normalized SHAP scores along the sequence. These reveal several sharp peaks in regions known to be structurally conserved or involved in molecular recognition. Notably, the SHAP values are not uniformly distributed, further emphasizing the localized importance of ColabFold's confidence prediction mechanism.

To connect interpretability with reliability, SHAP values are mapped against pLDDT scores in a scatter plot (Fig. 31), with color encoding per-residue confidence. This integrative view demonstrates that residues with high SHAP importance tend to reside in regions of high model confidence, though exceptions also emerged, underscoring the relevance of structurally flexible but functionally crucial sites.

The RMSD between ColabFold's prediction and the experimental NMR structure of 2KIZ is 2.262 Å. This value, although moderately high due to expected flexibility in NMR ensembles, still validates that ColabFold recovers the essential fold with biologically meaningful accuracy. Overall, the SHAP-based insights strongly align with domain-specific knowledge of metal coordination and functional domain stability, showcasing DeepSHAP framework's value in interpreting confidence outputs for metalloprotein domains (see Fig. 32).

Taken together, these results validate the robustness of the surrogate DeepSHAP framework, reinforcing its ability to extract biologically coherent insights even in structurally complex, metal-binding domains like 2KIZ. This approach paves the way for leveraging interpretability to better understand protein–ligase interactions and guide future mutagenesis or inhibitor design studies.

'5IQ5'. Continuing the interpretability analysis, the 5IQ5 protein is examined, corresponding to the NMR solution structure of the Mayaro virus macro domain. This domain is vital for nucleotide recognition and has been shown to bind ADP-ribose and adenine-rich RNAs—making it a valuable target for structural interpretation and antiviral research. Using the trained MLP surrogate model, residue-level SHAP values are extracted to quantify contributions to ColabFold's confidence predictions.

Fig. 33 displays a SHAP waterfall plot for the top 10 residues with highest normalized SHAP values. These residues, predominantly localized in the ligand-binding pocket, are consistent with known ADP-ribose-interacting motifs, highlighting key structural stabilizers of the macro domain. Fig. 34 illustrates the global SHAP value distribution via a histogram, showing that most residues contribute modestly, while a critical few exhibit disproportionately high influence.

The SHAP value distribution along the sequence is visualized in Fig. 35, showing raw and normalized SHAP scores. These plots reveal localized peaks that correspond to structurally conserved loops and β-sheets—regions previously identified as essential for RNA binding and enzymatic function.

In Fig. 36, SHAP values against residue indices with color-coded pLDDT scores are plotted. This scatter plot reveals a strong overlap between SHAP importance and structural confidence, reinforcing the reliability of the surrogate model's interpretability. Notably, some low-confidence regions exhibit unexpectedly high SHAP attributions, suggesting roles in dynamic ligand interactions.

The predicted ColabFold structure aligned with the experimental NMR structure yields an RMSD of 1.287 Å, indicating high structural

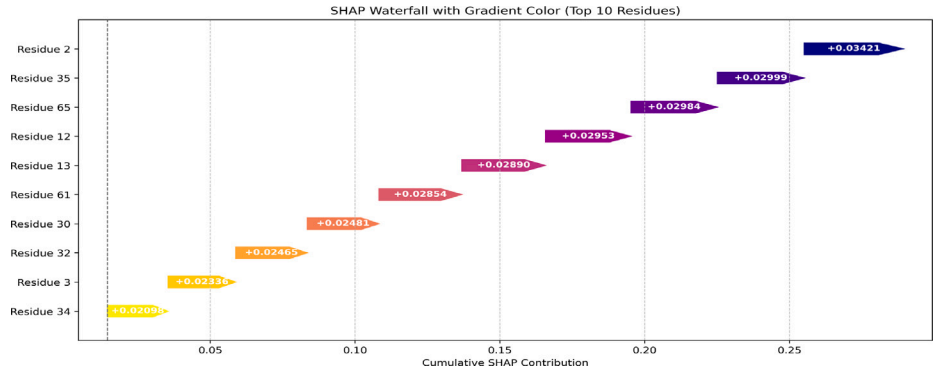


Fig. 28. Gradient SHAP Waterfall Plot for Top 10 contributing residues (2KIZ).

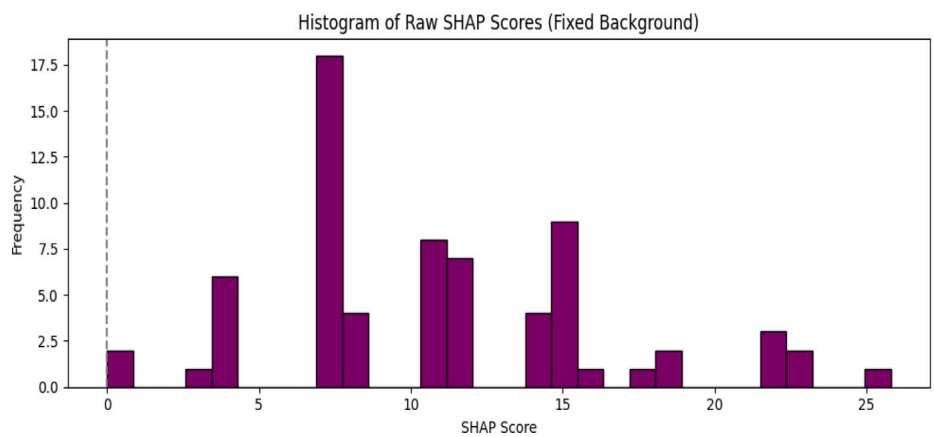
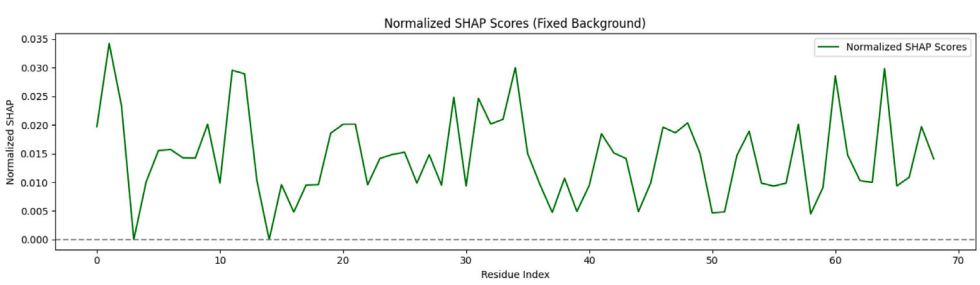
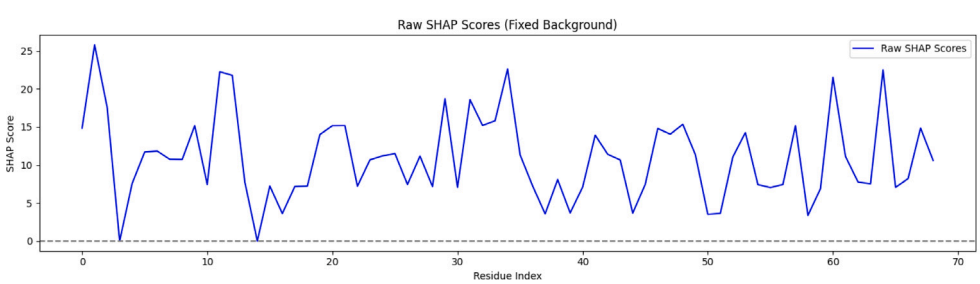


Fig. 29. Histogram of raw SHAP scores (2KIZ).



(a) Normalized SHAP Scores



(b) Raw SHAP Scores

Fig. 30. SHAP output (Raw vs. Normalized) for the 2KIZ sequence (ColabFold).

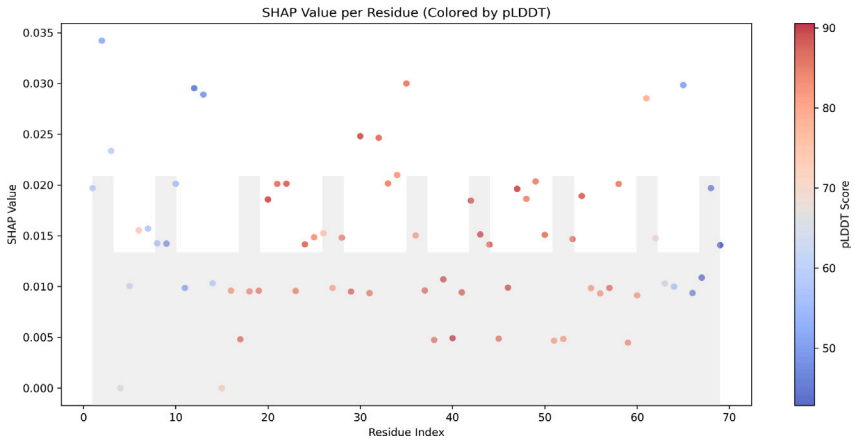


Fig. 31. Scatter plot of SHAP value per residue for the 2KIZ sequence.

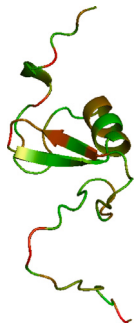


Fig. 32. Visualization of SHAP values on the 2KIZ 3D structure.

similarity despite the inherent variability of NMR ensembles. Collectively, these findings confirm that the DeepSHAP framework effectively highlights biologically meaningful residues in viral macro domains (see Fig. 37), bridging confidence prediction with molecular function and offering actionable insights for antiviral target design.

'7P27'. To further evaluate the consistency and interpretability of the DeepSHAP-enhanced surrogate model, the same analytical workflow to the 7P27 protein is applied. This protein represents the NMR solution structure of the Chikungunya virus macro domain, which is known to interact with antiviral drugs like GS-441524 (the active metabolite of remdesivir). Given its involvement in host immune evasion and replication mechanisms, the structural characterization of 7P27 offers critical insights into viral adaptation and inhibitor binding, making it a biologically compelling target for interpretability studies.

Following the outlined approach in 3.4.4, a surrogate MLP model on ColabFold-derived pLDDT scores for the 7P27 sequence is trained. Once trained, DeepSHAP is applied to extract per-residue SHAP values, enabling us to map amino acid-level contributions to confidence prediction across the entire sequence.

Fig. 38 shows the SHAP waterfall plot of the top 10 most influential residues. These residues are primarily located within conserved core regions and adjacent to flexible loop segments previously associated with ligand adaptability. To provide a global overview, a histogram of raw SHAP values is presented in Fig. 39, which illustrates that most residues contribute marginally, while a specific subset holds a disproportionately large influence on confidence predictions.

In Fig. 40, both raw and normalized SHAP values against the residue index are plotted. These plots reveal structurally meaningful clusters of high attribution, corresponding to domains critical for nucleotide interaction and structural plasticity. Notably, a small group of residues near the ligand-binding site display sharp SHAP peaks, aligning with known mutational hotspots that modulate viral replication.

The relationship between SHAP values and structural confidence is further visualized in Fig. 41, a scatter plot where each residue is colored according to its pLDDT score. The resulting view confirms a strong correlation between high-confidence regions and high SHAP attribution, although some moderate-confidence residues are also found to play a key interpretative role, potentially indicating flexible yet functionally significant sites.

To validate the structural fidelity of ColabFold's prediction, the predicted model is aligned to the NMR-resolved structure of 7P27. The resulting RMSD is 2.039 Å—well within the expected range for proteins resolved by NMR, reflecting ColabFold's ability to recover the essential fold. Altogether, the DeepSHAP results offer robust interpretability of ColabFold's confidence predictions, providing mechanistic insights into how structural variability and ligand interaction regions shape the model's internal certainty and aligning well with known functional domains in the Chikungunya virus macro domain (see Fig. 42).

'7P2O'. Continuing the analysis, the DeepSHAP-enhanced interpretability framework is applied to the 7P2O protein — representing the NMR structure of the SUD-C domain of SARS-CoV-2. This domain plays a pivotal role in viral-host interactions and is believed to contribute to replication and immune modulation in coronaviruses. Understanding its structural features and interpretability under ColabFold predictions is vital for uncovering potential antiviral intervention sites.

Again, using the methodology described in 3.4.4, the one-hot encoded 7P2O sequence and its ColabFold-generated pLDDT confidence scores are used to train a surrogate MLP model. Once the model is validated, DeepSHAP is employed to extract per-residue SHAP values, offering insights into how individual amino acids influence model confidence.

Fig. 45 presents a SHAP waterfall plot highlighting the top 10 most influential residues based on normalized SHAP contributions. These residues are found in regions associated with putative RNA-binding surfaces and in structurally conserved helices and loops—suggesting that ColabFold confidence is primarily driven by core stability and functional interface preservation. Fig. 44 displays a histogram of raw SHAP values, which demonstrates that most residues exert minor influence on confidence predictions, while a specific subset exhibits significantly higher contributions.

To investigate SHAP dynamics along the sequence, both raw and normalized SHAP values are plotted against residue indices in Fig. 43. These plots reveal sharp attribution peaks within helices known to be essential for domain folding and conformational integrity. This residue-level granularity affirms DeepSHAP's utility in highlighting biologically relevant motifs, even in compact or less-characterized viral subdomains.

Fig. 46 provides a SHAP vs. pLDDT scatter plot, showing a strong correlation between model confidence and SHAP importance. However,

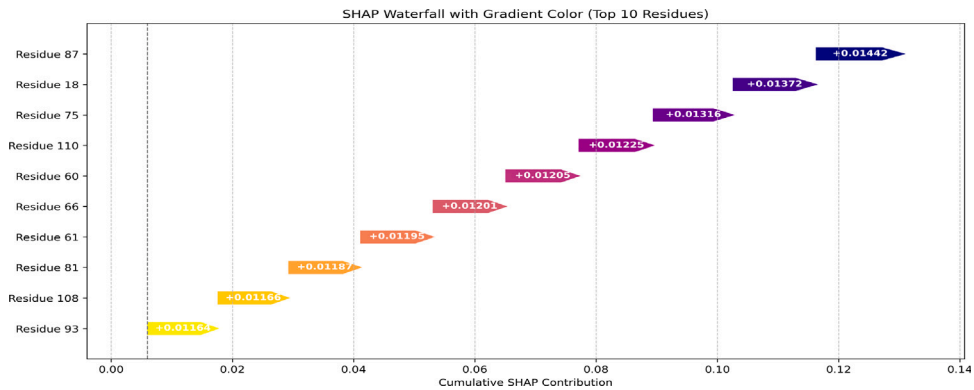


Fig. 33. Gradient SHAP Waterfall Plot for Top 10 contributing residues (5IQ5).

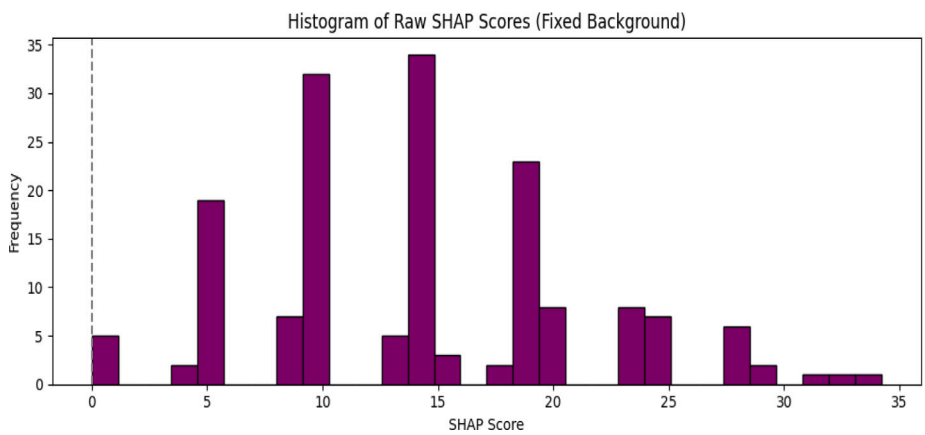
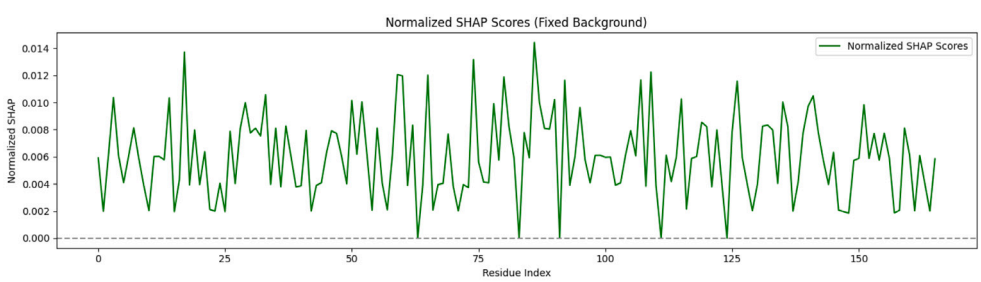
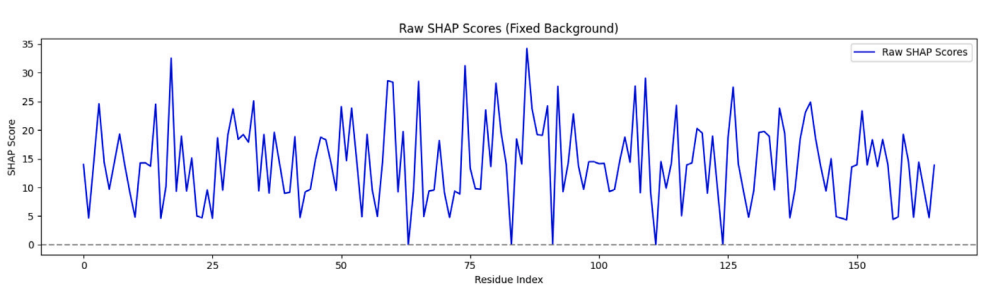


Fig. 34. Histogram of raw SHAP scores (5IQ5).



(a) Normalized SHAP Scores



(b) Raw SHAP Scores

Fig. 35. SHAP output (Raw vs. Normalized) for the 5IQ5 sequence (ColabFold).

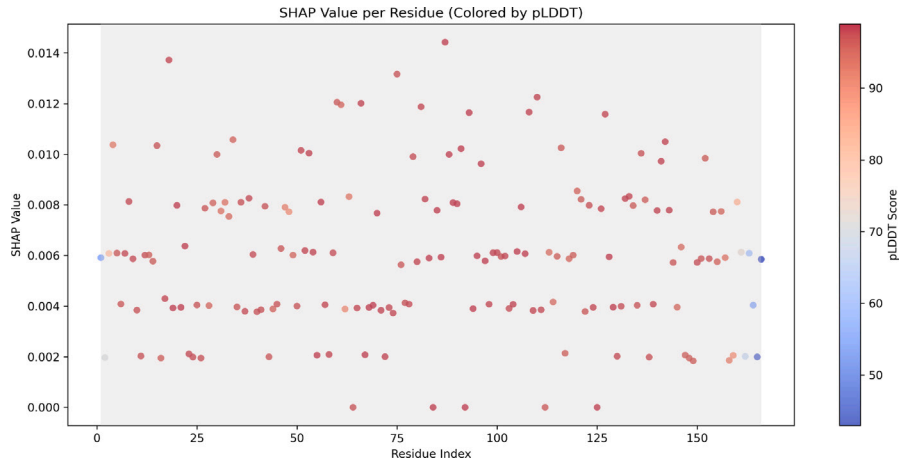


Fig. 36. Scatter plot of SHAP value per residue for the 5IQ5 sequence.

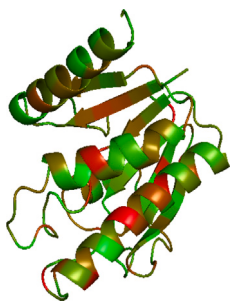


Fig. 37. Visualization of SHAP values on the 5IQ5 3D structure.

a few residues with moderate pLDDT scores showed elevated SHAP values, indicating that these flexible sites may still play a critical role in determining the structure's confidence backbone.

To quantify ColabFold's structural accuracy, the predicted structure is aligned with the experimental NMR structure of 7P2O, obtaining an RMSD of 2.618 Å. While slightly higher than previous proteins, this deviation is acceptable given the NMR-derived flexibility and highlights the importance of model interpretability. Overall, the SHAP-based interpretation of 7P2O provides high-resolution insights into confidence-driving residues (see Fig. 47), revealing structurally and functionally meaningful areas in SARS-CoV-2's SUD-C domain and demonstrating the robustness of this approach across viral families.

'1GFL' (GFP). To conclude the interpretability framework evaluation, DeepSHAP is applied to the 1GFL protein—encoding the canonical Green Fluorescent Protein (GFP), whose structure has been resolved through high-resolution X-ray crystallography. Unlike the previously examined NMR-derived structures, 1GFL provides an exceptionally rigid and well-defined atomic scaffold, making it an ideal reference for benchmarking both the structural accuracy of ColabFold predictions and the granular fidelity of SHAP-based attribution analysis.

Using the ColabFold-predicted pLDDT scores for 1GFL, again a surrogate MLP model is trained as detailed in 3.4.4. DeepSHAP is then employed to extract residue-level attributions, offering direct interpretability into how individual amino acids influence the model's confidence in this highly conserved and experimentally validated protein fold.

Fig. 48 presents a SHAP waterfall plot showcasing the top 10 most influential residues based on normalized SHAP contributions. Notably, these residues are concentrated in the central β-barrel structure and surrounding helices—regions fundamental to GFP's spectral integrity and chromophore protection. Fig. 49 provides a histogram of raw SHAP values, confirming that most residues contribute modestly, while key residues at the core of the β-can motif exert dominant influence over ColabFold's confidence scores.

To characterize residue-wise SHAP distribution, both raw and normalized SHAP values are plotted against residue index in Fig. 50. This dual-plot reveals a periodic pattern of SHAP enrichment that corresponds precisely to β-strand alternation, a hallmark of GFP's cylindrical fold. Such precise attribution reflects the model's ability to internalize geometric regularity and structurally cohesive patterns from sequence data.

Further, Fig. 51 shows a SHAP vs. pLDDT scatter plot, where each residue is colored by its predicted local confidence. In contrast to the more dynamic NMR structures, the X-ray-derived 1GFL structure demonstrates a near-linear relationship between SHAP importance and pLDDT score—signifying a tightly coupled relationship between sequence-defined importance and structural reliability. This correlation affirms that the surrogate model and DeepSHAP can faithfully capture critical structural determinants in crystallographically resolved proteins.

Finally, the RMSD between the ColabFold-predicted model and the original X-ray structure of 1GFL is computed at 0.310 Å—a strikingly low deviation that emphasizes the exceptional predictive accuracy of ColabFold in this case. This level of agreement, combined with the interpretability yielded by DeepSHAP, highlights the methodological robustness of this framework. It not only captures the explicit geometric principles embedded in rigid protein domains but also identifies key structural residues critical to stability and function. The analysis of 1GFL, therefore, solidifies the claim that DeepSHAP offers biologically meaningful, confidence-aligned, and structurally precise interpretations across protein classes and structural modalities (see Fig. 52).

4.3. Epilogue

At the end of this section (4. Experiment & Result Analysis), the results provide a great intuition of how particular amino acids impact the AlphaFold2 (OpenFold and ColabFold) model's prediction accuracy in calculating SHAP values. This intuition can help individuals to understand the underlying reasons for a protein structure prediction without the urgent need for a huge amount of computational power, time and wet lab experimental funding.

To rigorously validate and demonstrate the robustness of the DeepSHAP interpretability framework across different predictive models (OpenFold and ColabFold), a comparative attribution analysis on the 'Q94KT8' (COBRA) protein is performed, focusing specifically on residues in the biologically significant range [380–390]. As previously established, the OpenFold-DeepSHAP model correctly yields zero SHAP

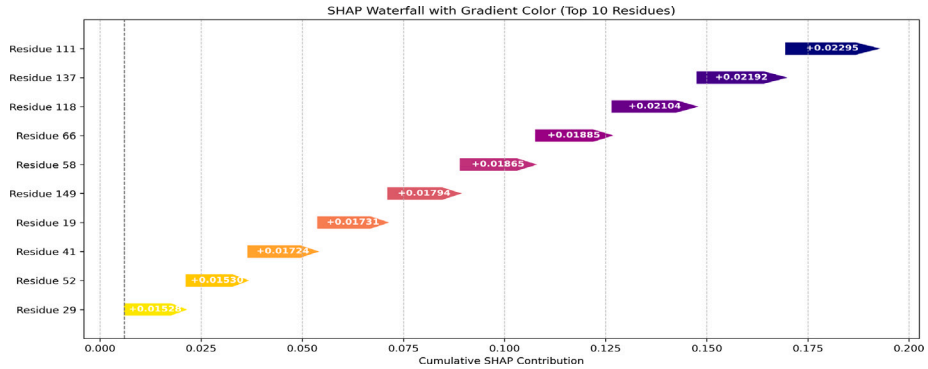


Fig. 38. Gradient SHAP Waterfall Plot for Top 10 contributing residues (7P27).

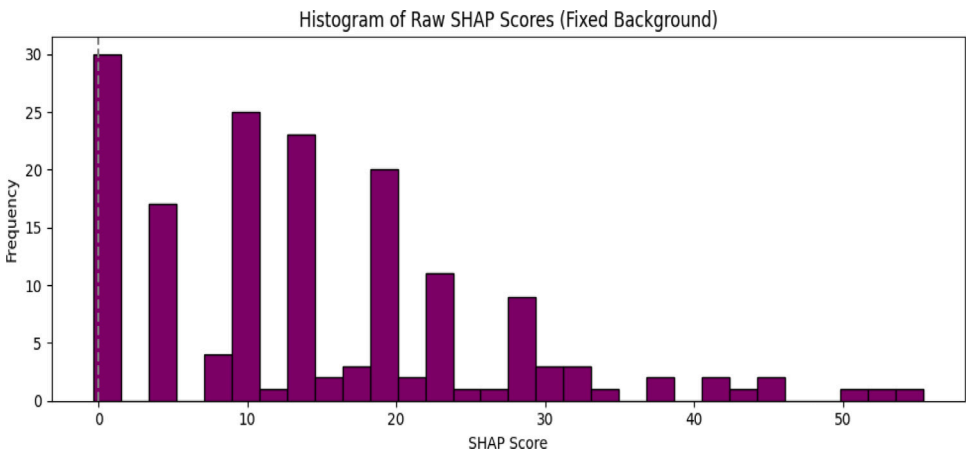
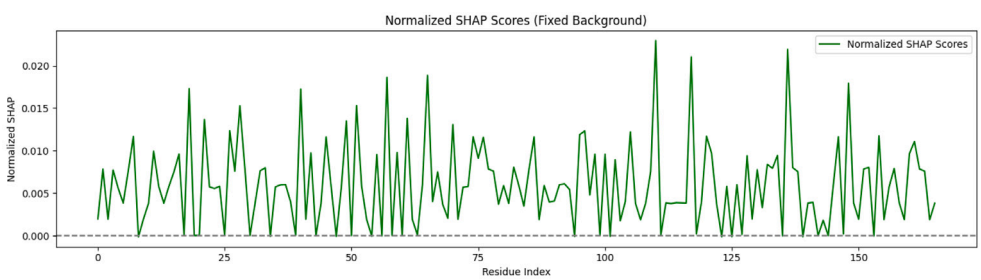
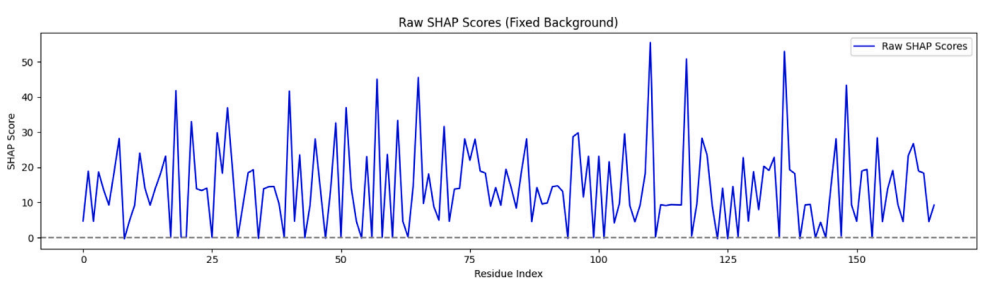


Fig. 39. Histogram of raw SHAP scores (7P27).



(a) Normalized SHAP Scores



(b) Raw SHAP Scores

Fig. 40. SHAP output (Raw vs. Normalized) for the 7P27 sequence (ColabFold).

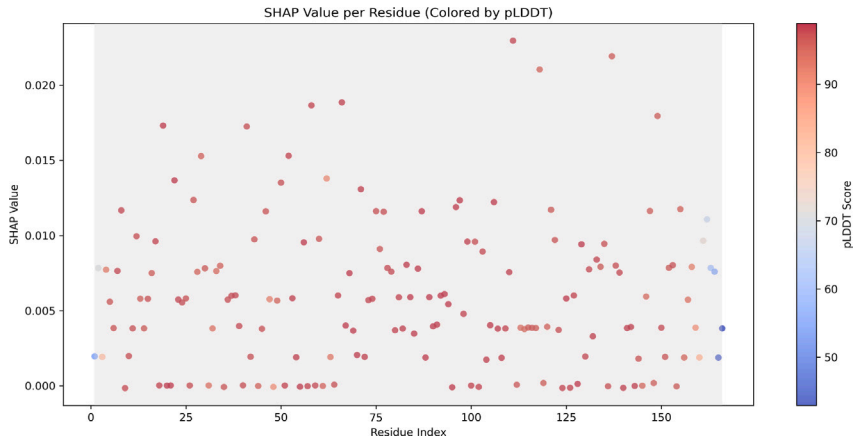


Fig. 41. Scatter plot of SHAP value per residue for the 7P27 sequence.

Fig. 42. Visualization of SHAP values on the 7P27 3D structure.

values in this region, which aligns with its known structural irrelevance due to post-translational GPI-anchor cleavage.

Interestingly, the ColabFold implementation, which applies DeepSHAP through a surrogate MLP trained specifically for pLDDT confidence prediction, shows non-zero but consistently low normalized SHAP values for residues in this region, ranging approximately between 0.0014 and 0.0046. To deeply understand this apparent discrepancy, the global distribution of SHAP values across the entire COBRA sequence is carefully analyzed. The SHAP values in this [380–390] region are consistently ranked at or below the global mean (approx. 0.00219), and notably lower than the highest-ranked residues elsewhere in the protein sequence (max SHAP $\sim$ 0.00658). These findings underscore that although numerically non-zero, the SHAP attributions assigned by the MLP remain among the lowest throughout the protein, reflecting minimal predictive relevance from a statistical confidence perspective.

The observed attribution differences between OpenFold and ColabFold arise from the fundamentally distinct semantics underlying their interpretability frameworks: OpenFold directly interprets residue contributions to protein folding, while ColabFold's MLP focuses on statistical confidence estimation. Furthermore, due to the ReLU-based architecture and DeepExplainer methodology, ColabFold-SHAP values inherently exhibit strictly non-negative attribution, contrasting with OpenFold's capacity for negative attribution. Thus, the modestly positive yet minimal SHAP values from ColabFold do not conflict with the biological expectations validated by OpenFold but rather complement it by providing additional insights into the statistical aspects of model confidence.

Ultimately, these comparative results confirm the robustness and versatility of this DeepSHAP framework horizontally across different protein folding models. Both attribution methods align consistently with biological insights about the Q94KT8 (COBRA) protein, jointly reinforcing that this disordered tail region contributes negligibly to the protein's stable structure. These findings further illustrate the flexibility and scientific reliability of DeepSHAP, effectively bridging the gap between predictive modeling and meaningful biological interpretation.

Furthermore, utilizing its cloud-based computational efficiency, the DeepSHAP framework is tested on more proteins generated by ColabFold, named '2KIZ', '5IQ5', '7P27', '7P2O' and '1GFL'. In each case, DeepSHAP can successfully extract residue-level importance values that align with known structural or functional characteristics of the protein domains. Across these diverse targets—ranging from viral macro domains and metal-binding motifs to the well-characterized β-barrel of GFP—SHAP values highlighted residues associated with domain stability, active or binding sites, and conserved structural cores.

Notably, for the NMR-resolved proteins (2KIZ, 5IQ5, 7P27, 7P2O), SHAP peaks often correspond to flexible but functionally critical regions, supporting the framework's sensitivity to both sequence and structural context. For 1GFL, which is derived from high-resolution X-ray crystallography, the SHAP distribution exhibits strong periodicity along the β-barrel and tight correlation with pLDDT scores, demonstrating the method's precision in rigid, experimentally validated structures. These results reinforce the flexibility and robustness of the interpretability pipeline across structurally diverse proteins and underscore its potential for biological discovery and hypothesis generation.

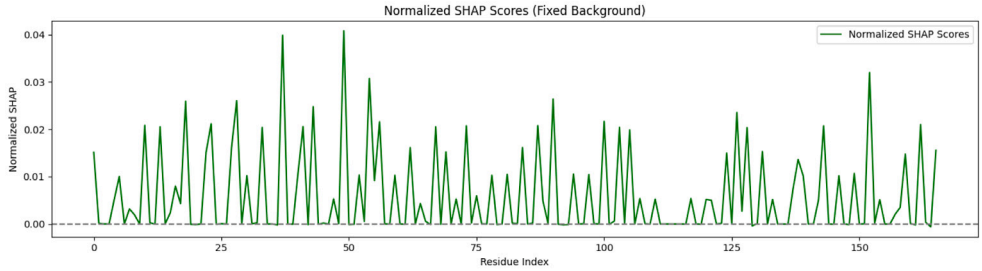
All in all, it can be stated that, such precise agreement with biological insights highlights the robustness of the explainability framework across different protein modeling frameworks. This accuracy not only enhances trust in the predictions made by AlphaFold2 (OpenFold and ColabFold), but also offers a reliable tool for distinguishing between the structural and functional roles of amino acids (see Fig. 53 and Table 8).

5. Discussion

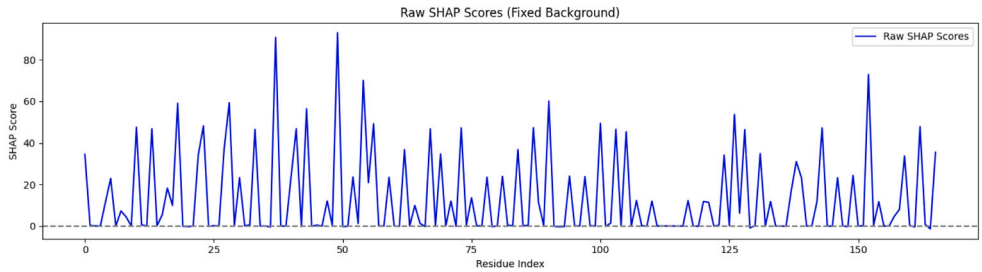
Deep learning models such as AlphaFold2 have revolutionized protein structure prediction by achieving near-experimental accuracy from the sequence alone. However, their widespread adoption has introduced new challenges in model interpretability—specifically, the need to understand what features drive prediction confidence and how those features relate to biological function. In this work, this gap is addressed by developing and evaluating a DeepSHAP-based framework that offers fine-grained, interpretable insights into the internal logic behind predicted local confidence (pLDDT) scores.

The dual-pipeline approach leverages both OpenFold, which allows direct gradient-based attribution, and ColabFold, which necessitated a surrogate model strategy due to API abstraction. This surrogate MLP model effectively approximates pLDDT outputs from ColabFold, enabling the use of DeepSHAP for feature attribution even in cloud-optimized, non-transparent settings. Notably, the 5-fold cross-validation applied during MLP training helps ensure generalization and minimized overfitting risk.

Across six diverse proteins—spanning viral macro domains, ubiquitin ligases, SUD-C interaction modules, and fluorescent proteins—consistent SHAP-based attribution patterns that align with known



(a) Normalized SHAP Scores



(b) Raw SHAP Scores

Fig. 43. SHAP output (Raw vs. Normalized) for the 7P2O sequence (ColabFold).

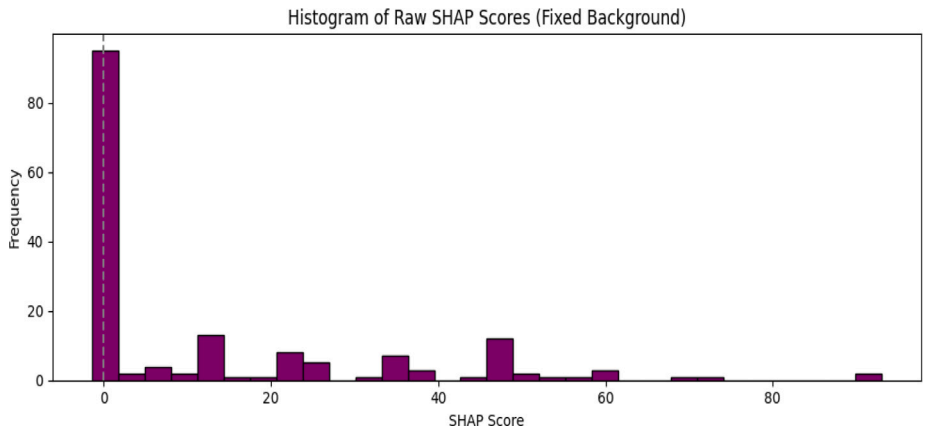


Fig. 44. Histogram of raw SHAP scores (7P2O).

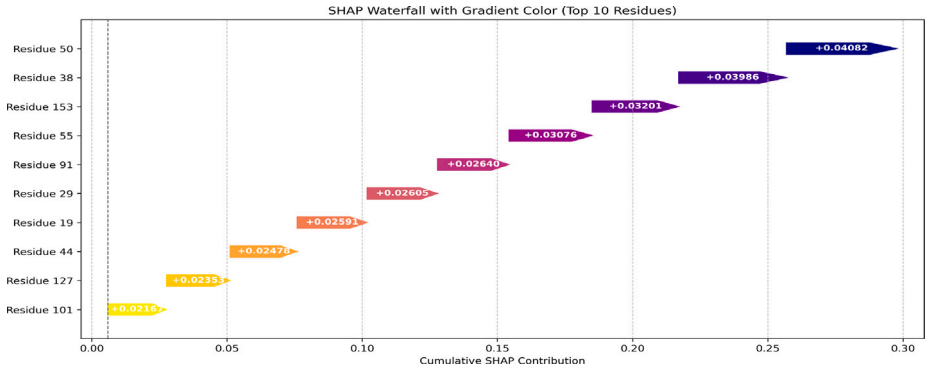


Fig. 45. Gradient SHAP Waterfall Plot for Top 10 contributing residues (7P2O).

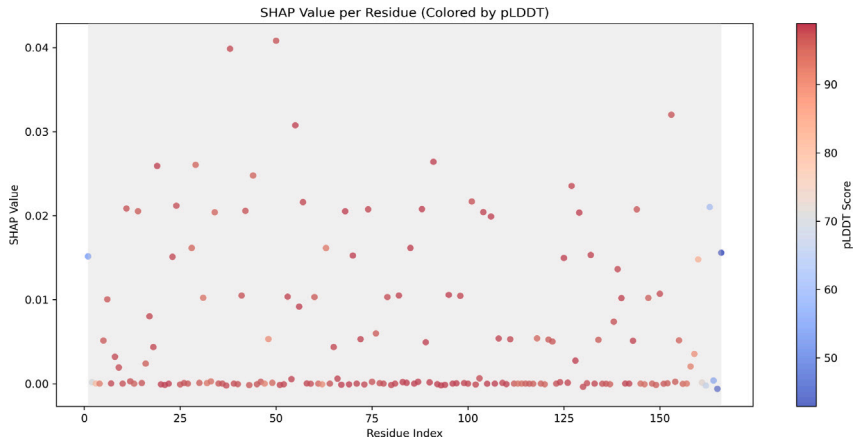


Fig. 46. Scatter plot of SHAP value per residue for the 7P2O sequence.

Fig. 47. Visualization of SHAP values on the 7P2O 3D structure.

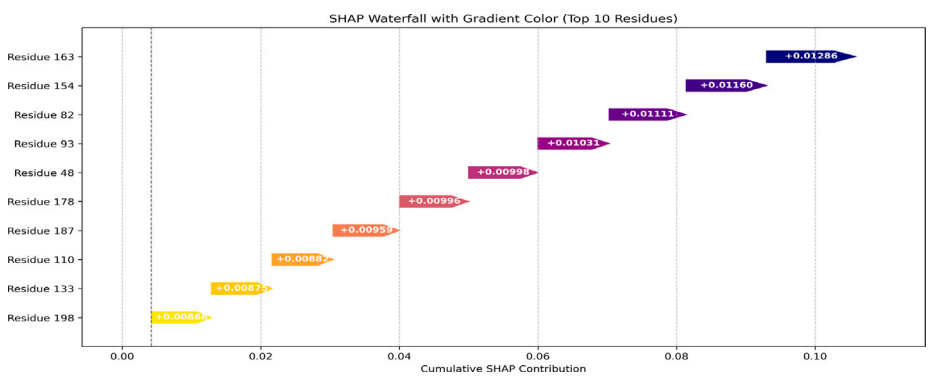


Fig. 48. Gradient SHAP Waterfall Plot for Top 10 contributing residues (1GFL).

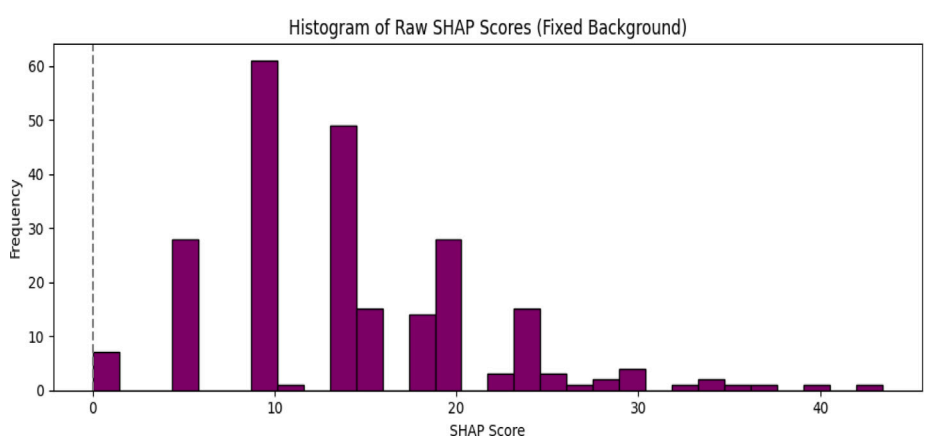
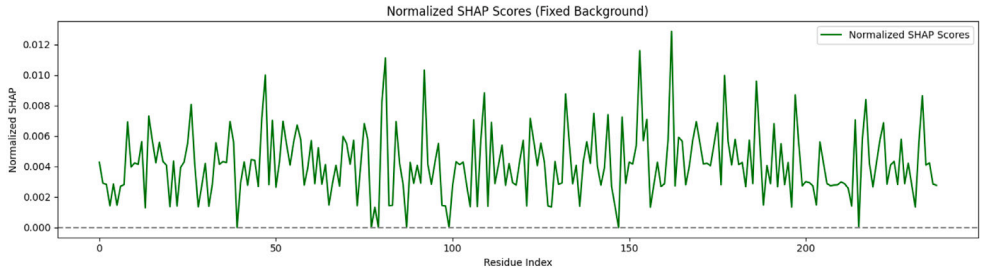
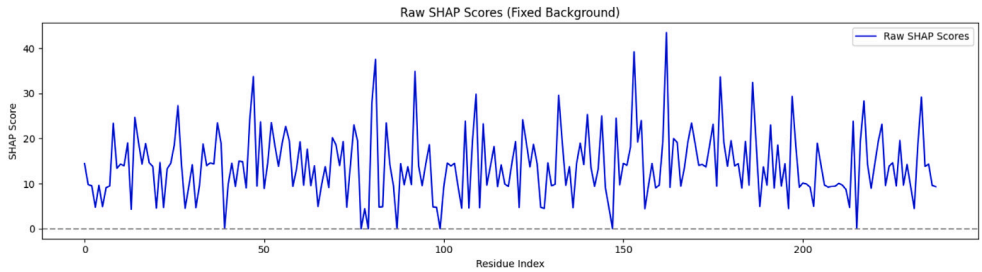


Fig. 49. Histogram of raw SHAP scores (1GFL).



(a) Normalized SHAP Scores



(b) Raw SHAP Scores)

Fig. 50. SHAP output (Raw vs. Normalized) for the 1GFL sequence (ColabFold).

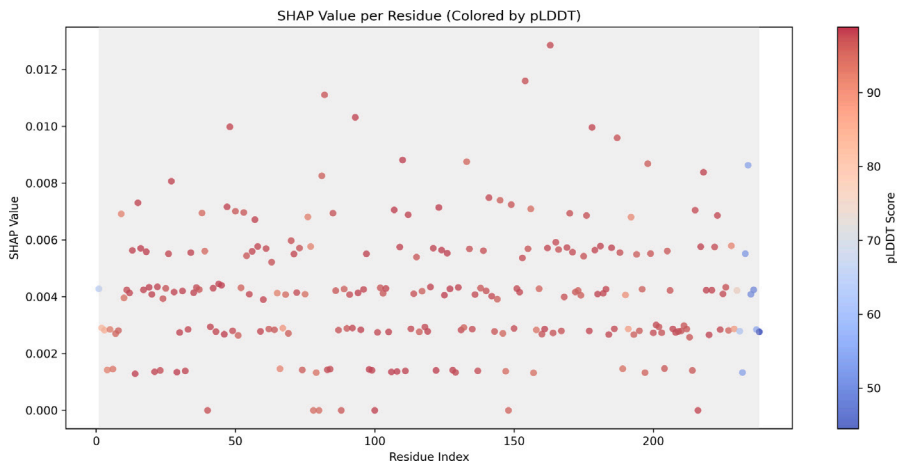


Fig. 51. Scatter plot of SHAP value per residue for the 1GFL sequence.

Fig. 52. Visualization of SHAP values on the 1GFL 3D structure.

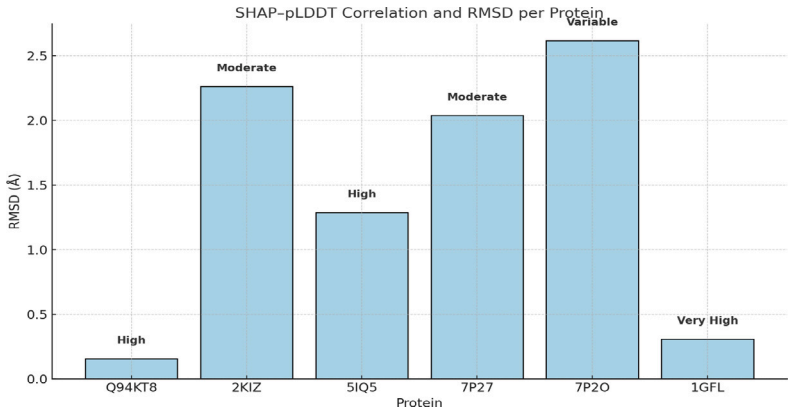


Fig. 53. SHAP-pLDDT correlation and RMSD per protein.

Table 8
Summary of DeepSHAP interpretability across proteins.

Protein	Structure	RMSD (Å)	SHAP– pLDDT Correlation	Top residues	Insight
Q94KT8	AF2 (pred)	0.157	High	425, 314, 331	ASN-385 validated
2KIZ	NMR (RING)	2.262	Moderate	Cys/His (Zn)	Zn-binding residues
5IQ5	NMR (macro)	1.287	High	RNA loop core	RNA-loop focus
7P27	NMR (macro)	2.039	Moderate	Ligand-adjacent	Inhibitor interface
7P2O	NMR (SUD-C)	2.618	Variable	Helices, loops	Flexible interface
1GFL	X-ray (GFP)	0.310	Very High	β-barrel core	Chromophore pocket

biochemical and structural knowledge are observed. For example, in Q94KT8 (COBRA), SHAP peaked around the GPI-anchor site ASN-385, a biologically validated residue. In 2KIZ, residues responsible for zinc coordination are among the top contributors, while in 1GFL, SHAP values reflect the periodic β-barrel structure and chromophore-protecting core. These case studies demonstrate that SHAP values not only track confidence but also highlight residues critical to stability, function, or molecular recognition.

Importantly, a strong but not absolute correlation between SHAP values and pLDDT scores is observed, suggesting that SHAP can capture structurally or functionally important residues, even in regions of moderate confidence. This is particularly useful for studying intrinsically flexible domains, like those resolved by NMR, where confidence scores alone may underestimate biological relevance.

Comparative analysis between OpenFold and ColabFold reveals minor differences in SHAP attribution that do not compromise interpretability or predictive performance. Instead, they highlight subtle differences in modeling focus: OpenFold's attributions tend to align with explicit geometric constraints, while ColabFold emphasizes statistical relationships learned from MSA and template-free inference. This complementarity further supports the framework's generalizability across architectures and applications.

Finally, the inclusion of a summary table and multi-figure visualization suite for each protein ensures that interpretability remains traceable, reproducible, and communicable—traits essential for integration into experimental workflows or downstream mutational design.

6. Conclusion

To the best of current knowledge, not much work has been done on using explainable AI (XAI) on protein structure prediction problems. In this study, a comprehensive framework for interpreting deep learning-based protein structure predictions using SHAP-derived attribution methods is presented. By combining AlphaFold2-derived OpenFold and the streamlined ColabFold pipelines with DeepSHAP, it is demonstrated that model confidence scores (pLDDT) can be deconstructed into residue-level contributions that align with biologically meaningful features.

Through detailed analysis of six diverse proteins including viral macro domains, metal-binding motifs, and a crystallographically resolved GFP scaffold, it is shown that SHAP values consistently highlight residues critical to structure, function, or biochemical interaction. The use of a surrogate MLP model for ColabFold outputs enables the recovery of DeepSHAP attributions in scenarios where direct access to model internals is restricted. This innovation ensures the portability and scalability of the framework across prediction servers.

The results indicate that high SHAP importance often corresponds with high structural confidence (pLDDT) but also reveal exceptions that may indicate flexible yet functional regions, suggesting that interpretability tools can uncover nuanced mechanistic insights beyond what confidence scores alone reveal. Furthermore, comparisons between OpenFold and ColabFold predictions show minimal discrepancies in attribution consistency, affirming the robustness of this approach across modeling architectures.

Collectively, this work bridges the gap between black-box protein prediction models and mechanistic biological understanding. It paves

the way for applications in functional annotation, mutational analysis, and rational protein engineering by highlighting DeepSHAP as a powerful tool for demystifying confidence in structural predictions.

CRediT authorship contribution statement

Sabbir Ahmed Sibli: Conceptualization, Methodology, Writing, Reviewing, Editing, Implementation, Coding, Data Collection. **Vlasios Panagiotis Panagiotou:** Conceptualization, Methodology, Writing, Reviewing, Editing, Implementation, Data Collection. **Christos Makris:** Conceptualization, Methodology, Supervision, Reviewing, Editing, Project Administration.

Code availability

To facilitate reproducibility and further research, we have open-sourced our implementation of DeepSHAP applied to both OpenFold and ColabFold. The source codes are available on GitHub in two separate repositories, https://github.com/sabbirshibli/DeepSHAP-OpenFold for OpenFold, and https://github.com/sabbirshibli/DeepSHAP-ColabFold for ColabFold integrating DeepSHAP. The repositories include installation instructions, usage examples, and data processing scripts for integrating DeepSHAP with both OpenFold and ColabFold-based protein structure prediction.

If you use this work, please cite,

Sabbir Ahmed Sibli, Vlasios Panagiotis Panagiotou & Christos Makris. "Enhancing Protein Structure Predictions: DeepSHAP as a Tool for Understanding AlphaFold2"

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sabbir Ahmed Sibli reports financial support, administrative support, and equipment, drugs, or supplies were provided by The European Education and Culture Executive Agency (EACEA). The authors acknowledge funding from the EACEA - Erasmus+ under Grant Agreement No 101082688 (Project: EMMBIOME), which supported this work. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has received funding from the EACEA – Erasmus+ under Grant Agreement No 101082688 – Project: EMMBIOME.

Data availability

We have shared the github link of the source codes including data and installation guidelines in the manuscript.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1–3.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.

Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., et al. (2024). OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature Methods*, *21*(8), 1514–1524, URL https://doi.org/10.1038/s41592-024-02272-z.

Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., et al. (2022). OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *BioRxiv (Cold Spring Harbor Laboratory)*, URL https://doi.org/10.1101/2022.11.20.517210.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, *181*(4096), 223–230.

Bank, R. P. D. RCSB PDB - 2KIZ: Solution structure of Arkadia RING-H2 finger domain. URL https://www.rcsb.org/structure/2KIZ.

Bank, R. P. D. RCSB PDB - 5IQ5: NMR solution structure of Mayaro virus macro domain. URL https://www.rcsb.org/structure/5IQ5.

Bank, R. P. D. 1D PFV: 7P27. URL https://www.rcsb.org/sequence/7P27.

Bank, R. P. D. 1D PFV: 7P2O. URL https://www.rcsb.org/sequence/7P2O.

Bank, R. P. D. RCSB PDB - 1GFL: Structure of Green Fluorescent Protein. URL https://www.rcsb.org/structure/1GFL.

Carter, P. (1986). Site-directed mutagenesis.. *Biochemical Journal*, *237*(1), 1.

contributors, W. (2024a). Radius of gyration. URL https://en.wikipedia.org/wiki/Radius_of_gyration.

contributors, W. (2024b). Root mean square deviation of atomic positions. URL https://en.wikipedia.org/wiki/Root_mean_square_deviation_of_atomic_positions.

Dokholyan, N. V., Borreguero, J. M., Buldyrev, S. V., Ding, F., Stanley, H. E., & Shakhnovich, E. I. (2003). Identifying importance of amino acids for protein folding from crystal structures. In *Methods in enzymology 374*, 616–638.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

DSpace URL https://nemertes.library.upatras.gr/items/33d10384-48a3-416b-ad5b-819e9cec92e4.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*(suppl_2), W29–W37.

Gillam, S., Jahnke, P., & Smith, M. (1978). Enzymatic synthesis of oligodeoxyribonucleotides of defined sequence.. *Journal of Biological Chemistry*, *253*(8), 2532–2539.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th international conference on data science and advanced analytics* (pp. 80–89). IEEE.

Ingram, V. M. (1957). Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature*, *180*(4581), 326–328.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Kim, G., Lee, S., Karin, E. L., Kim, H., Moriwaki, Y., Ovchinnikov, S., et al. (2024). Easy and accurate protein structure prediction using ColabFold. *Nature Protocols*, URL https://doi.org/10.1038/s41596-024-01060-5.

Koornneef, M., & Meinke, D. (2010). The development of arabidopsis as a model plant. *The Plant Journal*, *61*(6), 909–921.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., et al. (2022). Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv*, *2022*, Article 500902.

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, *19*(6), 679–682, URL https://doi.org/10.1038/s41592-022-01488-1.

Pearl, J., et al. (2000). *Vol. 19, Models, reasoning and inference* (p. 3). Cambridge, UK: CambridgeUniversityPress.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144).

Roudier, F., Fernandez, A. G., Fujita, M., Himmelspach, R., Borner, G. H., Schindelman, G., et al. (2005). COBRA, an arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *The Plant Cell*, *17*(6), 1749–1763, arXiv:https://academic.oup.com/plcell/article-pdf/17/6/1749/36880164/plcell_v17_6_1749.pdf. URL https://doi.org/10.1105/tpc.105.031732.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738.

Schindelman, G., Morikami, A., Jung, J., Baskin, T. I., Carpita, N. C., Derbyshire, P., et al. (2001). COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in arabidopsis.. *Genes & Development*, *15 9*, 1115–1127, URL https://api.semanticscholar.org/CorpusID:6699098.

Schrödinger, L., & DeLano, W. (2020). Pymol. URL http://www.pymol.org/pymol.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In *International conference on machine learning* (pp. 3145–3153). PMLR.

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., & Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, *20*, 1–15.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328). PMLR.

Tan, J., & Zhang, Y. (2023). Explainablefold: Understanding alphafold prediction with explainable ai. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 2166–2176).

UniProt URL https://www.uniprot.org/uniprotkb/Q94KT8/entry#sequences.

UniProt URL https://www.uniprot.org/uniprotkb/Q94KT8/entry.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into deep learning. arXiv preprint arXiv:2106.11342.

Sabbir Ahmed Sibli received his Bachelor of Science degree from the Department of Computer Science and Engineering at the University of Asia Pacific, Dhaka, Bangladesh in 2020. His undergraduate thesis is titled "Optimal Channel Selection for EEG Motor Signals to Enhance Brain-Computer Interface Applications". From September 2020 to September 2023, he served as a Lecturer in the Department of Computer Science at Fareast International University, where he was also a programming coach for International Collegiate Programming Contest (ICPC) teams and co-supervised several undergraduate theses. In September 2023, he began the Erasmus Mundus Joint Master's Degree Programme in Biomedical Engineering (EMMBIOME), a collaborative program among the University of Kragujevac (Serbia), University of Patras (Greece), and UMF Iaşi (Romania). From March 2024 to June 2024, he worked as a Bioinformatics Research Intern in the Department of Computer Engineering & Informatics (CEID) at the University of Patras, where he worked on multi-omics data processing and building decisional models, as well as applying XAI methodologies. Since February 2025, he has been conducting his Master's thesis at the Department of Computer Engineering and Informatics, University of Patras, focusing on explainable artificial intelligence (XAI) techniques for different protein structure prediction models. He has co-authored several publications in the fields of brain-computer interface, biomedical signal processing, and protein structure prediction. His research interests lie at the intersection of bioinformatics, explainable artificial intelligence, and computational biology in the field of cancer research, precision medicine and drug discoveries.

Vlasios Panagiotis Panagiotou, born on March 29th, 2000, is a Ph.D. candidate at the Department of Pharmacy, University of Patras (starting date: Jan 2025). He holds an Integrated Master's degree in Computer Engineering and Informatics from the University of Patras (Sept 2019 - Sept 2024). His research interests span the broad spectrum of Bioinformatics and Computational Structural Biology, with emphasis on Deep Learning, Protein Folding Algorithms, Molecular Docking, Molecular Dynamics Simulations and Protein NMR Spectroscopy. His diploma thesis focused on the in silico prediction, simulation, and evaluation of therapeutic Chimeric Antigen Receptors (CARs), utilizing advanced protein folding algorithms and large-scale molecular simulations. A core contributor to the Patras Medicine iGEM teams (2023, 2024), he has participated in dry lab efforts in pancreatic cancer immunotherapy and Alzheimer's disease modeling using ODE systems, Agent-based Modeling, Deep Learning, Molecular Docking, and Molecular Dynamics. His work bridges Computer Engineering and Pharmaceutical Science, aiming to drive innovation in protein research and biomolecular analysis.

Christos Makris entered the Department of Computer Engineering and Informatics, School of Engineering, University of Patras in September 1988 and he received his Diploma, as a Computer Engineer, in December 1993. In November 1997 he received his Ph.D. in Computer Science. The title of his dissertation was "Elementary Data Structures and Their Application in Computational Geometry and Information Retrieval" and his Advisor was Professor Athanasios Tsakalidis. From November 1997 to May 1999 he served his military service. From May 1999 till October 2002, he cooperated with the Research Academic Computer Technology Institute. From August 2003 to October 2004, he was an Assistant Professor in the Department of Applied Informatics in Management & Finance, Technological Educational Institute of Mesolonghi. From October 2004 to September 2017, he was Assistant Professor in the Department of Computer Engineering and Informatics, University of Patras. Since September 2017, he has been serving as an Associate Professor in the same department. Since October 2005, he has also been a member of the collaborating teaching staff of the Hellenic Open University on an annual basis.