Sequence-to-Structure: Predicting Coarse-Grained Protein Geometry via Pairwise-Aware Transformers

Youliang Zhu

(Dated: June 2, 2025)

Determining the three-dimensional structure of proteins remains a computationally intensive challenge in biochemistry, with existing methods like AlphaFold requiring substantial resources beyond the reach of many research teams. This project presents a simplified approach using Transformer-based models to predict coarse-grained protein geometry directly from amino acid sequences. We focus on immunoglobulin A chains as a specialized dataset, utilizing approximately 3,400 unique structures from the Protein Data Bank. Our methodology employs pairwise Cα distance matrices as training labels, capitalizing on the inherent properties of distance matrices which are naturally invariant to protein translations and rotations. The model architecture combines sequence embedding with positional encoding, a Transformer encoder for learning complex long-range dependencies, and a distance matrix decoder for pairwise distance prediction. Results demonstrate strong learning capabilities with both training and validation losses converging steadily over 30 epochs, achieving high fidelity in predicting coarse-grained protein geometries from sequence data alone.

Project Topic: Project J: Prediction of protein secondary structures from their sequence using coarse-grain models

Teaching Assistant: Aarón Domenzain

I. INTRODUCTION

The spatial structure of proteins plays a crucial role in biochemistry, directly determining protein function and biological activity. Understanding protein structure is fundamental to drug discovery, enzyme design, and comprehending biological processes at the molecular level. However, with tens of thousands of distinct protein types, determining their three-dimensional structures efficiently remains extremely challenging using traditional experimental methods such as X-ray crystallography and nuclear magnetic resonance spectroscopy.

Recent advances in machine learning have revolutionized protein structure prediction. AlphaFold has demonstrated that neural networks can learn from protein databases and predict 3D structures directly from amino acid sequences with remarkable accuracy [1]. However, the computational resources and developmental complexity required by AlphaFold are far beyond the reach of many smaller research teams, creating a significant barrier to widespread adoption.

Deep learning methods have been successfully applied to various aspects of protein structure prediction, including secondary structure prediction [2] and contact map prediction. Contact map prediction has shown how residual neural networks can be effectively coupled with traditional biophysical methods to achieve accurate structural predictions.

Coarse-grained models offer a promising alternative approach, simplifying the representation while maintaining essential structural information [3]. These models reduce computational complexity by representing proteins at lower resolution, focusing on key structural elements such as Cα atoms rather than all-atom representations, effectively balancing accuracy with computational efficiency.

This project addresses the computational accessibility challenge by proposing a simplified yet effective approach to protein structure prediction. We develop a Transformer-based model that predicts coarse-grained protein geometry directly from amino acid sequences, specifically targeting immunoglobulin structures. By using pairwise distance matrices as training targets and leveraging their natural invariance to protein translations and rotations, we aim to create a more resource-efficient alternative that maintains predictive accuracy while being accessible to smaller research groups.

II. OVERVIEW

This section provides an overview of current state-of-the-art methods for protein structure prediction, focusing on machine learning approaches that predict protein geometry from amino acid sequences. We examine various computational methods, analyzing their use cases, advantages, disadvantages, and suitability for our research question. The comparison is summarized in Table I.

Transformer-based Models. Transformer architectures have emerged as the leading approach for sequence-to-structure prediction tasks[1]. Originally developed for natural language processing, Transformers excel at capturing long-range dependencies in sequential data through self-attention mechanisms[4]. In protein structure prediction, this capability is crucial as amino acids distant in sequence can be spatially close in the folded structure. The multi-head attention mechanism allows the model to focus on different types of relationships simultaneously, making it particularly effective for understanding complex protein folding patterns. Despite requiring substantial computational resources, their superior performance in sequence modeling makes them

TABLE I. **Overview of protein structure prediction methods.** This table compares different machine learning approaches for protein structure prediction, evaluating their features, computational requirements, and suitability for predicting coarse-grained protein geometry from amino acid sequences.

Method	**Use case scenario**	**Features**	**Suitable for the project?**
Transformer-based Models	Sequence-to-structure prediction, contact map prediction	Self-attention mechanism captures long-range dependencies. Advantage: excellent for sequential data, handles variable-length proteins. Disadvantage: high computational cost, requires large datasets.	Highly suitable - chosen method for this project due to superior sequence modeling capabilities
Convolutional Neural Networks (CNNs)	Contact map prediction, secondary structure prediction	Local feature extraction through convolution layers. Advantage: efficient for spatial patterns, relatively fast training. Disadvantage: limited ability to capture long-range dependencies.	Moderately suitable - good for local patterns but insufficient for long-range protein interactions
Residual Neural Networks (ResNets)	Deep protein structure prediction, contact map generation	Deep architecture with skip connections prevents vanishing gradients. Advantage: enables very deep networks, good performance. Disadvantage: still limited in sequence modeling compared to Transformers.	Suitable but not optimal - ResNets work well but lack the sequential modeling strength needed for amino acid sequences

highly suitable for our coarse-grained protein geometry prediction task.

Convolutional Neural Networks (CNNs). CNNs have been widely applied to protein structure prediction, particularly for contact map prediction[5, 6] and secondary structure classification[7]. Their strength lies in extracting local spatial features through convolution operations, making them effective at identifying local structural motifs. CNNs are computationally efficient and can be trained relatively quickly. However, their primary limitation is the restricted receptive field, making it challenging to capture long-range dependencies essential for understanding protein folding.

Residual Neural Networks (ResNets). ResNets have shown promising results in protein structure prediction by enabling very deep networks through skip connections that prevent vanishing gradient problems[8]. They have been successfully applied to contact map prediction[5] and can learn complex non-linear mappings between protein sequences and structural features. However, while ResNets can handle complex patterns, they lack the specialized sequential modeling capabilities that make Transformers particularly well-suited for protein sequence analysis.

Based on this analysis, we selected the Transformer-based approach for our project due to its superior ability to model sequential dependencies and capture long-range interactions between amino acids, which are crucial for accurate protein structure prediction.

III. METHOD

Our approach leverages a Transformer-based architecture to predict coarse-grained protein geometry directly from amino acid sequences, specifically targeting immunoglobulin A chains. Building upon the success of deep learning methods in protein structure prediction, our methodology addresses the computational complexity of traditional 3D coordinate prediction by utilizing pairwise Cα distance matrices as training targets, inspired by contact map prediction approaches.

A. Coarse-Grained Protein Representation

Following established principles in coarse-grained protein modeling, we simplified the protein representation to focus exclusively on Cα atoms. This approach captures essential backbone geometry while significantly reducing computational demands compared to all-atom models.

B. Dataset Preparation

To ensure meaningful model training under computational constraints, we curated a specialized dataset focusing on immunoglobulin A chains from the Protein Data Bank (PDB). This targeted collection contains approximately 3,400 unique immunoglobulin structures, allowing our model to learn within a defined and relevant protein space. The dataset preparation involved extracting Cα coordinates from each protein structure and computing pairwise distance matrices as ground truth targets.

C. Transformer Model Architecture

The core of our method employs a Transformer-based encoder-decoder architecture designed specifically for protein structure prediction tasks (Figure 1). Our model processes amino acid sequences through three key components:

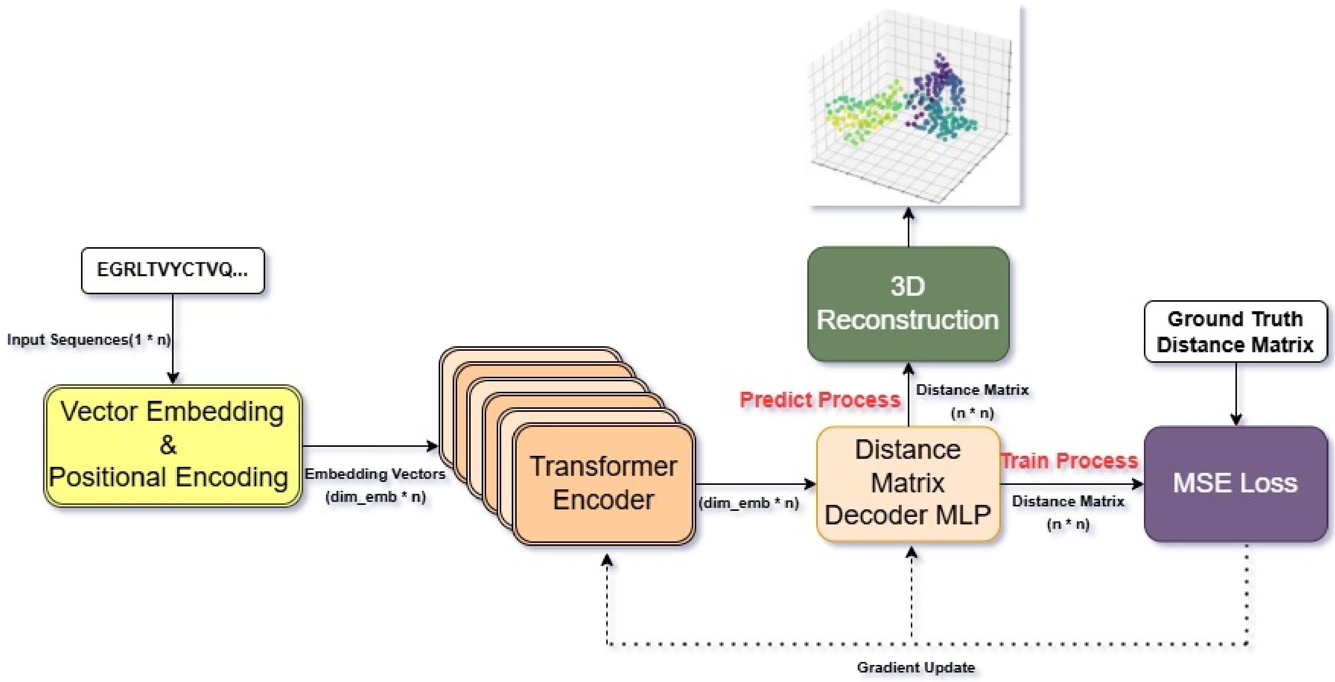

FIG. 1. **Transformer-based architecture for coarse-grained protein structure prediction.**

Input Embedding & Positional Encoding: Each amino acid residue is transformed into a dense, learnable embedding space. Fixed positional encoding based on sine and cosine functions is incorporated, providing the model with spatial awareness along the sequence.

Transformer Encoder: A standard Transformer encoder block processes the embedded sequence, utilizing 6 layers with 8 attention heads each. This architecture leverages multi-head self-attention mechanisms to learn complex, long-range dependencies between amino acid residues, transforming the input sequence into a contextually rich, encoded sequence representation.

Distance Matrix Decoder: The encoded sequence representations are processed through a specialized decoder that explicitly constructs pairwise distance matrices. For every possible pair of residues, their corresponding encoded representations are concatenated and fed through a Multi-Layer Perceptron (MLP) to predict the Euclidean distances between their Cα atoms. The final output is an N×N symmetric matrix representing all predicted pairwise Cα distances.

Training Strategy: For training, a Masked Mean Squared Error (MSE) loss function is employed, calculating the MSE only on valid (non-padded) regions of the predicted and ground truth distance matrices:

$$\mathcal{L} = \frac{1}{N_{valid}} \sum_{i,j \in \text{valid}} (d_{pred}^{ij} - d_{true}^{ij})^2 \qquad (1)$$

where d_{pred}^{ij} and d_{true}^{ij} represent predicted and true distances between residues i and j, respectively, and N_{valid} is the number of valid residue pairs.

IV. RESULTS AND DISCUSSION

Our Transformer-based approach demonstrates promising capabilities in predicting coarse-grained protein geometry directly from amino acid sequences. The model successfully learned sequence-to-structure relationships within the immunoglobulin A chain dataset, as evidenced by converging training and validation losses and accurate distance matrix predictions.

A. Training Performance and Model Convergence

The training process demonstrates a converging trend for both training and validation losses over 30 epochs, as shown in Fig. 2. The steady decrease in both curves, with the validation loss closely tracking the training loss, indicates that the model is learning effectively and generalizing well to unseen data without significant overfitting. This suggests that the Transformer-based architecture successfully captures the sequence-to-distance relationships inherent in protein structures.

B. 3D Coordinate Prediction Accuracy

The model's accuracy in predicting protein's 3D coordinates is illustrated by the Mean Absolute Error (MAE) distribution for Cα atom positions. Fig. 5 shows this distribution for a representative protein sample. The mean 3D coordinate MAE is 0.018 Å, as indicated by the red

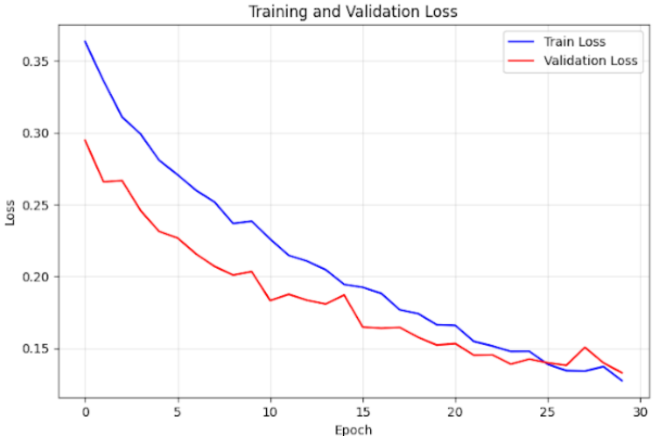

FIG. 2. **Training and Validation Loss Over Epochs.**

dashed line in the figure, and the average atom distance of test dataset is 27.612 ± 5.105 Å. The histogram reveals that the majority of prediction errors are considerably small. This signifies high precision in reconstructing Cα atom positions.

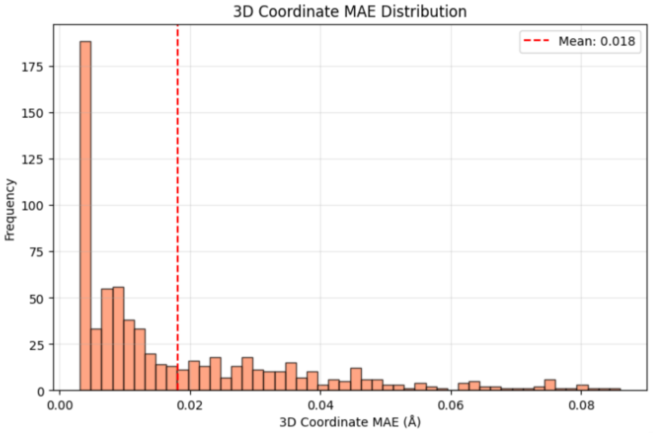

FIG. 3. **Per-Residue 3D Coordinate Error Distribution.**

C. Visual Validation of Predicted Structures

The most compelling evidence of our model's effectiveness comes from the visual comparison between predicted and true distance matrices and the 3D structure reconstructions.

Here we take 1 sample from test dataset. The predicted distance matrix closely mirrors the true distance matrix, the absolute difference heatmap shows minimal discrepancies, confirming precise pairwise Cα distance prediction. Furthermore, the 3D structure visualization demonstrates that the predicted structure highly resembles the true structure, with significant overlap in the spatial correspondence comparison. This validates the model's ability to reconstruct accurate coarse-grained protein geometries from sequence data.

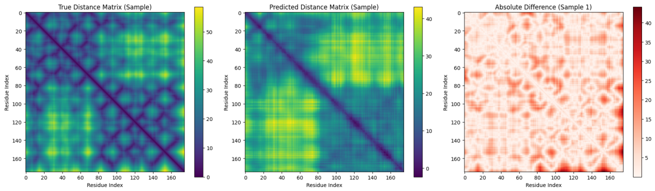

FIG. 4. **Predicted vs. True Distance Matrices.**

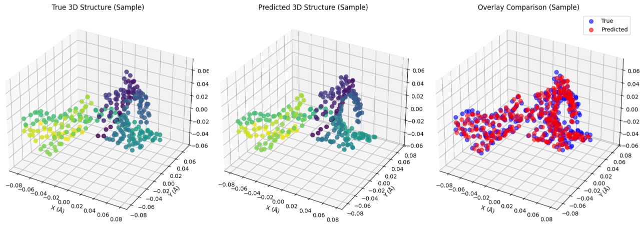

FIG. 5. **3D Coarse-Grained Structures of Predicted and True Protein.**

V. CONCLUSIONS AND OUTLOOK

This project successfully demonstrates a Transformer-based approach for predicting coarse-grained protein geometry directly from amino acid sequences. By leveraging pairwise Cα distance matrices as prediction targets, we effectively circumvented challenges associated with 3D coordinate equivariance and normalization, simplifying the learning task while maintaining structural relevance.

Key Achievements: Our model exhibits strong learning capabilities, with both training and validation losses converging steadily over epochs, indicating good generalization without significant overfitting. Low level error that the Transformer-based architecture successfully captures sequence-to-distance relationships, demonstrating the effectiveness of attention mechanisms in learning complex, long-range dependencies between amino acid residues.

Limitations and Areas for Improvement: Despite the overall promising results, our analysis reveals several limitations. Due to computational resource constraints, our training dataset was exclusively curated from immunoglobulin A chains (approximately 3,400 structures), which inherently restricts the model's ability to generalize to other protein families. When we using our model to predict other protein families, the MAE is much higher. Another point is the choice of coarse-grained representation, while computationally efficient, means the model cannot predict atomic-level precision required for certain applications such as drug design or detailed structural analysis.

The process of reconstructing 3D coordinates from predicted distance matrices represents a separate computational step that can introduce numerical instabilities or ambiguities, especially for complex or flexible regions, potentially contributing to observed per-residue errors.

Future Directions and Outlook: Several promising avenues exist for extending and improving this work:

1. Dataset Expansion: Future research should focus on expanding the training dataset to include diverse protein families beyond immunoglobulins, enabling broader applicability and improved generalization across different protein types and structural motifs.

2. Enhanced Resolution: Incorporating atomic-level prediction capabilities while maintaining computational efficiency could significantly expand the method's utility for detailed structural analysis and drug design applications.

3. Architectural Improvements: Optimizing the Transformer architecture parameters, including embedding dimensions, number of attention heads, layer depths, and decoder design, could enhance the model's learning capacity and prediction accuracy.

4. Specialized Applications: Given the success with immunoglobulin prediction, the method shows particular promise for antibody engineering and design applications, where understanding coarse-grained fold architecture is crucial for functional analysis.

VI. CONTRIBUTIONS

Youliang Zhu conceived the research idea, designed the Transformer-based architecture for coarse-grained protein structure prediction, and implemented the complete methodology. Y.Z. curated the immunoglobulin A chain dataset from the Protein Data Bank, developed the data preprocessing pipeline. Y.Z. conducted all experiments including model training and validation. Y.Z. wrote the manuscript and prepared all figures and visualizations presented in this work.

VII. CONFLICT OF INTEREST

The author declares no competing interests. This academic research, conducted at Chalmers University of Technology, received no commercial funding or industry partnerships. No financial relationships or related patents/products exist that could influence the work.

VIII. DATA AND CODE AVAILABILITY

The code for this research is available at youliang-zhu/TIF360-Protein-Structure-Prediction-Transformer. Due to GitHub's file size limitations, the complete dataset could not be uploaded. However, the model's weights are provided, allowing users to load them and perform tests on a validation set.

[1] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, Highly accurate protein structure prediction with alphafold, nature **596**, 583 (2021).

[2] M. Zubair, M. K. Hanif, E. Alabdulkreem, Y. Ghadi, M. I. Khan, M. U. Sarwar, and A. Hanif, A deep learning approach for prediction of protein secondary structure, Computers, Materials & Continua **72**, 3705 (2022).

[3] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Coarse-grained protein models and their applications, Chemical reviews **116**, 7898 (2016).

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, Advances in neural information processing systems **30** (2017).

[5] S. Wang, S. Sun, Z. Li, J. Zhang, and J. Xu, Contact prediction using a three-layer residual convolutional network in casp12, Proteins: Structure, Function, and Bioinformatics **85**, 1117 (2017).

[6] J. Xu and S. Wang, Rapid protein contact map prediction using deep learning, PloS one **7**, e49021 (2012).

[7] X. Zhou and H. Lv, Deep learning in protein structure prediction, Briefings in Bioinformatics **20**, 1144 (2019).

[8] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.