Quantifying the Role of OpenFold Components in Protein Structure Prediction

Tyler L. Hayes
College of Computing
Georgia Institute of Technology
thayes68@gatech.edu

Giri P. Krishnan
College of Computing
Georgia Institute of Technology
gkrishnan46@gatech.edu

Abstract

Models such as AlphaFold2 and OpenFold have transformed protein structure prediction, yet their inner workings remain poorly understood. We present a methodology to systematically evaluate the contribution of individual OpenFold components to structure prediction accuracy. We identify several components that are critical for most proteins, while others vary in importance across proteins. We further show that the contribution of several components is correlated with protein length. These findings provide insight into how OpenFold achieves accurate predictions and highlight directions for interpreting protein prediction networks more broadly.

1 Introduction

Recent protein structure prediction models such as AlphaFold2 [11] and OpenFold [2] have transformed biology, enabling breakthroughs in protein folding [23], drug discovery [12, 13], and protein synthesis [7, 23]. Yet, despite their success, it remains unclear how these models achieve such accuracy, i.e., which architectural components are most essential for prediction. At the core of several protein structure prediction models lies a transformer-based model that iteratively refines two key internal representations: the multiple sequence alignment (MSA) representations and the pairwise residue (Pair) representations. In AlphaFold2 and OpenFold, the main transformer model is called the Evoformer, which includes several components such as attention layers, transition MLPs, and triangular update operations. However, the relative importance of these components for structure prediction is not well understood. Previous ablations of AlphaFold2 and OpenFold focused mainly on auxiliary losses, training regimes, or coarse architectural changes (e.g., "no triangles, biasing, or gating"), leaving the role of individual Evoformer components largely unexplored. In this paper, we address this gap by systematically evaluating component-level contributions across proteins. Our study reveals which components are broadly critical, which are dispensable, and how their importance varies with properties such as protein length.

Our main contributions are as follows:

1. We propose a methodology to systematically examine the contribution of individual OpenFold components to structure prediction accuracy, revealing protein-specific reliance on different architectural components.
2. We identify several components that are critical for accurate predictions across most proteins, including MSA Column Attention, both MLP Transition layers, and the final Pair representation used for structure prediction, providing biological insights into *how* OpenFold achieves accurate predictions.
3. We analyze how these contributions vary with protein length and find that the importance of several components is statistically significantly correlated with sequence length.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: 2nd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences.

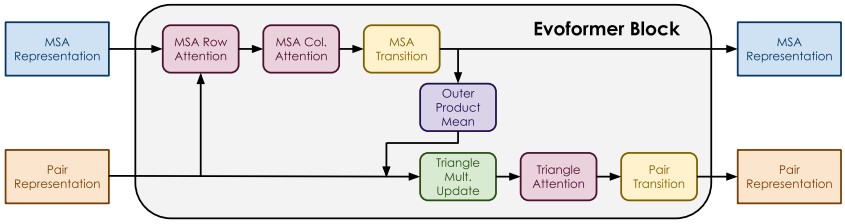


Figure 1: **One Evoformer Block in OpenFold.** Each block operates on the MSA and Pair representations via a series of attention, transition, and update operations. For clarity, residual connections are omitted, except for the Outer Product Mean connection from the MSA to the Pair representation.

2 Related Work

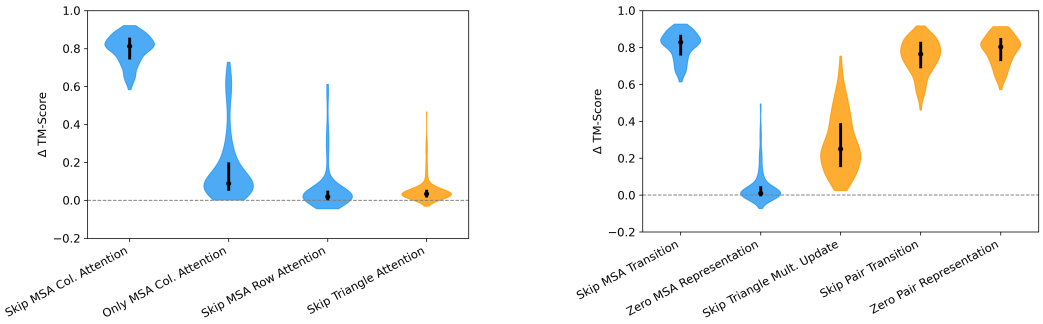
Several recent works have studied the inner workings of AlphaFold2. For example, [21] proposes an objective function to solve for a set of residue deletions or substitutions required to change the network's structure predictions. In [8], the impact of different templates and the recycling mechanism are studied in more detail. In [19], a framework is proposed to study the contribution of different amino acids to final structure prediction using SHAP. Our study differs from existing studies since we seek to characterize the contribution of architectural components to structure prediction, rather than the contribution of individual residues, templates, or recycling iterations. Beyond structure prediction models, there has also been interpretability work on protein language models (see [16] for a survey). In [6, 20], sparse auto-encoders are used to study the ESM-2 [15] representation space. There has also been work to study the intrinsic dimensionality of protein language model representation spaces [5, 22, 26]. In [24], a methodology is proposed to relate the attention maps of protein language models to protein properties such as structure or function. More broadly, the interpretability of protein models has been proposed as a promising future direction by several researchers [3, 4, 13, 16].

3 Methods

Model Components. OpenFold [2] is an open-source reproduction of AlphaFold2 [11], making it a suitable model for our study. Despite AlphaFold2's advances in protein structure prediction [18], many architectural contributions remain unclear [13]. Since recent models such as AlphaFold3 [1] and Boltz [17, 25] retain the same transformer-based architecture with triangular operations on Pair representations, our findings extend beyond OpenFold. OpenFold operates in three distinct phases to predict structures from amino acid sequences: i) a pre-processing phase that produces a Multiple Sequence Alignment (MSA) representation and a Pair representation; ii) Evoformer processing via 48 blocks to refine these representations; and iii) the Structure Module, which outputs a 3D structure from these representations. The MSA representation is produced by comparing the input sequence to existing sequences from nature, while the Pair representation is produced by comparing residue-residue pair relationships in the sequence.

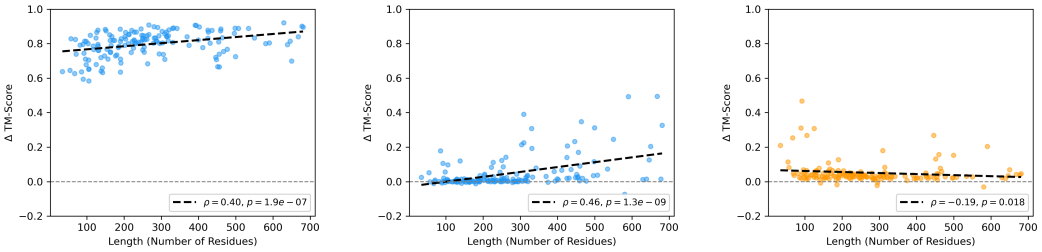
Here, we study the impact of various Evoformer components on structural accuracy (see Fig. 1 for an overview). Specifically, each Evoformer block consists of two pathways to operate on the MSA and Pair representations. Within the MSA pathway, MSA Row Attention integrates information across homologous sequences for each residue position, followed by MSA Column Attention, which correlates residues along each sequence. An MSA Transition (feedforward MLP) further transforms these features. The Outer Product Mean projects the MSA representation into the Pair representation, which is refined through Triangle Multiplicative Updates and Triangle Attention, designed to enforce the triangle inequality for geometric consistency among residue triplets. Finally, the Pair Transition applies a feedforward MLP before the final representations are passed to the next Evoformer block, or the Structure Module. There are also residual connections before and after each module.

Analysis Procedure. We perform three types of experiments: i) skipping attention modules, ii) skipping non-attention modules or zeroing out representations, and iii) evaluating the correlation between the change in performance due to component manipulation with (protein) sequence length. Skipping is implemented by bypassing the module's operations in all Evoformer blocks, while zeroing is implemented by setting all values of the final representation to zero before passing it to the



(a) Attention Experiments (b) Non-Attention Experiments

Figure 2: **Differences (Δ) Between Baseline TM-Scores and Component Studies Across Proteins.** Higher scores indicate more deviation from the Baseline. Studies involving MSA and Pair Representations are in blue and orange, respectively.



(a) Skip MSA Col. Attention (b) Zero MSA Representation (c) Skip Triangle Attention

Figure 3: **Comparison of Baseline Differences in TM-Scores (Δ) vs. Protein Length.** Higher scores indicate more deviation from the Baseline. We show the best-fit line, the Spearman correlation coefficient, and its associated p-value. Each point represents the results for one protein.

Structure Module. Component-level experiments identify which modules affect structure prediction the most, while representation-level experiments assess the usefulness of each representation. For length analyses, we plot the difference in TM-score between the baseline and component removal experiment as a function of protein length. We then fit a line to the data via linear regression and perform Spearman correlation analyses.

Data & Metrics. We use the three-month CAMEO [9] subset from OpenFold, restricted to proteins with fewer than 700 residues [2]. After filtering out targets with missing structural files and runs with baseline TM-scores below 0.7, we obtain 154 proteins. We run OpenFold `model_1_ptm` with the AlphaFold2 JAX weights and report results with unrelaxed structures. Our primary metric is the TM-score between predictions and experimental ground truth. To quantify the effect of each component, we compare the baseline TM-score to the score when a component is skipped or zeroed (Δ TM). For length correlation analyses, we report R^2, Spearman's ρ, and p-values. We run experiments with each protein three times and use the average performance across runs for our analyses. More details are in Supplemental Materials (Sec. S2), including raw TM-scores (Fig. S1).

4 Results: Component-Level Findings

Analyzing Attention Components. In Fig. 2a, we study the impact of skipping MSA Column Attention, MSA Row Attention, and Triangle Attention (around the starting and ending nodes) in the Evoformer. Most proteins are negatively impacted by skipping MSA Column Attention, the largest deviation from baseline. In contrast, bypassing MSA Row Attention has minor performance impacts for most proteins, and bypassing Triangle Attention has minimal impacts on most proteins. However, the broad variance of the violin plots for these experiments indicate that some proteins benefit from the inclusion of MSA Row or Triangle attention. Since MSA Column Attention was

Table 1: **Correlation Between ΔTM and Protein Length.** We report R^2, Spearman's ρ, and p-values for each experiment. We show $p < 0.05$ in **bold**.

ATTENTION STUDY	R^2	ρ	p-value	NON-ATTENTION STUDY	R^2	ρ	p-value
Skip MSA Col. Attention	0.13	0.40	**1.9e-7**	Skip MSA Transition	0.09	0.34	**1.2e-5**
Only MSA Col. Attention	0.02	-0.13	0.11	Zero MSA Representation	0.21	0.46	**1.3e-9**
Skip MSA Row Attention	0.01	-0.07	0.42	Skip Triangle Mult. Update	0.06	0.08	0.31
Skip Triangle Attention	0.02	-0.19	**0.018**	Skip Pair Transition	0.26	0.56	**3.8e-14**
				Zero Pair Representation	0.11	0.38	**1.1e-6**

most critical, we also tested models with *only* this attention component, i.e., skip MSA Row and Triangle attention. Surprisingly, we found that many proteins performed well with only MSA Column Attention, showing only slight deviations from the baseline (median ΔTM is 0.089). This implies that OpenFold heavily relies on evolutionary relationships for predicting structures. We also study the impact of using only MSA Row Attention or only Triangle Attention (Fig. S2a) and find that these attention types are not enough on their own for structure prediction.

Analyzing Non-Attention Components. In Fig. 2b, we evaluate the impact of the MSA Transition, the Pair Transition, the Triangle Multiplicative Update (outgoing and incoming), and the MSA and Pair representations on structure prediction. Zeroing the MSA representation had the smallest effect, while skipping transitions or zeroing the Pair representation caused the largest drops. The high variance of the violin plot for zeroing out the MSA representation indicates that some proteins benefit from the inclusion of this component. Removing transition layers results in large performance drops (median ΔTM is 0.765 and 0.829 for skipping Pair and MSA transitions, respectively), consistent with previous work suggesting that MLP layers in transformers contain crucial semantics [10, 14]. Skipping the Triangle Multiplicative Update produced highly variable outcomes across proteins, in contrast to Triangle Attention, which had little effect. This suggests that the update is more critical for performance. Removing both operations (Fig. S2b) led to poor accuracy, underscoring the heterogeneity in which components OpenFold relies on.

Correlating Component Removal with Protein Length. In Fig.3, we plot ΔTM-score versus protein length for three key experiments: skipping MSA Column Attention, zeroing the MSA representation, and skipping Triangle Attention (others are in Fig.S3). Summary statistics (R^2, Spearman's ρ, p-value) are in Table 1. Skipping MSA Column Attention led to increasing performance loss with length ($p = 1.9 \times 10^{-7}$), indicating stronger reliance in longer proteins. Zeroing the MSA representation showed little effect on short proteins but larger losses for longer proteins ($p = 1.3 \times 10^{-9}$). In contrast, skipping Triangle Attention correlated negatively with length ($p = 0.018$), suggesting greater reliance on Triangle Attention for predicting structures for shorter proteins. Other components (Pair representation, MSA/Pair Transitions) also showed length dependence, while skipping MSA Row Attention, only using MSA Column Attention, and skipping the Triangle Multiplicative Update did not. This suggests that length only partially explains the heterogeneity in component importance, and additional factors such as fold type could be studied in the future.

5 Discussion & Conclusion

We showed that OpenFold's structure prediction relies most heavily on MSA Column Attention, the MLP Transition layers, and the final Pair representation. Other components, including MSA Row Attention, Triangle Attention, the Triangle Multiplicative Update, and the final MSA representation, contributed variably across proteins, with the multiplicative update showing the widest variance. Remarkably, much of the baseline performance could be recovered using only MSA Column Attention, highlighting the reliance on evolutionary sequence information. Several components, including MSA Column and Triangle Attention, the Transition layers, and the MSA/Pair representations, exhibited length-dependent effects: longer proteins depended more on MSA-based features, whereas shorter proteins were more sensitive to Triangle Attention. In contrast, the Triangle Multiplicative Update showed no length dependence, indicating heterogeneity beyond sequence length. Overall, different proteins rely on different Evoformer components. Future work could test additional factors, such as fold type or function, to further explain this heterogeneity. Our findings provide key insights into the inner workings of AlphaFold-like models.

Acknowledgements

This work is supported in part by NSF award #2502793. This research is also supported by the National Artificial Intelligence Research Resource (NAIRR) Pilot and the Delta advanced computing and data resource, which is supported by the National Science Foundation (award NSF-OAC #2005572). The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements of any sponsor.

References

[1] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

[2] Gustaf Ahdritz, Nazim Bouatta, Christina Floristean, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J. O'Donnell, Daniel Berenberg, Ian Fisk, Niccolò Zanichelli, Bo Zhang, Arkadiusz Nowaczynski, Bei Wang, Marta M. Stepniewska-Dziubinska, Shang Zhang, Adegoke Ojewole, Murat Efe Guney, Stella Biderman, Andrew M. Watkins, Stephen Ra, Pablo Ribalta Lorenzo, Lucas Nivon, Brian Weitzner, Yih-En Andrew Ban, Shiyang Chen, Minjia Zhang, Conglong Li, Shuaiwen Leon Song, Yuxiong He, Peter K. Sorger, Emad Mostaque, Zhao Zhang, Richard Bonneau, and Mohammed AlQuraishi. Openfold: Retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature Methods*, 21(8):1514–1524, 2024.

[3] Shi-Jie Chen, Mubashir Hassan, Robert L Jernigan, Kejue Jia, Daisuke Kihara, Andrzej Kloczkowski, Sergei Kotelnikov, Dima Kozakov, Jie Liang, Adam Liwo, Silvina Matysiak, Jarek Meller, Cristian Micheletti, Julie C Mitchell, Sayantan Mondal, Ruth Nussinov, Kei-ichi Okazaki, Dzmitry Padhorny, Jeffrey Skolnick, Tobin R Sosnick, George Stan, Ilya Vakser, Xiaoqin Zou, and George D Rose. Protein folds vs. protein folding: Differing questions, different challenges. *Proceedings of the National Academy of Sciences*, 120(1):e2214423119, 2023.

[4] Valerie Chen, Muyu Yang, Wenbo Cui, Joon Sik Kim, Ameet Talwalkar, and Jian Ma. Applying interpretable machine learning in computational biology—pitfalls, recommendations and opportunities for new developments. *Nature Methods*, 21(8):1454–1461, 2024.

[5] Elena Facco, Andrea Pagnani, Elena Tea Russo, and Alessandro Laio. The intrinsic dimension of protein sequence evolution. *PLoS Computational Biology*, 15(4):e1006767, 2019.

[6] Edith Natalia Villegas Garcia and Alessio ansuini. Interpreting and steering protein language models through sparse autoencoders. In *ICLR 2025 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2025.

[7] Casper A Goverde, Benedict Wolf, Hamed Khakzad, Stéphane Rosset, and Bruno E Correia. De novo protein design by inversion of the alphafold structure prediction network. *Protein Science*, 32(6):e4653, 2023.

[8] Jannik Adrian Gut and Thomas Lemmin. Dissecting alphafold2's capabilities with limited sequence information. *Bioinformatics Advances*, 5(1):vbae187, 2025.

[9] Jürgen Haas, Alessandro Barbato, Dario Behringer, Gabriel Studer, Steven Roth, Martino Bertoni, Khaled Mostaguir, Rafal Gumienny, and Torsten Schwede. Continuous automated model evaluation (cameo) complementing the critical assessment of structure prediction in casp12. *Proteins: Structure, Function, and Bioinformatics*, 86:387–398, 2018.

[10] Shwai He, Guoheng Sun, Zheyu Shen, and Ang Li. What matters in transformers? not all attention is needed. *arXiv preprint arXiv:2406.15786*, 2024.

[11] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin

Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[12] Masha Karelina, Joseph J Noh, and Ron O Dror. How accurately can one predict drug binding modes using alphafold models? *Elife*, 12:RP89386, 2023.

[13] Oleg Kovalevskiy, Juan Mateos-Garcia, and Kathryn Tunyasuvunakool. Alphafold two years on: Validation and impact. *Proceedings of the National Academy of Sciences*, 121(34):e2315002121, 2024.

[14] Sihao Lin, Pumeng Lyu, Dongrui Liu, Tao Tang, Xiaodan Liang, Andy Song, and Xiaojun Chang. Mlp can be a good transformer learner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19489–19498, 2024.

[15] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[16] Viet Thanh Duy Nguyen and Truong-Son Hy. Advances in protein representation learning: Methods, applications, and future directions. *arXiv preprint arXiv:2503.16659*, 2025.

[17] Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025.

[18] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021.

[19] Sabbir Ahmed Sibli, Vlasios Panagiotis Panagiotou, and Christos Makris. Enhancing protein structure predictions: Deepshap as a tool for understanding alphafold2. *Expert Systems with Applications*, page 127853, 2025.

[20] Elana Simon and James Zou. Interplm: Discovering interpretable features in protein language models via sparse autoencoders. *bioRxiv*, pages 2024–11, 2024.

[21] Juntao Tan and Yongfeng Zhang. Explainablefold: Understanding alphafold prediction with explainable ai. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2166–2176, 2023.

[22] Lucrezia Valeriani, Diego Doimo, Francesca Cuturello, Alessandro Laio, Alessio Ansuini, and Alberto Cazzaniga. The geometry of hidden representations of large transformer models. *Advances in Neural Information Processing Systems*, 36:51234–51252, 2023.

[23] Mihaly Varadi, Damian Bertoni, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna, Malarvizhi Radhakrishnan, Maxim Tsenkov, Sreenath Nair, Milot Mirdita, Jingi Yeo, Oleg Kovalevskiy, Kathryn Tunyasuvunakool, Agata Laydon, Augustin Žídek, Hamish Tomlinson, Dhavanthi Hariharan, Josh Abrahamson, Tim Green, John Jumper, Ewan Birney, Martin Steinegger, Demis Hassabis, and Sameer Velankar. Alphafold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1):D368–D375, 11 2023.

[24] Jesse Vig, Ali Madani, Lav R. Varshney, Caiming Xiong, Richard Socher, and Nazneen Rajani. BERTology meets biology: Interpreting attention in protein language models. In *International Conference on Learning Representations*, 2021.

[25] Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, 2024.

[26] Soojung Yang, Juno Nam, Tynan Perez, Jinyeop Song, Xiaochen Du, and Rafael Gomez-Bombarelli. Probing the embedding space of protein foundation models through intrinsic dimension analysis. In *NeurIPS 2024 Workshop on AI for New Drug Modalities*, 2024.

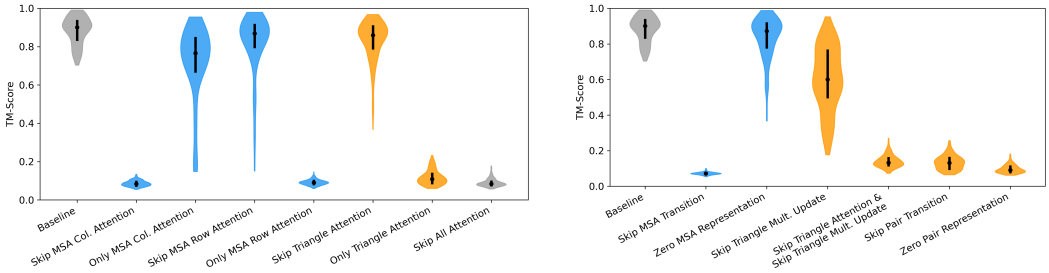
S1 Supplemental Materials

S2 Data & Metrics

We follow the OpenFold paper [2] and use a three-month subset of the CAMEO dataset (ending on January 16, 2022) of proteins with fewer than 700 residues [9]. This subset contains 185 proteins, of which we exclude one protein (7eqx_C) due to the removal of its mmCIF file (4uf4) from the RCSB website. This CAMEO subset is ideal for validating the impact of components on model performance as it was not used during training. We run OpenFold `model_1_ptm` with the suggested paper settings and the original AlphaFold2 JAX weights[1]. While three recycles are typically used, we found minor performance differences between using zero and three recycles (0.7% on the CAMEO subset), which is consistent with previous findings [8], so we use zero recycles here. Additionally, we use the unrelaxed structure predictions from the model, which have minimal performance differences with the relaxed predictions (0.003% difference). Our main performance metric is the TM-score between model predictions and experimental ground truth from the RCSB website. For all experiments, we filter proteins down to those that achieve a TM-Score of greater than 0.7 with a standard OpenFold run. This yields a set of 154 proteins and eliminates poor performance as a factor that could contribute to our component studies. To capture the influence of each component, regardless of model performance on specific proteins, we show the differences (Δ) between a Baseline run of OpenFold (without any components skipped) and the study with the component skipped or zeroed out. For our studies correlating performance with protein length, we report R^2 of the best-fit line, the Spearman correlation coefficient (ρ), and the Spearman p-value.

S3 Raw TM-Score Plots

We show the raw TM-Scores for each component study in Fig. S1, including the Baseline performance. We use Baseline differences for plots in our main paper (Fig. 2) to account for some protein structures being easier to predict than others. This allows us to focus directly on trends without model performance conflating the trends.



(a) Attention Experiments

(b) Non-Attention Experiments

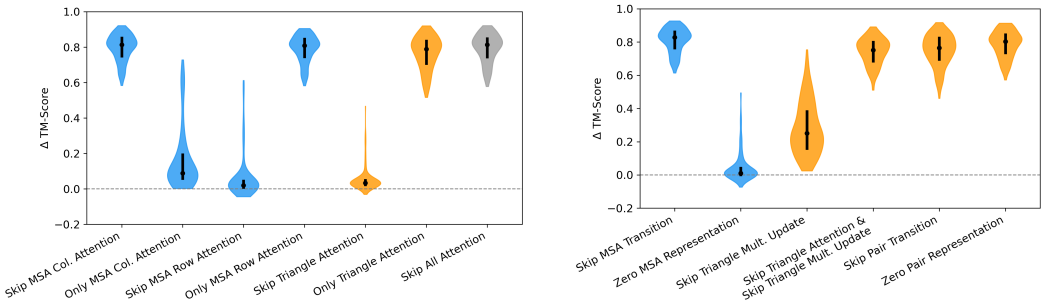
Figure S1: **Raw TM-Scores for Component Studies.** Higher scores indicate better performance. Studies with MSA and Pair Representations are in blue and orange, respectively.

S4 Analyzing the Impact of More Components

In Fig. S2, we show the differences between the Baseline and various component studies for a larger set of experiments than our main paper. In Fig. S2a, we additionally show the impact of using only MSA Row Attention, only Triangle Attention, or skipping attention altogether. We find that only using MSA Row Attention or only Triangle Attention is not sufficient for protein structure prediction with OpenFold as there is a large difference in performance with the Baseline. Moreover, in Fig. S2b, we additionally show the impact of skipping both Triangle Attention and the Triangle Multiplicative update. Unsurprisingly, this results in a large difference with the Baseline as these are the two main components contributing to the Pair representation, which we found to be critical to

[1] `https://github.com/aqlaboratory/openfold`

structure prediction performance (i.e., zeroing out Pair Representations results in large differences with the Baseline).



(a) Attention Experiments

(b) Non-Attention Experiments

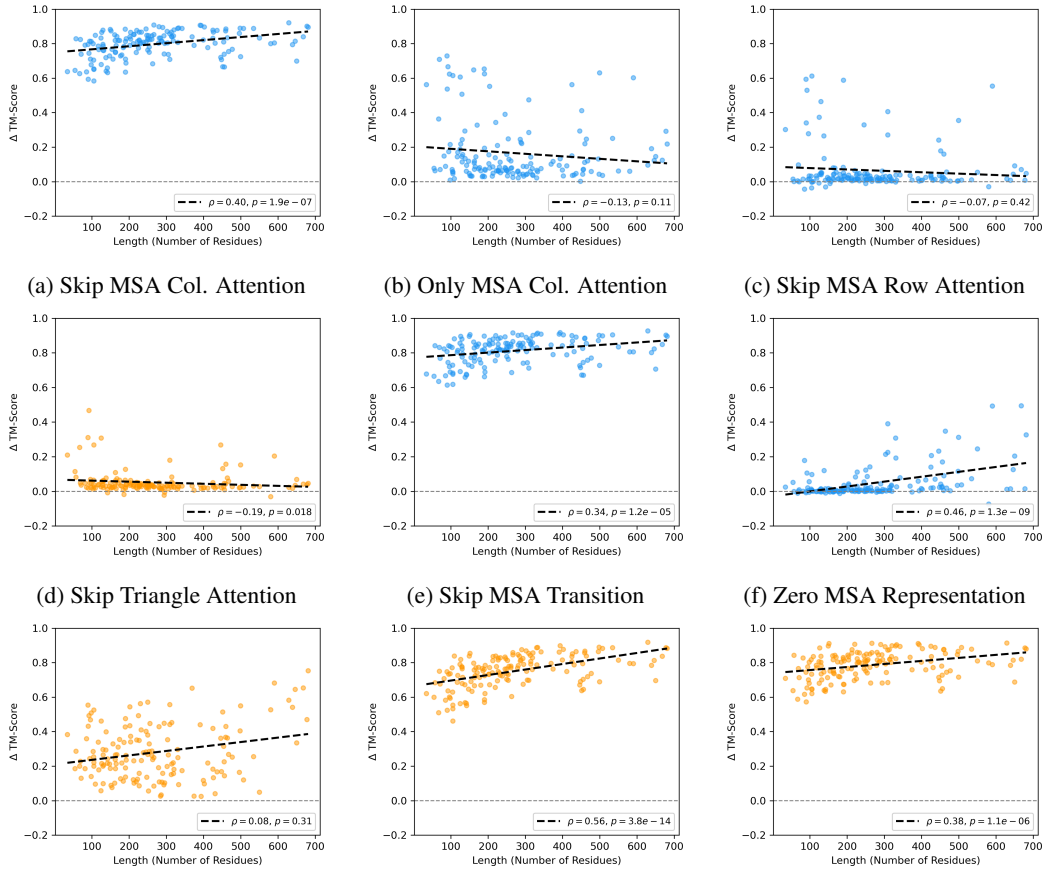
Figure S2: **Differences (△) Between Baseline TM-Scores and Component Studies Across Proteins.** Higher scores indicate more deviation from the Baseline. Studies involving MSA and Pair Representations are in blue and orange, respectively.

S5 Baseline Difference Versus Protein Length Plots

In Fig. S3, we show the Baseline Difference versus Protein Length plots for all component studies carried out in our main paper. The correlation values for these experiments are included in Table 1.

S6 Replacing Representations with Noise

In our main experiments, we evaluated the impact of the MSA representation and the Pair representation on structure prediction by zeroing out their respective tensors before the Structure Module. Here, we instead ablate these representations by replacing them with random noise while preserving tensor scale. Specifically, for a representation tensor $\mathbf{X}$, we compute the per-channel mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$ across the spatial dimensions (sequence or pair indices), and draw noise as $\boldsymbol{\mu} + \boldsymbol{\sigma} \odot \mathcal{N}(0, \mathbf{I})$. We then replace $\mathbf{X}$ with this noise. The results, shown in Fig. S4, mirror the zeroing experiments: replacing the MSA representation has minimal effect relative to the Baseline, while replacing the Pair representation leads to a large drop in TM-score.



(a) Skip MSA Col. Attention (b) Only MSA Col. Attention (c) Skip MSA Row Attention

(d) Skip Triangle Attention (e) Skip MSA Transition (f) Zero MSA Representation

(g) Skip Triangle Mult. Update (h) Skip Pair Transition (i) Zero Pair Representation

Figure S3: **Comparison of Baseline Differences in TM-Scores (Δ) vs. Protein Length.** Higher scores indicate more deviation from the Baseline. We show the best-fit line, the Spearman correlation coefficient, and its associated p-value. Each point represents the results for one protein.

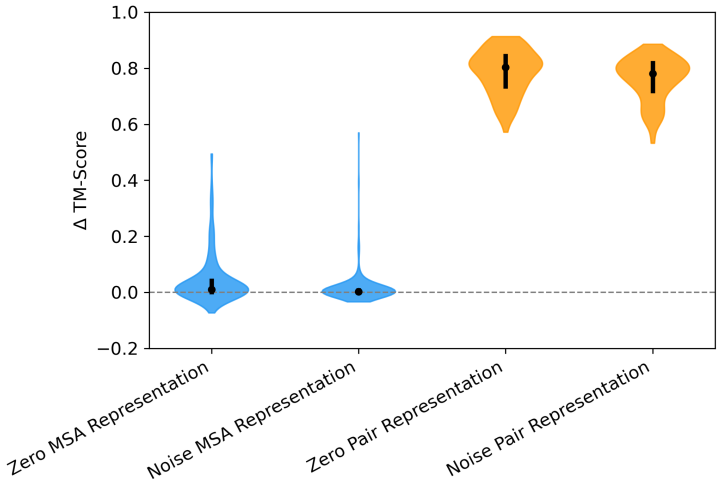


Figure S4: **Replacing MSA Representation and Pair Representation with Zeros vs. Noise.** We plot the differences (Δ) between Baseline TM-Scores and each study. Higher scores indicate more deviation from the Baseline. Studies involving MSA and Pair Representations are in blue and orange, respectively.