End-to-End Sidechain Modeling in AlphaFold2: Attention May or May Not Be All That You Need

Jonathan Edward King*
Comp. & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15213
jok120@pitt.edu

David Ryan Koes
Comp. & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15213
dkoes@pitt.edu

Abstract

AlphaFold2 (AF2) has made significant strides in computational structural biology and drug discovery. However, limitations remain, particularly for downstream tasks such as molecular docking. We propose inaccuracies in amino acid sidechain prediction could contribute to these limitations. To address this, we explored two simple and complementary strategies to improve sidechain accuracy in AF2: (1) substituting the default ResNet-based angle predictor in AlphaFold2 with a Transformer-like model, and (2) refining the angle predictor using an energy-like loss function. Our analysis indicates that ResNets and Transformers offer comparable performance. However, training with an energy-like loss can sometimes boost structural quality, especially when the entire model is finetuned. We suggest a holistic approach that looks beyond AF2's sidechain torsion angle predictor to improve sidechain modeling in future studies.

1 Introduction

1.1 The Significance of Sidechains in Drug Discovery

The discovery and design of novel drugs remain central to advancing medical science, with protein structures serving as critical components in structure-based drug discovery. While the unveiling of AlphaFold2 [1] (AF2) marked a monumental leap in computational protein structure prediction, it's important to delineate its capabilities and limitations.

One significant limitation of AF2 lies in the use of its predictions for molecular docking, a technique vital for drug discovery [2–4]. Molecular docking often necessitates highly accurate predictions of sidechain orientations to forecast how potential drug molecules may interact with target proteins. Although AF2 offers impressive accuracy overall, its predictions often contain differences (some subtle, some not) that adversely affect the performance of subsequent docking methods. For instance, a study by He et al. [4] examined four small molecule-GPCR binding complexes and found that, in three of these complexes, the sidechain conformations predicted by AF2 deviated sufficiently from experimental structures, leading to altered docking results for known active compounds.

Furthermore, sidechains are central to the differences observed between *apo* (without ligand) and *holo* (with ligand) protein structures. Accurate sidechain modeling is essential as these differences often arise from shifts in sidechain torsional angles, which in turn affect configurational entropy and the local chemical environment [5, 6]. These shifts in torsional angles are crucial to understanding the distinctions between *apo* and *holo* structures and are integral to the drug discovery process. Therefore, precise modeling of sidechains is imperative for developing effective drugs in various contexts.

*Carnegie Mellon University-University of Pittsburgh Joint PhD Program in Computational Biology

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

1.2 The Evolution of Sidechain Modeling Techniques

Sidechain modeling, along with protein structure prediction in general, has undergone significant development in recent years. There are several problem formulations of note: sidechain packing, simultaneous modeling of protein sidechain and backbone atoms, and lastly, extensions of AF2 to improve its sidechain modeling capabilities.

The Sidechain Packing Challenge The sidechain packing problem has historically been one of the prominent challenges in protein structure prediction. The issue revolves around determining the optimal sidechain orientations on a fixed backbone. Several algorithms and heuristics have been developed to address this challenge, with traditional approaches searching rotamer libraries for sidechain conformations with favorable energy [7, 8]. Several machine learning approaches have also been developed [9–11], taking advantage of diverse deep learning model architectures to achieve state-of-the-art accuracy and speed.

Unified Backbone and Sidechain Prediction While the sidechain packing approach provides valuable insights, it simplifies the real-world scenario where sidechain and backbone conformations are intrinsically linked. Recognizing this interdependence, researchers have aimed to jointly predict both [1, 12, 13]. Methods that adopt this strategy often may also benefit from iterative refinement, where backbone and sidechain predictions inform and refine each other in a cyclic manner [1]. AlphaFold2 is a prime example of both unified protein backbone and sidechain prediction as well as iterative refinement. It is worth noting that AF2's sidechain predictions, though accurate, are still not as accurate as some of the contemporary sidechain packing techniques described above.

1.3 Potential Avenues for Improved Sidechain Modeling

Given AF2's impact, efforts to enhance its sidechain accuracy have also emerged. These include traditional refinement methods using molecular dynamics (MD) [14] and docking pipelines [15], novel refinement methods [16, 17], enhancing AF2 s input [18, 19], or finetuning AF2 itself [20].

Building on these advancements, the integration of more sophisticated architectures and training techniques presents a potential avenue for improved sidechain modeling. For instance, considering the widespread success of the Transformer architecture across various domains, it stands as a promising replacement for the ResNet angle predictor in AF2. Adept at managing complex long-range dependencies, Transformers might be especially beneficial in modeling sidechain interactions and packing. In addition, an alternative strategy is implementing an energy-like loss function to guide models based on the energetics of protein structures.

2 Methods

Data One of our underlying hypotheses is that protein structures with lower potential energy and fewer atomic clashes are more desirable for practical applications like structure-based drug discovery. Therefore, we utilize an energy-minimized subset (about 32,000 protein chains) of the CASP12 iteration of the SidechainNet dataset [21, 22]. The protein chains in this dataset were minimized using an energy-like loss function that interprets forces computed by OpenMM software as gradients. A validation set from CAMEO [23] was selected to match the set used in the OpenFold paper [24]. A test set of 93 proteins was selected from a one-year window of CAMEO proteins ending on January 3, 2023. Validation and test sets were minimized similarly to our training set. Structures that failed minimization were excluded from training and evaluation.

Pretraining Procedure Training AlphaFold2 requires significant computational resources. To adapt to this challenge, we developed a pretraining method that enables the independent training of AF2's angle predictor component. To do this, we first performed inference for our dataset using the AF2 weights (`finetuning_5.pt` provided by OpenFold [24]). During inference, we recorded the inputs to the angle predictor and the target angle values from the true structure. Next, we trained each angle predictor model separately from the AF2 model on this data using the `supervised_chi` loss, a weighted combination of Mean Squared Error on predicted sin/cos angle values and on the magnitude of the sin/cos vectors to ensure they lie on the unit circle. After pretraining, angle predictors can be reconstituted in the complete AF2 model, replacing the default ResNet.

Model	Layers	Heads	Model Dim	Feed-Forward Dim	# Params
ResNet (all models)	2	-	128	-	166K
AngleTransformer	42	1	64	1024	6.4M

Table 1: Comparison of ResNet and AngleTransformer model architectures.

3 Results

We investigate the impact of two modifications to AlphaFold2 to improve its sidechain modeling capabilities: (1) replacing AF2's ResNet-based angle prediction with a Transformer-based model, and (2) finetuning AF2's angle predictors to predict structures with lower energy.

3.1 Replacing AF2's ResNet with the AngleTransformer

Because Transformer-like models for protein sequence modeling are highly performant, we posited that substituting them for AF2's ResNet could improve sidechain modeling. We executed a hyperparameter sweep over approximately 500 Transformer-like architectures and found several models comparable to the baseline ResNet from AF2. Models were first pretrained separately from the complete AF2 model, as described in Section 2.

We evaluate the performance on various metrics for four models: (1) ResNet (AF2) - the pretrained ResNet directly loaded from the OpenFold model, (2) ResNet (untrained) - the ResNet architecture from AF2 without the pretrained weights, (3) ResNet (retrained) - where the untrained ResNet is trained from scratch, and (4) AngleTransformer (AT) - the Transformer-like model from our hyperparameter sweep with the lowest Mean Absolute Error (MAE) on sidechain torsion angles from our validation set. ResNet (untrained) was included as a baseline to differentiate the impact of the angle predictor versus the rest of the AF2 model on structural accuracy. Model architectures are summarized in Table 1. Transformer models are encoder-only, use GeLU activation, and include linear and 1-D convolutional layers applied to their input before positional encoding and attention.

Angle Predictor Pretraining We first evaluated the models by pretraining them outside of the complete AF2 model, stopping training when the validation set loss angle MAE stopped improving. Results are seen in Figures 1 (upper half) and S1. Despite the theoretical shortcomings of pretraining angle predictors without using the full model, we find that the ResNet (retrained) can recover a large portion of the all-atom structural accuracy, measured by the All-Atom local Distance Difference Test (lDDT$_{AA}$). We also see that lDDT$_{AA}$ is mostly simlar across pretrained models (Figures S1b, S1c). Concerning sidechain torsion angle prediction MAE, ResNet (retrained) can match the performance of the original ResNet (AF2) (Figure S1d), though it is outperformed by the AngleTransformer (Figures S1e, S1f). Figures S1j and S1k summarize the performance of each model across the four sidechain torsional angles and by amino acid hydrophobicity.

Lastly, in Figures S1g, S1h, and S1i, we see that our pretraining procedure generally results in predictions with higher violation loss than the fully trained ResNet (AF2) model. However, ResNets have lower violation loss than AngleTransformers despite not being trained to minimize this component during pretraining. Violation loss is a component of AF2's composite loss function that reflects the number of structural violations (i.e., bond lengths or angles that differ from expected values).

Angle Predictor Pretraining and Finetuning We hypothesized that we would get a complete picture of the AngleTransformer and ResNet's strengths if we finetuned each further, reconstituting them into the full AF2 model while freezing the rest of AF2's weights. The results of this finetuning experiment are summarized in Figure 1 (lower half), Figure S2 and Table 2.

After finetuning the ResNet and AngleTransformer, we find that both increase in accuracy with respect to MAE and lDDT$_{AA}$, with the AngleTransformer outperforming the ResNet (retrained) on MAE by a few degrees (Figure S2e and Table 2, p=0.056) and the ResNet insignificantly more performant according to lDDT$_{AA}$ (Figure S2b and Table 2, p=0.306). Violation loss values have improved for both models after finetuning, with ResNets still having an advantage (Figures S2h, S2i). The AngleTransformer is slightly better at predicting hydrophilic residue torsion angles than the ResNet (Figures S2k and S3).

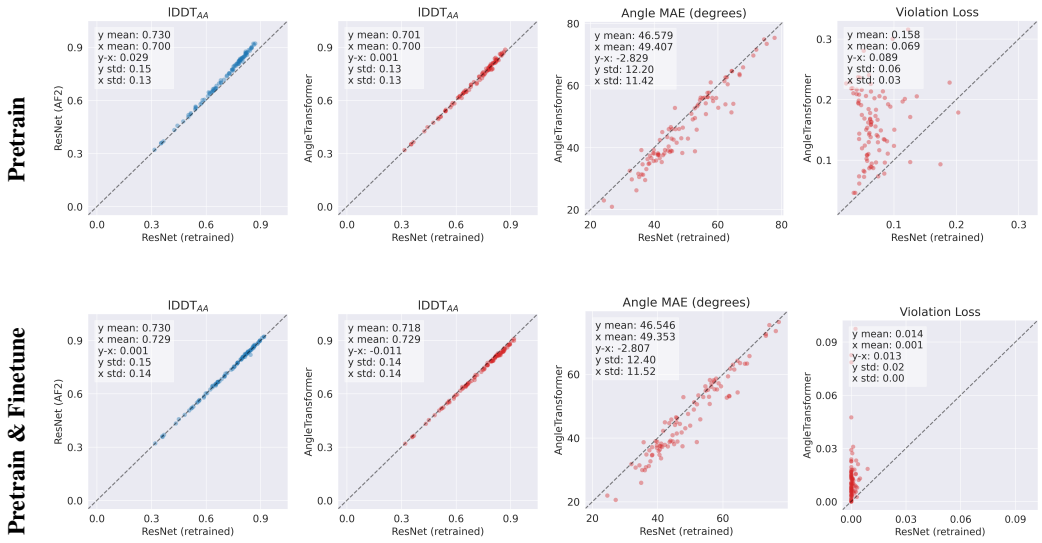

Figure 1: Comparing angle predictors before (top) and after (bottom) finetuning (other AF2 weights frozen) using the full AlphaFold loss. Higher lDDT$_{AA}$, lower violation loss, and lower angle MAE are better. Figure legends summarize x and y-axis statistics.

Model	lDDT$_{AA}$	MAE (degrees)	Violation Loss
ResNet (AF2)	**0.730**	49.177	**0.0004**
ResNet (untrained)	0.642	80.208	0.2051
ResNet (retrained)	0.729	49.352	0.0007
AngleTransformer	0.718	**46.546**	0.0136

Table 2: Accuracy on CAMEO test set structures after pretraining and finetuning procedures.

3.2 Energy-based Loss Functions

We also utilized an energy-based training procedure called OpenMM-Loss to improve the structural quality of predicted structures, introduced and described in detail in King and Koes [22]. At a high level, this method interprets the potential energy of predicted structures as a loss function and the atomic forces as gradients for end-to-end training.

To find the experimental conditions resulting in the highest accuracy, we evaluated several different variables in tandem: using the ResNet vs. AngleTransformer, OpenMM-Loss (OMM) vs. standard finetuning (no-OMM), and various finetuning procedures that finetuned only the angle predictor or the entire AF2 model.

Angle Predictor Finetuning With and Without OpenMM-Loss In our first set of experiments, we took the ResNet (retrained), hereafter referred to as "ResNet," and AngleTransformer models developed above and reconstituted them into the complete AF2 model. We finetuned the angle predictors only as in Section 3.1 while testing the effects of training with and without OMM. All models were trained for about one week using four A100 GPUs. ResNets and AngleTransformers significantly differed in their training behavior (Figure S4). ResNets, in general, benefitted from much faster training times and reached their maximal accuracy values in the shortest number of training steps. In the allotted compute budget, it is unclear if the AngleTransformer reached its maximum accuracy. Accuracy and structural quality metrics for four models (AT-OMM, AT-no-OMM, ResNet-OMM, and ResNet-no-OMM) are summarized in Figures S5, S6, and S7.

Entire Model Finetuning With and Without OpenMM-Loss Lastly, we investigated finetuning the angle predictor and the rest of the AF2 model with and without OpenMM-Loss (Figure 2).

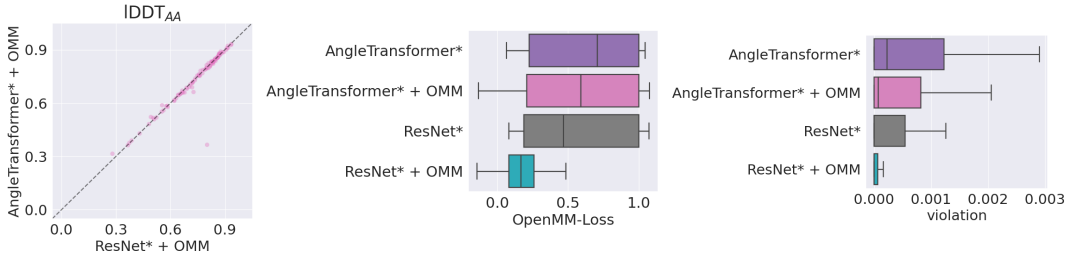

Figure 2: Comparing angle predictors after whole model finetuning with and without OpenMM-Loss.

3.3 Summary and Future Directions

We explored ResNets, Transformer-like architectures, and various training strategies for predicting sidechain torsional angles in AlphaFold2 (AF2). ResNets proved more efficient, achieving maximum accuracy in about ½ the number of steps (Figure S4). Eventually, both model types showed similar structural accuracy with several distinctions. AngleTransformers had lower angle MAE, with the biggest increases in accuracy for distal sidechain chi angles (χ_{2-4}) (Figure S6) and flexible residues like glutamic and aspartic acids (Figure S3). However, ResNets achieved similar or better lDDT$_{AA}$ and, surprisingly, lower violation loss scores even when pretrained to minimize angle loss alone (Figures 1 and 2).

Training with OpenMM Loss didn't significantly alter accuracy (Figures S5 and S6). OpenMM-Loss did, however, reduce the potential energy, violation loss values, and Clash Scores in both AngleTransformers and ResNets (Figures 2 and S7). Lastly, given the same time to train, ResNets achieve lower OpenMM-Loss and violation loss scores than AngleTransformers. We propose that the added model complexity of AngleTransformers, combined with our pretraining strategy to minimize angle MAE, may not offer enough overall benefits to warrant replacing AF2's ResNet.

Our experiments underscore the importance of a comprehensive approach to improving sidechain-level protein structures. While refining AF2's angle predictor can enhance sidechain modeling, finetuning the entire model yields the best accuracy and structural quality results. Future work could explore alternative sidechain prediction representations, possibly eschewing χ-angle prediction for atomistic modeling of sidechain-backbone interactions. Another worthwhile avenue might be implementing more efficient attention mechanisms, balancing improvement in expressiveness with cost-effectiveness. Alternatively, attention mechanisms might be supplemented by including a geometric prior, constructing attention weights based on distances or other geometric feature vectors measured from the data. Finally, since OpenMM-Loss does not have a significant impact on structural accuracy, it would be interesting to observe the effect of using predictions from models finetuned with OpenMM-Loss on downstream tasks like molecular docking.

Acknowledgments and Disclosure of Funding

This work is supported by R35GM140753 from the National Institute of General Medical Sciences and in part by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, which is supported by NSF award number OAC-2117681.

References

[1] 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, August 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2. URL `https://www.nature.com/articles/s41586-021-03819-2`. Number: 7873 Publisher: Nature Publishing Group.

[2] Karelina Masha, Noh Joseph J, and Dror Ron O. How accurately can one predict drug binding modes using AlphaFold models? *eLife*, 12, August 2023. doi: 10.7554/eLife.89386. URL https://elifesciences.org/reviewed-preprints/89386. Publisher: eLife Sciences Publications Limited.

[3] Valeria Scardino, Juan I. Di Filippo, and Claudio N. Cavasotto. How good are AlphaFold models for docking-based virtual screening? *iScience*, 26(1):105920, December 2022. ISSN 2589-0042. doi: 10.1016/j.isci.2022.105920. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9852548/.

[4] Xin-heng He, Chong-zhao You, Hua-liang Jiang, Yi Jiang, H. Eric Xu, and Xi Cheng. AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors. *Acta Pharmacologica Sinica*, 44(1): 1–7, January 2023. ISSN 1745-7254. doi: 10.1038/s41401-022-00938-y. URL https://www.nature.com/articles/s41401-022-00938-y. Number: 1 Publisher: Nature Publishing Group.

[5] Jordan J. Clark, Mark L. Benson, Richard D. Smith, and Heather A. Carlson. Inherent versus induced protein flexibility: Comparisons within and between apo and holo structures. *PLoS Computational Biology*, 15(1):e1006705, January 2019. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1006705. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6370239/.

[6] Francis Gaudreault, Matthieu Chartier, and Rafael Najmanovich. Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. 28(18):i423–i430. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts395. URL https://doi.org/10.1093/bioinformatics/bts395.

[7] Georgii G. Krivov, Maxim V. Shapovalov, and Roland L. Dunbrack. Improved prediction of protein side-chain conformations with SCWRL4. 77(4):778–795. ISSN 0887-3585. doi: 10.1002/prot.22488. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2885146/.

[8] Xiaoqiang Huang, Robin Pearce, and Yang Zhang. FASPR: an open-source tool for fast and accurate protein side-chain packing. 36(12):3758–3765. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa234. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7320614/.

[9] Matt McPartlon and Jinbo Xu. An end-to-end deep learning method for rotamer-free protein side-chain packing. URL https://www.biorxiv.org/content/10.1101/2022.03.11.483812v1. Pages: 2022.03.11.483812 Section: New Results.

[10] Gang Xu, Qinghua Wang, and Jianpeng Ma. OPUS-rota4: a gradient-based protein side-chain modeling framework assisted by deep learning-based predictors. 23(1):bbab529. ISSN 1477-4054. doi: 10.1093/bib/bbab529. URL https://doi.org/10.1093/bib/bbab529.

[11] Mikita Misiura, Raghav Shroff, Ross Thyer, and Anatoly B. Kolomeisky. DLPacker: Deep learning for prediction of amino acid side chain conformations in proteins. 90(6):1278–1290. ISSN 1097-0134. doi: 10.1002/prot.26311. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.26311. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26311.

[12] Alexander E. Chu, Lucy Cheng, Gina El Nesr, Minkai Xu, and Po-Ssu Huang. An all-atom protein generative model. URL https://www.biorxiv.org/content/10.1101/2023.05.24.542194v1. Pages: 2023.05.24.542194 Section: New Results.

[13] Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, Jianzhu Ma, and Jian Peng. High-resolution de novo structure prediction from primary sequence, . URL https://www.biorxiv.org/content/10.1101/2022.07.21.500999v1. Pages: 2022.07.21.500999 Section: New Results.

[14] Lim Heo, Giacomo Janson, and Michael Feig. Physics-based protein structure refinement in the era of artificial intelligence. 89(12):1870–1887. ISSN 1097-0134. doi: 10.1002/prot.26161.

[15] Yuqi Zhang, Marton Vass, Da Shi, Esam Abualrous, Jenny Chambers, Nikita Chopra, Chris Higgs, Koushik Kasavajhala, Hubert Li, Prajwal Nandekar, Hideyuki Sato, Edward Miller, Matt Repasky, and Steven Jerome. Benchmarking refined and unrefined AlphaFold2 structures for hit discovery. URL https://chemrxiv.org/engage/chemrxiv/article-details/62b41f0c0bbbc117477285a4.

[16] Tianqi Wu, Zhiye Guo, and Jianlin Cheng. Atomic protein structure refinement using all-atom graph representations and SE(3)-equivariant graph transformer. 39(5):btad298, . ISSN 1367-4811. doi: 10.1093/bioinformatics/btad298. URL https://doi.org/10.1093/bioinformatics/btad298.

[17] Recep Adiyaman, Nicholas S Edmunds, Ahmet G Genc, Shuaa M A Alharbi, and Liam J McGuffin. Improvement of protein tertiary and quaternary structure predictions using the ReFOLD refinement method and the AlphaFold2 recycling process. 3(1):vbad078. ISSN 2635-0041. doi: 10.1093/bioadv/vbad078. URL https://doi.org/10.1093/bioadv/vbad078.

[18] Jian Liu, Zhiye Guo, Tianqi Wu, Raj S. Roy, Chen Chen, and Jianlin Cheng. Improving AlphaFold2-based protein tertiary structure prediction with MULTICOM in CASP15. 6:188. ISSN 2399-3669. doi: 10.1038/s42004-023-00991-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10484931/.

[19] Fandi Wu, Xiaoyang Jing, Xiao Luo, and Jinbo Xu. Improving protein structure prediction using templates and sequence embedding. 39(1):btac723, . ISSN 1367-4811. doi: 10.1093/bioinformatics/btac723. URL https://doi.org/10.1093/bioinformatics/btac723.

[20] Toshiyuki Oda. Refinement of AlphaFold-multimer structures with single sequence input. URL https://www.biorxiv.org/content/10.1101/2022.12.27.521991v2. Pages: 2022.12.27.521991 Section: New Results.

[21] Jonathan Edward King and David Ryan Koes. SidechainNet: An all-atom protein structure dataset for machine learning. 89(11):1489–1496, . ISSN 1097-0134. doi: 10.1002/prot.26169. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.26169. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26169.

[22] Jonathan Edward King and David Ryan Koes. Interpreting molecular dynamics forces as deep learning gradients improves quality of predicted protein structures, . URL https://www.biorxiv.org/content/10.1101/2023.10.03.560775v1. Pages: 2023.10.03.560775 Section: New Results.

[23] Xavier Robin, Juergen Haas, Rafal Gumienny, Anna Smolinski, Gerardo Tauriello, and Torsten Schwede. Continuous Automated Model EvaluatiOn (CAMEO)—Perspectives on the future of fully automated evaluation of structure prediction methods. *Proteins: Structure, Function, and Bioinformatics*, 89(12):1977–1986, 2021. ISSN 1097-0134. doi: 10.1002/prot.26213. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.26213. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26213.

[24] Gustaf Ahdritz, Nazim Bouatta, 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, 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, November 2022. URL https://www.biorxiv.org/content/10.1101/2022.11.20.517210v2. Pages: 2022.11.20.517210 Section: New Results.

Supplementary Material

Code can be found at `https://github.com/jonathanking/angletransformer` and `https://github.com/jonathanking/openfold`.

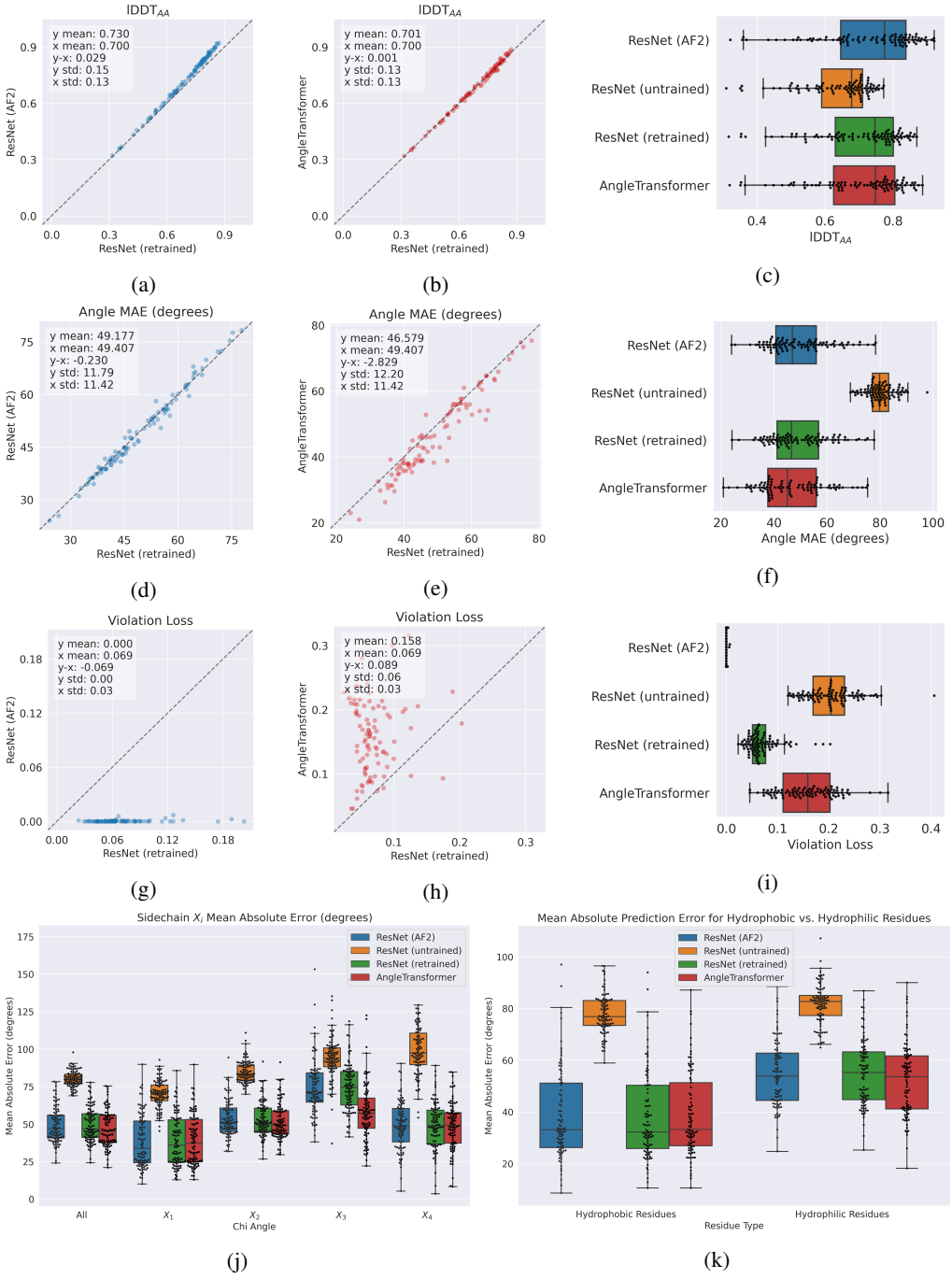

Figure S1: **Pretraining:** Comparing AF2 angle predictor structure accuracy on held-out CAMEO test set proteins after pretraining only. Higher lDDT$_{AA}$ is better. Lower violation loss and angle MAE are better.

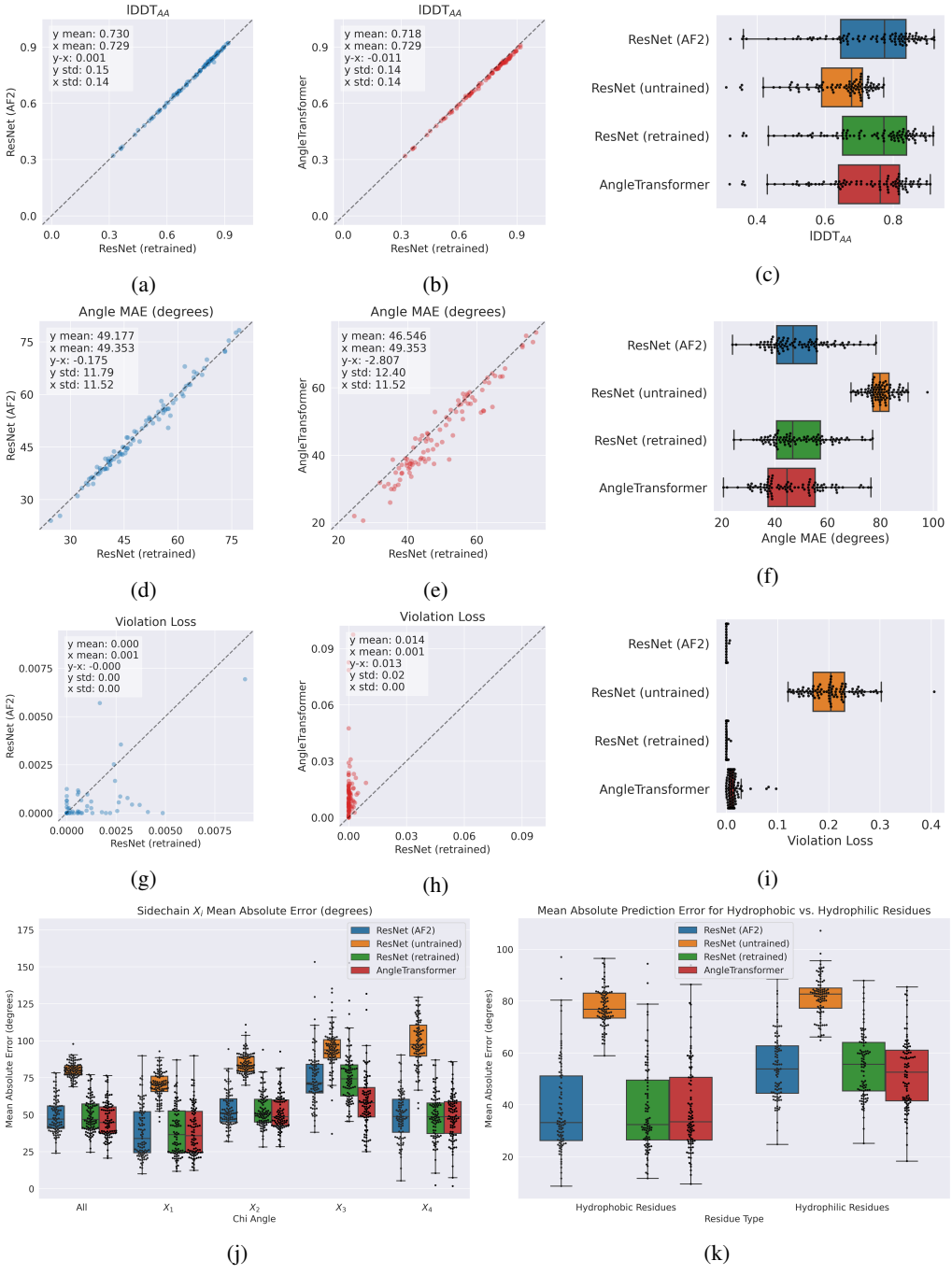

Figure S2: **Pretraining and Finetuning:** Comparing AF2 angle predictor structure accuracy after finetuning models from Figure S1. Higher lDDT$_{AA}$ is better. Lower violation loss and angle MAE are better.

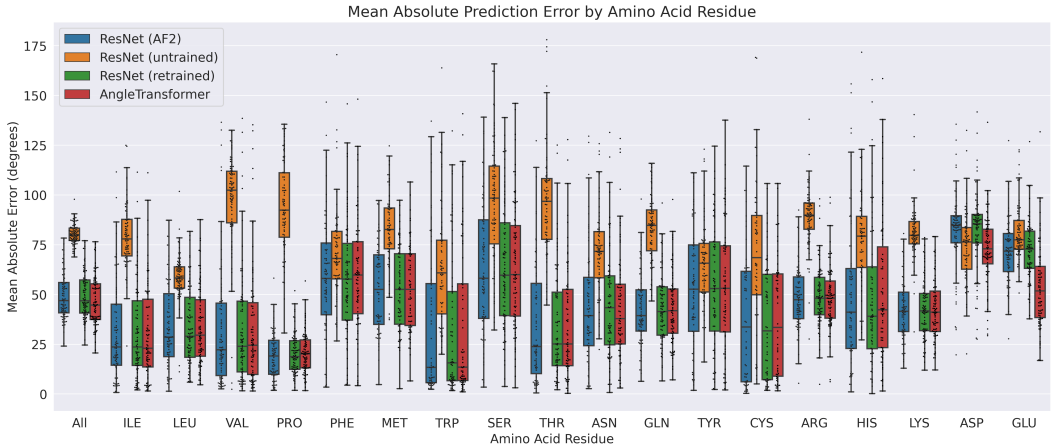

Figure S3: Comparing AF2 angle predictor angle accuracy by residue identity. Lower MAE is better.

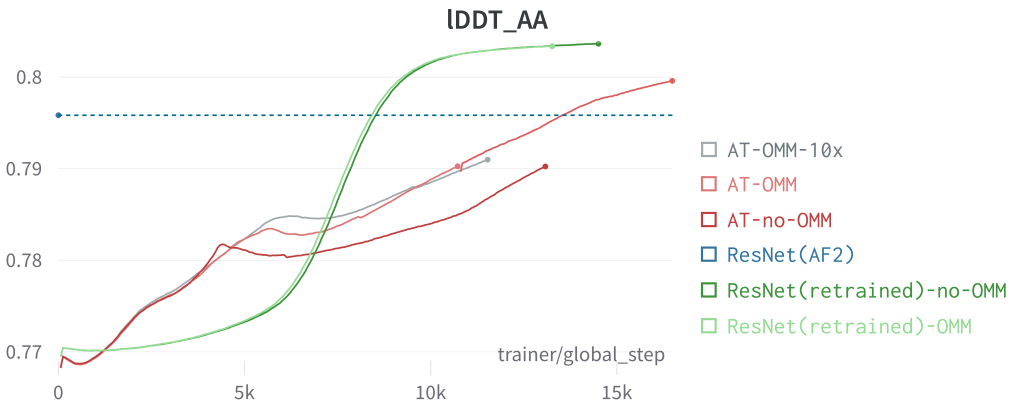

Figure S4: Evaluting the effect of OMM-Loss on training angle predictors. AT-OMM-10x is an AngleTransformer model identical to AT-OMM, but the weight of the OpenMM-Loss component is increased 10x (from 0.01 to 0.1) in the AlphaFold2 composite loss function. Higher lDDT$_{AA}$ is better.

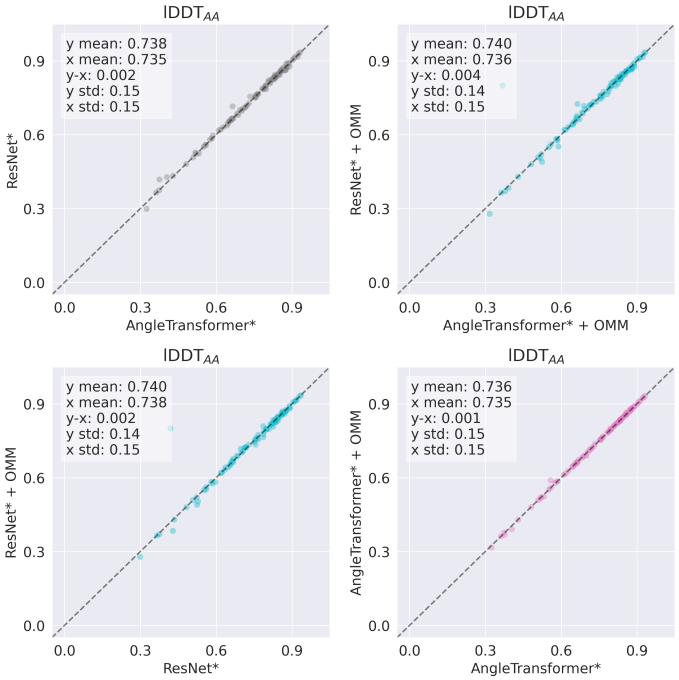

Figure S5: Comparing lDDT$_{AA}$ after finetuning full AF2 models across two variables: AngleTransformer vs ResNet, and OpenMM-Loss (OMM) vs standard AF2 loss. Higher lDDT$_{AA}$ is better.

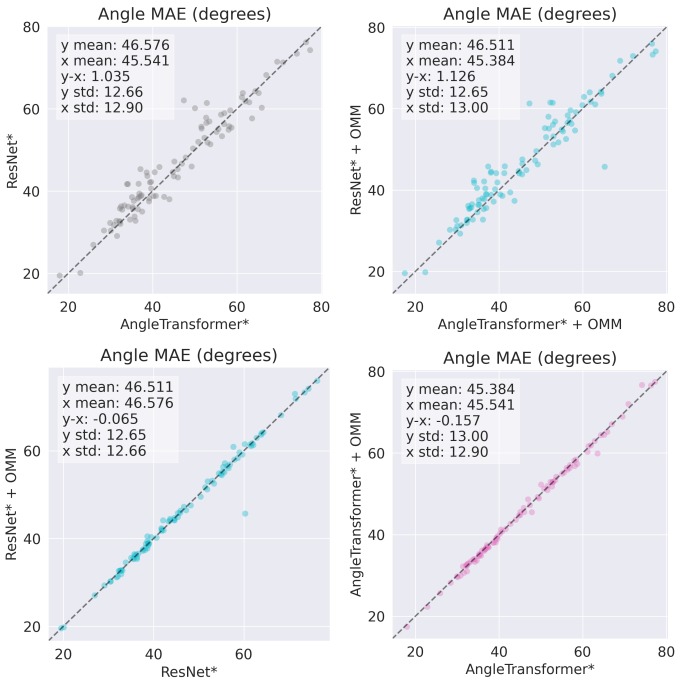

Figure S6: Comparing angle MAE after finetuning full AF2 models across two variables: AngleTransformer vs ResNet, and OpenMM-Loss (OMM) vs standard AF2 loss. For angle MAE, lower is better.

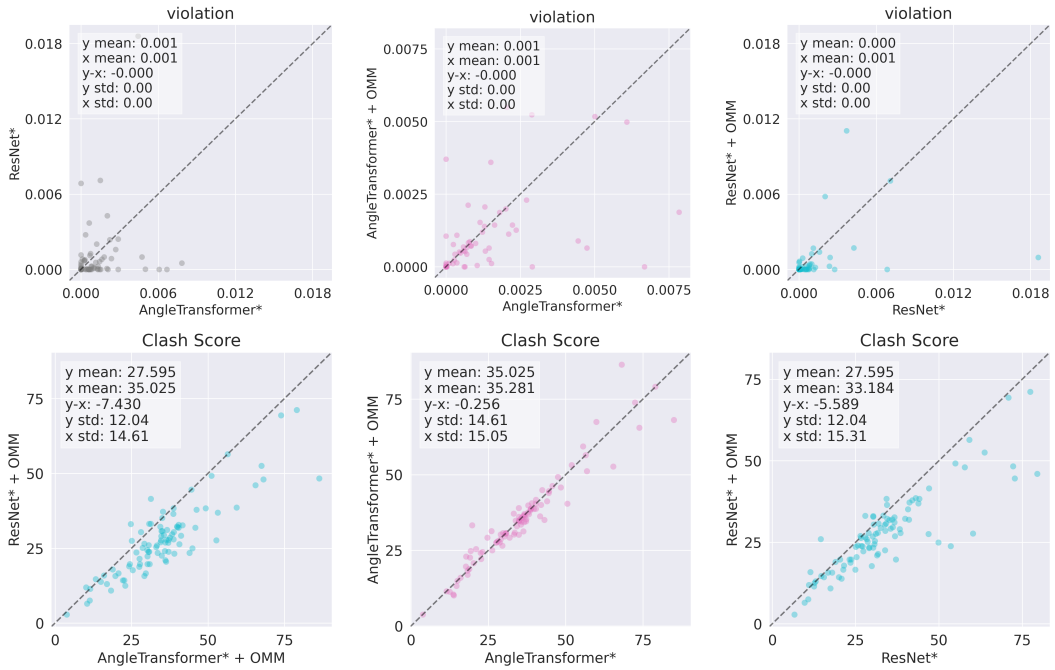

Figure S7: Comparing structural quality metrics after finetuning full AF2 models across two variables: AngleTransformer vs ResNet, and OpenMM-Loss (OMM) vs standard AF2 loss. For violation and clash score values, lower is better.