A-Prot: Protein structure modeling using MSA transformer

Yiyu Hong¹, Juyong Lee^{1,2} and Junsu Ko¹

1. Arontier co., Seoul, Republic of Korea

2. Department of Chemistry, Division of Chemistry and Biochemistry, Kangwon National University, Chuncheon, Republic of Korea

Abstract

In this study, we propose a new protein 3D structure modeling method, A-Prot, using MSA Transformer, one of the state-of-the-art protein language models. For a given MSA, an MSA feature tensor and row attention maps are extracted and converted into 2D residue-residue distance and dihedral angle predictions. We demonstrated that A-Prot predicts long-range contacts better than the existing methods. Additionally, we modeled the 3D structures of the free modeling and hard template-based modeling targets of CASP14. The assessment shows that the A-Prot models are more accurate than most top server groups of CASP14. These results imply that A-Prot captures evolutionary and structural information of proteins accurately with relatively low computational cost. Thus, A-Prot can provide a clue for the development of other protein property prediction methods.

Introduction

Modeling the 3D structure of a protein from its sequence has been one of the most critical problems in biophysics and biochemistry (Kwon et al., 2021; Pereira et al., 2021). The knowledge of the 3D structure of a protein facilitates the discovery of novel ligands, function annotation, and protein engineering. Due to its importance, the community of computational protein scientists have been developing various prediction methods and assessed their performance in the large-scale blind tests, CASPs, which have continued for over two decades (Pereira et al., 2021). In CASP14, Deepmind demonstrated that their deep learning-based model, AlphaFold2 (AF2), predicts the 3D structures of proteins from their sequences with extremely high accuracy, comparable to experimental accuracy (Jumper et al., 2021). The source code of AF2 became publicly available recently, and many model structures of genes of biologically important organisms have been released (Jumper et al., 2021).

The success of AF2 can be attributed to the accurate extraction of coevolutionary information from multiple sequence alignments (MSA). The idea of using coevolutionary information for contact prediction or structure modeling has been widely used (de Juan et al., 2013; Hopf et al., 2014; Jones et al., 2015; Lunt et al., 2010; Marks et al., 2012a, 2012b; Ovchinnikov et al., 2014; Seemayer et al., 2014; Tillier & Charlebois, 2009; Weigt et al., 2009). However, previous attempts were not successful enough because coevolutionary signals in MSAs were not strong enough or too noisy. Various statistical mechanics-based models were proposed, but their accuracy, discriminating true contacts from false contacts, was limited. In CASP13, AlphaFold and RaptorX used a neural network to extract coevolutionary information from MSAs (Senior et al., 2020; Xu, 2019; Xu & Wang, 2019). They achieved significant improvements than

previous CASPs, but their predictions were not accurate enough to obtain model structures comparable to experiments. Finally, in CASP14, with the help of the attention algorithm, truly accurate extraction of actual residue-residue contact signals from MSAs becomes available (Jumper et al., 2021).

Despite this remarkable achievement of AF2, certain limitations remain for this technology to be widely accessible and used for further development. First, the source code for training AF2 is not open-sourced yet. In the first release of AF2, only the production, model generation, part of AF2, and the model parameters are open. Second, training the AF2 architecture is computationally expensive. It is reported that Deepmind used 128 TPUv3 cores for approximately one week and four more days to train AF2. This amount of computational resources are not readily accessible for most academic groups. Thus, developing lighter and general models is still necessary.

To extract evolutionary information from MSAs, various protein language models have been proposed (Bepler & Berger, 2021; Madani et al., 2020; Rao, Liu, et al., 2021; Rao, Meier, et al., 2021; Rives et al., 2021; Strodthoff et al., 2020; Vig et al., 2020). Rao et al. proposed the MSA transformer model, an unsupervised protein language model using the MSA of a query sequence instead of a single query sequence. The model uses row and column attention of input MSAs and masked language modeling objectives. It is demonstrated that the model successfully predicted long-range contacts between residues. In addition, the model predicted other properties of proteins, such as secondary structure prediction and mutational effects, with high accuracy (Rives et al., 2021). These results indicate that the MSA Transformer model extracts the characteristics of proteins from their MSA profiles efficiently.

This study developed a new protein 3D structure prediction method, A-Prot, using MSA Transformer (Rao, Liu, et al., 2021). For a given MSA, we extracted evolutionary information with MSA Transformer. The extracted row attention map and input features were converted to a 2D residue-residue distance map and dihedral angle predictions. We benchmarked the 3D protein structure modeling performance using the FM/TBM-hard targets of CASP13 and 14 (Kinch et al., 2019, 2021). The results show that A-Prot outperforms most top server groups of CASP13 and 14 in terms of long-range contact predictions and 3D protein structure modeling.

Methods

Overview

The overall pipeline of the proposed protein 3D structure prediction method is shown in Figure. 1. We mainly combine the works of the MSA Transformer (Rao, Liu, et al., 2021) and the trRosetta (Yang et al., 2020). Given an MSA, it will input to the MSA Transformer to output MSA features and row attention maps. After a series of transformations like dimension reduction and concatenation etc., these two kinds of features will be transformed and combined into 2D feature maps, which is suitable for input to the trRosetta to finally output protein 3D structure.

The MSA Transformer used in this paper was an already pre-trained version that was learned from 26 million MSAs. It plays a role as a feature extractor that inputs an MSA and outputs its relative features.

trRosetta consists of a deep neural network part and protein structure modeling part. The deep neural network is modified to receive the MSA features from the MSA Transformer instead of the original manually engineered features generated by the statistic approach. Nevertheless, the protein structure modeling part remains the same.

Dataset and MSA Generation

We used the procedure with MSA transformer to generate the MSAs of target sequences (Rao, Liu, et al., 2021). The MSA of a query sequence was generated using HHblits ver. 3.3.0 together with the uncluster30_2017_10 (Mirdita et al., 2017) and BFD (Steinegger et al., 2019) databases. If the number of detected homologous sequences exceeds 256, up to 256 sequences were selected through diversity minimization (Rao, Liu, et al., 2021).

To perform contact predictions, a customized non-redundant protein structure database and we call it PDB30. The PDB30 dataset consists of protein structures deposited in PDB before Apr-30-2018 whose resolution is higher than 2.5 Å and sequence length is longer than 40 amino acids. Using 102,300 sequences that satisfy the condition, clustering analysis was performed using MMSeq2 (Steinegger & Söding, 2017) with a threshold of sequence identity of 30%, leading to 16,612 non-redundant sequences.

Network Architecture

Let us define an input MSA as an $r \times c$ character matrix, where r and c are rows (number of sequences) and columns (sequence length) in the MSA, respectively. Through the token and position embedding of the MSA Transformer, the input MSA is embedded into a $r \times c \times 768$ tensor which is the input and output of each attention block. The attention block is composed in the order of row attention layer, column attention layer, and feed-forward layer. A layer normalization operation is followed by each layer. And each attention layer has 12 attention heads. The MSA Transformer is a stack of 12 such attention blocks.

Two kinds of features were extracted from the MSA Transformer to construct 2D feature maps by a series of transformations. (1) One is the last attention block's output, a $r \times c \times 768$ tensor; we named it MSA features (Figure. 1). Only features corresponding to the query sequence are selected, which is a $1 \times c \times 768$ tensor. Then, the dimension of this feature is reduced to 128 by an MLP (multi-layer perceptron) consist of 3 linear layers with neuron sizes 384, 192, 128. The dimension reduced 1D feature is then outer concatenated (redundantly expanding horizontally and vertically and then stacked together) to form a query sequence feature of a $c \times c \times 256$ tensor. (2) The other one is row attention maps that are derived from each attention head of each row attention layer, totally $12 \times 12 = 144$ attention maps stacked to shape in $c \times c \times 144$. Then it is symmetrized by adding it to its transposed tensor to yield symmetrized row attention maps. The query sequence feature and the symmetrized row attention maps are concatenated in the feature map dimension to form a 2D feature map that is a $c \times c \times 400$ tensor.

The dimension of the 2D feature maps afterward reduced to 64 by three convolutional layers that consist of 256, 128, 64 kernels of size 1×1 (Lin et al., 2013), where each convolutional layer is followed by an instance normalization and a ReLU activation. Then this $c \times c \times 64$ tensor is input to the Dilated ResNet consist of 28 residual blocks (He et al., 2016), each having one instance normalization, two ReLU activations, and two convolutional layers each with 64 kernels of size 3×3 . Dilation is applied to the convolutional layers cycle through the residual blocks with rates of 1, 2, 4. After the last residual block, there are four independent convolutional layers with 25, 13, 25, 37 kernels of size 1×1 , each for predicting inter-residue geometries of θ , ϕ , ω , d , respectively. The values for the inter-residue geometries are all discretized into bins; each bin is treated as a classification label. Please refer to the reference (Yang et al., 2020) for more details about the inter-residue geometries as we used the same settings. Finally, the trRosetta Protein Structure Modeling module will predict and modeling the protein 3D structure based on the inter-residue geometries information.

Training and Inference

At the training stage, we fixed the parameters in the MSA Transformer. In contrast, parameters in the other deep neural networks were trained with a batch size of 16 with gradient accumulation steps, a learning rate of 1e-

A-Prot: protein structure modeling with MSA Transformer

Figure 1. Pipeline of the proposed protein 3D structure prediction method. Input an MSA to the MSA Transformer to extract MSA Features and row attention maps. Then, the MSA features corresponding to the query sequence and the row attention maps are combined to a 2D feature maps by a set of transformations. Next, the 2D feature maps are input to a Dilated ResNet after dimension reduction to output inter-residue geometries, which further input to the trRosetta Protein Structure Modeling to output a predicted protein 3D structure.

3, using the RAdam optimizer (Liu et al., 2020), the cross-entropy loss was calculated with equal weight for the four inter-residue geometry

objectives. The total model was trained end-to-end on an NVIDIA Quadro RTX 8000 GPU (48GB) for around 40 epochs which took about five days.

A-Prot: protein structure modeling with MSA Transformer

Table 2. The average TM-score and IDDT of the model structures of 25 CASP14 FM, FM/TBM, and TBM-hard domains

Server Group	TM-score	IDDT
FEIG-S	0.461	0.413
BAKER-ROSETTASERVER	0.517	0.455
Zhang-Server	0.595	0.489
QUARK	0.588	0.484
A-Prot	0.576	0.499

A head-to-head comparison with ROSETTASERVER

Because A-Prot uses trRosetta for modeling structure at the final stage, we performed a head-to-head comparison of A-Prot models with the ROSETTASERVER models to identify improvement in residue-residue distance predictions (Figure 2). In terms of TM-score, many predictions made by A-Prot are significantly better than BAKER-ROSETTASERVER. For instance, the model qualities of five targets that were predicted to have TM-score less than 0.4 by ROSETTASERVER were improved higher than 0.4, corresponding to a correct fold prediction. Four highly accurate models are depicted in Figure 3.

Similarly, in terms of the IDDT measure, prediction results of eight targets were improved significantly. On the contrary, only two targets deteriorated more than 0.05. Therefore, A-Prot predicts residue-residue distances and dihedral angles better than most server groups participated in CASP14 (Pereira et al., 2021).

On the other hand, nearly all worse predictions by A-Prot deteriorated only by a small margin except the T1026-D1 target. For T1026-D1, A-Prot did not predict its correct fold. This incorrect prediction is attributed to an incomplete MSA. T1026-D1 is a protein consisting of a virus capsid (PDB ID: 6S44). Our sequence search procedure found only less than 30 sequences, which appear to be not enough to extract correct evolutionary information. The failure of T1026-D1 suggests that having enough homologous sequences is critical in accurate 3D structure modeling using MSA-Transformer. In other words, a more extensive sequence search may improve the model accuracy of A-Pro. Except for T1026-D1, the deviations of all worse predictions than ROSETTASERVER were less than 0.1 TM-score, much smaller than the improvements.

Conclusion

In this study, we introduced a new protein structure prediction method, A-Prot, using MSA Transformer. Our benchmark results on the CASP13 TBM/FM and FM targets show that A-Prot predicts long-range residue-

Figure 3. Model comparison of four high-quality CASP14 models generated from our method versus their native structures. Brown: native structure; Blue: model.

residue contacts more accurately than the existing methods. We also assessed the quality of protein structure models based on the predicted residue-residue distance information. The model generated by A-Prot is more accurate than most of the server groups that participated in CASP14. The average IDDT of A-Prot models is higher than that of all server group models. In terms of TM-score, our model is slightly worse than QUARK and Zhang-Server. These results show that our approach yields highly accurate residue-residue distance predictions.

A-Prot requires less computational resources than the other state-of-the-art protein structure prediction models (Pereira et al., 2021). The source code of AlphaFold2 is only partially open (Jumper et al., 2021). Its model parameters are fixed, and only the structure modeling part is open. Thus, it is hard to tune AlphaFold2 for bespoke purposes. In addition, training the AlphaFold2 architecture requires a significant amount of computational resources, which is not assessable for most academic groups. On the other hand, A-Prot can be trained with a single GPU card. In summary, A-Prot will open new possibilities for training novel deep-learning-based models to predict various properties of proteins only using sequence information.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2018R1C1B600543513).

References

- Anishchenko, I., Baek, M., Park, H., Hiranuma, N., Kim, D. E., Dauparas, J., Mansoor, S., Humphreys, I. R., & Baker, D. (2021). Protein tertiary structure prediction and refinement using deep learning and Rosetta in CASP14. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.26194>
- Bepler, T., & Berger, B. (2021). Learning the protein language: Evolution, structure, and function. *Cell Systems*, 12(6), 654–669.e3. <https://doi.org/10.1016/j.cels.2021.05.017>
- de Juan, D., Pazos, F., & Valencia, A. (2013). Emerging methods in protein co-evolution. In *Nature Reviews Genetics* (Vol. 14, Issue 4, pp. 249–261). <https://doi.org/10.1038/nrg3414>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. In J. and S. N. and W. M. Leibe Bastian and Matas (Ed.), *Computer Vision – ECCV 2016* (pp. 630–645). Springer International Publishing.
- Heo, L., Janson, G., & Feig, M. (2021). Physics-based protein structure refinement in the era of artificial intelligence. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.26161>
- Hopf, T. A., Schärfe, C. P. I., Rodrigues, J. P. G. L. M., Green, A. G., Kohlbacher, O., Sander, C., Bonvin, A. M. J. J., & Marks, D. S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *ELife*, 3. <https://doi.org/10.7554/eLife.03430>
- Jones, D. T., Singh, T., Kosciółek, T., & Tetchner, S. (2015). MetaPSICOV: Combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics*, 31(7), 999–1006. <https://doi.org/10.1093/bioinformatics/btu791>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. <https://doi.org/10.1038/s41586-021-03819-2>
- Kinch, L. N., Kryshchak, A., Monastyrskyy, B., & Grishin, N. v. (2019). CASP13 target classification into tertiary structure prediction categories. *Proteins: Structure, Function and Bioinformatics*, 87(12), 1021–1036. <https://doi.org/10.1002/prot.25775>
- Kinch, L. N., Schaeffer, R. D., Kryshchak, A., & Grishin, N. v. (2021). Target classification in the 14th round of the critical assessment of protein structure prediction (CASP14). *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.26202>
- Kwon, S., Won, J., Kryshchak, A., & Seok, C. (2021). Assessment of protein model structure accuracy estimation in CASP14: Old and new challenges. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.26192>
- Lin, M., Chen, Q., & Yan, S. (2013). Network In Network. *Arxiv*, 1312.4400. <http://arxiv.org/abs/1312.4400>
- Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2020, April). On the Variance of the Adaptive Learning Rate and Beyond. *Eighth International Conference on Learning Representations (ICLR)*. <https://www.microsoft.com/en-us/research/publication/on-the-variance-of-the-adaptive-learning-rate-and-beyond/>
- Lunt, B., Szurmant, H., Procaccini, A., Hoch, J. A., Hwa, T., & Weigt, M. (2010). Inference of Direct Residue Contacts in Two-Component Signaling. In *Methods in Enzymology* (Vol. 471, Issue 10, pp. 17–41). [https://doi.org/10.1016/S0076-6879\(10\)71002-8](https://doi.org/10.1016/S0076-6879(10)71002-8)
- Madani, A., McCann, B., Naik, N., Keskar, N. S., Anand, N., Eguchi, R. R., Huang, P.-S., & Socher, R. (2020). *ProGen: Language Modeling for Protein Generation*. <https://doi.org/10.1101/2020.03.07.982272>
- Marks, D. S., Hopf, T. A., & Sander, C. (2012a). Protein structure prediction from sequence variation. *Nature Publishing Group*. <https://doi.org/10.1038/2419>
- Marks, D. S., Hopf, T. A., & Sander, C. (2012b). Protein structure prediction from sequence variation. *Nature Publishing Group*. <https://doi.org/10.1038/2419>
- Mirdita, M., von den Driesch, L., Galiez, C., Martin, M. J., Söding, J., & Steinegger, M. (2017). Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research*, 45(D1), D170–D176. <https://doi.org/10.1093/nar/gkw1081>
- Ovchinnikov, S., Kamisetty, H., & Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *ELife*, 2014, 1–21. <https://doi.org/10.7554/eLife.02030>
- Pereira, J., Simpkin, A. J., Hartmann, M. D., Rigden, D. J., Keegan, R. M., & Lupas, A. N. (2021). High-accuracy protein structure prediction in CASP14. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.26171>

A-Prot: protein structure modeling with MSA Transformer

- Rao, R., Liu, J., Verkuil, R., Meier, J., Canny, J. F., Abbeel, P., Sercu, T., & Rives, A. (2021). *MSA Transformer*. <https://doi.org/10.1101/2021.02.12.430858>
- Rao, R., Meier, J., Sercu, T., Ovchinnikov, S., & Rives, A. (2021). TRANSFORMER PROTEIN LANGUAGE MODELS ARE UNSUPERVISED STRUCTURE LEARNERS. *ICLR 2021 Conference*. <https://github.com/facebookresearch/esm>.
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Lawrence Zitnick, C., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15), e2016239118. <https://doi.org/10.1073/pnas.2016239118/-/DCSupplemental>
- Seemayer, S., Gruber, M., & Söding, J. (2014). CCMpred - Fast and precise prediction of protein residue-residue contacts from correlated mutations. *Bioinformatics*, 30(21), 3128–3130. <https://doi.org/10.1093/bioinformatics/btu500>
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A. W. R., Bridgland, A., Penadones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>
- Steinegger, M., Mirdita, M., & Söding, J. (2019). Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nature Methods*, 16(7), 603–606. <https://doi.org/10.1038/s41592-019-0437-4>
- Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028. <https://doi.org/10.1038/nbt.3988>
- Strodthoff, N., Wagner, P., Wenzel, M., & Samek, W. (2020). UDSMProt: Universal deep sequence models for protein classification. *Bioinformatics*, 36(8), 2401–2409. <https://doi.org/10.1093/bioinformatics/btaa003>
- Tillier, E. R. M., & Charlebois, R. L. (2009). The human protein coevolution network. *Genome Research*, 19(10), 1861–1871. <https://doi.org/10.1101/gr.092452.109>
- Vig, J., Madani, A., Varshney, L. R., Xiong, C., Socher, R., & Rajani, N. F. (2020). *BERTology Meets Biology: Interpreting Attention in Protein Language Models*. <http://arxiv.org/abs/2006.15222>
- Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., & Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1), 67–72. <https://doi.org/10.1073/pnas.0805923106>
- Wu, T., Guo, Z., Hou, J., & Cheng, J. (2021). DeepDist: real-value inter-residue distance prediction with deep residual convolutional network. *BMC Bioinformatics*, 22(1). <https://doi.org/10.1186/s12859-021-03960-9>
- Xu, J. (2019). Distance-based protein folding powered by deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, 116(34), 16856–16865. <https://doi.org/10.1073/pnas.1821309116>
- Xu, J., & Wang, S. (2019). Analysis of distance-based protein structure prediction by deep learning in CASP13. *Proteins: Structure, Function and Bioinformatics*, 87(12), 1069–1081. <https://doi.org/10.1002/prot.25810>
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, 117(3). <https://doi.org/10.1073/pnas.1914677117>
- Zheng, W., Li, Y., Zhang, C., Zhou, X., Pearce, R., Bell, E. W., Huang, X., & Zhang, Y. (2021). Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14. *Proteins: Structure, Function and Bioinformatics*. <https://doi.org/10.1002/prot.26193>