Web Server


# iQDeep: an integrated web server for protein scoring using multiscale deep learning models

**Md Hossain Shuvo [1], Mohimenul Karim [1] and Debswapna Bhattacharya [1*]**

*1 - **Department of Computer Science,** Virginia Tech, Blacksburg, VA 24061, United States*

***Correspondence to Debswapna Bhattacharya:*** *dbhattacharya@vt.edu (D. Bhattacharya)@mzs0149 (M.H. Shuvo), @CaptainRafi97 (M. Karim), @imDBhattacharya (D. Bhattacharya)*
https://doi.org/10.1016/j.jmb.2023.168057
***Edited by Rita Casadio***

## Abstract

The remarkable recent advances in protein structure prediction have enabled computational modeling of protein structures with considerably higher accuracy than ever before. While state-of-the-art structure prediction methods provide self-assessment confidence scores of their own predictions, an independent and open-access system for protein scoring is still needed that can be applied to a broad range of predictive modeling scenarios. Here, we present iQDeep, an integrated and highly customizable web server for protein scoring, freely available at http://fusion.cs.vt.edu/iQDeep. The underlying method of iQDeep employs multiscale deep residual neural networks (ResNets) to perform residue-level error classifications, and then probabilistically combines the error classifications for protein scoring. By adjusting the error resolutions, our method can reliably estimate the standard- or high-accuracy variants of the Global Distance Test metric for versatile protein scoring. The performance of the method has been extensively tested and compared against the state-of-the-art approaches in multiple rounds of Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments including benchmark assessment in CASP12 and CASP13 as well as blind evaluation in CASP14. The iQDeep web server offers a number of convenient features, including (i) the choice of individual and batch processing modes; (ii) an interactive and privacy-preserving web interface for automated job submission, tracking, and results retrieval; (iii) web-based quantitative and visual analyses of the results including overall estimated score and its residue-wise breakdown along with agreements between various sequence- and structural-level features; (iv) extensive help information on job submission and results interpretation via web-based tutorial and help tooltips.



## Introduction

Protein scoring is a critical component of protein structure prediction.[1–5] Protein scoring has been gaining noticeable attention in the Critical Assessment of Protein Structure Prediction (CASP) experiments under the accuracy estimation category.[6–9] Promising progress has been made in the CASP14 accuracy estimation category[7] with various deep learning-based methods performing well. Among them is our previously published method QDeep,[2] introducing several new advances for the first time including the incorporation of the deep residual neural networks (ResNets) architectures for protein scoring, effective integration of predicted inter-residue interaction with other sequential and structural features, and the use of ensemble learning. With rapid new developments in the field of protein structure prediction,[10–12] however, the scoring resolution used in our original QDeep method no longer represents the state of the art. Recent advances in deep learning-based protein structure prediction methods such as AlphaFold2[10] have enabled computational prediction of protein with considerably higher accuracy than what has been achieved before. As such, the development of high-resolution protein scoring methods commensurate with the increasing accuracy of structure prediction methods is of critical importance.

While our original QDeep method uses the Global Distance Test Total Score (GDT-TS)[13] as the ground truth metric, the recent advances in structure prediction[11–12,14–16] and refinement[17–21] make the high-accuracy variant of the Global Distance Test (GDT-HA)[22] more suitable as the ground truth metric. State-of-the-art methods such as AlphaFold2[10] and RoseTTAFold[11] attain very high accuracy in terms of GDT-TS, but the GDT-HA metric reveals subtle performance differences between the methods (**Supplementary Figure S1**) due to the sensitivity of GDT-HA to minor structural deviations. As such, a high-resolution scoring function that can estimate the GDT-HA metric with high fidelity can capture minute structural differences in structural models predicted from the state-of-the-art protein structure prediction methods, enabling improved model selection and ranking (**Supplementary Text S2**). Furthermore, an integrated protein scoring framework that can alternate between the standard and high accuracy variants of the Global Distance Test on-demand in order to control the scoring resolution, can enhance the versatility of protein scoring by covering a broad range of predictive modeling scenarios. Open availability of such a versatile method via a publicly accessible web server has the potential for broad dissemination and a field-wide impact.

Here we present iQDeep, an integrated and fully configurable web server for protein scoring using multiscale deep learning models. iQDeep employs multiscale deep residual neural networks (ResNets) to perform residue-level error classifications at multiple predefined error resolutions, and then probabilistically combines the predictions from the multiscale error classifiers for protein scoring. Building on the original QDeep method, we train a new set of ResNet classifiers to perform residue-level ensemble error classifications at finer-grained error resolutions explicitly targeting the GDT-HA metric. The residue-level error classifications from the multiscale ResNets can then be probabilistically combined for quantitating the accuracy of a predicted protein model. By adjusting the resolutions and probabilistic combination of the multiscale ResNets, our method can reliably estimate the standard- or high-accuracy variants of the Global Distance Test metric for protein scoring. The interactive and privacy-preserving web interface of iQDeep allows customizable job submission, tracking, and results retrieval with quantitative and visual analysis along with extensive help information on job processing and results interpretation. The performance of the underlying methods has been rigorously tested and compared against the state-of-the-art approaches in multiple rounds of CASP experiments including benchmark assessment in CASP12 and CASP13 and blind evaluation in CASP14. The iQDeep web server is freely available at https://fusion.cs.vt.edu/iQDeep.

## Results and discussion

Our previously published method QDeep[2] performed quite well on CASP12 and CASP13 benchmarking datasets. Additionally, it has been tested in a strict blind mode in CASP14 under the group name "Bhattacharya-QDeep". We use the CASP14 dataset to evaluate the performance of iQDeep and compare it against 25 groups participating in the CASP14 accuracy estimation category, including our own group "Bhattacharya-QDeep". The CASP14 benchmarking dataset consists of 10,494 models for 70 targets submitted by the CASP14 tertiary structure predictors. Given the progress made in protein structure prediction, we use the high-accuracy variant of the Global Distance Test (GDT-HA) as the ground truth metric for evaluation. For performance assessment, we use twofold evaluation criteria: (1) the ability to reproduce the ground truth scores, and (2) the ability to distinguish acceptable from incorrect models.

### Reproducing ground truth scores

To evaluate the ability to reproduce the ground truth scores, we calculate the global Pearson correlation (Pearson r) and the average absolute difference ($\Delta$GDT-HA) between the estimated scores and the ground truth GDT-HA scores for all models in the CASP model pool. Meanwhile, higher Pearson r and lower $\Delta$GDT-HA indicate an enhanced ability to reproduce the ground truth scores. Figure 1 shows the performance of iQDeep and the groups participating in the CASP14 accuracy estimation category, including our own group "Bhattacharya-QDeep" employing the original QDeep method. The results demonstrate that iQDeep achieves state-of-the-art performance. For example, iQDeep attains a Pearson r of 0.646, outperforming "Bhattacharya-QDeep", and better than most of the CASP14 predictors except "BAKER-ROSETTASERVER", "BAKER-experimental", "3DCNN_prof", and "RaptorX-QA" groups.[17,23–24] Remarkably, iQDeep attains the lowest $\Delta$GDT-HA of 0.119, which is significantly better than second best performer

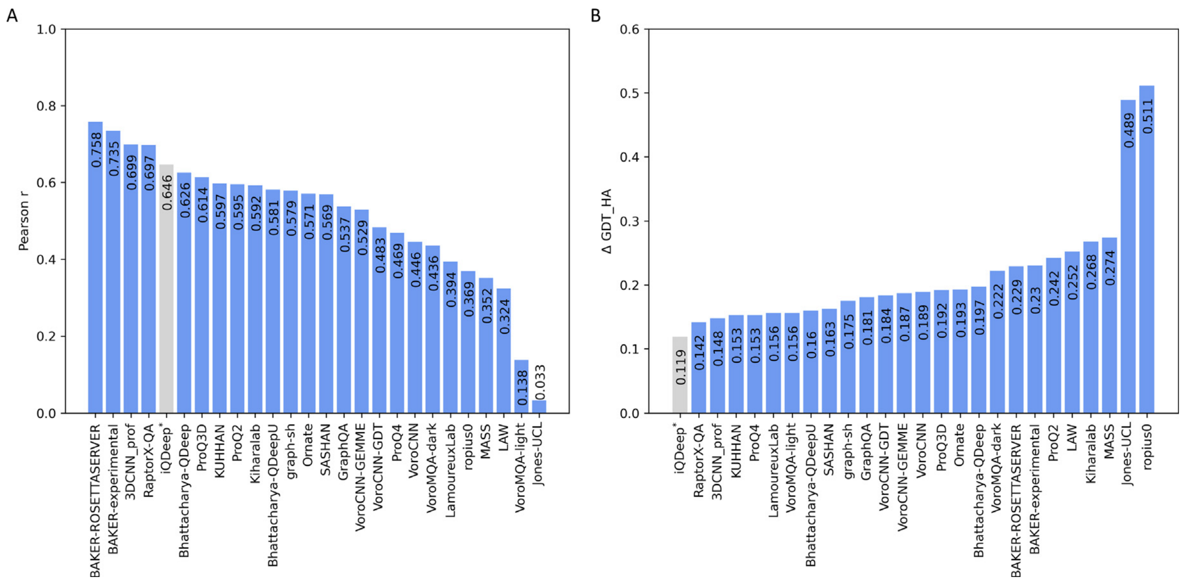


**Figure 1.** Reproducibility of ground truth scores for iQDeep and 25 groups participating in the CASP14 accuracy estimation category in terms of (A) global Pearson correlation coefficient and (B) average absolute difference between the estimated scores and the ground truth GDT-HA scores in the CASP14 dataset.* iQDeep is not a participating group in CASP14.

"RaptorX-QA", let alone our previously published method QDeep implemented in "Bhattacharya-QDeep". It is interesting to note that "BAKER-ROSETTASERVER" and "BAKER-experimental", despite attaining high Pearson r, underperform in terms of ΔGDT-HA. In contrast, iQDeep and two CASP14 groups "RaptorX-QA" and "3DCNN_prof" deliver consistent performance. In summary, iQDeep exhibits an excellent all-round ability to reproduce ground truth scores.

### Distinguishing acceptable from incorrect models

In addition to reproducing the ground truth scores with high fidelity, the ability to discriminate between acceptable and non-acceptable prediction is critically important. We measure the separation between acceptable (GDT-HA $\geq$ 0.3[25]) and incorrect models using a receiver operating characteristic (ROC) curve. As shown in Figure 2, the Area under the ROC Curve (AUROC) of iQDeep and our CASP predictor "Bhattacharya-QDeep" employing the original QDeep method achieved AUROC of 0.826 and 0.812, respectively. Once again, iQDeep outperforms our previously published method QDeep. The AUROC of iQDeep is only slightly lower than the top performing CASP predictor "BAKER-ROSETTASERVER", and better than multiple well-known CASP groups including "ProQ2",[4,26] "ProQ3",[6] "ProQ3D", and "VoroMQA".[1] Overall, iQDeep can reliably discriminate between acceptable and incorrect predictions.

## Materials and methods

### Overview of iQDeep pipeline

Figure 3 shows the flowchart of iQDeep webserver consisting of front-end and back-end modules. The front-end module performs the input validation, allows users to select customizable job parameters, monitors the job processing, and enables results retrieval with quantitative and visual analysis. The back-end module processes the job, manages the job queue, and ensures web storage management and retrieval of job-related metadata. Users are able to protect the privacy of their jobs by selecting a private mode of job processing and results retrieval.

***Front-end module.*** The front-end of the web server provides an interface for job submission, performs input validation, and returns the results back to the user. While the default job submission mode requires minimum user input that includes a valid protein structure and a job name, users can perform on-demand customization of the job parameters including selecting the scoring mode, scoring resolution, and job privacy. Users may select "single" or "batch" scoring mode for processing just a single or a pool of models, respectively. The scoring resolution parameter provides users the choice of selecting a specific set of multiscale deep learning models trained at error resolutions targeting the GDT-TS metric as implemented in our original QDeep work (scoring resolution "standard"), or the newly introduced

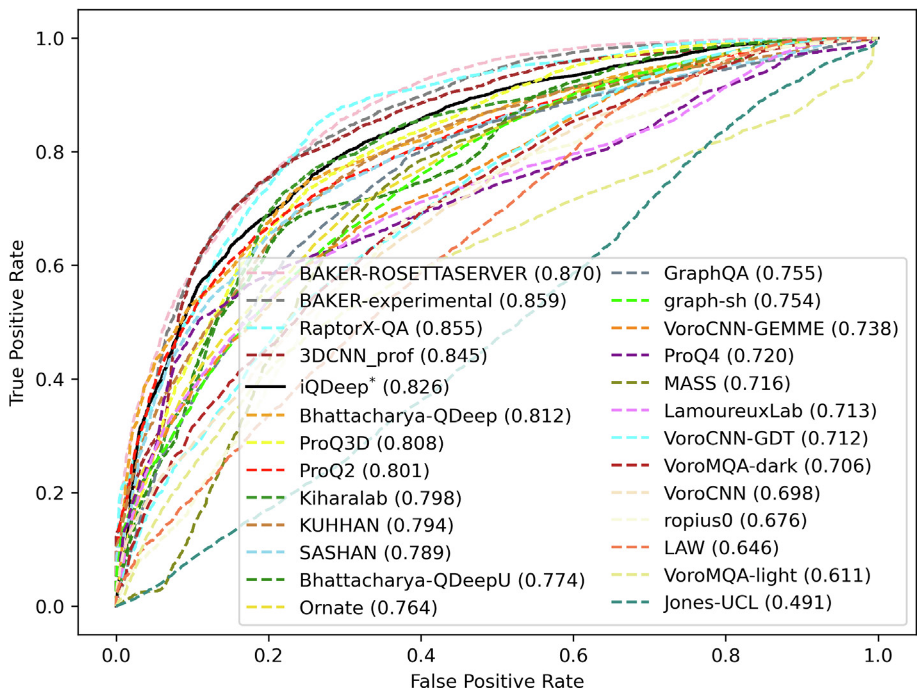


**Figure 2.** Distinguishability of acceptable vs. incorrect models for iQDeep and 25 groups participating in the CASP14 accuracy estimation category using a receiver operating characteristic (ROC) curve. The numbers reported are the Area under the ROC Curve (AUROC) values.* iQDeep is not a participating group in CASP14.

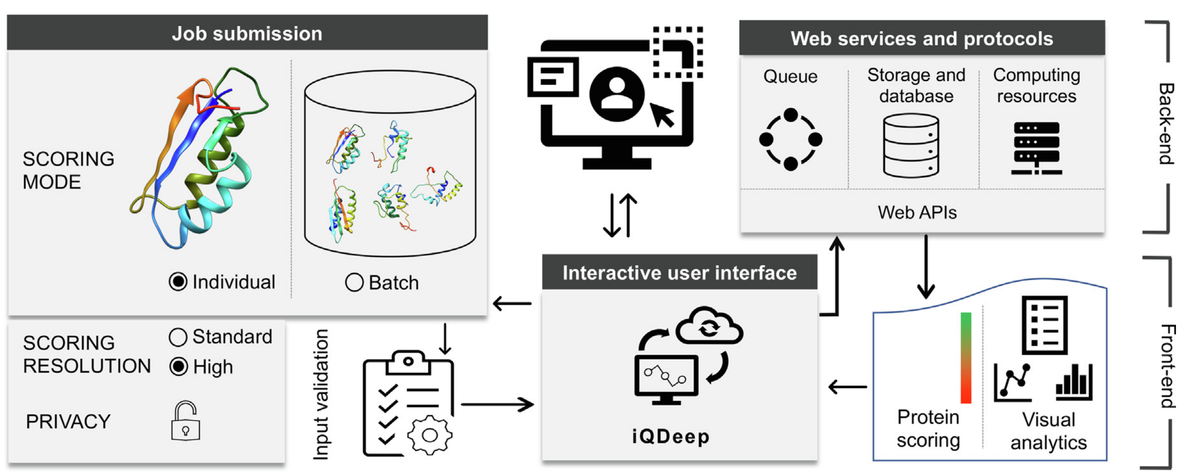


**Figure 3.** The flowchart of iQDeep web server for protein scoring consisting of the front-end module for automated job submission with input validation and interactive results analytics and the back-end module for processing and managing jobs.

finer-grained error resolutions explicitly targeting the GDT-HA metric (scoring resolution “high”). The privacy parameter allows the users to configure the privacy of their job. The input validation ensures the correctness of all the inputs to ensure seamless job execution. The front-end module also offers a web-based analysis of the scoring results through easily interpretable and interactive web-based plots and three-dimensional molecular visualizations. An optional email address can also be provided for automated status updates of the job via email.

***Back-end module.*** The back-end module of iQDeep performs job management using the queuing layer, controls the database and web storages, processes protein scoring jobs in our in-house compute cluster, and maintains client–

server communication. The back-end module actively communicates with the front-end module to accept validated jobs and return the results upon successful job completion. After a job is successfully submitted, the queuing layer of the back-end module starts preparing the job for processing using a first-in-first-out (FIFO) job scheduler. During the job processing, the queuing layer first stores the job-related information and parameters in the web storage and the corresponding job metadata in the database. It then sends the job to the computational node for processing based on the selected job parameters, while continuously monitoring the computational resource availability in our in-house compute cluster. The back-end module maintains the client–server protocol by establishing a two-way communication with the front-end module during the entire job processing phase to periodically send the job updates to the front-end. Upon completion of the job, the back-end module sends the results back to the front-end for interactive and interpretable web-based visual and quantitative analytics, with an optional email notification sent to the user, if an email address is provided.

### The architecture of multiscale deep learning models for protein scoring

iQDeep employs multiscale deep residual neural networks (ResNets) to perform multi-resolution protein scoring. The standard scoring resolution represents our original distance-based protein scoring method QDeep.[2] It employs an ensemble of four deep ResNets, each independently trained to classify the residue level errors at four different error thresholds of 1, 2, 4, and 8Å to target the GDT-TS[13] metric. The method leverages several distance-based alignment features, sequence profile information, consistency between the predicted and observed structural properties, and several biophysical energy terms, as described in **Supplementary Text S1** and in the published article of QDeep.[2] The residue-level ensemble error estimates are then combined to estimate an accuracy score of a protein model, thereby estimating a probabilistic equivalent of the GDT-TS metric.

Building on our prior work, in iQDeep we train a new set of ResNet classifiers having an identical architecture, as described in **Supplementary Text S1**, to perform residue-level ensemble error classifications at finer-grained error resolutions of 0.5, 1, 2, and 4Å to explicitly target the GDT-HA metric. The predictions from the finer-grained error classifiers can then be probabilistically combined for high-resolution protein scoring, as follows,

$$\boldsymbol{iQDeep\ score} = \frac{\boldsymbol{N}0.5 + \boldsymbol{N}1 + \boldsymbol{N}2 + \boldsymbol{N}4}{4}$$

where $\boldsymbol{N}0.5$, $\boldsymbol{N}1$, $\boldsymbol{N}2$, and $\boldsymbol{N}4$ represent the fraction of residues estimated to be within 0.5, 1, 2, and 4Å from the corresponding residues in the experimental structure. The estimated score ranges between 0 and 1 with a higher score indicating better model accuracy.

### Web server

***Hardware and software.*** The iQDeep web server runs on a Linux x86_64 cluster with 2.50GHz Intel Xeon Silver 32-core processors. The underlying client–server architecture of iQDeep is implemented using server-side PHP and client-side JavaScript scripting languages and deployed through an Apache web server with Common Gateway Interface (CGI). iQDeep uses MySQL relational database management system (RDBMS) for storing, retrieving, and updating job-related information. Additionally, it uses the WebGL-based JavaScript library 3Dmol.js[27] for interactive 3D molecular visualization. The iQDeep pipeline for protein scoring is implemented using Python. The webserver is compatible with most modern web browsers including Google Chrome, Mozilla Firefox, Safari, and Microsoft Edge.

***Input and output.*** The iQDeep web server allows users to submit a job using only two required fields including a job name and a valid protein structure for scoring. However, users can customize several job parameters including scoring mode, scoring resolutions, and job privacy. An optional email address can also be provided for automated status updates of the job via email. The web server offers interactive results update using easy-to-interpret quantitative and visual analytics. In addition to providing global estimated accuracy scores, the web server provides the residue-level local accuracy estimation to identify reliable and unreliable regions of an input protein model along with interactive residue-wise visualization of estimated local accuracies. A 3D representation of the protein structure with an option to download the corresponding structure file is also provided. Users can also analyze several structural properties through easily interpretable web-based graphical alignment between the observed and predicted structural properties including secondary structure, solvent accessibility, and residue-residue contact maps at various distance thresholds. The full set of results, including the global and local accuracy scores and text files containing the additional analyses, can be downloaded as a compressed zipped archive.

***CRediT authorship contribution statement.*** **Md Hossain Shuvo:** Methodology, Data curation, Writing – original draft, Software. **Mohimenul Karim:** Software, Validation, Writing – review & editing. **Debswapna Bhattacharya:** Conceptualization, Investigation, Methodology,

Writing – original draft, Writing – review & editing, Supervision, Funding acquisition, Project administration.

## Acknowledgments

This work was partially supported by the National Institute of General Medical Sciences [R35GM138146 to D.B.] and the National Science Foundation [DBI2208679 to D.B.].

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2023.168057.

*Received 14 December 2022;*
*Accepted 17 March 2023;*
*Available online 23 March 2023*

***Keywords***:
protein scoring;
accuracy estimation;
deep learning;
protein structure prediction

## References

1. Won, J., Baek, M., Monastyrskyy, B., Kryshtafovych, A., Seok, C., (2019). Assessment of protein model structure accuracy estimation in CASP13: Challenges in the era of deep learning. *Proteins: Struct., Funct., Bioinformatics* **87**, 1351–1360. https://doi.org/10.1002/prot.25804.
2. Shuvo, M.H., Bhattacharya, S., Bhattacharya, D., (2020). QDeep: distance-based protein model quality estimation by residue-level ensemble error classifications using stacked deep residual neural networks. *Bioinformatics* **36**, i285–i291. https://doi.org/10.1093/bioinformatics/btaa455.
3. Wallner, B., Elofsson, A., (2003). Can correct protein models be identified?. *Protein Sci.* **12**, 1073–1086. https://doi.org/10.1110/ps.0236803.
4. Kryshtafovych, A., Barbato, A., Monastyrskyy, B., Fidelis, K., Schwede, T., Tramontano, A., (2016). Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. *Proteins* **84** (Suppl 1), 349–369. https://doi.org/10.1002/prot.24919.
5. Alapati, R., Bhattacharya, D., (2018). clustQ: Efficient protein decoy clustering using superposition-free weighted internal distance comparisons. *In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. Association for Computing Machinery, New York, NY, USA, pp. 307–314. https://doi.org/10.1145/3233547.3233570.
6. Kryshtafovych, A., Monastyrskyy, B., Fidelis, K., Schwede, T., Tramontano, A., (2018). Assessment of model accuracy estimations in CASP12. *Proteins: Struct., Funct., Bioinformatics* **86**, 345–360. https://doi.org/10.1002/prot.25371.
7. Kwon, S., Won, J., Kryshtafovych, A., Seok, C., (2021). Assessment of protein model structure accuracy estimation in CASP14: Old and new challenges. *Proteins* **89**, 1940–1948. https://doi.org/10.1002/prot.26192.
8. Kryshtafovych, A., Barbato, A., Fidelis, K., Monastyrskyy, B., Schwede, T., Tramontano, A., (2014). Assessment of the assessment: evaluation of the model quality estimates in CASP10. *Proteins* **82** (Suppl 2), 112–126. https://doi.org/10.1002/prot.24347.
9. Elofsson, A., Joo, K., Keasar, C., Lee, J., Maghrabi, A.H.A., Manavalan, B., McGuffin, L.J., Hurtado, D.M., et al., (2018). Methods for estimation of model accuracy in CASP12. *Proteins: Struct., Funct., Bioinformatics* **86**, 361–373. https://doi.org/10.1002/prot.25395.
10. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., et al., (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
11. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., et al., (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876. https://doi.org/10.1126/science.abj8754.
12. G. Ahdritz, N. Bouatta, S. Kadyan, Q. Xia, W. Gerecke, T.J. O'Donnell, D. Berenberg, I. Fisk, et al., 2022. OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization, 2022.11.20.517210. doi: 10.1101/2022.11.20.517210.
13. Zemla, A., (2003). LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31**, 3370–3374. https://doi.org/10.1093/nar/gkg571.
14. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., et al., (2022). Evolutionary-scale prediction of atomic level protein structure with a language model, 2022.07.20.500902. https://doi.org/10.1101/2022.07.20.500902.
15. Wu, R., Ding, F., Wang, R., Shen, R., Zhang, X., Luo, S., Su, C., Wu, Z., et al., (2022). High-resolution de novo structure prediction from primary sequence, 2022.07.21.500999. https://doi.org/10.1101/2022.07.21.500999.
16. Jumper, J., Hassabis, D., (2022). Protein structure predictions to atomic accuracy with AlphaFold. *Nat. Methods* **19**, 11–12. https://doi.org/10.1038/s41592-021-01362-6.
17. Hiranuma, N., Park, H., Anishchanka, I., Baek, M., Baker, D., (2020). Improved protein structure refinement guided by deep learning based accuracy estimation. *Bioinformatics*. https://doi.org/10.1101/2020.07.17.209643.
18. Shuvo, M.H., Gulfam, M., Bhattacharya, D., (2021). DeepRefiner: high-accuracy protein structure refinement by deep network calibration. *Nucleic Acids Res.* **49**, W147–W152. https://doi.org/10.1093/nar/gkab361.
19. Jing, X., Xu, J., (2021). Fast and effective protein model refinement using deep graph neural networks. *Nat.*

*Comput. Sci.* **1**, 462–469. https://doi.org/10.1038/s43588-021-00098-9.
20. Heo, L., Feig, M., (2020). High-accuracy protein structures by combining machine-learning with physics-based refinement. *Proteins: Struct., Funct., Bioinformatics* **88**, 637–642. https://doi.org/10.1002/prot.25847.
21. Bhattacharya, D., (2019). refineD: improved protein structure refinement using machine learning based restrained relaxation. *Bioinformatics* **35**, 3320–3328. https://doi.org/10.1093/bioinformatics/btz101.
22. Kopp, J., Bordoli, L., Battey, J.N.D., Kiefer, F., Schwede, T., (2007). Assessment of CASP7 predictions for template-based modeling targets. *Proteins* **69** (Suppl 8), 38–56. https://doi.org/10.1002/prot.21753.
23. Jing, X., Xu, J., (2020). Improved Protein Model Quality Assessment By Integrating Sequential And Pairwise Features Using Deep Learning. *Bioinformatics* **36**, 5361–5367. https://doi.org/10.1093/bioinformatics/btaa1037.
24. Sato, R., Ishida, T., (2019). Protein model accuracy estimation based on local structure quality assessment using 3D convolutional neural network. *PLoS One* **14**, e0221347. https://doi.org/10.1371/journal.pone.0221347.
25. MacCallum, J.L., Pérez, A., Schnieders, M.J., Hua, L., Jacobson, M.P., Dill, K.A., (2011). Assessment of protein structure refinement in CASP9. *Proteins* **79** (Suppl 10), 74–90. https://doi.org/10.1002/prot.23131.
26. Kryshtafovych, A., Fidelis, K., Tramontano, A., (2011). Evaluation of model quality predictions in CASP9. *Proteins* **79** (Suppl 10), 91–106. https://doi.org/10.1002/prot.23180.
27. Rego, N., Koes, D., (2015). 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324. https://doi.org/10.1093/bioinformatics/btu829.