



OPEN

MULTICOM2 open-source protein structure prediction system powered by deep learning and distance prediction

Tianqi Wu¹, Jian Liu¹, Zhiye Guo¹, Jie Hou² & Jianlin Cheng¹✉

Protein structure prediction is an important problem in bioinformatics and has been studied for decades. However, there are still few open-source comprehensive protein structure prediction packages publicly available in the field. In this paper, we present our latest open-source protein tertiary structure prediction system—MULTICOM2, an integration of template-based modeling (TBM) and template-free modeling (FM) methods. The template-based modeling uses sequence alignment tools with deep multiple sequence alignments to search for structural templates, which are much faster and more accurate than MULTICOM1. The template-free (ab initio or de novo) modeling uses the inter-residue distances predicted by DeepDist to reconstruct tertiary structure models without using any known structure as template. In the blind CASP14 experiment, the average TM-score of the models predicted by our server predictor based on the MULTICOM2 system is 0.720 for 58 TBM (regular) domains and 0.514 for 38 FM and FM/TBM (hard) domains, indicating that MULTICOM2 is capable of predicting good tertiary structures across the board. It can predict the correct fold for 76 CASP14 domains (95% regular domains and 55% hard domains) if only one prediction is made for a domain. The success rate is increased to 3% for both regular and hard domains if five predictions are made per domain. Moreover, the prediction accuracy of the pure template-free structure modeling method on both TBM and FM targets is very close to the combination of template-based and template-free modeling methods. This demonstrates that the distance-based template-free modeling method powered by deep learning can largely replace the traditional template-based modeling method even on TBM targets that TBM methods used to dominate and therefore provides a uniform structure modeling approach to any protein. Finally, on the 38 CASP14 FM and FM/TBM hard domains, MULTICOM2 server predictors (MULTICOM-HYBRID, MULTICOM-DEEP, MULTICOM-DIST) were ranked among the top 20 automated server predictors in the CASP14 experiment. After combining multiple predictors from the same research group as one entry, MULTICOM-HYBRID was ranked no. 5. The source code of MULTICOM2 is freely available at https://github.com/multicom-toolbox/multicom/tree/multicom_v2.0.

Building the high-quality structure of a protein from its sequence is important for studying protein function and has important applications in protein engineering, protein design and drug design. Because the expensive, time-consuming and low-throughput experimental methods for determining protein structures (e.g., X-ray crystallography, nuclear magnetic resonance spectroscopy-NMR, or cryo-electron microscopy) can only be used to solve the structures of a small portion of proteins, fast high-throughput computational protein structure prediction is necessary for constructing structures of the majority of millions of proteins in the nature^{1,2}. Various computational methods for protein structure prediction have been proposed, which can be largely classified as template-free modeling (FM, also called ab initio/de novo prediction) and template-based modeling (TBM, also called homology/comparative modeling) methods. When significant structural templates and good template-target sequence alignments are available for a target, template-based modeling methods can generate accurate models for the target. But when there are no good templates, template-free modeling methods are the only viable choice for constructing good structural models without referring to known protein structure templates.

¹Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA. ²Department of Computer Science, Saint Louis University, St. Louis, MO 63103, USA. ✉email: chengji@missouri.edu

Method	Mean TM-score (first model)	Mean TM-score (best of five models)	TM-score > 0.5 (first model)	TM-score > 0.5 (best of top five models)
(A) On CASP14 58 TBM domains				
MULTICOM-HYBRID	0.720	0.751	55	57
MULTICOM-DEEP	0.730	0.757	55	57
MULTICOM-DIST	0.702	0.722	53	56
(B) On CASP14 38 FM and FM/TBM domains				
MULTICOM-HYBRID	0.514	0.540	21	22
MULTICOM-DEEP	0.512	0.542	20	22
MULTICOM-DIST	0.513	0.554	20	23

Table 1. TM-scores of models generated by each MULTICOM2 server predictor on CASP14 domains.

Template-free modeling had been studied for decades and progressed very slowly until some breakthroughs in the several few years. The breakthroughs were enabled by two major technical advances, leading to the accurate model prediction for both template-based and template-free targets by template-free modeling methods such as AlphaFold³. One advance is the residue-residue co-evolution analysis that provides informative features to improve residue-residue contact/distance predictions for guiding template-free structure reconstruction. Another one is the application of deep learning to protein contact/distance prediction, which dramatically improves the accuracy of contact/distance prediction and therefore substantially enhances the distance-guided template-free modeling^{3–6}.

Most recently, the attention mechanism integrated with deep learning was applied to predict residue-residue distances and even 3D coordinates of atoms, which further enhanced the prediction accuracy and interpretability^{7,8}. Despite the substantial advances in the field, there are still very few state-of-the-art automated open-source template-based and template-free modeling system available for the community to use^{6,9}, hindering the development of new methods and application of protein structure prediction in biomedical research and technology development. Here we introduce our latest open-source protein structure modelling system MULTICOM2 that was recently benchmarked in the 14th Critical Assessment of Techniques for Protein Structure Prediction (CASP14) in 2020. Compared to its previous version (MULTICOM1) tested in CASP13¹⁰, its template-based modeling method is leaner, more accurate and much faster. Its distance-based template-free modeling method uses our newly developed deep-learning-based inter-residue distance prediction method—DeepDist¹¹ and is much more accurate than the contact-based tertiary structure modeling based on CONFOLD2¹² in MULTICOM1¹⁰. The server predictors based on MULTICOM2 generated structural models with correct topologies for almost all TBM targets and most FM and FM/TBM targets in the CASP14 experiment and were ranked among the top CASP14 server predictors.

Results

Performance of MULTICOM2 in CASP14. The global distance test score (GDT_TS)¹³ and TM-score¹⁴ are the two standard metrics to evaluate the model quality. The value of TM-score ranges from 0 to 1. A TM-score greater than or equal to 0.5 indicates the predicted model and the native structure have the same fold topology, while a TM-score less than 0.17 means no structural similarity between the predicted model and the native structure. GDT_TS score ranges from 0 to 100% (or simply from 0 to 1), a higher value indicating better model accuracy. Both GDT_TS and TM-score measure the global backbone similarity between a model and the native structure without considering side chain atoms.

Table 1 reports the average TM-score of the top 1 model and the best of top 5 models predicted by MULTICOM-HYBRID, MULTICOM-DEEP, and MULTICOM-DIST for the CASP14 domains, including 58 TBM (regular) domains and 38 FM and FM/TBM (hard) domains. For 58 CASP14 TBM domains, MULTICOM-DEEP ranks highest among three MULTICOM2 sever predictors. It has the average TM-score of 0.730 if it only makes one prediction (one trial) for each domain. Since a model with TM-score > 0.5 is considered to have the correct fold (or topology), we define the success rate as the number of domains for which a correct fold is predicted divided by the total number of domains in consideration. MULTICOM-DEEP predicts correct folds (TM-score > 0.5) for 55 out of 58 domains (i.e., ~ 95% of domains) if it predicts one model domain, and therefore its success rate of one trial is 95% for TBM domains. If the best of the top five models predicted by MULTICOM-DEEP (five trials) is considered, its success rate increases to 98% on the TBM domains. The other two predictors (MULTICOM-HYBRID and MULTICOM-DIST) has the similar performance.

The average TM-score of the pure ab initio predictor—MULTICOM-DIST on the TBM domains is 0.702 for one trial, close to that of MULTICOM-HYBRID and MULTICOM-DEEP, indicating that the performance of the distance-based template-free modeling method has reached the performance of the template-based modeling method on the TBM domains. It is remarkable progress considering the substantial gap between the two just a few years ago. Moreover, on the 38 very hard FM and FM/TBM domains, the average TM-score of the three predictors is very close and ranges from 0.512 to 0.514 for one trial. The success rate of MULTICOM-HYBRID on the FM and FM/TBM domains is 55% for one trial. For five trials, its success rate on the FM and FM/TBM domains increases to 58%. MULTICOM-DIST and MULTICOM-DEEP's performance on the hard domains is similar to MULTICOM-HYBRID, which is not surprising because they are based on the same template-free modeling pipeline of MULTICOM2 for FM and FM/TBM domains.

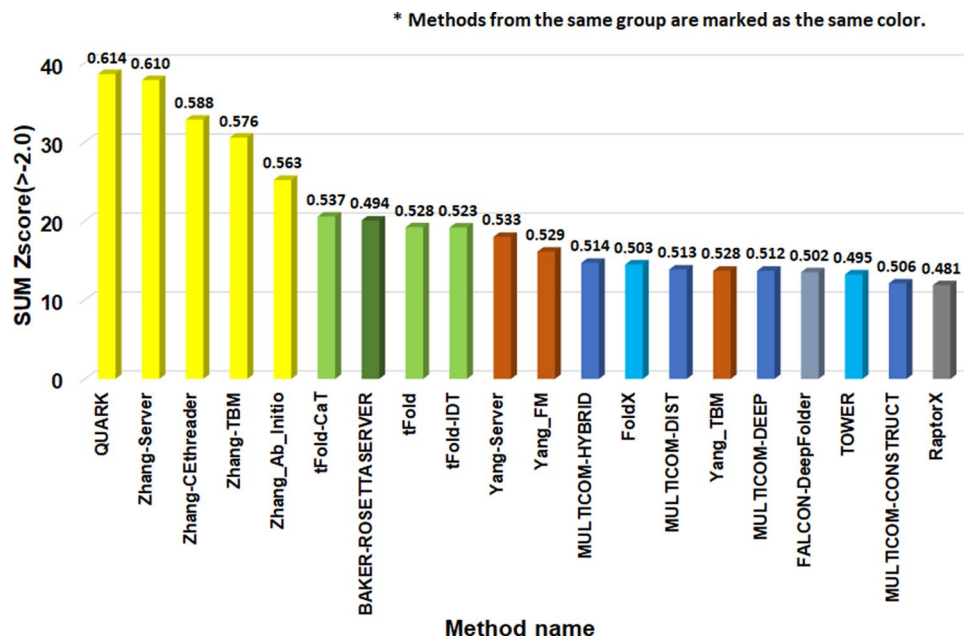


Figure 1. The top 20 server predictors on the 38 CASP14 FM and FM/TBM hard domains ranked by the sum of Z-score calculated according to the CASP14 assessor's evaluation. The predictors are ranked based on the first models they predicted for 38 CASP14 domains. The predictors from the same group are marked with the same color. The Y-axis denote the sum of the Z-score. The average TM-score of each predictor is reported on top of the bar representing each predictor.

To fairly compare different predictors that may perform well according to different evaluation metrics, the CASP14 assessors used the summation of Z-scores over the CASP14 targets to rank them¹⁵, considering three complementary metrics including GDT_TS¹³, Quality Control Score(QCS)¹⁶ that measures the assembly of secondary structure elements and Molprobit¹⁷ that evaluates the quality of all the atoms of a model including side chain atoms. The raw score of a model is the weighted average of the three metrics (i.e., $GDT_TS + QCS + 0.1 \times Molprobit$). And it is converted into a Z-score based on the raw scores of all the models for a target. A Z-score less than -2 is discarded in calculating the summation of Z-scores on all the targets¹⁸. Since MULTICOM2 is an automated prediction system, we evaluated its performance together with other CASP14 automated server predictors on 38 FM and FM/TBM (hard) domains, excluding CASP14 human predictors involving human intervention in prediction.

On the 38 CASP14 hard (FM and FM/TBM) domains, all the three MULTICOM2 server predictors, MULTICOM-HYBRID, MULTICOM-DEEP, and MULTICOM-DIST are ranked among the top 20 automated server predictors based on the CASP14's official sum of Z-scores (Fig. 1). The three MULTICOM2 predictors also performed better than MULTICOM-CONSTRUCT and MULTICOM-CLUSTER—the enhanced version of MULTICOM1¹⁰ (results not shown). If the multiple predictors from the same research group (e.g., QUARK, ZhangServer, ZhangCEThreader, ZhangTBM, and Zhang_Ab_Initio from Zhang Group) are combined into one entry represented by one predictor with the best performance (e.g., QUARK), MULTICOM-HYBRID representing the MULTICOM2 Group is ranked No. 5. Although the difference in the relatively Z-score of the top five groups is pronounced, the absolute average quality score (e.g., TM-score) of these predictors is much closer. As shown in Fig. 1, the difference of Z-score between MULTICOM-HYBRID and the best predictor of the top four groups—QUARK is 23.9 on the 38 CASP14 FM and FM/TBM hard domains, while the difference of TM-score between them is 0.1.

Impact of several factors on the performance of the MULTICOM2 system. From the MULTICOM2 results on 58 CASP14 TBM domains in Table 1(A), the pure template-free modeling method of MULTICOM2 (i.e., MULTICOM-DIST) can predict high-quality structures for TBM targets (i.e., average TM-score > 0.7), which is a significant improvement over MULTICOM1. The two integrated MULTICOM2 methods (MULTICOM-HYBRID and MULTICOM-DEEP) that combine the template-free modeling and the traditional template-based homology modeling still outperform MULTICOM-DIST. However, the difference in their performance is not statistically significant.

To analyze the impact of the template-based modeling branch of MULTICOM2, we compared the top-1 ranked models from the two integrated methods (MULTICOM-HYBRID and MULTICOM-DEEP) and top-1 ranked template-based models built by their templated-based modeling branch. We found that E-value (i.e., a measure of the significance) of the top template from homology search could be one simple criterion to assess if the templated-based modeling branch can contribute to the final prediction. According to the results on CASP14 TBM domains in Fig. 2 and Table 2, when the top template hits have E-value $\leq 10^{-50}$ (i.e., highly significant),

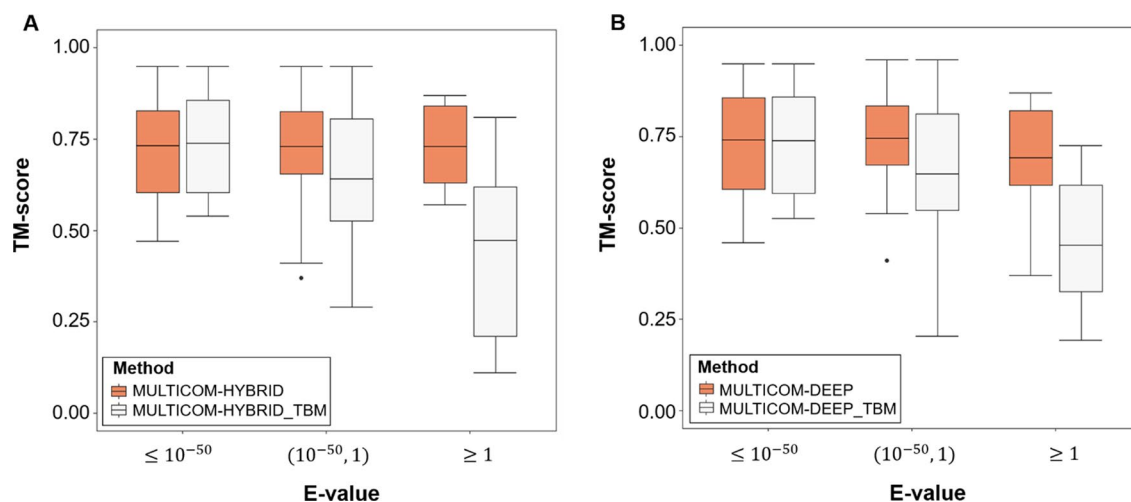


Figure 2. The box plots of the quality of top-1 models from MULTICOM2 integrated server predictors and top-1 models from their templated-based modeling branches on CASP14 TBM domains. **(A)** Comparison between MULTICOM-HYBRID and its templated-based models (MULTICOM-HYBRID_TBM). **(B)** Comparison between MULTICOM-DEEP and its templated-based models (MULTICOM-DEEP_TBM). Top1 templated-based models are selected based on the same model selection methods of MULTICOM-HYBRID and MULTICOM-DEEP mentioned in the “Methods” section.

Method	MULTICOM-HYBRID	MULTICOM-HYBRID_TBM	MULTICOM-DEEP	MULTICOM-DEEP_TBM
E-value $\leq 10^{-50}$	0.731	0.738	0.740	0.739
$10^{-50} < \text{E-value} < 1$	0.730	0.642	0.746	0.648
E-value ≥ 1	0.730	0.473	0.692	0.452

Table 2. Comparison of the quality of top-1 models from MULTICOM2 integrated server predictors (MULTICOM-HYBRID and MULTICOM-DEEP) and top-1 models from their templated-based modeling branches (MULTICOM-HYBRID_TBM and MULTICOM-DEEP_TBM) on 54 CASP14 TBM domains with native structures available for this analysis. The central line in the box marks the average TM-score values of top 1 models. Top-1 templated-based models are selected based on the same model selection method of MULTICOM-HYBRID and MULTICOM-DEEP described in the “Methods” section.

the top templated-based models from the template-based modeling branch have the quality comparable to that of the final top models predicted by the integrated system. For MULTICOM-HYBRID, the average TM-score of top 1 models from the template-based branch and the integrated system are 0.738 and 0.731, respectively. For MULTICOM-DEEP, the average TM-score of top 1 models from the template-based branch and the final integrated system are 0.739 and 0.740, respectively. In this situation, the template-based models sometimes are selected as top-1 models of the integrated system. When the top template hits have E-value between 10^{-50} and 1, the average TM-score of top templated-based models from the template-based modeling branch is 0.642 for MULTICOM-HYBRID_TBM and 0.648 for MULTICOM-DEEP_TBM, which is much lower than the scores (0.730 and 0.746) of top models of the integrated systems (MULTICOM-HYBRID and MULTICOM-DEEP), indicating that the template-based modeling has a much less impact in this situation. When the top template hits have E-value ≥ 1 (i.e., insignificant), the average TM-score of the top-1 models from MULTICOM-HYBRID and MULTICOM-DEEP is 0.730 and 0.692, much higher than 0.473 and 0.452 of the top-1 template models from the template-based modeling branch, indicating that the template-based models are not useful in this situation. The detailed per-domain comparison is shown in the Supplementary Table S1. Overall, the performance of template-based modeling method on TBM domains quickly decreases as templates become less significant, while the integrated prediction system such as MULTICOM-HYBRID still maintains a high prediction accuracy, indicating that the template-free modeling method can achieve a rather good accuracy on TBM domains across the board regardless of their difficulty. In fact, according to the analysis on 74 full-length CASP14 targets, MULTICOM-HYBRID selected template-free modeling models as top-1 models for 55 targets including ones having highly significant templates (see Table S2 for details).

While the significance of templates determines the quality of template-based models and their contribution to the whole system, the number of effective sequence (Neff)¹⁹ or more precisely quality of multiple sequence alignments (MSA) and the accuracy of the inter-residue distance prediction are critical for the template-free modeling branch and the entire MULTICOM2 prediction system. Fig. 3A shows the comparison between the quality of MSA roughly measured by the logarithm of Neff and the quality (i.e., TM-score) of the top-1 models built by the three MULTICOM2 server predictors. There is a moderate correlation between the two on all 91

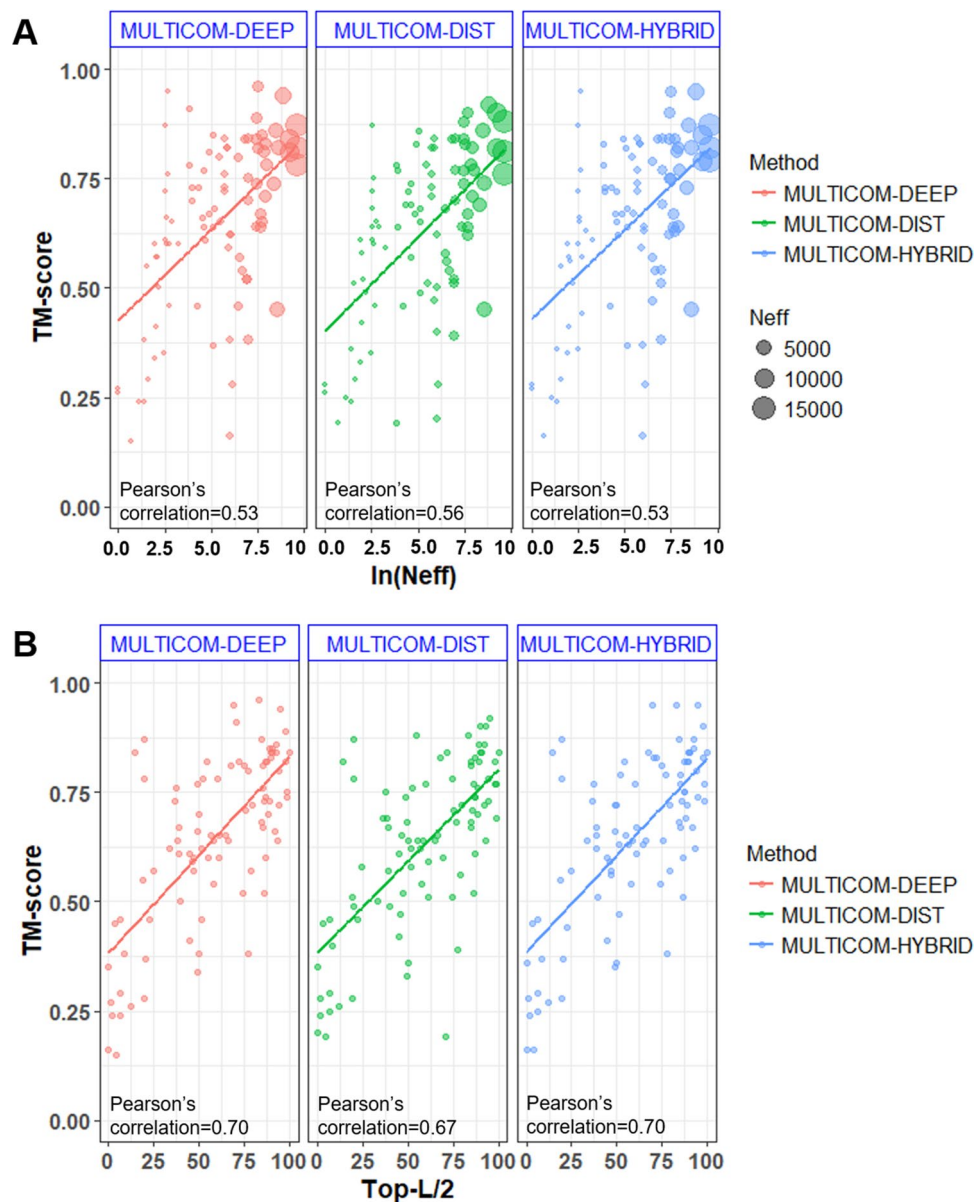


Figure 3. Impact of Neff and the accuracy of the inter-residue distance prediction on the model quality of the MULTICOM2 system on 91 CASP14 domains whose experimental structures are available for analysis. **(A)** Logarithm of Neff of MSA vs. the quality of models built from three MULTICOM server predictors (MULTICOM-DEEP, MULTICOM-DIST and MULTICOM-HYBRID). The size of a dot is proportional to the value of Neff. **(B)** The precision of top L/2 long-range contact predictions vs. the quality of models.

CASP14 domains whose native structures are available. The Pearson's correlation coefficients of MULTICOM-DIST, MULTICOM-HYBRID and MULTICOM-DEEP are 0.56, 0.53 and 0.53, respectively. Figure 3B shows a stronger correlation between the distance prediction accuracy (measured as the precision of top L/2 long-range contact predictions²⁰; L: sequence length) and the quality of the top-1 models. Pearson's correlation coefficient for MULTICOM-DIST, MULTICOM-HYBRID, and MULTICOM-DEEP is 0.67, 0.70, and 0.70, respectively. The detailed results for each CASP domain are shown in the Supplementary Table S3.

Runtime of MULTICOM2. It typically requires 12 hours for a protein of typical length (e.g., 300 residues) for MULTICOM2 to complete the entire modeling process on an Intel(R) Xeon(R) CPU E5-2660 v3 10-core processor, which is about two times faster than MULTICOM1. With hundreds of cores available on a computer cluster, MULTICOM2 can be used to predict structures of many proteins per day and therefore is applicable to the genome-level protein structure modeling.

Conclusion

We develop and release our latest automated protein structure prediction system (MULTICOM2) as an open-source software package for the community to use. MULTICOM2 is leaner, more accurate, and much faster than MULTICOM1. It was ranked among top automated server predictors in the latest 2020 CASP14 experiment. It can predict correct folds for almost all regular targets and a majority of very hard targets. Its template-free modeling without using any known structures as templates can work well on both regular template-based modeling targets and hard template-free modeling targets, which provides a uniform, reliable approach to modeling the structure of any protein. Therefore, it can be a useful tool for both the current development and application of protein structure prediction methods. While the inter-residue distance prediction can be converted to tertiary structures, significant modeling effort is still needed to do the conversion. The recent success of AlphaFold2⁸ in CASP14 demonstrates that an end-to-end deep learning system with the attention mechanism can directly predict the tertiary structure of a protein from its multiple sequence alignment, skipping the distance-to-structure conversion. In the future, we plan to develop deep learning-based end-to-end protein structure prediction methods and add them into the next release of MULTICOM.

Methods

Figure 4 illustrates the flowchart of the MULTICOM2 system. It is an automated and integrated protein modeling system that combines template-based modeling, template-free modeling, and model quality assessment.

Given a target sequence, MULTICOM2 searches it against the non-redundant protein sequence databases to build target sequence profiles such as position-specific scoring matrices (PSSM) and hidden Markov models (HMMs) using HH-suite²¹, PSI-BLAST¹, and HMMER²². Each profile is searched against a template library in order to identify a list of structurally similar templates and their sequence alignments with the target sequence. The pairwise target-template alignments are combined into multi-template alignments between the target and the multiple templates if the structures of the templates are consistent^{23,24}. Multi-template alignments along with the template structures are fed into Modeller²⁵ to build the structural models for the target protein. Different from MULTICOM1¹⁰ that uses more than a dozen sequence alignment tools and relatively slow threading tools (e.g. COMPASS²⁶, FFAS²⁷, SAM²⁸, PRC²⁹, HH-suite²¹, PSI-BLAST¹, HMMER²², RaptorX³⁰, I-TASSER/MUSTER^{9,31}), MULTICOM2 only uses HH-suite²¹, PSI-BLAST¹, and HMMER²² to build profiles, identify protein templates and generate target-template sequence alignments. Among those tools, hhblits and hsearch in HH-suite are much more sensitive than PSI-BLAST and HMMER and play an essential role in the template identification. Therefore, the template search process in MULTICOM2 is much faster than MULTICOM1. In order to further improve the sensitivity of template identification, MULTICOM2 applies DeepMSA³² to search the target sequence against large sequence databases (Uniclust30³³, Uniref90, metagenomics sequence database³⁴) for generating the deep multiple sequence alignments, which are used to build profiles to identify templates for template-based modeling.

In parallel to the template-based modeling, the target sequence is also used as input for a deep learning distance predictor—DeepDist³⁵ to predict inter-residue distances. The predicted distance maps are used by DFOLD^{35,36} (our in-house distance-guided ab initio modeling tool based on Crystallography and NMR System—CNS³⁷) and trRosetta⁶ to generate template-free models. DFOLD differs from CONFOLD2 (also based on CNS) that takes binary contacts as input to build tertiary structures. To use trRosetta, the distance maps predicted by DeepDist are used to substitute the default distance maps generated by trRosetta from the multiple sequence alignments before ab initio tertiary structure modeling. The DeepDist's distance prediction is also used to select templates based on their matching with the predicted distance maps for template-based modeling. By default, about 100 template-free and template-based models are constructed for a target. The models are ranked by three different model quality assessment methods: APOLLO³⁸ of ranking models based on their pairwise structural similarity, SBROD³⁹—a single model energy function of ranking models, and the distance-based ranking based on the similarity between the predicted distance maps and the distance maps of a model⁴⁰. Any one of the three rankings or their consensus can be used to select top models for the target.

Based on the target-template sequence alignment generated by HH-suite, MULTICOM2 also predicts if a target needs to be split into multiple domains or modeling units. If some region (> 40 residues) of a target does not have significant templates but other regions have, it will be split into multiple template-based and template-free domains according to the alignment. The structural models for each domain are then predicted by the same pipeline above. The top five ranked models from each domain are joined into full-length models for the multi-domain target by Modeller²⁵ or AIDA⁴¹.

In order to evaluate different modeling options of MULTICOM2, three predictors based on MULTICOM2—MULTICOM-HYBRID, MULTICOM-DEEP, and MULTICOM-DIST participated in the CASP14 experiment. MULTICOM-HYBRID and MULTICOM-DEEP used both the template-based and template-free modeling and differed only in model ranking. MULTICOM-HYBRID primarily used APOLLO³⁸ to rank models, whereas MULTICOM-DEEP primarily used the average ranking of SBROD³⁹ and the distance-based ranking (the matching between a model and the predicted distance map) to select final models. MULTICOM-DIST skipped template-based tertiary structure modeling entirely and only used the template-free modeling to generate tertiary structure models and used SBROD³⁹ to rank them, even though it still used template-target alignments to identify domain boundaries if needed.

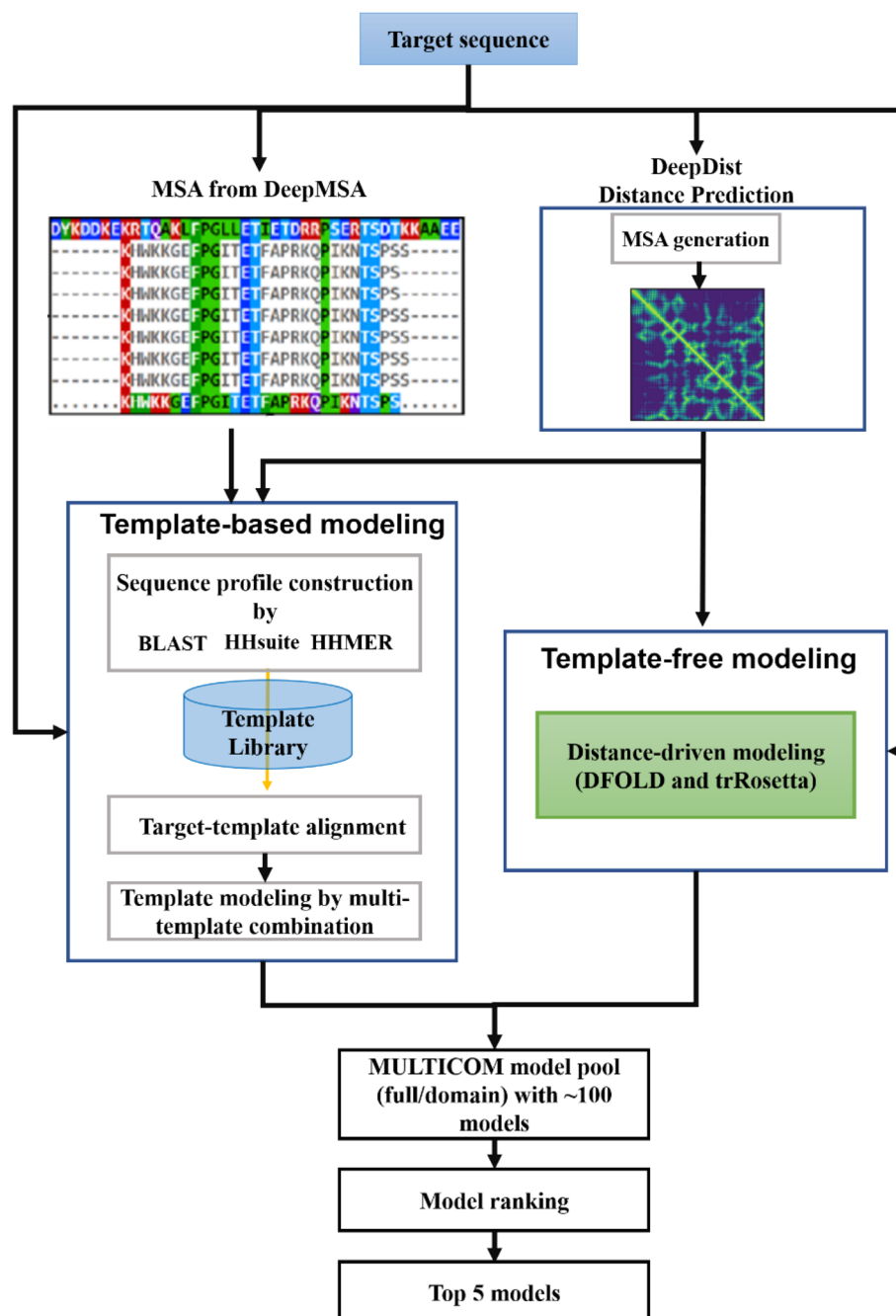


Figure 4. Flowchart of the MULTICOM2 system consisting of template-based, template-free modeling methods and model ranking (quality assessment).

Received: 18 March 2021; Accepted: 9 June 2021
Published online: 23 June 2021

References

- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389> (1997).
- Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007> (2018).
- Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710. <https://doi.org/10.1038/s41586-019-1923-7> (2020).
- Xu, J. Distance-based protein folding powered by deep learning. *Proc. Natl. Acad. Sci.* **116**, 16856–16865. <https://doi.org/10.1073/pnas.1821309116> (2019).
- Greener, J. G., Kandathil, S. M. & Jones, D. T. Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. *Nat. Commun.* **10**, 3977. <https://doi.org/10.1038/s41467-019-11994-0> (2019).

6. Yang, J. *et al.* Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci.* **117**, 1496–1503. <https://doi.org/10.1073/pnas.1914677117> (2020).
7. Chen, C., Wu, T., Guo, Z. & Cheng, J. Combination of deep neural network with attention mechanism enhances the explainability of protein contact prediction. *BioRxiv* **87**, 1058. <https://doi.org/10.1101/2020.09.04.283937> (2020).
8. Jumper, J. *et al.* High accuracy protein structure prediction using deep learning. In *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction* (Abstract Book), pp. 22 (2020).
9. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738. <https://doi.org/10.1038/nprot.2010.5> (2010).
10. Hou, J., Wu, T., Cao, R. & Cheng, J. Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. *Proteins Struct. Funct. Bioinform.* **87**, 1165–1178. <https://doi.org/10.1002/prot.25697> (2019).
11. Wu, T., Guo, Z., Hou, J. & Cheng, J. DeepDist: Real-value inter-residue distance prediction with deep residual convolutional network. *BMC Bioinform.* **22**, 30. <https://doi.org/10.1186/s12859-021-03960-9> (2021).
12. Adhikari, B. & Cheng, J. CONFOLD2: Improved contact-driven ab initio protein structure modeling. *BMC Bioinform.* **19**, 22. <https://doi.org/10.1186/s12859-018-2032-6> (2018).
13. Zemla, A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31**, 3370–3374. <https://doi.org/10.1093/nar/gkg571> (2003).
14. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710. <https://doi.org/10.1002/prot.20264> (2004).
15. Kinch, L. N., Li, W., Monastyrskyy, B., Kryshtafovych, A. & Grishin, N. V. Evaluation of free modeling targets in CASP11 and ROLL. *Proteins* **84**(Suppl 1), 51–66. <https://doi.org/10.1002/prot.24973> (2016).
16. Cong, Q. *et al.* An automatic method for CASP9 free modeling structure prediction assessment. *Bioinformatics (Oxford, England)* **27**, 3371–3378. <https://doi.org/10.1093/bioinformatics/btr572> (2011).
17. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21. <https://doi.org/10.1107/s0907444909042073> (2010).
18. Lisa Kinch, J. P., Schaeffer, D. & Grishin, N. *CASP14 Tertiary Structure Prediction Assessment Topology (FM) Category*. https://predictioncenter.org/casp14/doc/presentations/2020_11_30_Topology_assessment1_Kinch_Updated.pdf (2020). Accessed 30 Nov 2020.
19. Jones, D. T., Singh, T., Kosciolk, T. & Tetchner, S. MetaPSCOV: Combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics* **31**, 999–1006. <https://doi.org/10.1093/bioinformatics/btu791> (2014).
20. Graña, O. *et al.* CASP6 assessment of contact prediction. *Proteins Struct. Funct. Bioinform.* **61**, 214–224. <https://doi.org/10.1002/prot.20739> (2005).
21. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform.* **20**, 473. <https://doi.org/10.1186/s12859-019-3019-7> (2019).
22. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37. <https://doi.org/10.1093/nar/gkr367> (2011).
23. Cheng, J. A multi-template combination algorithm for protein comparative modeling. *BMC Struct. Biol.* **8**, 18. <https://doi.org/10.1186/1472-6807-8-18> (2008).
24. Cao, R., Wang, Z. & Cheng, J. Designing and evaluating the MULTICOM protein local and global model quality prediction methods in the CASP10 experiment. *BMC Struct. Biol.* **14**, 13. <https://doi.org/10.1186/1472-6807-14-13> (2014).
25. Webb, B. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **1137**, 1–15. https://doi.org/10.1007/978-1-4939-0366-5_1 (2014).
26. Sadreyev, R. & Grishin, N. COMPASS: A tool for comparison of multiple protein alignments with assessment of statistical significance. *J. Mol. Biol.* **326**, 317–336. [https://doi.org/10.1016/S0022-2836\(02\)01371-2](https://doi.org/10.1016/S0022-2836(02)01371-2) (2003).
27. Xu, D., Jaroszewski, L., Li, Z. & Godzik, A. FFAS-3D: Improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* **30**, 660–667. <https://doi.org/10.1093/bioinformatics/btt578> (2013).
28. Hughey, R. & Krogh, A. *SAM: Sequence alignment and modeling software system* (University of California at Santa Cruz, 1995).
29. Madera, M. Profile COMPARE: A program for scoring and aligning profile hidden Markov models. *Bioinformatics* **24**, 2630–2631. <https://doi.org/10.1093/bioinformatics/btn504> (2008).
30. Källberg, M., Margaryan, G., Wang, S., Ma, J. & Xu, J. RaptorX server: A resource for template-based protein structure modeling. *Methods Mol. Biol. (Clifton)* **1137**, 17–27. https://doi.org/10.1007/978-1-4939-0366-5_2 (2014).
31. Wu, S. & Zhang, Y. MUSTER: Improving protein sequence profile–profile alignments by using multiple sources of structure information. *Proteins Struct. Funct. Bioinform.* **72**, 547–556. <https://doi.org/10.1002/prot.21945> (2008).
32. Zhang, C., Zheng, W., Mortuza, S. M., Li, Y. & Zhang, Y. DeepMSA: Constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins. *Bioinformatics* **36**, 2105–2112. <https://doi.org/10.1093/bioinformatics/btz863> (2020).
33. Mirdita, M. *et al.* Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176. <https://doi.org/10.1093/nar/gkw1081> (2016).
34. Mitchell, A. L. *et al.* MGnify: The microbiome analysis resource in 2020. *Nucleic Acids Res.* **48**, D570–D578. <https://doi.org/10.1093/nar/gkz1035> (2019).
35. Wu, T., Guo, Z., Hou, J. & Cheng, J. DeepDist: real-value inter-residue distance prediction with deep residual convolutional network. *BMC Bioinform.* **22**, 1–17 (2021).
36. DFOLD. <https://github.com/jianlin-cheng/DFOLD> (2019). Accessed 22 Oct 2020.
37. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733. <https://doi.org/10.1038/nprot.2007.406> (2007).
38. Wang, Z., Eickholt, J. & Cheng, J. APOLLO: A quality assessment service for single and multiple protein models. *Bioinformatics (Oxford, England)* **27**, 1715–1716. <https://doi.org/10.1093/bioinformatics/btr268> (2011).
39. Karasikov, M., Pagès, G. & Grudin, S. Smooth orientation-dependent scoring function for coarse-grained protein quality assessment. *Bioinformatics* **35**, 2801–2808. <https://doi.org/10.1093/bioinformatics/bty1037> (2019).
40. https://predictioncenter.org/casp14/doc/CASP14_Abstracts.pdf (2020). Accessed 21 Nov 2020.
41. Xu, D., Jaroszewski, L., Li, Z. & Godzik, A. AIDA: Ab initio domain assembly server. *Nucleic Acids Res.* **42**, W308–W313. <https://doi.org/10.1093/nar/gku369> (2014).

Acknowledgements

We wish to thank CASP organizers and predictors for sharing the data used in this work.

Author contributions

J.C. conceived the project. T.W., J.L., Z.G., J.H. and J.C. designed the method. T.W. and J.C. implemented the method. T.W. gathered the results. T.W., J.L., Z.G. and J.C. analyzed the results. T.W. and J.C. wrote the manuscript. All authors reviewed the manuscript.

Funding

The funding was provided by National Science Foundation (DBI1759934 and IIS1763246), National Institutes of Health (GM093123) and also by Department of Energy, USA (DE-AR0001213, DE-SC0020400 and DE-SC0021303).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92395-6>.

Correspondence and requests for materials should be addressed to J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021