# Quality Estimation of Protein Modeling Techniques in CASP_Commons (SARSCoV-2) Using a Structure-Based Functional Prediction Approach: ResiRole

**Geoffrey J. Huang[1†] and William A. McLaughlin[1]**
[1]Geisinger Commonwealth School of Medicine, Scranton, PA 18509
[†]Doctor of Medicine Program
Correspondence: ghuang@som.geisinger.edu

## Abstract

Model quality estimation is essential to the field of protein structure prediction through computational methods. Our technique, ResiRole, rates model quality principally based on preservation of the structural characteristics of protein functional sites between the model and the target. ResiRole is a method that uses basic statistical techniques to allow for an unbiased, quantitative measure of structure model quality by implementing a machine learning program called FEATURE. CASP is a biannual protein modeling competition where various research teams in the field of protein structure prediction compete against each other to produce three-dimensional (3D) models of protein structure given only the primary sequence. In this study, we applied ResiRole to the protein structure models available through CASP_Commons related to the SARS-CoV-2 proteome. A statistical analysis was conducted by calculating the difference in the cumulative probabilities for each functional prediction model (SeqFEATURE model) with the corresponding cumulative probability of the same functional site prediction model in the target protein to generate a "difference" score. We then globally averaged the difference scores obtained for each structure prediction technique, using the value obtained as a metric to rank the accuracy of these techniques. On a limited dataset, we showed that average difference score (ADS) correlates well with existing methods of protein quality assessment used for CASP. We also showed that ADS is sensitive enough to detect quality differences between the primary and secondary structure prediction attempts by each structure prediction technique submitted to CASP_Commons. A similar study is currently being conducted on CASP15 (2022) to provide a larger dataset to achieve more generalizable results.

## Introduction

Protein structure prediction by computational methods requires accurate means to assess the quality of modeled protein structures. Using techniques such as X-ray crystallography, which can solve protein structures up to as high as 1.5 Å resolutions, many quality estimation techniques that use the experimentally determined protein structure to benchmark modeling accuracy have been developed. A non-exhaustive list of common methods of protein quality estimation currently used include root-mean-square deviation (14), template modeling score (15), global distance test total score (GDT-TS) (17), contact area difference score (16), Local Distance Difference Test (LDDT) (13), and SphereGrinder (5). Many of these techniques rely on a 3D superposition of the modeled structure and the reference structure, which is also called the "target" or "experimentally determined" structure. Simplified, subsequent calculation of a distance difference between backbone atoms (i.e., $\alpha$-carbons) or sidechain atoms is often used as a proxy for measuring protein structure model quality. Other techniques, such as ResiRole (10), approach quality estimation of protein models from a different perspective. ResiRole approaches quality estimation from the perspective of preserving protein function using a program called FEATURE [see Altman et al. (2)], which is a program that uses data with information containing hydrogen bonding patterns, ionic bonding interactions, hydrophobicity, and other variables to determine the probability of existence of a functional motif, such as a calcium binding site, in the given protein. ResiRole differs from such conventional techniques by incorporating a structure-based functional prediction approach, where model quality is judged based on the tendency of a model to preserve their functional motifs compared to the target protein. Other quality estimation techniques, such as QMEANDisCo (9), measure model quality without requiring knowledge of the experimental structure, but we do not discuss their impact here. In the biannual CASP experiment, the experimental structures and quality analysis of these proteins are not given to the researchers until after all submissions have been received to ensure the fidelity of the competition. There is also a comparable weekly competition that ResiRole has already been applied to where the accuracy of protein models is evaluated via the Continuous Automated Model EvaluatiOn (CAMEO) project (7). Here we conduct a study of the comparison of the quality of protein modeling techniques used in CASP Commons to build SARS-CoV-2 protein models using ResiRole, analyzing only the proteins whose experimental structures are currently available through the Protein Data Bank (PDB). Only two of the seven released whole protein targets had solved experimental structures, and we limited our analysis to these two proteins. These two proteins are the ORF3a ion channel and the ORF8 accessory protein, possibly involved in immune evasion (1).

In the biannual CASP experiment, the experimental structures and quality analyses of the proteins are not given to the researchers until after all submissions have been received to ensure the fidelity of the competition. Five modeling attempts are allotted for each team participating in CASP per protein structure. There is also a comparable weekly competition that ResiRole has already been applied to where the accuracy of protein models is evaluated via the Continuous Automated Model EvaluatiOn (CAMEO) project (7). In this study, we conducted a comparison of the quality of protein modeling techniques that were used in CASP_Commons (3), which was an extension of the CASP14 experiment. The target sequences of CASP_Commons are all from the SARS-CoV-2 proteome.

We used the ResiRole method to analyze only the protein targets whose experimental structures are currently available in the Protein Data Bank (PDB) (12). At the time of our study, only two of the seven released whole protein targets had solved experimental structures, and we limited our analysis to these two proteins as test cases. These two proteins are the ORF3a ion channel and the ORF8 accessory protein, which is a protein possibly involved in immune evasion (1). As part of the ResiRole method, the FEATURE program was used to extract and analyze hydrogen bonding patterns, ionic bonding interactions, hydrophobicity, and other variables to determine the probability of existence of a functional motif, such as a calcium binding site, in the given protein. FEATURE uses machine learning techniques trained on known protein data sets to create submodels that capture spherical microenvironments around selected atoms in a protein to predict the presence or absence of various functional site motifs. Using FEATURE, ResiRole differs from conventional techniques by incorporating a structure-based functional site prediction approach, where model quality is judged based on the tendency of a model to preserve the structural characteristics of the predicted functional sites in the models compared to the reference structures. Other quality estimation techniques, such as QMEANDisCo (9), measure model quality without requiring knowledge of the experimental (reference) structure, but we do not discuss their roles in quality estimation here.

## Methods

Many of the scripts and programs used in this study were designed using the CentOS or Rocky Linux operating system, and they were implemented to run on CAMEO data [see Toth et al. (10)]. Modifications and additions to the programs were made for the purposes of restructuring the code to account for different input data from CASP. Data containing information about each protein model and their corresponding experimental structure first needed to be downloaded from the CASP public website and sorted into a library of directories using Python. Then, we ran the FEATURE program on all the 3D coordinate files. The FEATURE program works by first deciphering atoms of interest for a particular functional site prediction model. FEATURE then outputs a raw score based on how likely a spherical microenvironment around the atom of interest describes a particular SeqFEATURE functional site (11). FEATURE selects atoms and constructs their functional site prediction models using data from established sequence-based functional motif mapping initiatives such as PROSITE (8). Using an expansive library of data, covering an extensive portion of the PDB, Toth et al. (10) calculated the mean and standard deviation of the raw FEATURE scores for each functional prediction model. Assuming that these scores are normally distributed, we converted our FEATURE raw scores to Z-scores. Only those functional site predictions that had a Z-score greater than the Z-scores corresponding to specificity levels of at least 90% for the SeqFEATURE models for the reference structures were included based on the receiver operating characteristic curve analysis by Toth et al. (10). Using Z-scores, we calculated a measure of cumulative probability for the existence of a particular functional site in both the modeled protein structure and the experimentally determined protein structure. The difference between those cumulative probabilities was then averaged across all functional prediction models of to achieve an ADS, which we used as a measure of the quality of the corresponding protein structure prediction technique. We used Open-PBS, an open-source job scheduler to manage the large volume of script-runs needed to perform our data analysis. CASP publicly provides quality estimation and assessment results on their website (https://predictioncenter.org/) for the solved SARS-CoV-2 protein targets which we obtained to compare ResiRole with other established methods of protein structure quality estimation.

## Results

An empirical demonstration of the value of our functional site prediction-based protein quality estimation technique was conducted using the protein visualization software Chimera (6). By identifying an atom selected by a SeqFEATURE model with a relatively high difference score, we can visualize the structural differences between a model and the reference structure by superposing the model and reference structure. In the image produced, we zoomed in around the atom or residue of interest. Figure 1 illustrates the result of such a demonstration. Here we identified a region corresponding to a relatively large difference score. In the illustration, we then see a difference in both 3D and hydrogen bond structure illustrating ResiRole's ability to identify and properly rank areas of structural mismatch.

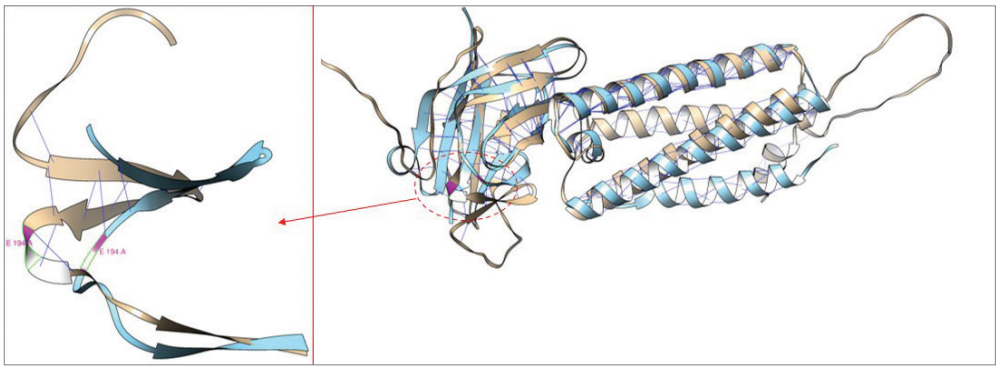

**Figure 1.** The figure depicts a ribbon-style superposition by RMSD of AlphaFold's protein structure model (brown) onto the experimentally determined structure of chain A of SARS-CoV-2 ORF3a (blue) (PDB ID: 6XDC). The figure focuses on a 20-residue range around GLU 194 (purple) in both structures. The blue lines indicate paths containing hydrogen bonds. GLU 194 was identified by ResiRole to have a high magnitude difference score (83.73%) by FEATURE functional prediction model BETA AMYLASE 2.5.GLU.OE2. Note the absence of the helical motif in the target structure that is present only in the model. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (6).

To show the relationship between ResiRole and other already established methods of protein quality estimation, we analyzed the trend between ADS and LDDT. LDDT is a widely used measure of protein model quality that is a superposition-free score that evaluates local distance differences of all atoms in a model, including validation of stereochemical plausibility, with respect to a reference structure (4). Figure 2 demonstrates that ADS has a significant correlation with LDDT, with AlphaFold ranking best from the perspective of ADS. A similar analysis and result were obtained when comparing ADS and GDT-TS, but not shown here.

Because the CASP experiment allows for the submission of multiple protein structure prediction attempts, we performed an analysis comparing the quality of models submitted in each group's first attempt with each group's corresponding second attempt. Under the assumption that groups submitted their best models for their first attempt, we demonstrate both ADS and LDDT averages are consistent with this trend. Notably, a t-test conducted at $\alpha$=.05 between mean average difference score (MEADS) of groups' attempt 1 and attempt 2 submissions show a statistically significant difference, shown in Figure 3A. However, the same test does not show a statistically significant difference in the mean LDDT (MLDDT) scores of groups' attempt 1 and attempt 2 submissions, shown in Figure 3B. If our original assumption is valid, we may infer that ADS performed better than LDDT in discerning the relative intentional quality of protein models between each group's attempt 1 vs. attempt 2 submissions made to CASP_Commons. That is, if we assume the structure prediction group intended the model for attempt one to be of higher accuracy than model submitted for attempt 2, ADS could detect that quality difference. ADS can thereby inform CASP modeling groups about the degree to which their accuracy differed or was improved for their first attempt model compared to their second attempt model.

## Discussion

There is a need in the field of protein quality estimation to generate techniques that evaluate the quality of protein models based on metrics that are representative of protein similarity. Such techniques need not be purely distance dependent. A comparison of the efficacy of ResiRole with that of traditional protein quality estimation methods would be useful to guide the development of improved protein modeling techniques, especially since many protein modeling groups modify their programs to optimize quality estimation scores. The QR code provided in Figure 4 provides a link to our website showcasing ResiRole's application to CAMEO. To further establish the validity of our technique in assessing protein model quality, we are currently conducting an extensive application of ResiRole to CASP15 data. CASP15 data contains experimental information from the most up-to-date modeling techniques as applied to a large variety of proteins found in nature. This larger set of data will allow us to test and rank the estimated accuracies of various protein structure prediction techniques more rigorously. Additionally, we will perform an outlier analysis and a head-to-head comparison between modeling groups regarding the accuracies of their techniques using ResiRole.

## Conclusion

We intend to help identify the most accurate modeling techniques, as well as apply our methods to evaluate existing forms of protein quality estimation. Properly ranking the accuracy of each structure prediction technique will inform studies that require knowledge of which models of protein targets are the most reliable and accurate. These models could then be used in applications that include structure-based drug design as the structure model with the highest estimated accuracy would be identified for the research team.

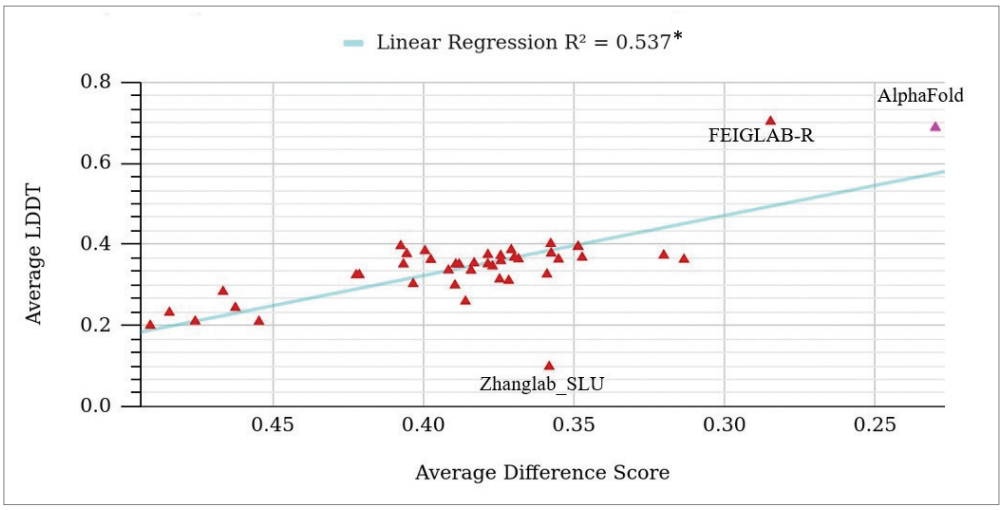


Figure 2. Average Difference Score vs. Local Distance Difference Test (LDDT). The figure depicts a scatter plot of the average difference scores across all SeqFEATURE models compared to the average LDDT scores for all protein structure model submission attempts. Each point corresponds to a unique protein prediction group (not all group names shown). The average difference score is calculated as $|Prob_{model} - Prob_{target}|$. Note the descending values of the x-axis, with a lower ADS corresponding to a better model. Three group names which likely represent outliers to the relationship are shown. *The p-value of the F-statistic for the linear regression was calculated to be $2.30441 * 10^{-8}$, despite the shown value of $R^2$.

Figure 4. QR code accessing the ResiRole website: https://protein.som.geisinger.edu/ResiRole

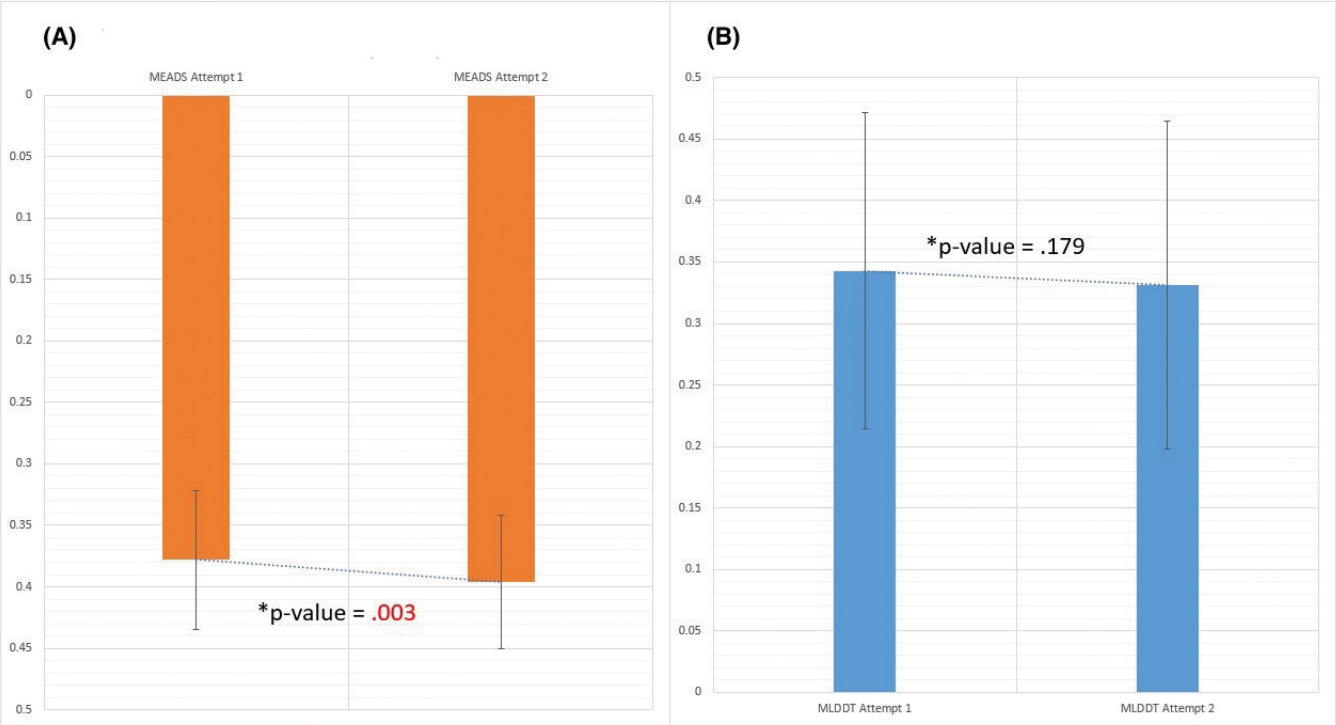


Figure 3. (A) Attempt 1 vs. Attempt 2 Model Quality by Mean Average Distance Score (MEADS). The graph in (A) depicts a comparison between MEADS across all attempt 1 models vs. all attempt 2 models across all modeling groups for both SARS- CoV-2 protein targets addressed in the study. (B) Attempt 1 vs. Attempt 2 Model Quality by Mean Local Distance Difference Test (MLDDT). The graph in (B) depicts a comparison between mean local distance difference test (MLDDT) scores across all attempt 1 models vs. all attempt 2 models across all modeling groups for both SARS-CoV-2 protein targets. Only groups that submitted at least 2 modeling attempts for each SARS-CoV-2 protein target were considered for calculating MLDDT. Error bars show one standard deviation around the mean. Standard deviation for the MEADS was calculated from ADS data, and it does not show the high variance of individual difference scores that occur as a result of the nature of the FEATURE program (10). *The p-values shown are calculated from a two-tailed paired t- Test at α = .05.

## Acknowledgments

We thank Thomas K. Parry for preparing an updated pipeline of the data analyses scripts as applied to CAMEO data which enabled the project to use the current computing environment. This research was funded in part by the National Institute of General Medical Sciences, grant number 5U01GM093324-02.

## References

1. Flower TG, Buffalo CZ, Hooy RM, Allaire M, Ren X, Hurley JH. Structure of sars-cov-2 orf8, a rapidly evolving immune evasion protein, *Proceedings of the National Academy of Sciences*, 118 (2021), p. e2021785118.
2. Halperin I, Glazer DS, Wu S, Altman RB. The feature framework for protein function annotation: modeling new functions, improving performance, and extending to novel applications, *BMC Genomics*, 9 (2008), pp. 1–14.
3. Kryshtafovych A, Moult J, Billings WM, Della Corte D, Fidelis K, Kwon S, Olechnovič K, Seok C, Venclovas Č, Won J, et al., Modeling sars-cov-2 proteins in the casp-commons experiment, *Proteins: Structure, Function, and Bioinformatics*, 89 (2021), pp. 1987–1996.
4. Mariani V, Biasini M, Barbato A, Schwede T. lddt: a local superposition-free score for comparing protein structures and models using distance difference tests, *Bioinformatics*, 29 (2013), pp. 2722–2728.
5. Olechnovič K, Monastyrskyy V, Kryshtafovych A, Venclovas Č. Comparative analysis of methods for evaluation of protein models against native structures, *Bioinformatics*, 35 (2018), pp. 937–944.
6. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. Ucsf chimera—a visualization system for exploratory research and analysis, *Journal of Computational Chemistry*, 25 (2004), pp. 1605–1612.
7. Robin X, Haas J, Gumienny R, Smolinski A, Tauriello G, Schwede T. Continuous automated model evaluation (cameo)—perspectives on the future of fully automated evaluation of structure prediction methods, *Proteins: Structure, Function, and Bioinformatics*, 89 (2021), pp. 1977–1986.
8. Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. PROSITE, a protein domain database for functional characterization and annotation, *Nucleic Acids Research*, 38 (2009), pp. D161–D166.

9. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—distance constraints applied on model quality estimation, *Bioinformatics*, 36 (2019), pp. 1765–1771.
10. Toth JM, DePietro PJ, Haas J, McLaughlin WA, Resirole: residue-level functional site predictions to gauge the accuracies of protein structure pre- diction techniques, *Bioinformatics*, 37 (2021), pp. 351–359.
11. Wu S, Liang MP, Altman RB, The seqfeature library of 3d functional site models: comparison to existing methods and applications to protein function annotation, *Genome Biology*, 9 (2008), pp. 1–19.
12. Berman HM, et al. The protein data bank. *Nucleic Acids Research* 28.1 (2000): 235-242.
13. Mariani V, et al. LDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 29.21 (2013): 2722-2728.
14. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Science*. 2001 Jul;10(7):1470-3.
15. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*. 2004 Dec 1;57(4):702-10.
16. Olechnovič K, Kulberkytė E, Venclovas Č. CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins: Structure, Function, and Bioinformatics*. 2013 Jan;81(1):149-62.
17. Zemla A. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Research*. 2003 Jul 1;31(13):3370-4