PREDICTION REPORT

Prediction of global and local quality of CASP8 models by MULTICOM series

Jianlin Cheng,[1,2,3]* Zheng Wang,[1] Allison N. Tegge,[2] and Jesse Eickholt[1]

[1] Computer Science Department, University of Missouri, Columbia, Missouri

[2] Informatics Institute, University of Missouri, Columbia, Missouri

[3] Bond Life Science Center, University of Missouri, Columbia, Missouri

ABSTRACT

Evaluating the quality of protein structure models is important for selecting and using models. Here, we describe the MULTICOM series of model quality predictors which contains three predictors tested in the CASP8 experiments. We evaluated these predictors on 120 CASP8 targets. The average correlations between predicted and real GDT-TS scores of the two semi-clustering methods (MULTICOM and MULTICOM-CLUSTER) and the one single-model *ab initio* method (MULTICOM-CMFR) are 0.90, 0.89, and 0.74, respectively; and their average difference (or GDT-TS loss) between the global GDT-TS scores of the top-ranked models and the best models are 0.05, 0.06, and 0.07, respectively. The average correlation between predicted and real local quality scores of the semi-clustering methods is above 0.64. Our results show that the novel semi-clustering approach that compares a model with top ranked reference models can improve initial quality scores generated by the *ab initio* method and a simple meta approach.

INTRODUCTION

In recent years, protein structure prediction has become time efficient and hundreds of alternative models of varying quality levels (accuracies) can be generated in a relatively short time.[1] As a result, Model Quality Assurance Programs (MQAP) are needed to assess, refine, rank, and select the highest quality models. Furthermore, an accurate MQAP can ensure the correct application of a model.[2,3]

MQAP methods can be divided into two categories: global model quality predictions and local (residue specific) model quality predictions. Amongst the global quality predictors, most of the methods output relative scores that can be used to discriminate native or near-native structures from decoys and a few methods output absolute scores that directly indicate the similarities between the models and the native structures.[3] The techniques frequently used by QA predictors include clustering (multiple-model) methods[1,4–8] and single-model techniques.[3,7,9–13] Clustering methods assume that models which are highly similar to others have better quality. Single-model techniques make predictions by analyzing various sequence alignment features[14] or structural features. These features include solvent exposure, secondary structure contact probability map, and probability map of β-strand residue pairing.

We participated in CASP8 (the eighth Critical Assessment of Techniques for Protein Structure Prediction) quality assessment experiments with the MULTICOM series. The MULTICOM series is a set of predictors incorporating various techniques for quality assessment, such as semi-clustering approaches, single-model machine learning approaches, and meta and hybrid approaches which combine two or more single approaches.

METHODS

Single-model quality assessment by MULTICOM-CMFR server

MULTICOM-CMFR is an *ab initio*, single model, structure-based model quality assessment method. It predicts the absolute quality score of a single protein model from its structural features as in ModelEvaluator.[3] Given a model, it compared the secondary structure, solvent accessibility, β-sheet topology, and contact map to those that the SCRATCH suite predicted from the primary sequence.[15] The comparison resulted in a number of fitness scores, such as secondary structure matching. These fitness scores were fed into a support vector machine trained on CASP6 and CASP7 models to predict the quality score (i.e. GDT-TS) of the model.

Hybrid semi-clustering quality assessment by MULTICOM-CLUSTER server

MULTICOM-CLUSTER uses a novel hybrid semi-clustering approach to assess both global and local model quality of CASP8 server models. It first used MULTICOM-CMFR[3] to predict the GDT-TS score for each model and subsequently ranked them according to the predicted GDT-TS score. The top five models were chosen as reference models and each model was then superimposed on each of the top five models using the structure comparison tool TM-Score,[16] which resulted in a GDT-TS score. The average GDT-TS score between each model and the five reference models was the predicted global quality of the model. This method is a hybrid combination of the single-model evaluation method and the model comparison approach and was partially inspired by Lee's approach as described in the CASP7 quality assessment work.[2] During the structure comparison between a model and each reference model, a superimposition of the model and the reference model was generated and the distance between the position of a residue in the model and its counterpart in the reference model was calculated. The average distance over the five reference models was calculated and used as the predicted local quality of the residue.

Meta and hybrid model quality assessment by MULTICOM

The MULTICOM model quality assessment procedure has two fully automated steps. It first downloaded all CASP8 QA server predictions. The predicted scores of the models in these predictions were averaged together to generate a consensus prediction (referred to as *simple_meta*). The consensus predicted quality score of the models was then used to rank all the models. The top five ranked models were selected as reference models for model comparison. As in the MULTICOM-CLUSTER approach, all the models are compared and superimposed

Table I
The Results of MULTICOM Series on CASP8

Predictor	Ave Corr.	Over Corr.	Ave Loss	Mean Diff.	RMSE
MULTICOM-CMFR	0.74	0.76	0.073	0.126	0.161
MULTICOM-CLUSTER	0.89	0.90	0.062	0.076	0.115
simple_meta	0.80	0.86	0.054	0.100	0.129
MULTICOM	0.90	0.92	0.048	0.077	0.117

The six columns represent average correlation, overall correlation, average GDT-TS loss, absolute mean difference and root mean square difference/error (RMSE) between real GDT-TS scores and predicted GDT-TS scores on all 120 targets.

with the reference models to generate both global and local quality scores.

RESULTS

Data preparation

We downloaded the predicted models and experimental structures of 120 valid CASP8 targets from the CASP8 website and the residues without coordinates in experimental structures were removed from the models. The models and experimental structures were compared by TM-Score[16] to calculate the real GDT-TS score of the models. The real GDT-TS scores were then used to evaluate the predicted GDT-TS scores from global quality predictors. To evaluate the local quality predictions, we superimposed the models and experimental structures using TM-Score. According to superimposition, the Euclidean distance between the positions of a residue in the model and in the experimental structure was calculated as the real local quality of the residue. The real local quality scores were compared with local quality scores (position deviation) predicted by local quality predictors.

Evaluations of global quality predictions

We used the following criteria to evaluate the global quality predictions of our methods: the average Pearson correlation, the overall correlation, the GDT-TS loss, the absolute mean difference, and the root mean square difference/error (RMSE, i.e. standard deviation error) between real GDT-TS and predicted GDT-TS scores (Table 1).

The average correlation is the average of the correlations of predicted and real GDT-TS scores of models of each target. MULTICOM-CLUSTER (resp. MULTICOM) achieved an average correlation of 0.89 (resp. 0.90), significantly higher than 0.74 (resp. 0.80) of its counterpart MULTICOM-CMFR (resp. simple_meta) without using structure comparison with reference models. The overall correlation is the Pearson correlation between the real GDT-TS scores and the predicted GDT-TS score on all the CASP8 models of all the targets pooled together. Similarly, MULTICOM_CLUSTER and MULTICOM
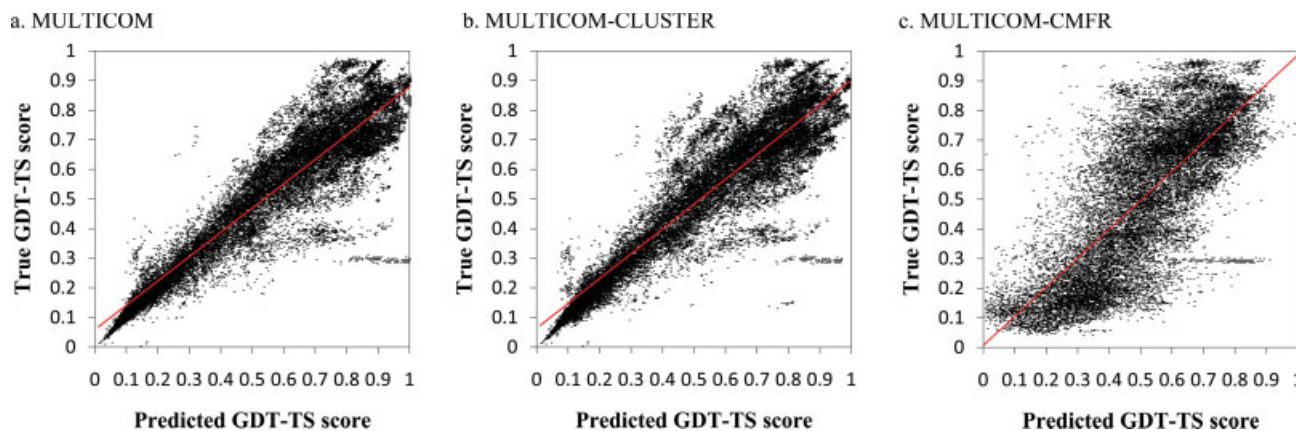
Figure 1
Plots of predicted GDT-TS scores against the true GDT-TS scores on all CASP8 models

achieved relatively high overall correlations, 0.90 and 0.92, respectively, better than 0.76 and 0.86 of MULTI-COM-CMFR and simple_meta. Figure 1 shows the plots of the real and predicted GDT-TS scores of the three methods. The good overall correlation indicates that the quality scores of models from different proteins are comparable. GDT-TS loss is the difference between the real GDT-TS score of the top model ranked by predicted GDT-TS scores and the real GDT-TS score of the best model for a target, where the best model is the model with the highest real GDT-TS score. The GDT-TS losses of MULTICOM_CLUSTER and MULTICOM are 0.062 and 0.048, lower than 0.073 and 0.054 of their counterparts. The absolute mean differences between predicted GDT-TS scores and real GDT-TS scores of all models for MULTICOM_CLUSTER and MULTICOM are 0.076 and 0.077, lower than 0.126 and 0.1 of their counterparts. The two semi-clustering methods also have lower RMSE than their counterparts.

Overall, the results show that MULTICOM-CLUSTER (resp. MULTICOM) consistently performed better than MULTICOM-CMFR (resp. simple_meta) that was used to generate initial quality scores. This indicates that the score refinement by structural comparison with reference models can improve both ranking and correlation. The fact that MULTICOM performs best of all demonstrates that the hybrid meta approach is very effective.

Evaluations of local quality predictions

MULTICOM and MULTICOM-CLUSTER also predict the local quality of each residue in a model. They predict the distance (position deviation) between the positions of a residue in a model and its experimental structure. We use the Pearson correlation to evaluate the predicted distances. We calculated the correlation between the predicted distances and the real distances of all the residues for each target. The average correlations of MULTICOM and MULTICOM-CLUSTER on 120 targets are 0.65 and 0.64, respectively. For many good quality models (e.g. T0404), the predicted local quality scores are almost the same as real quality scores.

CONCLUSIONS

We developed the MULTICOM series of model quality assessment predictors. The *ab initio* single-model method and the simple meta method can predict good initial quality scores. The semi-clustering approach can further refine and improve the correlation and GDT-TS loss of initial quality scores. The predicted absolute GDT-TS scores are well correlated with the real GDT-TS scores and can be used to compare the quality of models from different proteins. The low GDT-TS loss indicates that our predictors are good at ranking and selecting good models. The semi-clustering methods can also predict the local quality of a residue rather reliably.

REFERENCES

1. Wallner B, Elofsson A. Prediction of global and local model quality in CASP7 using Pcons and ProQ. Proteins 2007;69:184–193.
2. Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A. Assessment of predictions in the model quality assessment category. Proteins 2007;69:175–183.
3. Wang Z, Tegge A, Cheng J. Evaluating the absolute quality of a single protein model using structural features and support vector machines. Proteins 2009;75:638–647.
4. McGuffin L. Benchmarking consensus model quality assessment for protein fold recognition. BMC Bioinformatics 2007;8:345.
5. Zhang Y, Skolnick J. SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 2004;25:865–871.
6. Paluszewski M, Karplus K. Model quality assessment using distance constraints from alignments. Proteins, in press.

7. McGuffin L. The ModFOLD server for the quality assessment of protein structural models. Bioinformatics 2008;24:586.

8. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions, Vol.19. London: Oxford University Press; 2003. pp 1015–1018.

9. Wallner B, Elofsson A. Can correct protein models be identified? Protein Sci 2003;12:1073–1086.

10. Cline M, Hughey R, Karplus K. Predicting reliable regions in protein sequence alignments, Vol. 18. London: Oxford University Press; 2002. pp 306–314.

11. Qiu J, Sheffler W, Baker D, Noble W. Ranking predicted protein structures with support vector regression. Proteins 2008;71:1175.

12. Archie J, Karplus K. Applying undertaker cost functions to model quality assessment. Proteins, in press.

13. Benkert P, Tosatto S, Schomburg D. QMEAN: a comprehensive scoring function for model quality assessment. Proteins 2008;71:261–277.

14. Wallner B, Elofsson A. Identification of correct regions in protein models using structural, alignment, and consensus information. Protein Sci 2006;15:900–913.

15. Cheng J, Randall A, Sweredoski M, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 2005;33:W72.

16. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2007;68:1020.