**BIOINFORMATICS APPLICATIONS NOTE**

Vol. 27 no. 12 2011, pages 1715–1716
doi:10.1093/bioinformatics/btr268

Structural bioinformatics

APOLLO: a quality assessment service for single and multiple protein models

Zheng Wang¹, Jesse Eickholt¹ and Jianlin Cheng¹,²,³,*

¹Department of Computer Science, ²Informatics Institute and ³C. Bond Life Science Center, University of Missouri, Columbia, MO 65211, USA

Received on February 13, 2011; revised on April 3, 2011; accepted on April 18, 2011

1 INTRODUCTION

Protein model quality assessment plays an important role in protein structure prediction and application. Assessing the quality of protein models is essential for ranking models, refining models and using models (Cheng, 2008). Model Quality Assessment Programs (MQAPs) predict model qualities from two perspectives: the global quality of the entire model and the residue-specific local qualities. The techniques often used by MQAPs include multiple-model (clustering) methods (Ginalski et al., 2003; McGuffin, 2007, 2008; Paluszewski and Karpplus, 2008; Wallner and Elofsson, 2007; Zhang and Skolnick, 2004a), single model methods (Archie and Karpplus, 2009; Benkert et al., 2008; Cline et al., 2002; Qiu et al., 2008; Wallner and Elofsson, 2003; Wang et al., 2008) and hybrid methods (Cheng et al., 2009; McGuffin, 2009).

According to the CASP experiments, multiple-model clustering methods are currently more accurate than single model methods. However, they cannot work well if only a small number of models are available. A hybrid quality assessment method (Cheng et al., 2009) was recently developed to combine the two approaches and integrate their respective strengths. Here, we build a web server to provide the community with access to all three model quality assessment approaches (i.e. single, clustering and hybrid).

2 METHODS

2.1 Input and output

Users only need to upload or paste a single model file in Protein Data Bank (PDB) format or a zipped file containing multiple models.

If a single model is submitted, APOLLO predicts the absolute global and local qualities. If multiple models are submitted, APOLLO outputs the absolute global qualities, average pair-wise GDT-TS scores, refined average pair-wise Q-scores, refined absolute scores and pair-wise local qualities. All the global qualities range between (0, 1), where 1 indicates a perfect model and 0 indicates the worst case.

2.2 Algorithms

The absolute global quality score is generated based on our single model QA predictor—ModelEvaluator (Wang et al., 2008). Given a single model, ModelEvaluator (as MULTICOM-NOVEL server in CASP9) extracts secondary structure, solvent accessibility, beta-sheet topology and a contact map from the model, and then compares these items with those predicted from the primary sequence using the SCRATCH program (Cheng et al., 2005). These comparisons generate match scores which are then fed into an SVM model trained on CASP8 and CASP7 data to predict the absolute global quality of the model in terms of GDT-TS scores. To predict absolute local quality score of a residue, the secondary structure and solvent accessibility predicted from the sequence are compared with the ones parsed from the model in a 15-residue window around the residue. For each residue in the window, we also gather its contact residues that are ≥ 8 Å away in sequence and have an Euclidean distance ≤ 8 Å in the model. Their probabilities of being in contact according to the predicted contact probability map are averaged. The averaged contact probabilities, the match scores of secondary structure and solvent accessibility comparison and the residue encoding are fed into an SVM to predict local quality. The SVM are trained on the models of 30 CASP8 single domain targets.

The average pair-wise GDT-TS score is generated using our latest implementation (as MULTICOM-CLUSTER server in CASP9) of the widely used pair-wise comparison approach (Larsson et al., 2009). Taking a pool of models as an input, it first filters out illegal characters and chain-break characters in their corresponding PDB files. It then uses TM-Score (Zhang and Skolnick, 2004b) to perform a full pair-wise comparison between these models. The average GDT-TS score between a model and all other models is used as the predicted GDT-TS score of the model. One caveat is that the GDT-TS score of a partial model is scaled down by the ratio of its length divided by the full target length.

The refined global and local quality scores are generated using a hybrid approach (as MULTICOM-REFINE server in CASP9) (Cheng et al., 2009) that integrates single model ranking methods with structural comparison-based methods. It first selects several top models (i.e. top five or top ten) as reference models. Each model in the ranking list is superposed with the reference models by the TM-Score. The average GDT-TS score of these superpositions is considered as the predicted quality score. The superposition

© The Author(s) 2011. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
These PDB files were preprocessed in order to select correct chains. We assessed most of the methods used by APOLLO on 107 valid CASP9 targets. We also conducted a blind test of the absolute local quality predictor (trained on the CASP9 dataset) on the CASP9 models of 92 CASP9 single domain proteins. On the residues whose actual distances to the native are ≤ 10 and 20 Å, the average absolute difference between our predicted distances and the actual distances is 2.60 and 3.18 Å, respectively.

**Conflict of Interest:** none declared.

### RESULTS

We assessed most of the methods used by APOLLO on 107 valid CASP9 targets. We downloaded all the CASP9 models from CASP9 (http://predictioncenter.org/download_area/CASP9/) and the experimental structures from the PDB (Berman et al., 2000). These PDB files were preprocessed in order to select correct chains and residues that match the CASP9 target sequences. TM-Score was used to align each model with the corresponding native structure and generate its real quality score (GDT-TS). The CASP9 QA predictions made by our methods were evaluated against the actual quality scores by four criteria: average per-target correlation (Cuzzetto et al., 2009), the average sum of the GDT-TS scores of the top one ranked models, the overall correlation on all targets and the average loss—the difference in GDT-TS score between the top ranked model and the best model (Cuzzetto et al., 2009) (Table 1). The results show that the average correlation can be as high as 0.92 (respectively, 0.67) and the average loss can be as low as 0.057 (respectively, 0.095) for multiple model (respectively, single model). Our multiple- and single-model global QA methods were ranked among the most accurate QA methods of their respective kind according to the CASP9 official assessment (http://www.predictioncenter.org/casp9doc/presentations/CASP9_QA.pdf). The average per-target correlation of our pair-wise local quality predictions is ~0.53, which is also among the top local quality predictors in CASP9.

### Table 1. Results of global quality assessment methods used by APOLLO server on 107 CASP9 targets

| Methods | Average correlation | Average top 1 correlation | Overall correlation | Average loss |
|---------|-------------------|--------------------------|-------------------|-------------|
| Absolute score | 0.671 | 0.552 | 0.767 | 0.095 |
| Average pair-wise GDT-TS | 0.917 | 0.591 | 0.943 | 0.057 |
| Refined absolute score | 0.870 | 0.567 | 0.928 | 0.081 |
| Refined pair-wise Q-score | 0.835 | 0.572 | 0.904 | 0.076 |

**REFERENCES**

To see full references, please consult the original paper. Z.Wang et al.