KP233 FAMSD_QA: Model quality assessment using the side chain environment consensus score

Kazuhiko Kanou[1], Tomoko Hirata[1], Genki Terashi[1], Hiroko Sakai, Mayuko Takeda-Shitaka[1] and Hideaki Umeyama[1]

[1]School of Pharmaceutical Sciences, Kitasato University

INTRODUCTION

Selecting the best quality model from a set of predicted structures is one of the most important parts of protein structure prediction. In the Seventh Critical Assessment of Techniques for Protein Structure Prediction (CASP7) experiment, the new prediction category called "QA" (Quality Assessment) was implemented. In this category participants assess the quality of the models which were predicted automatically by other teams. Each QA participant gives a "reliability score" to these models from 0 (indicates 'bad model') to 1 (indicates 'good model'). After the experimental structures of CASP7 targets became available, the GDT_TS score was calculated for all predicted models and was compared with the "reliability score". Several values such as the Pearson correlation coefficient between GDT_TS and "reliability score" of each QA method were calculated to measure the quality of each QA method.

As a result of QA category in CASP7, we found that the most powerful method was consensus method like 3D-Jury [1]. 3D-Jury score is the summation of the number of residues (Cα atom) within 3.5Å from each predicted model. This method can select "good backbone" models but the quality of the side chain of selected models is not so good.

Thus we developed a new consensus method which considers side chain environment for the purpose of selecting good side chain models, and participated in the latest CASP experiment (CASP8) [2] using this method. We describe the algorithm of this method and the results of CASP8.

METHODS

First, we calculated the side chain environment composed of 'fraction buried' and 'fraction polar' for each residue of predicted model. 'Fraction buried' is the fraction of buried area within the surrounding side chain atoms, and 'fraction polar' is the fraction of buried area within the surrounding polar atoms. These values range from 0 to 1.0 per residue. When the model A was assessed, for each residue of model A, the side chain environment was calculated and is compared with the other models. If the Euclidian distance between the side chain environment ('fraction buried' and 'fraction polar') of one residue of model A and that of corresponding residue of another model was within 0.2, we considered that the two residues were in the same environment. For each model, we counted the number of residues in the same environment and the side chain environment score is the summation of those numbers. The threshold of 0.2 was determined using CASP7 models as a training set.

In CASP8, we participated in QA category as a team 'FAMSD_QA'. We had refined all predicted models by FAMS [3] and had assessed quality of these models using following combined score.

$$score = env\_con + w * SSscore$$

Here, *env_con* represents the side chain environment consensus score and *SSscore* represents the degree of match between the secondary structure of a predicted model and the secondary structure predicted from the given sequence with PSIPRED [4]. *w* is the weighting factor for *SSscore* and ranges from 0 to 1. In the case of difficult targets, more weight is given to *SSscore* than easy targets. This value was optimized using CASP7 models.

RESULTS and DISCCUSSION

Correlation coefficients

103 experimental structures of 128 CASP8 targets became available by September 2008. We calculated GDT_TS (accuracy score of backbone geometry) of all predicted models for 103 structure available targets, and calculated Pearson and

kanouk@pharm.kitasato-u.ac.jp

Spearman correlation coefficients between GDT_TS and FAMSD_QA score. As a result, average Pearson and Spearman correlation coefficients for all targets were 0.85 and 0.75, respectively. Furthermore the averages for 75 relatively easy targets were 0.91 and 0.79, and for 28 relatively difficult targets were 0.69 and 0.67, respectively. (Target classification is referred to on Robetta evaluation page [5].) Given this, it can be considered that FAMSD_QA scoring is more effective for easy targets than for difficult targets. The reason for the difference between Pearson and Spearman correlation coefficients for easy targets is that some targets of in the easy category have the bipolar distribution (there are both moderately good models and extremely bad models), that is, non normal distribution. The target that has the biggest difference between Pearson and Spearman correlation coefficients was T0444 (PDB code is 2VUX), these coefficients were 0.857 and 0.289, respectively. Figure 1 shows the scatter plot of FAMSD_QA score versus GDT_TS. In this case FASMD_QA scoring could judge the moderately good models (GDT_TS > 50) as "good model" and could judge the extremely bad models (GDT_TS < 30) as "bad model". Therefore Pearson correlation coefficient was very high (0.857). But among the moderately good models, FASMD_QA scoring couldn't distinguish relatively good models from relatively bad models, so Pearson correlation coefficient calculated with only these models was 0.416.

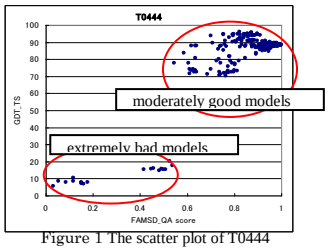

Figure 1 The scatter plot of T0444

This is not so good, but the GDT_TS of the first ranked model by FASMD_QA score is 88.6 and the highest GDT_TS among all models is 96.0. The ratio of the GDT_TS of the first ranked model to the highest GDT_TS, we call MGR (Max GDT_TS Ratio), is 92.3 (88.6/96.0) %.
The average MGR value for all targets, easy targets and difficult targets were 89.6, 93.8 and 79.0 %, respectively.

Evaluate the first raked model

We calculated the cumulative GDT_TS score of the first ranked models by FAMSD_QA score and compared with that of other automatic servers (Table 1). FAMSD_QA ranked at second following Zhang-Server.

Rank	team name	Sum of GDT_TS	average
1	Zhang-Server	6464.83	62.77
2	FAMSD_QA	6358.29	61.73
3	pro-sp3-TASSER	6226.73	60.45
4	RAPTOR	6206.66	60.26
5	METATASSER	6159.68	59.8

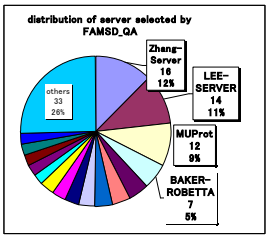

Table 1. Server ranking by the cumulative GDT_TS score

Furthermore, we evaluated the accuracy of side chain torsion angles for easy targets. We calculated the cumulative number of residues that have sufficiently accurate chi1 angle (within 30 degrees from native). As a result, FAMSD_QA ranked first of all server teams and 3D-Jury ranked fourth (Figure 2). This shows that FAMSD_QA scoring can select good models in terms of not only backbone geometry but also side chain torsion angles.

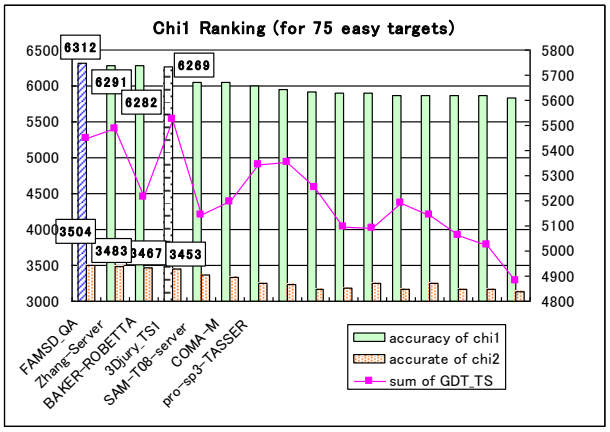

Figure 2. Server ranking by the accuracy side chain.

CONCLUSION

We developed an alternative consensus score for the purpose of selecting good models that have accurate side chain atoms. The new consensus score considers the side chain environment. We participated in recent CASP8 experiment and evaluated our method. As a result, side chain accuracy of the first ranked models by our new method was the best of all servers including 3D-Jury. It was proved that our consensus method using the side chain environment can select better side chain models.

REFERENCE

[1] Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics. 2003; 19:1015–8
[2] http://www.predictioncenter.org/casp8/index.cgi
[3] Ogata, K. and Umeyama, H. (2000). An automatic homology modeling method consisting of database searches and simulated annealing J Mol Graph Model/J Mol Graph Model 18, 258-272, 305-256.
[4] Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.
[5] http://robetta.bakerlab.org/CASP8_eval/index.html