KP229

Fams-ace2: Structure evaluation program using the combination of local consensus method and circle quality assessment method

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

1) School of Pharmacy, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641 JAPAN
terashig@pharm.kitasato-u.ac.jp

1. Introduction

The Critical Assessment of Protein Structure Prediction (CASP) is carried out in order to survey the capabilities and limitations of current methods of modeling protein structure from sequence. The methods are assessed on the basis of the analysis of a large number of blind predictions of protein structures. In the eighth round (CASP8 [1]), our fams-ace2 server participated in the 3D coordinate prediction category as a human expert group. We applied two different scoring functions for the fully automated model prediction server, fams-ace2: (1) the local consensus score; and (2) the model quality score based on classification of the side-chain environment for each residue. The local consensus score was used as a filter to select the models which have locally similar structures comparing with the set of models. The model quality score was then used for the final selection of the best model. This model quality score was calculated by our model quality assessment program CIRCLE [2] (see KP230).

2. Method

The procedure of fams-ace2 can be summarized as the following 4 steps:

(1) Obviously incorrect models which have serious physical clashes or broken main-chain structures were removed. (2) The top 10% (an optimized parameter of fams-ace2) of server models were selected in the order of the local consensus score; the local consensus score is calculated as the equation (1).

$$LocatCons_m = \frac{\sum_{n}^{N}\sum_{i}^{R_m} MAXSUB\left(LOC_{m,i}, LOC_{n,i}\right)}{N}$$

Equation 1

N is the number of server models. $LOC_{m,i}$ is a set of C-alpha coordinates which exist within 10Å from the *i* th residue of model *m*. *MAXSUB(a,b)* is a maximum number of C-alpha coordinates (subset *a*) which superimpose well (within 3Å) upon their corresponding C-alpha coordinates in subset *b*. The values of 10 and 3 Å are optimized parameters of fams-ace2. (3) All of the server models, selected in step (2), were refined and rebuilt utilizing our homology modeling program FAMS[3]. (4) The top 5 structures were selected, according to a model quality evaluation based on their CIRCLE score. The coefficients of *SSscore* in the KP230 which do not use the consensus method were changed in the fams-ace2 from 0.35 and 0.75 to 0.30 and 0.30, respectively. The fams-ace2 is a fully automated server and does not require human intervention.

The parameters of fams-ace2 were optimized by the data set of previous CASP7. We used the Global Distance Test Total Score (GDT_TS) as the quality of model compared to native. When we applied optimized fams-ace2 to CASP7 targets, fams-ace2 obtained the best results over all server groups (Fig.1). Moreover, in Template Based Modeling Targets, fams-ace2 also achieved best results over all groups including human groups.

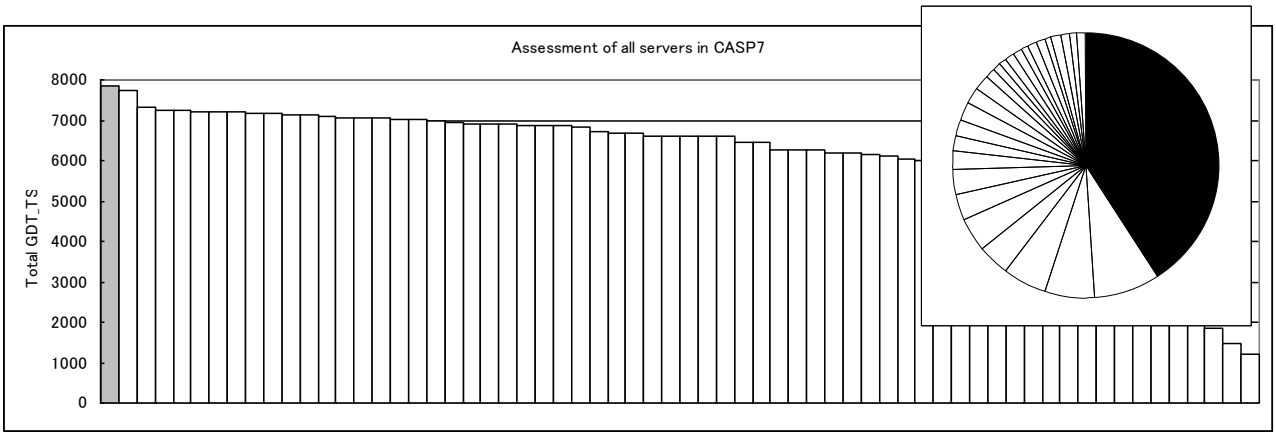

Fig. 1 Results of fams-ace2 (gray bar) and distribution of selected servers (pie graph) in CASP7

3. Results and Discussion

The 103 native protein structures of CASP8 128 targets were published in CASP8 web site (Sep 03 2008). We calculated GDT_TS of all models submitted by servers and fams-ace2 (Fig. 2). The total GDT_TS of fams-ace2 (gray in Fig.2) were obviously better than almost all of the other servers. The fams-ace2 selected models of the best server (Zhang-server) among 40% and 34% of targets in CASP7 and CASP8, respectively (The black area in pie graph of Fig.1 and Fig.2).

Although the advantage of fams-ace2 over other servers is slightly smaller than the results applying for CASP7 (123 domains, Fig.1), the intended results are accomplished. This small difference between CASP7 and CASP8 might be caused by the change of the distribution of target difficulty and performance of servers. When we calculate GDT_TS of CASP8 models, we did not consider the domain regions. Therefore the results of some targets will be changed. The advantages of fams-ace2 are the fully automated process, the low calculation costs and a high accuracy similar to the top of human groups. We are planning to optimize fams-ace2 according to the target difficulty and performance of each server by using much huger data set.

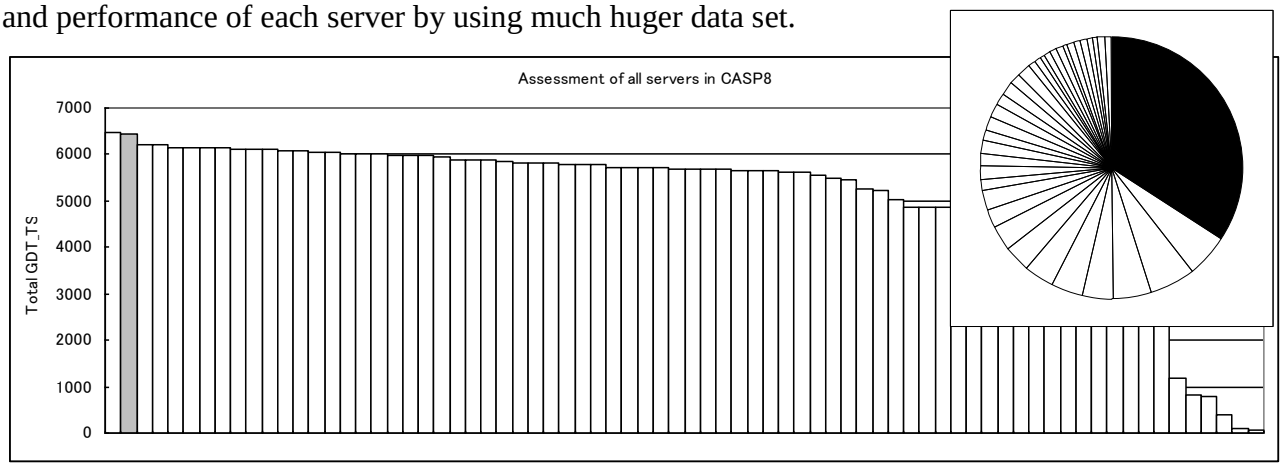

Fig. 2 Results of fams-ace2 (gray bar) and distribution of selected servers (pie graph) in CASP8

References

1. http://predictioncenter.org/casp8/index.cgi
2. Terashi G, Takeda-Shitaka M, Kanou K, Iwadate M, Takaya D, Hosoi A, Ohta K, Umeyama H. Fams-ace: a combined method to select the best model after remodeling all server models. *Proteins*. 2007;69 Suppl 8:98-107.
3. Ogata, K. and Umeyama, H. An automatic homology modeling method consisting of database searches and simulated annealing. *J. Mol. Graphics Mod*. 2000 Jun;18(3):258-72, 305-6.