KP231

FAMS_multi: Automated homology modeling based upon multiple reference proteins using better pairwise alignments in CASP8

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

[1]) School of Pharmacy, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641 JAPAN

INTRODUCTION

We developed an automated method of protein structure prediction called FAMS (Full Automatic Modeling System) [1,2]. FAMS is a homology modeling program consisting of database search and simulated annealing, and can construct high accuracy model when appropriate reference protein was detected. For predicting more accurate model, especially of loop structure and side chain torsion angles, we developed a new version of FAMS, called FAMS-multi, which uses multiple reference proteins.

For the purpose of assessing this method, we participated in CASP8 (8th Critical Assessment of Techniques for Protein Structure Prediction) experiment (our team name is '*FAMS_multi*'). CASP is a world-wide experiment for protein structure prediction held every two years since 1994. CASP provides participants with more than 100 protein sequences, and the each of participants must submit the predicted structures within 76 hours and more 2 weeks as an automatic predictor and a non-automatic one, respectively. Non-automatic predictors can use models which have been predicted by automatic predictors. We participated as a non-automatic predictor for the purpose of using automatically predicted models, but all processes were performed automatically. Models which were predicted by other automatic predictors were used to generate better alignments, and we rebuilt models by using FAMS-multi program which uses multiple reference proteins. In the following, we describe the scheme of this method and our results for CASP8.

METHODS

1. Generation of better pairwise alignments

We used the predicted models by other teams to generate better pairwise alignments between the target and its template in the PDB. First, we rebuilt these models by using FAMS program for the purpose of removing collisions. These rebuilt models were used to generate pairwise sequence alignments between the target and its template. The pairwise alignments were generated by structural superposition between each refined model and the its template using CE program [3]. When the superposition of the model and its template was not performed with the criteria of Z-score > 3.7, the alignment was not used.

Next, we constructed Cα models from these alignments using FAMS-multi program, and calculated 3D-jury scores of these Cα models which is Cα consensus score. Some alignments whose Cα model has a high 3D-jury score were used to construct full atom models using FAMS-multi program, and these models were evaluated using fams-ace2 method. Figure 1 shows the distribution of teams whose alignment was used to construct models that were finally selected by fams-ace2 method.

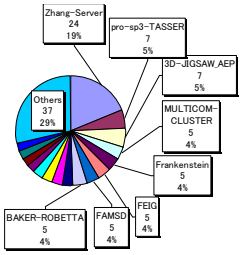

Figure 1

2. Construction of models by FAMS-multi

Some reference proteins were chosen based on the sequence and structural similarity with template. Next, a multiple structural alignment based on the superposition of Cα atoms was performed among the reference proteins. The target sequence was put on for this alignment based on the piarwise alignment between target and template mentioned in the preceding section. Thus, we get a result of multiple alignment between a target protein and reference proteins.

Using this alignment, tertiary structures were constructed mainly with next three steps, Cα construction, main chain construction, side chain construction. In each step, optimization was executed by the simulated annealing method.

Cα construction step: For the initial Cα coordinates,

kanouk@pharm.kitasato-u.ac.jp

first, the weighted average of Cα coordinates and the average distance were obtained from pairwise structural alignment based on the superposition of Cα atoms of the target and reference proteins. The weight factor of Cα coordinates for each reference proteins was decided based on Local Space Homology (LSH) calculated for each secondary structure segment. Next, the coordinates of Cα atoms were optimized by simulated annealing.

Main chain construction step: Initial coordinates of main chain atoms were constructed with the same method as FAMS. In the simulated annealing step, the potential function, which is consisting of (1) the weighted average of the coordinates of main chain atoms, (2) the average of distance and (3) the pair of N and O atoms forming the hydrogen bond as structural information, was used.

Side chain construction step: For the generated main chain atoms, conserved side chain torsion angles were obtained from homologous proteins. The coordinates of side chain atoms consisting of conserved side chain torsion angles were placed in relation to the fixed main chain atoms. The structural information such as the weighted average of the coordinates, average of distance, and the pair of N and O atoms forming the hydrogen bond, was derived from homologous proteins, and this information was used in optimization procedure.

3. Evaluate models (fams-ace2 method)

Thus, some full atom models were constructed. These models were evaluated using fams-ace2 selecting method (combined Cα consensus and Circle score [4]). Consequently top five models were selected.

4. Refine models

Five selected models were refined using Energy minimize & Molecular dynamics. With this procedure, hydrogen bonds, main chain torsion angles and side chain torsion angles were refined slightly and collisions of hydrophobic atoms were decreased.

All the procedures were implemented automatically.

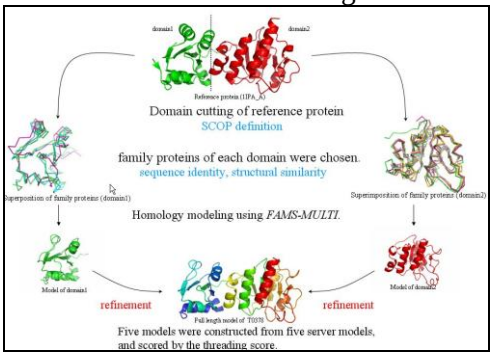

Figure 2. The scheme of *FAMS_multi*.

RESULTS & DISCUSSION

103 experimental structures of 128 CASP8 targets became available by September 3, 2008. We evaluated the accuracy of *FAMS_multi* models and that of the other server models, and compared them. The accuracy of backbone geometry was assessed by GDT_TS score, and the accuracy of side chain was assessed by the number of residues which have a sufficiently accurate side chain (chi1 torsion angle within 30 degrees from native structures or chi2 torsion angle within 60 degrees from native). Figure 3 shows the server ranking with the cumulative GDT_TS score of 103 targets (bar graph). Line graphs of square and triangle point is the cumulative number of accurate Chi1 torsion angles and Chi2 torsion angle, respectively.

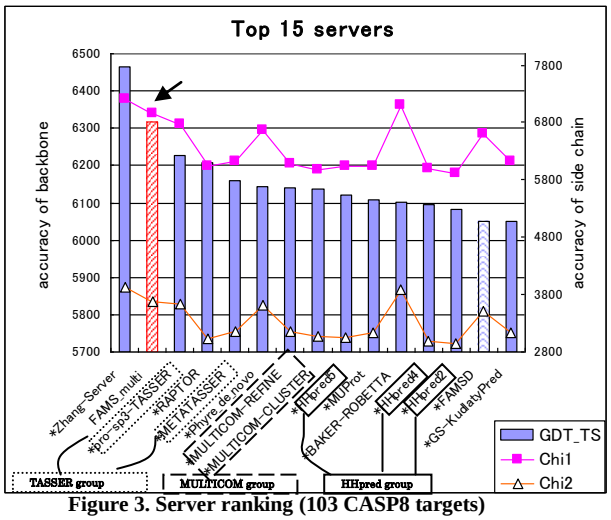

Figure 3. Server ranking (103 CASP8 targets)

As the results, *FAMS_multi* ranked second following Zhang-Server with GDT_TS score. *FAMS_multi* also ranked second following Zhang-Server with side chain accuracy. *FAMS_multi* could construct good models in terms of backbone geometry and side chain conformation.

REFERENCE

[1] Ogata, K. and Umeyama, H. J Mol Graph Model 2000; 18, 258-272.

[2] Ogata K, Umeyama H. Proteins. 1998; 31(4):355-69.

[3] Shindyalov IN, Bourne PE. Protein Engineering 1998; 11(9) 739-747.

[4] Terashi G, Takeda-Shitaka M, Kanou K, Iwadate M, Takaya D, Hosoi A, Ohta K, Umeyama H. Proteins. 2007;69 Suppl 8:98-107.