KP232 FAMSD: Individual Comparative Modeling server using SP3, FAMS and CIRCLE.

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

Recently, the number of proteins whose three-dimensional structures are already solved is increasing more and more. In September 2008, more than 52,000 structures are available on the Protein Data Bank (PDB) website. But the sequences whose structure has not been solved yet are more than 100 times as many as the sequence which were solved structurally. Therefore some approaches for protein structure prediction are required for implementing the structure based drug design and so on. Some effective approaches have been developed all over the world. Among those approaches, the most effective one is the Comparative Modeling when suitable template structures which have high sequence identities are detected. Our comparative modeling consists of following four steps: (1) making sequence alignments between target protein and template structures, (2) constructing three-dimensional structures based upon each alignment, (3) selecting the best structure model and (4) refinement of the selected model. We have developed an automatic protein structure prediction server called FASMD. Programs such as SP3 [1], FAMS (Full Automatic Modeling System) [2], CIRCLE [3] and Molecular dynamics were used at the each step (1) ~ (4), respectively.

We had participated in the past Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments. CASP is held once in 2 years, and each participant receives more than 100 protein sequences whose structure was unknown, and returns the predicted three-dimensional structures. After prediction season, the organizers of CASP assess the quality of all predicted models using its experimental structures. From April to August 2008, the 8th CASP (CASP8) experiment was held and 128 protein sequences were released totally. We participated in CASP8 as an automatic predictor using FAMSD. We describe the algorithm of FAMSD and our results for CASP8.

METHODS

(1) Making sequence alignments

8 kinds of alignment programs, BLAST, PSI-BLAST [4], PSF-BLAST, RPS-BLAST, IMPALA, Pfam-BLAST, SPARKS2 and SP3 were executed for each target protein sequence. Various alignments were generated and were filtered with its alignment score. The alignment scores for 6 kinds of methods except SPARKS2 and SP3 were calculated with following equation,

$$score = f(k_i, Hom, Len, SS) \qquad (1)$$

Here *Len* is the number of residues of a predicted model. *Hom* indicates sequence identity % value, *SS* is the degree of secondary structure agreement between the secondary structures predicted one from sequence using PSI-PRED [5] and one calculated from model using STRIDE. k_i is a coefficients for each alignment method.

And as the alignment score for SPARKS2 and SP3, Z-score of their output was used.

When the alignment score was more than the maximum score of all alignments * X, these alignments were used to construct model. A parameter X is a cut-off value which was decided using CASP7 targets as a training set.

(2) Constructing three-dimensional structures

We constructed three-dimensional structures using FAMS program based on each selected alignment which was mentioned in the preceding section.

kanouk@pharm.kitasato-u.ac.jp

(3) Selecting the best structure

All constructed models were evaluated using following scoring function,

score = *CIRCLE* + *w* * *SSscore*

Here, *Circle* represents the 3D1D score which was improved based on verify3D and *SSscore* represents the degree of secondary structure agreement. *w* is the weighting factor for *SSscore* which was optimized using CASP7 models as a training set.

Figure 1 shows the distribution of alignment method of finally ranked first models by above scoring function.

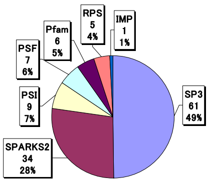

Figure 1.

(4) Refinement of the selected 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.

RESULTS & DISCUSSION

103 experimental structures of 128 CASP8 targets became available by September 3, 2008. We evaluated the quality of all server models, and compared GDT_TS of FAMSD model and the average of all server models (Figure 2). As a result, in almost of all targets GDT_TS of FAMSD model is higher than the average of all server models. Figure 2 shows that two targets were failed to prediction and these targets are in the difficult category.

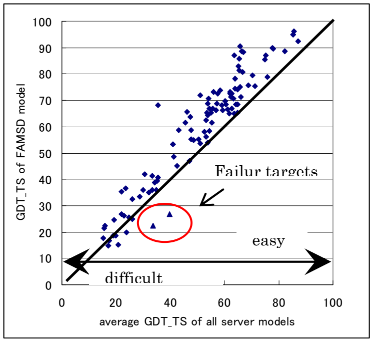

Figure 2.

Figure 3 shows top 15 of 126 server predictors sorted by the cumulative GDT_TS score of all 103 targets. Then FAMSD ranked at 13th. The accuracy of side chain was also assessed with the number of residues in the case that each model have a sufficiently accurate side chain, i.e., chi1 and chi2 torsion angle which is within 30 and 60 degrees, respectively, from native structure. In Figure 3, line graphs of square and triangle point is the cumulative number of accurate chi1 torsion angles and chi2 torsion angle, respectively.

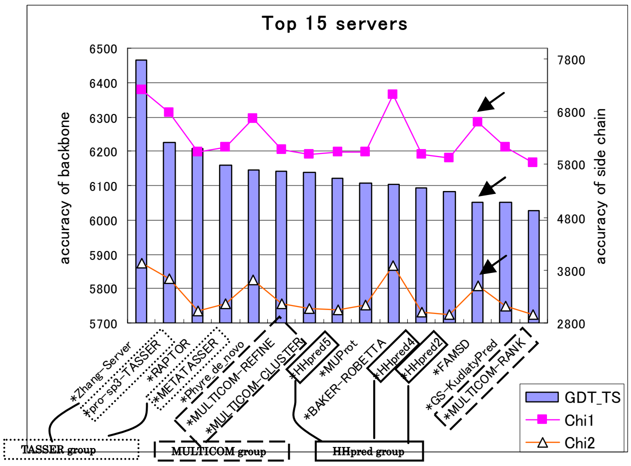

Figure 3 Top 15 servers for all 103 targets

Furthermore we calculated the cumulative score of GDT_TS, chi1 and chi2 for only 75 targets in the relatively easy category. Target classification is referred to on Robetta evaluation page [6]. As the results, the rank of FAMSD with GDT_TS, chi1 and chi2 were 11th, 7th and 11th, respectively. The six servers (Zhang-Server, Phyre_de_novo, pro-sp3-TASSER, FAMSD, BAKER-ROBETTA and COMA-M) predicted high quality models in terms of not only backbone geometry but also side chain conformation.

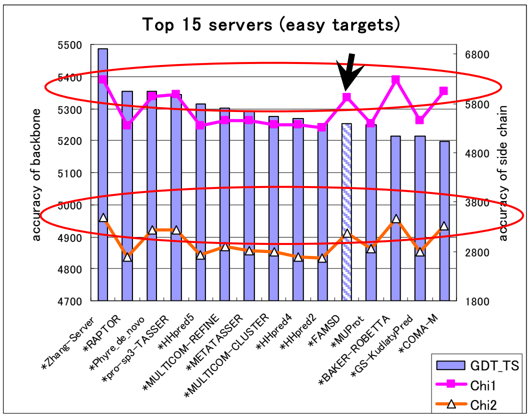

Figure 4 Top 15 servers for 75 easy targets

REFERENCE

[1] Zhou H, Zhou Y. Proteins. 2005;61 Suppl 7:152-6.

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

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

[4] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. Nucleic Acids Res. 1997; 25, 3389-3402.

[5] McGuffin LJ, Bryson K, Jones DT. Bioinformatics. 2000; Apr;16(4):404-5.

[6] http://robetta.bakerlab.org/CASP8_eval/index.html