High Quality Protein Structure Prediction Server,  
SKE-FAMSD

Kazuhiko Kanou  
kanouk@pharm.kitasato-u.ac.jp

Mitsu. Iwadate  
iwadatem@pharm.kitasato-u.ac.jp

Genki Terashi  
terashig@pharm.kitasato-u.ac.jp

Daisuke Takaya  
p99150@st.pharm.kitasato-u.ac.jp

Mayuko Takeda-Shitaka  
shitakam@pharm.kitasato-u.ac.jp

Hideaki Umeyama  
umeyamah@pharm.kitasato-u.ac.jp

Department of Biomolecular Design, School of Pharmaceutical Sciences, Kitasato University,  
5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

Keywords: homology modeling, fully automated protein structure prediction

1 Introduction

In accordance with increasing PDB experimental structure on recent 10 years, the availability of model structure has been rapidly enhanced in amino acid sequences of many genomes. Structural models are available from alignment software, for example PSI-BLAST or FASTA, and homology modeling tools, for example FAMS.

We had participated in the CASP6 (Critical Assessment of Techniques for Protein Structure Prediction) experiment in fully-automatic section, and all target protein structures were predicted using "SKE-FAMSD" server within 48 hours. The details of our method as a high quality protein structure prediction server, SKE-FAMSD are described.

2 Method

To perform high-quality predictions of protein structures, fully-automatic homology-modeling server "SKE-FAMSD" calculate some alignments for amino acid sequences using 7 kinds of methods, BLAST [1], PSI-BLAST, PSF-BLAST, RPS-BLAST, IMPALA, FASTA and Pfam. PSF-BLAST is PSI-BLAST whose profile sequence group of PSSM construction process is revised, and the selection criterion is E-value <= 0.001 from template PDB sequence on PSI-BLAST search. For selecting the "best alignment" in all alignments calculated by 7 kinds of methods, the score-function that was constructed by model length, homology% and degree of secondary structure agreement between PSI-PRED and STRIDE was defined:

\[ score = f(k_i, Hom, Len, SS) \]

Len is residue length of model protein. Hom indicates homology % value, the ratio between the number of match residues and Len. SS is so called Q3 value, degree of secondary structure agreement between PSI-PRED and STRIDE. ki are coefficients. The subscript number "i" indicate kind of alignment method, 0 is PSI-BLAST, 1 is BLAST, 2 is RPS-BLAST, 3 is Family-BLAST, 4 is IMPALA, 5 is FASTA, 7 is Pfam.

The next step is the construction of structures using best scored alignment by homology-modeling program FAMS [2]. In the gap-regions in the modeling process, the fragment selection procedure of FAMS was modified to select fragment of same protein family. The selection criteria in the gap-regions were RMSD of fitting and degree of SCOP [3] ID agreement between template PDB and fragment.

Finally, for refining constructed models by FAMS, both of Energy Minimization and Molecular Dynamics (MD) are applied. The condition for MD calculation (temperature, time, position constraint, torsion angle constraint etc.) was determined in the optimization process not to separate from the referred high-resolution X-ray structures.
Input: **Query sequence**

I. Various alignments by 7 kinds of methods.

II. Selecting the best alignment by original scoring-function.

III. Automatic Modeling by new version FAMS.

IV. Refinement of FAMS output structures by energy minimization(MM) & molecular dynamics (MD).

Output: **Tertiary structure**

Figure 1: Flowchart of SKE-FAMSD system.

### 3 Results and Discussion

#### 3.1 Scoring-Function

We evaluated our scoring-function for all CASP6 targets in order to investigate whether the best alignment was selected, using a MGR value as follows. MGR (Max GDT_TS Ratio) is the ratio between GDT_TS (the selected alignment used in the CASP6) and GDT_TS (the best alignment prepared by the SKE-FAMSD). Here GDT_TS is an index that shows the accuracy of the backbone structure. The higher this value is, the better. If MGR value is 100.0, scoring-function succeeded to select the best alignment. As a result, this scoring-function succeeded selecting alignment (MGR >= 95.0) in 22/29 "PDB-Blast hits" targets. Otherwise in the "no PDB-Blast" hit targets (more difficult targets), succeeded only 7/34. This scoring-function was able to select good alignments on the easy targets, but was not able to select on the difficult targets.

#### 3.2 Refinement Model by MM and MD

The accuracy were almost similar between MD models and non MD models on RMSD or GDT_TS value, but the hydrogen bonds (especially Main chain - Main chain hydrogen bond) and dihedral angles are improved by MD calculation. Furthermore the collisions between hydrophobic carbons of side-chain decreased clearly.

### References


