

September 28, 2006

RIKEN Genomic Sciences Center
Protein Function Team, Protein Research Group
Project Director: Dr. Shigeyuki Yokoyama
Senior Visiting Scientist: Dr. Hideaki Umeyama

Open the modeling database of three-dimensional protein structures to the public in the web on a worldwide level

- The acceleration of drug design and the design of novel functional biomolecules -

Particular notes

- Approximately 6,000,000 models at present in the RIKEN FAMSBASE, which are opened to the public in the RIKEN web site.
- Homology modeling as application of the determination of 3D structures by X-ray crystallography or NMR spectroscopy in the "Protein 3000 Project"
- Prediction of 3D structures dependent upon homology modeling which expands the possibility of drug design

The modeling database of 3-dimensional protein structures for which drug-design researchers aiming at the design of novel functional biomolecules are able to use easily in the fields of medical science, pharmaceutical science, biological science, science and engineering is opened to the public in the web on a worldwide level on September 28, 2006. This work was done by the members of the Protein research Group (Dr. Shigeyuki Yokoyama, Project Director, Dr. Hideaki Umeyama, Senior Visiting Scientist, et al.) at the Genomic Sciences Center (Dr. Yoshiyuki Sakaki, Director). This research has been supported by the "Protein 3000 Project", a national project in Japan.

In the present study, the research group predicted protein models of the whole publicly available species including human and laboratory animals (rat and mouse) that are recognized as the most important species for drug design predicted by homology modeling method^{*1}, and developed a database, RIKEN FAMSBASE, which contains 6,000,000 models together with the useful information: the evaluation of model structures, the candidates of the

protein-ligand binding sites and the explanation of proteins. RIKEN FAMSBASE is now available at RIKEN (<http://famshelp.gsc.riken.jp/famsbase/>), and is automatically updated for its regulatory to maintain the database fine.

We released a model database of neuraminidases of influenza viruses last January. The newly released RIKEN FAMSBASE exceeds the previous database in both quantity and quality. RIKEN FAMSBASE will be the first detailed database that is widely used in various research fields. RIKEN FAMSBASE is expected to activate drug design. Moreover, it is expected that novel functional biomolecules will be efficiently designed based on functional changes of proteins using the data in RIKEN FAMSBASE.

1. Background

The human genome and other genome sequencing projects have generated huge amounts of protein sequence information. To efficiently carry out drug design, a precise understanding of the biological function of the target protein based on the 3D structures is indispensable. However, the determination of 3D structures by X-ray crystallography or NMR spectroscopy proceeds more slowly than protein sequencing (39,000 structures are deposited in the Protein Data Bank² as of September, 2006). Therefore, there are many important proteins for which the sequence is available but the 3D structure has not been solved. In such cases, homology modeling that can reliably predict 3D protein structure is useful. Previously, Dr. Umeyama (Senior Visiting Scientist at RIKEN/ Professor at Kitasato University) et al. developed FAMSBASE, a database of about 1,400,000 homology models of 277 species, which is available at Nagahama Institute of Bio-Science and Technology (<http://daisy.nagahama-i-bio.ac.jp/Famsbase/index.html>). In the present study, Dr. Umeyama et al. newly predicted protein models of the whole publicly available genomes including human and laboratory animals (rat and mouse) that are recognized as the most important species for drug design. RIKEN FAMSBASE contains not only these newly predicted models but also the previously predicted models in FAMSBASE, and is now available at RIKEN (<http://famshelp.gsc.riken.jp/famsbase/>).

2. Methods

A homology modeling soft FAMS (Full Automatic Modeling System) developed by Dr. Umeyama et al. was used for protein structure prediction. FAMS was

developed based on homology modeling methods. Homology modeling is a method that constructs 3D model structures of unknown proteins (target proteins) based on known homologous 3D structures (reference proteins) arising from the concepts that homologous proteins have similar structures. At present, homology modeling is the most reliable tool for protein structure prediction using computer. The accuracy of FAMS was demonstrated by the Critical Assessment of Techniques for Proteins structure Prediction (CASP), a blind contest of protein structure prediction. In this study, we developed a database of homology models of the whole publicly available genomes constructed by FAMS.

3. Results

RIKEN FAMSBASE contains the previously predicted protein models of 277 species in FAMSBASE that is available at Nagahama Institute of Bio-Science and Technology, and the newly predicted 793,612 protein models of human, 505,628 protein models of laboratory animals (rat and mouse) and 3,368,709 protein models of the whole species^{*3}. As shown in Fig. 1, RIKEN FAMSBASE is freely available at RIKEN website (<http://famshelp.gsc.riken.jp/famsbase/>). The number of models in the RIKEN FAMSBASE is approximately 6,000,000 at present. RIKEN FAMSBASE is full of useful information: the evaluation of model structures, the candidates of the protein-ligand binding sites and the explanation of proteins. Moreover, RIKEN FAMSBASE is automatically updated for its regulatory to maintain the database fine.

Previously, we developed a database of 1,603 models of neuraminidases that play an essential role in the life cycle of the influenza viruses and released the database at RIKEN (http://protein.gsc.riken.jp/Research/index_na.html) on January 20, 2006 (press release: <http://www.riken.go.jp/engn/r-world/info/release/press/2006/060120/index.html>). The newly released RIKEN FAMSBASE including protein models of human, laboratory animals (rat and mouse), 277 species and the whole species will be the first detailed database that is widely used in various research fields.

4. Prospects

It is expected that RIKEN FAMSBASE activates and accelerates drug design. RIKEN has already started collaboration with various laboratories on drug designs against SARS (Severe Acute Respiratory Syndrome) coronavirus (press

release on September 8, 2004:

<http://www.gsc.riken.go.jp/eng/press/press/040908-2.html>), influenza virus and HCV (Hepatitis C Virus), and achieved some important results. Moreover, it is expected that novel functional biomolecules will be efficiently designed based on functional changes of proteins.

For more information, please contact:

Protein research group, Genomic Sciences Center, RIKEN
Shigeyuki Yokoyama, Project Director
Tel: +81-45-503-9196 (RIKEN Genomic Sciences Center)
Fax: +81-45-503-9195 (RIKEN Genomic Sciences Center)
e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

School of Pharmaceutical Sciences, Kitasato University
Hideaki Umeyama, Senior Visiting Scientist
Tel: +81-3-5791-6330
Fax: +81-3-3446-9553
e-mail: umeyamah@pharm.kitasato-u.ac.jp

Protein research group, Genomic Sciences Center, RIKEN
Takehisa Matsumoto
Tel: +81-45-508-7471
Fax: +81-45-508-7468
e-mail: takemats@gsc.riken.go.jp

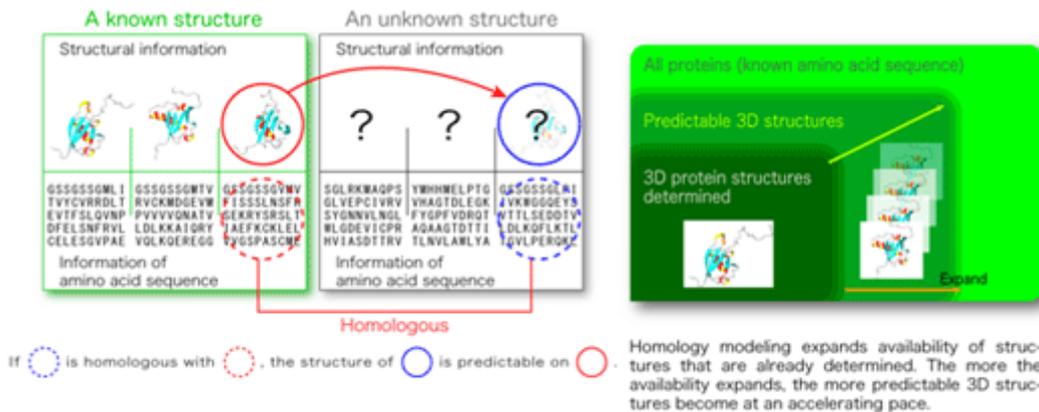
Yokohama Research Promotion Division, RIKEN
Rei Mizobe
Tel: +81-45-503-9117
Fax: +81-45-503-9113

RIKEN Public Relations Office
Email: koho@riken.jp

<Notes>

* 1 Homology modeling method

Homology modeling is a method that constructs 3D model structures of unknown proteins (target proteins) based on known homologous 3D structures (reference proteins) arising from the concept that homologous proteins have similar structures (left figure). At present, homology modeling is the most reliable and fastest tool for protein structure prediction. The structural genomics project that aims to determine experimentally at least one representative 3D structure from every protein family such as Protein 3000 Project seeks to determine approximately 10000 structures. Based on these representative structures, 3D structures of the majority of the proteins encoded in genomes might be predicted by homology modeling. For this reason, homology modeling has become an essential tool (right figure).

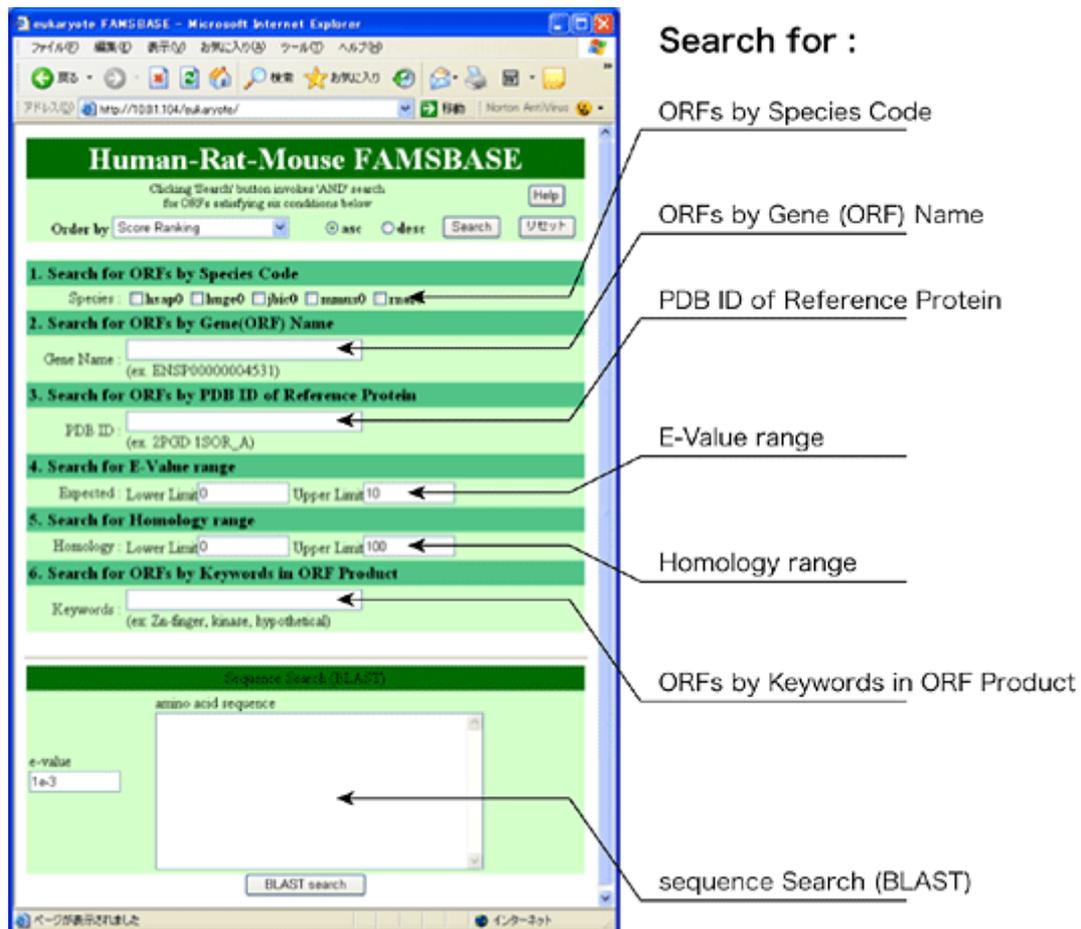


* 2 Protein Data Bank

The Protein Data Bank (PDB; <http://www.rcsb.org/pdb/>) is the single worldwide archive of structural data of biological macromolecules solved by the techniques of X-ray crystal structure determination, NMR, cryoelectron microscopy and theoretical modeling (Nucleic Acids Research, 28, 235-242, 2000).

* 3 Protein models of the whole species

Protein models predicted by FAMS for the amino acid sequence of the whole species in the nonredundant protein sequence database (nr) provided by National Center for Biotechnology Information (NCBI).



Search for :

ORFs by Species Code

ORFs by Gene (ORF) Name

PDB ID of Reference Protein

E-Value range

Homology range

ORFs by Keywords in ORF Product

sequence Search (BLAST)

Fig. 1 The website of Human-Rat-Mouse FAMSBASE in RIKEN FAMSBASE

A protein 3D structure model of a certain ORF can be found by gene name, PDB ID of the reference protein, keywords, amino acid sequence, etc.

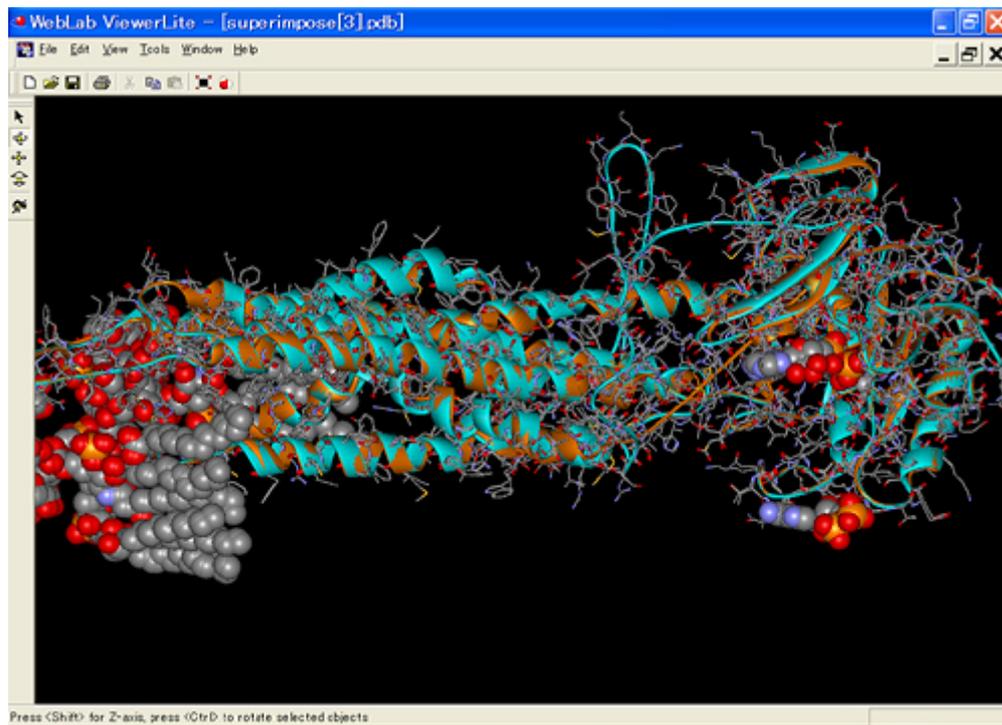


Fig. 2 An example of a 3D protein model structure in RIKEN FAMSBASE