Creation and Management of Information Platform in Targeted Proteins Research Program

hmizutan@lab.nig.ac.jp

hsugawar@genes.nig.ac.jp

[^1]: Targeted Proteins Research Program, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan
[^2]: Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[^3]: Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan
[^4]: Central Research Laboratory, Hitachi, Ltd., 1-280, Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan

Keywords: Information platform, database, targeted proteins research program

1 Introduction

The Targeted Proteins Research Program (TPRP) is a national project promoted by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. The program aims to reveal the structure and function of key proteins in three categories: fundamental biology, medicine and pharmacology, and food production and environment protection. The program also aims at the development of protein technologies for the production, structural analysis and regulation by chemicals. The program is relatively huge national project and the sharing and timely diffusion of accomplishments are essential to the success. Therefore, the program includes from the beginning the subject of Information Platform (TPRP IP). The structural biology project in US is also aware of the need and has maintained the knowledgebase [1]

Figure 1 shows the information flow in the program and modules of TPRP IP.

2 Method and Results

2.1 Website of TPRP

The public portal site of TPRP (http://www.tanpaku.org) was opened on October 1st, 2007. It is designed to announce goals, plans and research results in TPRP. The list of published papers and news from TPRP are frequently updated. For biological researchers, useful bioinformatics tools were supplied. In the meantime, for the purpose of information sharing and information management of tasks in TPRP, the project member's site was launched on June, 2008.
2.2 Bioinformatics tools

At present, the portal site provides (Comprehensive Amino Acids Sequences Annotated database (CASA db) [2], Functionally Annotated Japanese Protein Structural Information (FUJI) DB [3], Protein Experimental Information Management System (PREIMS) [4] and eSOL [5]) (Figure 2).

CASA db is a non-redundant amino acids sequences database. From three major public databases of amino acids, which are DDBJ DAD in Japan, NCBI NR in USA and UniRef in Europe, we have gathered over 10 million data of amino acids, over 22 million of species and over 62 million of cross-reference IDs. User can see the detailed protein data such as basic information like amino acid composition, structural data or cross reference ID. The details of FUJI database will be published elsewhere.

PREIMS is a database of experimental protocols for the study of proteins. It is to be noted the ontology to describe the protocol to promote the reuse of methods once developed and also to improve the reproducibility.

The database of eSOL is on the solubility of entire ensemble E.coli proteins [1] individually synthesized by PURE system that is chaperon free. The database is a collaborative product of a protein production team and TPRP IP. TPRP IP will expand specialized databases of this kind in collaboration with research groups in TPRP.

3. Discussion

National projects should be accountable to research communities and the public as well. Large scale projects especially have been requested to improve the visibility and the availability the data, information and knowledge acquired in the projects. TPRP IP is a model of the mechanisms to satisfy the requirement.

References