Publication of experimental data in crystallographic analysis of bacterial proteins by integrated database project

Naoki Kunishima, Akina Ozawa, Michihiro Sugahara,

Akeo Shinkai, Seiki Kuramitsu, Yukuhiko Asada

(RIKEN Harima SPring-8 Center)

e-mail: kunisima@spring8.or.jp

In the RIKEN SPring-8 Center (RSC), a large-scale crystallographic analysis of bacterial proteins was performed as a contribution to the national structural genomics project of Japan's "Protein 3000" initiative (FY 2002-2006), using SPring-8 X-rays in combination with a high-throughput platform for structural studies. After that, the useful experimental data from Protein 3000 are being compiled and published as a contribution to the integrated database project funded by the MEXT of Japan (FY 2007-2010), using a database publication platform "Scientists' Networking System (RIKEN SciNeS; http://scines.org)" developed by the RIKEN Bioinformatics and Systems Engineering Division (BASE). Contents of the RSC protein crystallography database (released tentatively in July 2009) are: 1) experimental data from nine species of bacteria that covers a large variety of protein molecules in terms of both evolution and property (http://scines.org/item/rib220i); 2) experimental data of mutant proteins that were designed systematically to study the influence of mutation to the diffraction quality of protein crystals (http://scines.org/item/rib220i); 3) experimental data of heavy-atom labeled proteins from which the user interface *HATODAS* (http://hatodas.harima.riken.jp) suggests potential compounds suitable for the preparation of a haevy-atom derivatized protein crystal (http://scines.org/item/rib108i). Potential applications of the RSC database are as follows. The bacteria data 1) will provide reference information for the structural study of homologous proteins. The mutant data 2) have a favorable characteristic for a detailed structural comparison between the mutant proteins because they share the same crystallization condition, which allows a software development of high-performance homology modeling, for instance. The heavy atom data 3) may be applied to the imaging of bio-molecules with a heavy-atom labeling. The database integration of RIKEN SciNeS adopts the semantic web format that is suitable for data reuse and automatic processing, thereby allowing a batch download of full data and a data reconstruction to make new databases. In FY 2010, the last year of the integrated database project, we will release additional data and a comprehensible user interface *Bacpedia* (http://bacpedia.harima.riken.jp), to complete the establishment of a useful data platform for general researchers in biosciences. In this meeting, current progress and the introduction of contents will be presented for the integrated database project.