CrossMark

# Special issue: big data analyses in structural and functional genomics

**Shigeyuki Yokoyama[1] · Kei Yura[2,3,4]**

"Big data" is a buzz term in all scientific disciplines. The amounts of data that modern measurements produce are so massive that we have both expectations for finding something new and problems in storing, transferring and interpreting these data. The expectations are based on the aphorism "Knowledge is power." Once we obtain huge amounts of data, we believe that we have great knowledge that enables us to deduce new rules and theories. These new ideas may play the roles of hypotheses that are tenable by experiments. The process itself has been conducted in human brains for millennia, but the amounts of data we currently have are so huge that big data science is expected to change the way we conduct scientific exploration. Traditionally, in our brains, we stored data and experiences, connected these data and then deduced ideas. In the big data era, these processes are massive challenges in all scientific disciplines, including the field of structural and functional genomics.

The "Platform for Drug Discovery, Informatics and Structural Life Science," one of the life science national projects in Japan run from fiscal year 2012 through 2016, has been tackling these challenges. The project aims for improvements in methods for protein structure determination, ligand screening and information retrieval from structural life science big data. The newly developed methods are applied to the promotion of structural life science in Japan. The efforts to improve the methods for information retrieval from genome and protein data in this project have achieved some helpful results, not only for the project but also for the entire scientific community in this field. Hence we organized this special issue of Big Data Analysis in Structural and Functional Genomics.

This special issue includes six papers concerning methods for big data analysis and applications of the methods to life science research. K. Yura et al. reported a new type of computer application, named VaProS, which searches numerous life science databases on the Internet for relevant data and visualizes the integrated results. T. Kawabata introduced a web tool, called HOMCOS, which systematically connects both genome information and protein structure information. K. Kinoshita et al. described the Natural Ligand Database (NLDB), which connects the modified ligands in the PDB, the protein structural database, and the natural ligands in organisms. K. Nagata et al. applied new information retrieval methods to the data analysis of G protein-coupled receptors (GPCRs) and analyzed the relationship between orphan class-A GPCR proteins and diseases. T. Shirai et al. reported an improvement in the method to compare ligands in the PDB. K. Tomii et al. demonstrated the importance of improving the substitution matrix for highly sensitive homology searches against the current huge sequence databases.

Using big data in structural and functional genomics through sophisticated retrieval methods should provide new perspectives on problems to be solved. We hope that the databases and tools described here are the ones that pave the way for building novel hypotheses and conducting new experiments in this field.

✉ Shigeyuki Yokoyama
yokoyama@riken.jp

[1] RIKEN Structural Biology Laboratory, Yokohama 230-0045, Japan

[2] Graduate School of Humanities and Sciences, Ochanomizu University, Otsuka, Bunkyo, Tokyo 112-8610, Japan

[3] Center for Informational Biology, Ochanomizu University, Otsuka, Bunkyo, Tokyo 112-8610, Japan

[4] National Institute of Genetics, Yata, Mishima, Shizuoka 411-8540, Japan