Research Activities of the Comprehensive Analysis Program in the National Project on Protein Structural and Functional Analyses (Protein 3000 Project)

RIKEN Structural Genomics/Proteomics Initiative (RSGI) has been serving as a core group to pursue Japan's national project, "the Protein 3000 Project," under the supervision of the MEXT. The comprehensive analysis program, a part of the project, was founded to solve 2,500 important protein structures from bacteria, archaea, and higher eukaryotes. Construction of state-of-the-art facilities, such as high-throughput beamlines at the SPring-8 (Harima) and the large-scale NMR facility (Yokohama), in conjunction with technical developments such as automated and efficient protein production techniques, synergistically facilitates the evolution of structural proteomics.

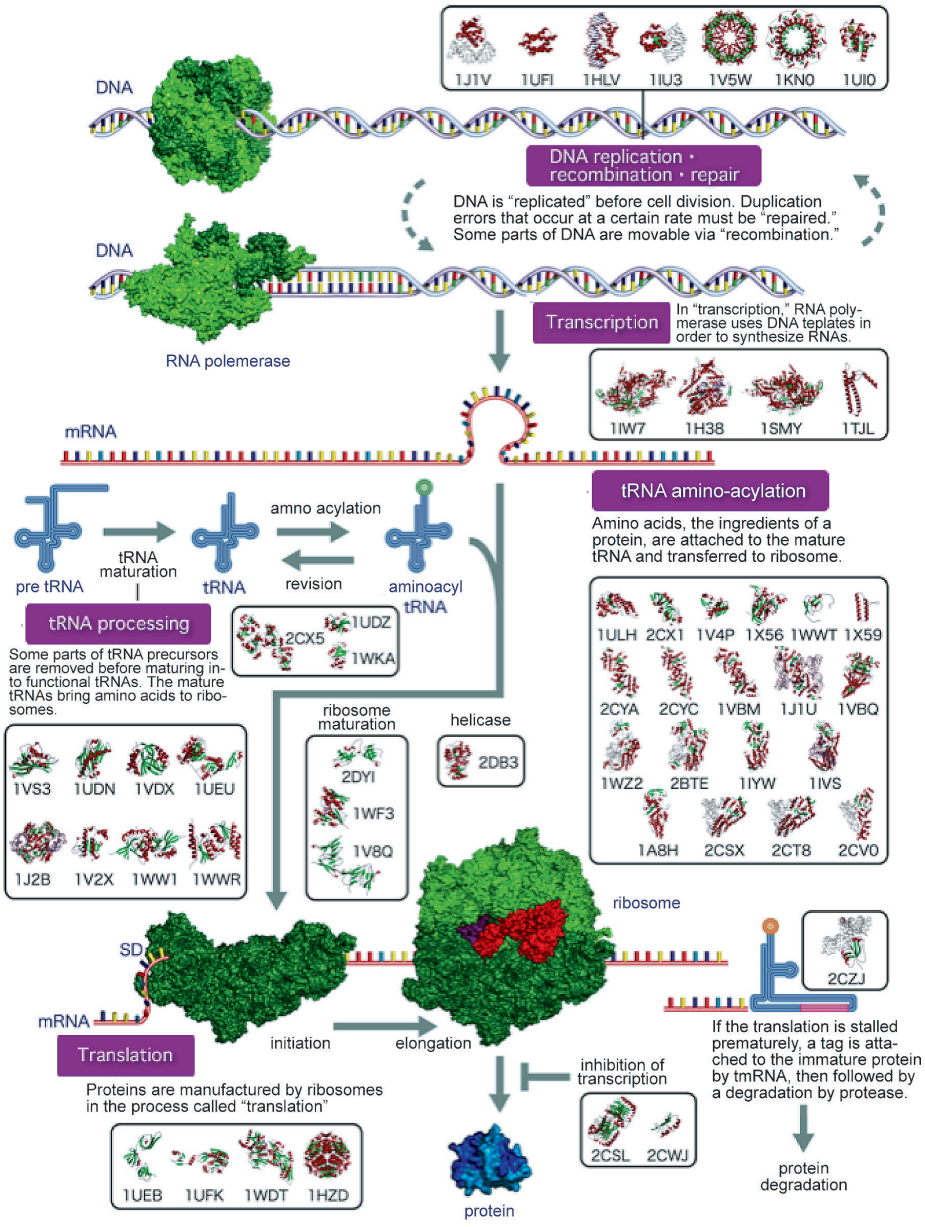

Fig. 1. Comprehensive structural analyses of the DNA replication/repair, transcription, and translation systems. Almost all structures are determined by X-ray crystallography.

Progress Report	Proteins
(1) Plasmids for overexpression	2059
(2) Overproduction in *E. coli*	1450
(3) Purification	944
(4) Crystallization	682
(5) Data collection	460
(6) 3D structure	360
360 + (106) =	466

Fig. 2. Progress of the *Thermus thermophilus* whole-cell project, and the examples of the crystal structures. (data provided by Dr. Seiki Kuramitsu)

For such large-scale structural/functional protein studies, efficient, high-throughput protein production is essential. Flexible and automated protein preparation protocols have been adopted. Proteins that cannot be produced by conventional methods are earmarked for large-scale production by alternative methods. In RSGI, therefore, both cell-based and cell-free systems are adopted. Generally, the structures of large proteins (above 20 kDa) are determined by X-ray crystallography using SPring-8 synchrotron radiation, whereas those of the smaller proteins (below 20 kDa) are determined by NMR spectroscopy using the NMR facility. When NMR spectroscopy or X-ray crystallography alone is not sufficient for the satisfactory structure determination, the two methods can be combined.

RSGI has put enormous effort on the automation and refinement of the system. This innovation have decreased the amount of manual labor, eliminated human error, improved data accuracy and maximized the efficiency and output of each experiment. We now can determine more than 800 protein structures per year. In this 5-year project, we successfully determined 1,333 X-ray structures and 1,342 NMR structures (total 2,675).

Our future challenge would be the technological development so that we can determine more medically-valuable proteins, membrane-spanning proteins, and macromolecular complexes.

Computer-based approaches, such as homology modeling and *in silico* screening, complement experimental approaches. These computer-based approaches are useful tools for the investigation of inter-molecular interactions, and molecular networks (ex. transcription, translation, signaling etc.). The marriage between experimental and computational approaches maximizes our capacity and makes us best prepared for the drug development efforts, for example, against emerging and reemerging infectious diseases. For future research for the drug development, 1-Petaflops computer specialized for molecular dynamics (MDGRAPE-3) has been developed.

Our research system and results will provide a good basis for various areas of life science researches.

Shigeyuki Yokoyama[a,b,c]

[a] Protein Research Group, Genomic Sciences Center, Yokoyama Institute, RIKEN
[b] SPring-8 / RIKEN
[c] Graduate School of Science, The University of Tokyo

E-mail: yokoyama@riken.jp