

Whole-Cell Project of *Thermus thermophilus* HB8 toward Atomic-Resolution Biology

Seiki Kuramitsu^{a,b,c}, Akio Ebihara^a, Mayumi Kanagawa^a, Noriko Nakagawa^{a,c}, Ryoji Masui^{a,c}, Kazutaka Murayama^b, Takaho Terada^{a,b}, Mikako Shirouzu^{a,b}, Kunio Miki^{a,d}, Shigeyuki Yokoyama^{a,e}, ^a*RIKEN Harima Institute at SPring-8*. ^b*RIKEN Genomic Sciences Center*. ^c*Department of Biology, Graduate School of Science, Osaka University*. ^d*Department of Chemistry, Graduate School of Science, Kyoto University*. ^e*Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Japan*. E-mail: kuramitu@bio.sci.osaka-u.ac.jp

In order to interpret the whole biological phenomena of the cell, we selected a model organism, *Thermus thermophilus* HB8. Its genome size is about 2 Mbp, and the number of its open reading frames (ORFs) is about 2,200 (<http://www.thermus.org>). Two-thirds of the ORFs are common to the genomes of most organisms including the human, and one-third of the ORFs are hypothetical proteins.

Plasmid construction for protein production has been completed for 2,000 ORFs. *In vivo* protein-production system of *E. coli* could successfully overproduced about 81% of the proteins. More than 85% of the purified proteins were successfully crystallized. For approximately 40% of the purified proteins, their diffraction data sets are of sufficient quality such that structural analysis is possible.

For the hypothetical proteins, structural analysis predicted the function with the success rate of about 60%. The function of the rest of the hypothetical proteins was estimated from transcriptome analysis, metabolome analysis, gene disruption experiments, and other methods as necessary.

Keywords: structural genomics, functional genomics, model organism