Genetics, Mishima, Japan. [b]*Graduate School of Science, Hokkaido University, Sapporo, Japan.* E-mail: hitou@lab.nig.ac.jp

Genomic-DNA analysis on number of organisms is now accomplished. Using these information, comprehensive structure analysis of transcription factors for their structure-based functional understanding is in progress. Our research target is transcription factors including putative ones from hyper-thermophilic archaeon *P.horikoshii*, mesophilic bacteria *C.glutamicum* and vertebrates *H.sapience.* Target proteins were cloned, over-expressed, purified and purified proteins were tried crystallization for X-ray crystal structure analysis. We had already succeeded in structure analysis of three of them, PH1161 [1] and PH1932 from *P.horikoshii* and CGL2612 from *C.glutamicum*. PH1161 protein is a homologue of bacterial transcriptional activator TenA, and PH1932 and CGL2612 are homologue of transcriptional repressor protein ArsR and QacR, respectively. As a further functional analysis, recognition DNA sequences of PH1932 and CGL2612 proteins were analyzed using the SELEX (Systematic Evolution of Ligand by EXponential enrichment) method. The SELEX suggested several consensus sequence of DNA recognized by these proteins, providing indispensable information to reveal their biological functions.

[1] Itou H., Yao M., Watanabe N., Tanaka I., *Acta. Cryst*, 2004, **D60**, 1094.

Keywords: structural genomics, transcription factor structure, X-ray crystal structure detemination

P.04.22.16

Acta Cryst. (2005). A**61**, C259

TIMomics: Genome-wide Search for Evolutionary Relationships among TIM (triose-phosphate isomerase) Fold Proteins *via* Structural Genomics Approaches

Xiao-Dong Su, *Department of Biochemistry and Molecular Biology, College of Life Sciences, Peking University, Beijing 100871, China*. E-mail: su-xd@pku.edu.cn

With more and more protein structures determined *via* world-wide efforts of structural genomics (SG), it becomes a common theme that many sequence unrelated proteins adopt the same folds. What are the origin and evolutionary pathways of these structure folds? How can we use this sort of information to predict protein structures with unrelated sequences? To answer these questions, we are trying to solve all possible TIM barrel proteins from a given genome. By using different methods and starting from SCOP TIM barrel PDB sets, we have searched exhaustively all potential TIM fold proteins from several complete genomes.

With the help of a high-efficiency and low-cost structural genomics platform set up at Peking University, China, we have chosen 288 (3x96) potential TIM fold genes from *B. subtilis* since 2005 Jan. as a pilot project for TIMomics. So far, we have got 259 genes PCR amplified and ready for subsquent cloning; a few dozen genes have already been cloned and expressed in *E. coli*, about 20 proteins have been purified and about 10 crystallized. We anticipate that we will be able to solve several dozens of protein structrures from the selected genes in the near future, in order to test our hypothesis and to study their structure, function and evolutionary relationships, and to answer the questions we proposed above.

Keywords: structural genomics, TIM barrel, TIMomics

P.04.22.17

Acta Cryst. (2005). A**61**, C259

Functional Discoveries from Crystal Structures of Proteins from *M. tuberculosis*

Edward N. Baker, Vickery L. Arcus, Kristina Backbro, Graeme L. Card, Jodie M. Johnston, Nayden Koon, J. Shaun Lott, Andrew A. McCarthy, Neil A. Peterson. *School of Biological Sciences, University of Auckland, New Zealand.* E-mail: ted.baker@auckland.ac.nz

Less than 50% of the gene products encoded in complete genome sequences can be annotated with firm biochemical functions. A primary goal of structural genomics then is to use protein structures for the discovery of function. Here we present some of the varied outcomes from crystal structure analyses of a selection of proteins from *Mycobacterium tuberculosis* which we have undertaken in the context of a laboratory-scale structural genomics project.

For two proteins, Rv1170 (MshB) and Rv3710 (LeuA), functions were known, but the crystal structures revealed metal and substrate binding sites from adventitious binding of ions or small molecules in the crystal. For Rv3853, which was annotated as the methyltransferase MenG, the crystal structure showed clearly that this function was incorrect. A fourth protein, Rv1347c, proved to be a CoA-dependent acyltransferase of the GCN5 family, but the crystal structure and associated bioinformatic analyses suggested a role in siderophore biosynthesis instead of the annotated function of antibiotic resistance. Finally, PAE2754, of previously unknown function, was found to be a metal-dependent nuclease that was representative of a large family of related proteins with major implications for TB biology [1].

[1] Arcus V.L., Backbro K, Roos A., Daniels E.L., Baker E.N., *J. Biol. Chem*., 2004, **279**, 16471.

Keywords: structural genomics, *Mycobacterium tuberculosis*, protein function

P.04.22.18

Acta Cryst. (2005). A**61**, C259-C260

Establishing High-throughput Protein Structure Determination Pipeline for Structural Genomics

Andrzej Joachimiak[1], Rongguang Zhang[1], Youngchang Kim[1], Jerzy Osipiuk[1], Marianne Cuff[1], Changsoo Chang[1], Boguslaw Nocek[1], Andrew Binkowski[1], Marcin Cymborowski[2], Krzysztof Lazarski[1], Maksymilian Chruszcz[2], Roman Laskowski[3], Janet Thornton[3], Norma Duke[1], Frank Rotella[1], Zbyszek Otwinowski[4], Alexei Savchenko[5], Aled Edwards[5,6], Wladek Minor[2], [1]*Midwest Center for Structural Genomics and Structural Biology Center, Biosciences, Argonne National Laboratory, 9700 South Cass Ave. Argonne, IL 60439.* [2]*University of Virginia, Charlottesville, VA 22908, USA.* [3]*European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.* [4]*University of Texas, Southwestern Medical Center, Dallas, TX 75390 USA.* [5]*Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario M5G, Canada.* [6]*Clinical Genomics Centre/Proteomics, University Health Network, 101 College St., Toronto, Ontario, M5G 1L7*. E-mail: andrzejj@anl.gov

Genome projects provide comprehensive access to genomic sequence information. The accumulation of sequence data has accelerated significantly, currently 1,386 genome projects are under way, sequences of 256 genomes have been completed, annotated, and available to the public. Many aspects of protein function, including molecular recognition, assembly and catalysis, depend on the 3D atomic structure. Protein structural analysis also contributes to an understanding of the evolutionary and functional relationships among protein families that are not apparent from the genome sequences. However, the structural coverage of proteins coded by new genomes remains low. The structural genomics efforts were initiated to increase structural coverage of proteomes in a rapid and cost-effective manner. Structural genomics programs contribute several tools: (1) comprehensive dictionary of high-resolution protein structures determined experimentally by x-ray crystallography and nuclear magnetic resonance (NMR); (2) comprehensive library of recombinant protein expression clones representing protein structures and functions; (3) methods for automated, HTP protocols of molecular and structural biology; and (4) functional information derived from structure.

Toward these goals the Midwest Center for Structural Genomics (MCSG) has established a protein structure determination pipeline using x-ray crystallography and synchrotron radiation. The current MCSG pipeline integrates all essential experimental and computational processes. Public databases of genomic sequences are being analyzed and targets are selected for structural studies. The MCSG pipeline generates well-characterized protein target expression strains, produces milligram quantities of proteins and heavy-atom labeled crystals. The cryoprotected crystals of x-ray quality are used for data collection at the synchrotron beamlines and structure determination using semi-automated SAD or MAD approach.