structure. The large domain shows (βα)8 barrel motif and the small domain suggests structural similarity to cyclophilin A.

Keywords: structural genomics, high-resolution protein structures, domain structure

P.04.22.11

Acta Cryst. (2005). A**61**, C258

A High-efficiency, Low-cost Platform for Structural Genomics Studies at Peking University

Yu-He Liang[a], Lanfen Li[a], Xiaoyan Zhang[a], Dan Li[a], Xiaofeng Zheng[a], Yuhui Dong[c], Peng Liu[c], Dingchang Xian[c], Jie Nan[a], Jia-Wei Wang[b], Hai-Fu Fan[a,b], Xiao-Dong Su[a], [a]*Department of Biochemistry and Molecular Biology, College of Life Sciences, Peking University, Beijing 100871, China.* [b]*Institute of Physics.* [c]*Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China*. E-mail: su-xd@pku.edu.cn; liangyh@pku.edu.cn

A high-throughput, high-efficiency and low-cost platform based on Beckman-Coulter robotic system Biomek FX for structural genomics has been set up. Several projects of structural genomics are in processing. Now, the platform has a capacity to process more than 1000 genes/year for structural and functional analyses. *B. Subtilis,* a model ogranism for Gram-positive bacteria and *S. Mutans,* the primary pathogen of dental caries were selected as our main target sources. So far, more than 450 *B. subtilis* and 250 *S. mutans* proteins and some proteins from other sources were selected as targets for this platform, the selected targets are mainly related to important metabolism pathways, and/or of potential for drug design. Up to 2005 Jan., 20 protein structures from the selected targets were determined, among them, eight structures were determined ab-initio. The application of beamline at **BSRF** (Beijing Synchrotron Radiation Facilities) and the ***OASIS-2004*** program have been crucial components for the operation of our platform. The use of SAD (single-wavelength anomalous diffraction) phasing methods combined with direct methods in ***OASIS-2004*** has increased the efficiency significantly, 5 out of 8 ab-initio determined structures have been solved this way.

Keywords: structural genomics, BSRF, OASIS-2004

P.04.22.12

Acta Cryst. (2005). A**61**, C258

Progress in the Whole Cell Project of a Model Organism, *Thermus thermophilus* HB8

Akio Ebihara[a], Noriko Nakagawa[a,b], Mayumi Kanagawa[a], Shinya Satoh[a], Yoshihiro Agari[a], Nobuko Maoka[a], Hitoshi Iino[a], Aiko Kashihara[a], Chizu Kuroishi[a], Ryoji Masui[a,b], Mikako Shirouzu[a,c], Takaho Terada[a,c], Kunio Miki[a,d], Shigeyuki Yokoyama[a,c,e], Seiki Kuramitsu[a,b,c], [a]*RIKEN Harima Institute at SPring-8.* [b]*Department of Biology, Graduate School of Science, Osaka University.* [c]*RIKEN Genomic Sciences Center.* [d]*Department of Chemistry, Graduate School of Science, Kyoto University.* [e]*Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Japan.* E-mail: ebihara@spring8.or.jp

One of the long-term goals of structural and functional genomics is the interpretation of all fundamental biological phenomena at atomic resolution. An extremely thermophilic bacterium, *Thermus thermophilus* HB8, is a promising model organism for structural and functional studies, because of the small genome size, the availability of genetic tools for functional analysis, and the thermostability of its proteins. Toward this aim, the "Whole Cell Project" of this bacterium is currently in progress (http://www.thermus.org/). The complete genome sequence identifies approximately 2,200 ORFs, and about 2,000 expression plasmids have been constructed. The target proteins were overproduced in *E. coli*, purified, crystallized, and characterized by X-ray crystallography, through which about 200 protein structures have been solved. As part of functional studies, we have constructed the gene disruption plasmids using the thermostable selective marker (kanamycin resistance) and analyzed mRNA by the DNA microarray system.

Keywords: structural genomics, functional genomics, *Thermus thermophilus*

P.04.22.13

Acta Cryst. (2005). A**61**, C258

Structural Proteomics of Proteins Coded by the *cag* PAI of *Helicobacter pylori*

Giuseppe Zanotti, Laura Cendron, Anke Seydel, Alessandro Angelini, Roberto Battistutta, *Department of Chemistry and VIMM, University of Padua, Italy.* E-mail: giuseppe.zanotti@unipd.it

H. pilory is a Gram-negative bacterium that colonizes the stomach of probably half of the human population. It is associated with gastritis, peptic ulcers and mucosa-associated lymphoid tissue lymphomas. Many factors contribute to the virulence of *H. pylori* [1]. Among them, the enzyme urease, the Neutrophil Activating Protein, NAP [2] and the secreted protein toxin VacA. However, the major genetic difference in HP isolates is the presence or absence of a specific pathogenicity island, named *cag*-PAI. It is a 40-kb locus that contains about 30 ORFs, whose function is unknown, with few exceptions.

We have cloned, expressed, and purified several proteins of the *cag* pathogenicity island of *H. pylori*. They all have been expressed in *E. coli.* We have already solved the structure of CagZ, using the Se-Met method [3] and the structure will be described in detail. We have also obtained crystals of a second protein, CagS, and its structure determination is in progress, along with crystallization tests on other *cag* proteins. Our final goal is to determine, in collaboration with other groups [4], most of the proteins coded by the *cag*-PAI island.

[1] Covacci, et al., *Science*, 1999, **284**, 1328. [2] Zanotti, et al., *J. Mol. Biol*., 2002, 323, 125-130. [3] Cendron L., Seydel A., Angelini A., Battistutta R., Zanotti G., *J. Mol. Biol.*, 2004, **340**, 881. [4] *The Helicopter Structural and Molecular Biology Consortium.*

Keywords: structural genomics, bacterial pathogenesis, MAD phasing

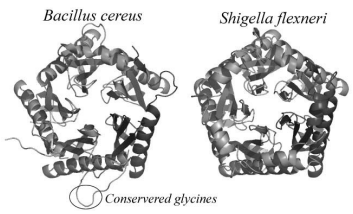
P.04.22.14

Acta Cryst. (2005). A**61**, C258

Trying To Deduce Function From Structural Variability And Conservation

Joseph S. Brunzelle[a], Xaiojing Yang[b], George Minasov[b], Ludmilla Shuvalova[b], Wayne F. Anderson[b], [a]*LS-CAT, APS, Argonne National Laboratory.* [b]*Dept of Mol Pharm and Bio Chem, Northwestern University.* E-mail: j-brunzelle@northwestern.edu

A comparative case is presented for two similar proteins from *Shigella flexneri* and *Bacillus cereus* which are homologs of the *E. coli* protein ybjQ. These two proteins are members of the COG0393, a widely conserved family of proteins in bacteria and archaea that are functionally uncharacterized. All members of the sequence family are about 100 residues. The two examples presented are both homopentamers and have 54% sequence identity. Despite the high sequence identity, the *B. cereus* protein contains regions of structural variation. A sequence alignment of the protein family reveals a pair of conserved glycines at residues 44 and 45. These conserved glycines are located in a loop that it is a region of structural variation in the *B. cereus* protein. This area of structural variation has been predicted as a region of disorder from the DisEMBL server which may be important to the function of these proteins.



Keywords: structural genomics, conformational change, macromolecular structure

P.04.22.15

Acta Cryst. (2005). A**61**, C258-C259

Comprehensive Structure-based Functional Analysis on Transcription Factors

Hiroshi Itou[a], Ui Okada[b], Min Yao[b], Nobuhisa Watanabe[b], Isao Tanaka[b], [a]*Biomolecular Structure Laboratory, National Institute of*

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.