MS15 THE BIG QUESTION IN STRUCTURAL GENOMICS: GETTING FUNCTION FROM STRUCTURE
Chairpersons: Sung-Hou Kim, Edward Neill Baker

MS15.25.1
Acta Cryst. (2005). A**61**, C25
Crystal Structure of Rv0813c Reveals a New Family of Putative FABPs
William Shepard[a], A. Haouz[b], A. Buschiazzo[b], J.M. Betton[b], S.T. Cole[b], P. Alzari[b], [a]*ESRF, 6 rue Jules Horowitz, BP 220, 38043 Grenoble cedex 9, France,* [b]*Institut Pasteur 25 rue du Dr. Roux, 75724 Paris cedex 15, France.* E-mail: shepard@esrf.fr

Rv0813c is a protein of unknown function that we have selected as a target for crystallographic studies in the context of a structural genomics effort on tuberculosis. The crystal structure of Rv0813c, a conserved protein in *M. tuberculosis*, reveals a new family of putative fatty acid binding proteins (FABPs). Rv0813c adopts a 10-stranded beta barrel fold, which closely resembles those of the FABPs found in eukaryotes. This is in fact the first FABP-like protein to be found in prokaryotes. However, Rv0813c lacks the double helix insert of FABPs that covers the entry to the binding site. The beta barrel forms a deep cavity, where a small ligand, which appears to be a morpholine, binds to the phenol hydroxyl group of Tyr192. This tyrosine corresponds to a RxY motif, which forms part of the binding site in FABPs. Furthermore, a network of H bonds, hydrophobic residues and an internal salt bridge surround the binding site and define the shape of the cavity. Most of these residues are well conserved in homologous proteins. Phylogenetic studies show that this family of FABP-like proteins is represented in GC-rich prokaryotes. The structural analysis of Rv0813c suggests that this cytoplasmic protein may have a role in fatty acid transport, storage or signaling. This work supports the notion that high resolution structural studies can provide strong leads as to the biochemical function(s) of the protein.
Keywords: structural genomics, mycobacteria, fatty acid binding protein

MS15.25.2
Acta Cryst. (2005). A**61**, C25
Chasing the Function of Hypothetical Proteins
Osnat Herzberg[a], Gary Gilliland[b], John Orban[a], Andrew Howard[c], John Moult[a], [a]*Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, MD, USA.* [b]*Center for Advanced Research in Biotechnology, National Institute for Standards and technology, Rockville, MD, USA.* [c]*Biological, Chemical, and Physical Sciences, Illinois Institute of Technology, Chicago, IL, USA*. E-mail: osnat@carb.nist.gov

The focus of the Structure2Function program at CARB during the first five years was on prokaryotic proteins of unknown function, mostly from *Heamophilus influenzae* and *E. coli*. The goal was to leverage the 3-D structural information to gain insight into the biochemical function. Approximately 75% of the new structures exhibit folds that have been seen before, and in many cases we were able to make new discoveries about function by utilizing standard assays and mining genome context. The remaining 25% structures defined new folds, a somewhat higher fraction of new folds compared with structures entering the PDB. In some of these cases, the function became known around the time the structure was determined, while in other cases we were able to make broad predictions about the function.
Keywords: structural genomics, hypothetical proteins, crystallography

MS15.25.3
Acta Cryst. (2005). A**61**, C25
Functional & Structural Proteomics of SARS: Defining a Rational Response to Emerging Diseases
Jeremiah Joseph[a], Alexei Brooun[a], Benjamin Neuman[a], Enrique Abola[a], James Stevens[a], Kumar Saikatendu[a], Margaret Johnson[a], Michael Recht[b], Michelle Kraus[a], Mike Nelson[a], Renaud Burrer[a], Sophie Coon[a], Vanitha Subramanian[a], Weizhong Li[c], Adam Godzik[c], Ian Wilson[a], Kurt Wuthrich[a], Mike Buchmeier[a], Raymond Stevens[a], Richard Bruce[b], Ron Milligan[a], Peter Kuhn[a], [a]*The Scripps Research Institute, La Jolla, CA, USA.* [b]*Palo Alto Research Center, Palo Alto, CA, USA.* [c]*The Burnham Institute, La Jolla, CA, USA.* E-mail: jjoseph@scripps.edu

Rapid rational therapeutic and prophylactic responses are crucial when faced with new infectious diseases. The emergence of the coronavirus responsible for the Severe Acute Respiratory Syndrome (SARS) tested the utility of post-genomic technologies to characterize and combat this virus. While virus identification and complete genome sequencing took mere weeks, they have been tough acts to follow for drug and vaccine development. We have undertaken a multi-pronged initiative to understand and address precisely this bottleneck using a structural and functional proteomics approach involving bioinformatics, structural biology (X-ray crystallography, NMR, cryo-electron microscopy), genetic approaches (site-directed mutagenesis, antisense functional mapping, microarray-based functional mapping), and macromolecular interaction studies (nano-calorimetry, ligand-fishing techniques, mass spectrometry) to generate a structure-function-interaction map of the entire proteome of the SARS-CoV and its interactions with the host cell. This presents an exciting and comprehensive set of targets for rational, structure-based drug and vaccine design, defining a paradigm adoptable for *any* emerging infectious disease.

We designed multiple constructs of the 28 SARS-CoV ORFS for expression in *E. coli*, baculovirus and mammalian systems. Over 150 constructs have been processed by high-throughout protein expression and purification. Of the 31 expressing constructs, nano-volume crystallization has produced crystal hits for 5. One crystal structure and one NMR structure has been determined. Cryo-electron microscopy has characterized the packing arrangement of the S, M and N proteins in the virion. The challenge is that most SARS proteins are involved in intimate protein-protein, protein-membrane, or protein-RNA interactions which must be understood for a complete description of its biology.
Keywords: structural genomics, viral structure and function, intermolecular interactions

MS15.25.4
Acta Cryst. (2005). A**61**, C25-C26
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 theromphilus* 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