MS02 O1

The Oxford Protein Production Facility; a biomedically driven structural proteomics project
R.J.Owens, K. Harlos, RE Esnouf, J.Grimes, DK Stammers, EY Jones, DI Stuart *Structural Biology Division, Henry Wellcome Building for Genomic Medicine, University of Oxford;*
E-mail: ray@strubi.ox.ac.uk

Keywords: high throughput, protein production, Crystallization

High throughput sequencing of eukaryotic, viral and bacterial genomes is providing a huge database of proteins with potential for structure-function analysis. In response to this opportunity, structural proteomics projects, including the Oxford Protein Production Facility have been established focused on high throughput (HTP) structure determination. Crucial to this effort has been the development of technologies for HTP protein production and crystallization. For all projects, there has been an emphasis on parallel processing driven by the need to accommodate relatively large numbers of potential targets at an acceptable cost. The OPPF, in common with most other groups has set up semi-automated liquid handling systems to carry out some of the protocols for protein production [1] and crystallization [2]. However, many of the methods can equally well be carried out manually with appropriate equipment e.g. multi-channel pipette dispensers. The motivation to implement automation is largely to enable processes to be scaleable and sustainable as error-free operations. Target selection in the OPPF is focused on human proteins and those of human pathogens, both viral and bacterial, selected for their direct biomedical relevance. Recent technical developments in OPPF to address some of these challenging targets will be reviewed.

[1] Berrow N, Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, DI, Owens, RJ*, Nucleic Acids Res*. 2007 March 12, doi:10.1093/nar/gkm047.
[2] Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC,Berrow NS, Brown J, Berry IM, Stewart-Jones GB, Grimes JM, Stammers DK, EsnoufRM, Jones EY, Owens RJ, Stuart DI, Harlos K. (2005) Acta Crystallogr D Biol Crystallogr. 61: 651-7.

MS02 O2

Application of high throughput ligand screening to promote protein crystallization
Abdellah Allali-Hassani, Aled M. Edwards, Cheryl Arrowsmith and Masoud Vedadi**.** *Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada*

Keywords: Structural genomics, thermal stability, ligand screening.

Structural genomics efforts have led to the expression of thousands of proteins, many of which have not been purified or characterized previously. Identification of small molecules that bind to and stabilize these proteins can promote their crystallization as well as provide valuable functional information. We have employed differential static light scattering (DSLS) and differential scanning fluorimetry (DSF) to investigate the thermostability and ligand binding specificity of our protein targets. Optimum buffer conditions which further stabilized the aggregating or hard to concentrate proteins often resulted in more soluble proteins which were further concentrated. Presence of identified ligands in many cases resulted in crystallizing those hard to crystallize proteins and improving crystal quality leading to structure determination. Screening different members of families of proteins against customized libraries of compounds also resulted in binding profiles for each protein and providing the opportunity to compare small molecule binding specificity of different members of each family of protein. The screening results facilitated the comparison of substrate specificities and also identified compounds which appeared to be general inhibitors for each of these protein families. Moreover, other compounds were discovered that only bind to a subset of proteins in each family of proteins and thus appear to discriminate among different members of the family.

MS02 O3

Split-GFP as a Tool for Finding Soluble Protein Domains
Jean-Denis Pédelacq, Stéphanie Cabantous, Thomas C.Terwilliger, Geoffrey S. Waldo, *Structural Biology Group, Bioscience Division, Los Alamos National Laboratory, Los Alamos NM 87545. USA.*.
E-mail: jpdlcq@lanl.gov

Keywords: solubility reporter, domain trapping, structural genomics.

The high-throughput detection and quantification of proteins *in vivo* and *in vitro* remains challenging. To help address these issues, we have developed a quantitative, fluorescent universal protein tagging and detection system based on self complementing fragments of the Green Fluorescent Protein (GFP) [1,2,3]. These fragments originally poorly folded have been engineered for improved solubility and complementation. Using a set of control proteins, we validated the ability of the system to quantify soluble and total protein expression *in vitro* and *in vivo*.
As part of our structural genomics initiatives on *Mycobacterium tuberculosis**, we applied this system to the identification of soluble domains of large multi-domain proteins that failed to crystallize. We have developed a new domain trapping strategy which includes several selection steps, enabling to reduce progressively the throughput and increase the fidelity of our screening process. The first step consists of selecting DNA fragments containing in frame inserts regardless of protein solubility. This is achieved using a DHFR insertion construct which confers bacteria survival in the presence of selecting agent only when “in frame” fragments are cloned in the DHFR scaffolding. We applied the split-GFP tagging system to select soluble clones from the selected pool of “in frame” variants. The most soluble clones are sorted by a first solubility screening *in vivo.* A second screening *in vitro* enables the selection of a smaller subset of domains that will be further characterized by DNA sequencing to determine precise boundaries. We demonstrate the ability of the ORF-selector combined with the split GFP solubility screen for creating repertoires of soluble protein fragments from several intractable proteins and genomes.

TB Structural Genomics Consortium (http://www.doe-mbi.ucla.edu/TB/) & The Integrated Center for Structure and Function Innovation: a PSI-2 Specialized Technology Center (http://techcenter.mbi.ucla.edu/).

[1] Cabantous S., Terwilliger T.C., Waldo G.S., & Pédelacq JD, *BIOforum Europe*, 2006, 10, 26-28.
[2] Cabantous S., Pédelacq J.D., Mark B.L., Naranjo C., Terwilliger T.C. & Waldo G.S., *Journal of Structural and Functional Genomics,* 2005*,* 6 (2-3): 113-119.
[3] Cabantous S., Terwilliger T.C., & Waldo G.S., *Nature Biotechnology*, 2005, 23 (1): 102-7.

MS02 O4

High throughput structural genomics and its general applicability to structural biology Ian A. Wilson, *Joint Center for Structural Genomics (JCSG), La Jolla, CA 92037 USA*. E-mail: wilson@scripps.edu

Keywords: Structural Genomics; Metagenomics; Microbiome

Over the past 7 years, the JCSG (www.jcsg.org) has evaluated, developed, adapted and integrated various methodologies and technologies into a high throughput (HT) production pipeline for all steps from target selection, cloning, expression, crystallization to structure determination. The pipeline which was initially developed using the full proteome screen of *T. maritima* as our main source of targets, now forms the base of our current production pipeline. As one of the 4 PSI production centers, we have focused on large Pfam families without structural coverage, as well as on proposing and validating new protein families, that have recently been add to Pfam. Our biomedical theme project revolves around, The Central Machinery of Life, proteins that are conserved in all kingdoms of life. New and exciting projects in our target portfolio are focused on metagenomes, in particular, the Global Ocean Sampling (GOS) and human microbiomes. To date. we have processed over 11.000 targets, solved over 435 structures and deposited over 405 in the PDB. The *T. maritima* effort has resulted in 163 structures (total unique structures in PDB: 282) that now account for a 15% direct structural coverage of its predicted soluble proteome. Moreover, of its 1877 predicted ORFs, only 95 proteins (5% of the predicted soluble) are currently without fold predictions. As a result of processing such a large number and variety of targets through our HT pipeline, we have been able to test and validate various strategies to increase efficiency, improve yield, and lower cost per structure. This large scale focused studies have resulted in adding a number of routine salvage pathways to our pipeline. Many of these advances in technologies and methodologies can easily be implemented on a smaller scale in individual structural biology labs. The JCSG, located at The Scripps Research Institute, Genomic Institute of the Novartis Research Foundation, U.C. San Diego, Burnham Institute, and the Stanford Synchrotron Radiation Laboratory/Stanford University, is supported through the NIH Protein Structure Initiative (U54-GM074898), (www.nigms.nih.gov/psi).

MS02 O5

High throughput crystallography in fragment based drug discovery Mladen Vinković, *Astex Therapeutics, Cambridge,UK*.
E-mail:m.vinkovic@astex-therapeutics.com

Keywords: automated crystallization, automated ligand refinement, fragment screening

Fragment based drug discovery is becoming increasingly popular among pharmaceutical and biotech companies. Astex is a pioneer in that field and in the use of protein X-ray crystallography for fragment screening. The advantage of fragments, small molecules with 100<Mr<300, is that they have a better chance of achieving high ligand efficiency than the larger molecules used in traditional HTS. In addition they can be grown to larger and potent molecules while keeping physical, chemical and pharmacological properties in check.
X-ray crystallography has been traditionally used in the lead optimization stage of the drug discovery process to clarify SAR and aid ligand design. Introduction of protein crystallography into the screening phase, where many hundreds of data sets are collected and processed, requires a reduction in the time for solution and refinement of protein-ligand complexes from days to hours. That has been achieved through integration of commercially available crystallization [1] and data collection robotics with our corporate database, as well as by development of an *in house* software platform for automated interpretation of diffraction data (AutoSolve) [2]. A web based GUI enables necessary process flexibility together with model inspection and rebuilding through Astex Viewer [3].
Such a high throughput platform enables the generation of 10s to 100s of hits (protein-fragment crystal structures) and structural protein-ligand elucidation of protein complexes for every relevant compound synthesized in subsequent hit and lead optimization steps, therefore greatly aiding ligand design. Rapid progress from mM hits to nM leads and subsequently drug candidates has been achieved on a number of targets.

[1] Vinković M., *Insights*, 2005, 19, 15.
[2] Mooij W.T.M., Hartshorn M.J., Tickle I.J., Sharff A.J., Verdonk M.L., Jhoti H., *Chem. Med. Chem.*, 2006, 8, 827.
[3] Hartshorn M.J., *J. Comput. Aided Mol. Des.*, 2002, 16, 871