

Meeting report

## Structural genomics: a new era for pharmaceutical research

Yang Liu, Nicholas M Luscombe, Vadim Alexandrov, Paul Bertone, Paul Harrison, Zhaolei Zhang and Mark Gerstein

Address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA.

Correspondence: Mark Gerstein. E-mail: mark.gerstein@yale.edu

Published: 14 January 2002

*Genome Biology* 2002, **3**(2):reports4004.1–4004.3

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/2/reports/4004>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

---

A report on the 15th Annual Center for Advanced Biotechnology and Medicine Symposium on structural genomics in pharmaceutical design, Princeton, USA, 24-25 October 2001.

---

The 15th Annual Center for Advanced Biotechnology and Medicine (CABM) symposium on structural genomics in pharmaceutical design gathered 20 keynote speakers, who presented the most recent advances in their structural genomics projects carried out in both academic institutions and companies. Here, we will give an overview of some of the most interesting talks.

### High-throughput structural and functional genomics

After recent successes in genome-sequencing projects, initiatives in structural genomics aim to understand fully the biological role of proteins by determining representative structures for protein families on a genomic scale. Several investigators, including Stephen Burley (Rockefeller University, New York, USA), Andrzej Joachimiak (Argonne National Laboratories, USA), Gaetano Montelione (Rutgers University, Piscataway, USA and CABM), Wim Hol (University of Washington, Seattle, USA) and Sean Buchanan (Structural GenomiX Inc., San Diego, USA), talked about their ongoing structural genomic projects, most of which are funded by the US National Institutes of Health (NIH). Automation appears to be a critical factor for all of the large-scale structure determination projects discussed here, as protein-structure determinations traditionally rely on very labor-intensive methods of gene cloning, protein expression and purification, crystallization, and structure determination and refinement.

High-throughput protocols are now changing this aspect of structural biology.

Joachimiak described an automated protocol for data collection and structure determination. Using the advanced photon source (APS), the third-generation synchrotron radiation facility, his group has made great advances in accelerating progress in structure determination. The procedure they have automated includes: firstly, mounting crystals and adjusting positions on a beamline; secondly, collecting diffraction data; and thirdly, structure determination and refinement. By following this protocol, most structures are solved to high resolution within a very short time period.

Hol described the on-going efforts in his lab towards structure-based drug-design, with a focus on drugs for diseases caused by tropical parasites. He also reported the approval and funding of the latest NIH-sponsored structural genomics initiative, the Structural Genomics of Pathogenic Protozoa (SGPP); it will focus on microbial pathogens such as *Trypanosoma brucei* and *Plasmodium falciparum*, which cause sleeping sickness and malaria, respectively. In contrast to the other structural genomics centers, SGPP selects potential targets on the basis of function, which should accelerate the determination of pathogenic protein structures for drug design.

As the central repository of protein structures, the Protein Data Bank (PDB) [<http://www.rcsb.org/pdb/>] plays a critical role in structural genomics initiatives. Helen Berman (Rutgers University, Piscataway, USA) described on-going developments at the PDB. The database now includes more than 16,000 structures, and new entries are being added at a rate of 50-70 structures per week. About 90% of the new entries are protein structures, and the average sizes of these structures have gradually increased over time. One major

improvement to the database is the automated processing of submissions to ensure uniformity in the data files and quality of the structures, and a future goal is to allow seamless deposition of data generated by structural genomics consortia. Berman also mentioned the development of the Ligand Depot database; it contains information for about 250,000 small molecules, many of which are found in complexes in PDB structures. Combined with the protein structure data in PDB, the Ligand Depot will be very useful for drug-design research in the future.

### Predicting structure and function

In addition to structure determination, structure prediction provides alternative ways to look at protein structures and understand the biological functions of proteins. Jeff Skolnick (Donald Danforth Plant Science Center, St. Louis, USA) described a novel unified method for structural prediction that combines in one the three most used methods - homology modeling, threading and *ab initio* prediction. The unified method has the potential to predict active sites from *ab initio* models of protein structure, using a library of three-dimensional active-site descriptors. Furthermore, this method has been extended to the problems of multimer threading (the prediction of protein complexes) and even the prediction of metabolic pathways.

Ming-Ming Zhou (Mount Sinai School of Medicine, New York, USA) detailed progress in designing ligands for bromodomains, protein motifs involved in chromatin re-modeling and binding to acetylated histones. A procedure for assessing the efficacy of drug binding on a large scale was presented by Ray Salemme (3-Dimensional Pharmaceuticals Inc., Exton, USA): the change in melting temperature upon a ligand binding to a protein target was measured in a high-throughput screen.

Ronald Levy (Rutgers University, Piscataway, USA) then presented his laboratory's recent progress in *ab initio* tertiary structure prediction. His work incorporates protein structural data to guide an all-atom molecular dynamics simulation, using residual dipolar couplings (a method in nuclear magnetic resonance structure determination) to generate a model of the target protein backbone. A key feature of this method involves searching the SCOP database [<http://scop.mrc-lmb.cam.ac.uk/scop>] with seven-residue subsets of the target protein, in order to find a set of dipolar couplings that most closely match each query fragment. The predicted backbone is then assembled from candidate proteins according to their local fit to the dipolar coupling data. Side-chain placement is accomplished using an energy-minimization routine to select optimal angles iteratively from a rotamer library, until the algorithm converges. Because the prediction strategy makes use of experimental data to narrow the conformational search, there are opportunities to integrate this approach into high-throughput structural genomics efforts.

### The role of bioinformatics in structural genomics

In a field involving large quantities of biological data, bioinformatics provides important guidance for structural genomics projects, from target selection to data analysis. Two bioinformatics talks given at the symposium are worthy of mention. One was by M.G., who described the bioinformatics studies being conducted by our group, and in particular research on pseudogenes. Although typically there are many pseudogenes in eukaryotic genomes, very little is known about them compared with the functional genes. Research on pseudogenes investigates the occurrence of genes that have lost their function through the acquisition of a premature stop codon. Combined with the recent development of a mathematical model that describes the process by which genomes evolved to their present state, the research provides a glimpse into the evolutionary history of the organism.

The second bioinformatics talk was by Cyrus Chothia (MRC Laboratory of Molecular Biology, Cambridge, UK), who discussed several mechanisms of molecular evolution with respect to the observed repertoire of protein families. A protein family is defined as a group of proteins that share a common evolutionary ancestor. The recent sequencing of entire genomes enables researchers to identify the members of these families, and to reveal the biological processes that contributed to their divergence. Two processes that appear to be instrumental in protein evolution are gene duplication and combination. A goal of Chothia's work is the elucidation of relationships in sequence and structure between protein families in a genomic context, which is expected to shed light on the molecular basis of organismal evolution.

### The role of structural genomics in pharmaceutical drug design

Recent advances in structural genomics not only help us to understand protein functions but also have a big impact on the pharmaceutical industry. In the last few years, the use of protein structural information in drug discovery research has matured, and it is now used at all levels, ranging from genomics-derived target identification and selection to the final design of suitable drug candidates. An especially powerful methodology has arisen from the synergy of target structural information with combinatorial chemistry.

An excellent example of how structural genomics and combinatorial chemistry can be used in rational drug design was presented by Edward Arnold (Rutgers University, Piscataway, USA and CABM), who discussed the optimization of multiple factors in developing potent inhibitors for the reverse transcriptase (RT) enzyme of human immunodeficiency virus (HIV), as potential drugs to prevent the development of acquired immunodeficiency syndrome (AIDS). HIV RT is vital for replication of the virus and is therefore rightly considered as a primary target for developing drugs. On the basis of crystal structures of HIV RT and its ligand-bound

forms, two primary types of target site were identified: the dNTP-binding site (used when RT carries out its polymerase function) and the non-nucleoside RT-inhibiting site. Non-nucleoside RT-inhibitors (NNRTIs) were found to be especially suitable for disrupting HIV RT function. NNRTI-targeted drugs should be chemically very diverse and should not compete with nucleotide-binding substrates, while inhibiting RT in nanomolar concentrations. To identify the most potent NNRTI drugs, Arnold and colleagues used molecular modeling to calculate the interaction energy of a potential substrate bound at the non-nucleoside RT-inhibiting site, which consists of a hydrophobic pocket. It turns out that the lower energy conformations usually produce the more potent drug. Selected targets are synthesized and then assessed through antiviral screening experiments. This procedure allowed faster design of a number of powerful NNRTI drugs. The most successful candidates, such as TMC120 and TMC125, appear to have low toxicity (in the micromolar range) and high selectivity. Monotherapy experiments with each of these produce up to 30-fold drop in the virus population *in vitro*.

There were many interesting talks in this symposium covering topics such as structural genomics, bioinformatics, computational biology, and drug design methods. Several NIH-funded structural genomics pilot projects have shown promising preliminary results in their first year of funding, with about 100 structures solved. These projects will do even better in years to come and will accelerate the pace of structure determination, and are sure to have great impact on structural biology research and on drug design in the future.