Title

Cambridge Healthtech Institute's Fifth Annual Functional Genomics Conference
"Using a Systems Biological Approach to Develop Novel Therapeutics"
Organizer: Cambridge Health Institute
Oct. 9-10, Cambridge, MA

Author

Christian V. Forst
Mailstop M888
Los Alamos National Laboratory
Los Alamos, NM 875454
Tel: +1 (505) 665-5268
FAX: +1 (505) 665-3024
E-mail: chris@lanl.gov

Introduction

With the human genome sequence available and with an exponentially growing number of completely sequenced genomes from various model organisms, the research paradigm is shifted towards the analysis of the function of interacting genes and gene-product. The reduced estimate of 30,000-40,000 genes in the human genome strongly indicates that the functional complexity expected for an organism such as human may originate in locations and processes beyond the identified genes. Functional genomics emerges as valuable research area for the identification of key-players in the complex processes within a living organism. It is complemented by Systems Biology [7,103,104], a new research area that combines data obtained from Functional Genomics analysis with biocomputation and simulation of cellular networks. Both Functional Genomics and Systems Biology outline the analysis, prediction and control of complex biological systems.

Highlights

Systems biology has been the common thread that covers the specific research topics of the conference. High-throughput identification of cellular component, their individual function and interaction with other components on the one hand as well as prediction and inferences of networks on the other hand cover two important areas within Systems Biology. In the following, representative examples of presentations with respect to a system biological research approach will be presented.

Highlights - Computational Systems Biology

M Thompson (Protein Pathways, Los Angeles, CA, USA) introduced several computational techniques to annotate and reconstruct pathways within genomes [8]. These techniques exploit comparative genomics data as well as expression and literature data. Predictive methods based on genomic context information, such as phylogenetic profiles/co-occurrence of genes in genomes, conservation of local gene neighborhood and gene-fusion events as well as validative methods utilizing literature information are employed to identify *protein functional links*. Protein functional links, i.e., protein-protein interactions, protein complexes or enzyme members in subsequent steps of

pathways then serve as building-blocks for protein functional networks, i.e., inferred completely connected networks that assemble an upper estimate of a cellular network. Thompson has shown that protein functional networks aid in the annotation of unannotated genes as well as help in finding missing genes by identifying linkages with proteins that themselves link to alternative steps in a pathway. A database of interacting proteins based on experimental data has been established and is publicly accessible [101]. A step further into prediction of particular pathways from a given metabolic network has been pursued by BO Palsson (Genomatica and University of California, San Diego, CA, USA). Although Palsson was unable to attend the conference his research covers the identification of particular pathways within a metabolic networks with maximum metabolic flux [10]. By a linear algebra approach combined with combinatorics, extreme pathways, i.e., pathways with the maximal stationary throughput of metabolites are identified. Palsson and coworkers are able to identify and predict different metabolic phenotypes based on environmental and nutritional constraints [1]. CV Forst (Los Alamos National Laboratory, Los Alamos, NM, USA) developed a method, *Comparative Network Genomics*, for comparative analysis of cellular networks. This approach combines network connectivity information with sequence information of the cellular components involved [2]. Interesting aspect of this method is the identifications of similar pathways between organisms and the relationship with operon organization in microbial genomes. By Comparative Network Genomics between closely related pathogens specific, metabolic host-pathogen interactions can be identified [3].

Highlights – Functional Proteomics

Parallel to computational/theoretical efforts towards Systems Biology, experimental techniques emerge that provide high-throughput data for the analysis of biological systems. New developments include combination of biochemical assays with DNA-array chips or novel protein array techniques. A particular interesting approach has been presented by M Snyder (Yale University, School of Medicine, Newhaven, CT, USA). Snyder and coworkers developed a method to identify transcription factors (TFs) and their binding sites and to characterize the biochemical activities of the TFs in a yeast model on genomic scale. To identify TF targets Snyder *et al.* combined chromatin immunoprecipitation (ChIP) with microarray hybridization [6]. For the cell-cycle TF *Swi4*, Snyder identified 163 intergenic regions and 181 potential gene targets. Biochemical properties of gene-product are then analyzed by protein chips. Snyder outlined the benefit of protein chips by their capabilities to identify protein function, to track protein modification and regulation and for drug discovery. He also discussed limitations of protein array due to problems in the generation of high-quality expression clones. He also identified bottlenecks in the making of a sufficient amounts of proteins for the chips as well as in the array technology itself.

Also worthwhile to mention are presentations with respect to alternative splicing. WS Agnew (Johns Hopkins University, School of Medicine, Baltimore, MA, USA) discussed combinatorial problems that arise in the analysis of alternative spliced genes. In the case of an inositol-1,4,5-triphosphate receptor, a 63 exon gene, 19 different splice variants have been identified. C Case (Sangamo Biosciences, Richmond, CA, USA) presented a method to control endogenous genes expression and study alternative splicing by engineered transcription factors. Designed Zinc finger proteins are employed to activate and repress specific chromosomal loci in a broad range of cells, animals and plants. None of the latter two cases are capable for high-throughput analysis though.

Highlights – Systems Biology

A strategy for visualization of biochemical networks in living cells has been presented by SW Michnick (Université de Montréal, Montréal, QC, Canada). This strategy is based on Protein fragment Complementation Assays (PCA) [9] and allows quantitative probing of constitutive or induced protein–protein interactions in living cells. In PCA, designed fragments of an enzyme are fused to two interacting proteins. When the interacting proteins come together in space the fragments fold/reassemble into an active form, whose activity can be monitored time–resolved and in different subcellular locations. Michnick and coworkers applied this technique to identify signaling networks and to establish pharmacological profiles by differential screens of different inhibitors [11]. Michnick outlined that in combination with knowledge systems such as the Gene Ontology Consortium [4,102] his approach can be applied to derive models of cellular pathways and to identify functions of novel genes.

A highly enjoyable and top–notch presentation has been performed by D Lewin (CuraGen, New Haven, CT). CuraGen's approach in target discovery takes full advantage of Systems Biology [5]. With CuraGen's proprietary platform "GeneScape" integrative drug discovery is facilitated by a combination of sequencing, transcription profiling, SNP genotyping and pathway knowledge/analysis. On computational aspects, CuraGen's PathCalling System for pathway prediction and utilization strives on metabolic network data as well as protein interaction data. The latter has been assembled from yeast two–hybrid screens as well as from literature. Apparently, CuraGen pursues the generation of a human protein interaction map. Cooperative interests of CuraGen cover various therapeutic targets including obesity and cancer. Lewin demonstrated the causative relationship between human papillomavirus infection and development of cervical cancer by employing CuraGen's GeneScape System.

Highlights – Summary

- Regulation of alternative splicing is still in an early stage of research.
- After resolving technological bottlenecks, protein–chips will serve as valuable high–throughput techniques for fast and parallel analysis of protein function, protein modification and regulation as well as for drug discovery.
- "Intelligent" experimental strategies are capable to identify and visualize cellular networks in living cells.
- Knowledge of cellular networks provide information for new and more sophisticated drug targets. Emphasize is laid on high–throughput identification of interactions between cellular components as building–blocks for the reconstruction of cellular networks.
- Systems Biology serves as a new paradigm for system wide analysis of biological systems.

Expert opinion and Outlook

There is no doubt that Systems Biology will become of increasingly importance for the analysis of complex pharmacological scenarios. No longer one can focus on one gene one phenotype problems ignoring important interactions with other components of the biological system. It has been impressively shown by some speakers that Systems Biological approaches are not only necessary but provide essential insights for

accelerated drug discovery. Successful techniques combine both experimental data from various sources such as gene-expression profiles, proteomic data, SNP genotyping analysis and phenotypic profiling with theoretical and biocomputational methods such as gene-context analysis and protein interaction analysis, maximal flux analysis and comparative network genomics. Interdisciplinary research that include the full spectrum from discrete mathematics and graph theory to cell biology and pharmacology is essential in this new discipline of functional genomics/systems biology.

The advent of post-genomic research and the exponentially growing number of completely sequenced genomes suggests a driving need and opportunity for the use of such combined experimental and theoretical/computational strategies. Hierarchical approaches that extend the long prolonged one-gene one-phenotype paradigm towards general gene networks will become increasingly important. In the sense of Systems Biology the initial steps of the assembly of a parts list of an organism by whole genome sequencing efforts will be accompanied by whole-organism, high-throughput methods such as DNA or protein-chip arrays. Further steps into a system wide analysis of biological systems are the identification of interactions between components of the parts list and the (re)construction of biological networks.

Future challenges of pharmacogenomics will include not only the discovery of components and functional relationships but also a predictive, hypothesis driven approach for faster, cheaper and more accurate drug design. Such a strategy would essentially rely on the recruitment of novel theoretical approaches and massive computational power. Prerequisites for this endeavor not only include the identification of components and their interactions but also the accurately determination of parameters for further modeling and simulation of such complex biological systems. This need would translate in the development of high-throughput measurements of biochemical properties and parameters such as kinetic rate constants, binding and dissociation constants, response rates for various types of cellular components and for inter-cellular interaction. Having such a modeling/simulation capability in place one future perspective could be the systems analysis of the microbial ecology in the human gastro-intestinal tract. The identification of specific systems, such as quorum sensing, that triggers a toxic reaction of otherwise harmless microbes, the analysis of specific host-pathogen interactions and the response to drug treatment is a challenging research area for the future. A farfetched and speculative viewpoint but nevertheless critical milestone for pharmacology and the development of novel therapeutics would ensemble a "virtual human", a model system for predictive analysis of drug responses under changing physiological conditions. Although such a "virtual human" may still be science fiction for an extended time, pharmacogenomics and Systems Biology show the potential for such new and exciting developments.

References

- [1]Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO: Metabolic modeling of microbial strains in silico. *Trends Biochem.*, 26(3) 179–186 (2001).
- [2]Forst CV, Schulten K: Evolution of Metabolisms: A new Method for the Comparison of Metabolic Pathways using Genomics Information. *J. Comput. Biol.*, 64, 343–360 (1999).

- Initial publication on Comparative Network Genomics. This method has been applied to electron transport pathways, and the Krebs citric acid cycle.
- [3]Forst CV, Schulten K: Phylogenetic Analysis of Metabolic Pathways. *J. Mol. Evol.*, 52, 471–489 (2001).
 * Extension of [1]. This paper includes the evolutionary analysis of terminal oxidase complexes as well as the connection between pathway similarity and operon conservation in the case of the Tryptophan/Serin biosynthesis network.
- [4]The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genet.*, 25, 25–29 (2000).
 ** A very useful concept for the classification of biological concepts. Critics may state that the Gene Ontology is more a hierarchical nomenclature than an ontology in a strict sense. See also the web-link [xxx].
- [5]Green CD, Simons JF, Taillon BE, Lewin DA: Open systems: Panoramic views of gene expression. *J Immunol. Methods*, 250(1–2), 67–79 (2001).
 This article provides an overview of CuraGen's GeneCalling strategy.
- [6]Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO: Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, 409, 533–538 (2001).
 ** Systems biology research on DNA protein interaction with respect to cell cycle regulation.
- [7]Kitano H (Ed.): Foundations of Systems Biology. The MIT Press, Cambridge, MA, USA (2001).
 * This book provides an overview of the state of the art of the field. It covers the central topics of systems biology: comprehensive and automated measurements, reverse engineering of genes and metabolic networks from experimental data, software issues, modeling and simulation, and system-level analysis.
- [8]Marcotte E, Pellegrini M, Thompson M, Yeates T, Eisenberg D: A combined algorithm for genome-wide prediction of protein function. *Nature*, 402, 83–86 (1999).
 ** A very nice summary of non-homology based protein function prediction and application for construction of protein functional networks
- [9]Ostermeier M, Benkovic SJ: Finding Cinderella's slipper –proteins that fit. *Nature Biotech.*, 17, 639–640 (1999).
 * Analysis article on Protein fragment Complementation Assays.
- [10]Reder C: Metabolic Control Theory: A Structural Approach. *J. theor. Biol.*, 135, 175–201 (1988).
 Seminal paper on metabolic control theory and metabolic flux analysis.
- [11]Remy I, Michnick SW: Visualization of biochemical networks in living cells. *PNAS*, 98, 7678–7683 (2001).

Websites

- [101]<http://dip.doe-mbi.ucla.edu>
 Database of Interacting Proteins, David Eisenberg, UCLA, Los Angeles, CA, USA
- [102]<http://www.geneontology.org>
 The Gene Ontology Consortium

[103]<http://www.systemsbiology.org>

Institute of Systems Biology, Seattle, WA, USA; Leroy Hood, Director

[104]<http://www.systems-biology.org>

ERATO Hiroaki Kitano's Systems Biology Group at the California Institute of Technology, Pasadena, CA, USA