04281 Abstracts Collection
Integrative Bioinformatics - Aspects of the Virtual Cell
— Dagstuhl Seminar —

Julio Collado-Vides[1], Ralf Hofestädt[2] and Christoph W. Sensen[3]

[1] UNAM Mexico, MX
[2] Univ. Bielefeld, DE
[3] Univ. Calgary, CA
csensen@ucalgary.ca

Abstract. From 04.07.04 to 09.07.04, the Dagstuhl Seminar 04281 "Integrative Bioinformatics - Aspects of the Virtual Cell" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Integrative system biology approach for molecular medicine: The case study of Cyclin dependent kinase5 in neurodegenerative pathologies

The development of system biology approach for molecular medicine is a very complex task. The integration of genomic and proteomic data is very difficult in the case of multifactorial diseases. The pathology depends by the multiple effect of genetic and environmental factors. The currently available methods allow to evaluate a single aspect,but in term of therapeutic target selection we need to combine molecular and cellular approaches. An integration of Knowledge discovery and system biology can give a great effort in this research field. Here we present an application of this philosophy to the evaluation of the CDK5 effects in the neurodegeneration.

Keywords: system biology pathological pathways, text mining neurodegenerative diseases

Joint work of: Fattore, Matteo; Arrigo, Patrizio

Transcription Profiling - beyond Simple Comparisons of two Situations

Lothar Altschmied (IPK Gatersleben, D)

Interpretaion of cDNA array data which goes beyond the comparison of two contrasting situations or the analysis of a time series will be presented. Emphasis is put on the development of hypotheses which can be tested to increase the understanding of biological systems. Two examples, one from the model plant Arabidopsis thaliana with a fully sequenced genome and the other from the crop plant barley for which ESTs are the main source of sequence information will be discussed.

cDNA arrays for Arabidopsis were used to analyse early events during the greening of etiolated seedlings under white light illumination for wildtype plants and several mutants with altered photomorphogenic behavior or defective photorecptors. A group of co-expressed genes was identified that shows light-induced mRNA-levels in wildtype seedlings. The promotors of these genes seem to harbor a common sequence element, closely related to the well known G-box motif. For the G-box motif it had been shown previously that it is involved in light-regulation via the red-light receptor phytochrome and the DNA-binding protein PIF3. In contrast to those results, the unaltered expression behavior of that group of co-expressed genes in phytochrome and in cryptochrome (blue-light receptor) mutants with respect to wildtype plants suggests that these photreceptors do not play a major role during the early phase of greening under white light illumination for the regulation of these genes. Furthermore, this group of genes shows no light-induction in the det1 mutant which undergoes constitutive photomorphogenesis in darkness. This indicates that a signalling pathway independent of the red and blue light photoreceptors exists, which is destroyed by the mutation in the DET1 gene. This result is in contrast to the previous genetical data which had placed the DET1 gene somewhere in the signal transduction pathway of phytochrome. Studies are under way to define properties of this alternative signalling pathway.

Based on a large resource of ESTs from barley at the IPK - currently >180,000 sequences from more than 20 different cDNA libraries - cDNA arrays were constructed and used for expression profiling during spike, flower and seed development. Originally intented for the identificacion of candidate genes for the isolation of tissue-specific promoters, analysis of the data set revealed a group of genes which is expressed during early spike and early seed development at a similar level and with a similar developmental gradient. Genes of this group are clearly distinct from housekeeping genes, since they show strong differential expression between some of the analysed tissue samples. Functional annotation of this group of genes indicates that cell division plays a major role during early stages of spike and seed development. The identification of this group raises the question, if they constitute a regulon which is used twice in different organs of the plant. Even more general one might ask, if regulons can be identified which

are repeatedly used during ontogenesis of an organism and if major aspects of development can be described as a succession of such regulons.

Keywords: Transcription profiling, Arabidopsis, barley, promoter element, development

Joint work of: Haehnel, Urs; Siefken, Martina; Sreenivasulu, Nese; Altschmied, Lothar

Integrative system biology approach for molecular medicine: The case study of Cyclin dependent kinase5 in neurodegenerative pathologies

Patrizio Arrigo (CNR - Genova, I)

The identification of the overall interaction that drive the multifactorial diseases is a very complex task because it is important to evaluate different molecules that act at different level.

In order to screen molecular target an integrative bioinformatic approach can help to face this problem. In this perspective the system biology costitutes a fundamental step. The development of the disease specific pathway requires the integration of the different basic bioinformatic fields with textual analysis and dynamical systems. In the framework of Italian Neuroinformatic portal development we have developed a small integrated tool for the evaluation of the infleunce of CDK5 on neurodegenerative diseases

Keywords: System biology, multifactorial diseases, Alzheimer

Joint work of: Fattore, Matteo; Arrigo, Patrizio CNR ISMAC, Section of Genova

Metabolites and Pathway Flexibility

Thomas Dandekar (Universität Würzburg, D)

The flexibility of life allows it to survive under harshest conditions.

Beautiful adaptations allow for instance archaebacteria to survive under hottest conditions or very halophilic environments.

However, the dark-side of such adaptation potential is the unexpected flexibility of parasites and pathogens to adapt to any antibiotic condition or other effort of the host. This is one motivation for the present paper, another is the desire of bioinformatics to measure and model biological phenomena not only in extreme or medical circumstances but for all organisms - enzyme and metabolite flexibility are excellent indicators of evolution and adaptation. Three levels have to be considered. For our purpose we have to consider three different levels where these two ingredients of adaptation and evolution are acting:

There is the individual metabolite and enzyme flexibility, resulting from this there are different possibilities for an individual pathway and the combined effects of both allows on the system level also for a wide and often underestimated network flexibility. The talk will cover all three aspects and bioinformatical tools applicable to these levels as well as important mechanisms both for enzyme and for metabolite flexibility on these different levels.

References
1) Schmidt S, Sunyaev S, Bork P, Dandekar T. (2003) Metabolites: a helping hand for pathway evolution? Trends Biochem Sci. 28(6):336-341.
2) Dandekar T, Sauerborn R. (2002) Comparative genome analysis and pathway reconstruction. Pharmacogenomics. 3(2):245-256

Keywords: Metabolites, enzymes, pathways, recruitment, network

Bio-chemical Organization Theory for Integrated Dynamical Networks

Peter Dittrich (Universität Jena, D)

Integration of biological data leads to increasingly complex network models that reflect more and more the self-sustaining property characteristic for living systems. In order to understand such complex dynamical reaction systems (especially those where new components can appear and disappear) a theory of bio-chemical organization is presented. The theory allows to view reaction networks from a different, more global perspective by decomposing the network into a overlaping hierarchy. The central element of such decomposition is a so called organization, which we define simply as a closed and self-maintaining set of components.

First applications are shown for artificial chemistries, planetary photo chemistries, and a metabolic network. Finally the dynamical part of the theory, which is currently under development, is briefly presented.

Keywords: Constructive dynamical system, stochastic simulation, network analysis

Simulating and Analysing Biochemical Networks using a Concurrency Approach

David Roger Gilbert (University of Glasgow, GB)

The behaviour of cells is governed and coordinated by biochemical signalling networks that translate external cues (hormones, growth factors, stress etc) into adequate biological responses such as cell proliferation, differentiatoin or death, and metabolic control. Regulatory malfunction underlies many diseases such as

cancer, and therefore a deep understanding is crucial for drug development and other therapies. Biochemical pathways govern fundamental biological control networks; while individual pathways are understood, the combinatorial complexity of the biological control apparatus is too high to be fully understood through "wet" analysis alone.

Aims of project

The aim of this interdisciplinary project is to model diverse biochemical networks and develop an associated computational system to facilitate the analysis of the behaviour of these networks. Specifically we plan to

• model diverse biochemical networks, taking the growth factor activated kinase (MAPK) signalling pathway and apoptosis (programmed cell death) as exemplars

• develop a computational system, based on process algebra, to analyse network behaviour A strategic aspect of this work will be a continuous cross check between modelling and real experimental data.

Predicted outcomes

The tools we will develop will facilitate basic research by enabling a deeper understanding of the effects of mutations, and applied research by helping to predict drug effects.

Overview

As paradigms we are using the regulation of the MAPK network and apoptosis (programmed cell death). Both networks are at the focus of current drug discovery efforts in important disease areas including cancer, arteriosclerosis, stroke, heart disease, chronic inflammatory and degenerative diseases. Specifically, for signal transduction networks we will investigate the role of threshold effects, enzyme processing, and positive and negative regulatory (feedback) mechanisms of signal propagation in the ERK (MAPK) pathway. The model will be informed by in vitro experiments using core component proteins of the ERK pathway in order to obtain behavioural parameters. Data on apoptosis will be mainly recruited from on-line literature using a text-mining tool.

A concurrent model

A key component of this research project is to devise a model of interacting biological processes based on concurrency theory. Our approach is thus quite distinct from the traditional simulation approach based on differential and/or difference equations. We hope that an approach based on the appropriate concurrency theory will give us two important aspects: i) the ability to compose/decompose processes, that is, to model, refine and define interactions between components; and ii) the ability to make logical judgements about all, or some possible behaviours. In the first year we have concentrated on surveying the field and developing prototype models.

Keywords: Sinal transduction networks, biochemical networks, pathways, process algebra, modelling

Joint work of: Gilbert, David Roger; Sturm, Oliver; Vyshemirsky, Vladislav

Model Checking of Biochemical Networks

Monika Heiner (BTU Cottbus, D)

Biochemical Pathways, i.e. networks of chemical reactions can be modelled by Place/Transition Nets in a straightforward manner. What you get in the first place, are bounded, but not live models.

To bring these models to live, i.e. to animate them with infinite behavior, we need to model the environment of the considered pathways, too. Three styles of environment behavior and there related models are introduced, each establishing a validation criterion for the pathways model. The validation criteria rely on well-known basic Petri net analysis techniques as place and transition invariants. The third style of environment model makes the model bounded and live, while preserving all essential behavior, i.e. pathways, of the same model under weaker environment assumptions. These models can now be checked against temporal logic formulae. This approach is demonstrated for a basic model for apoptosis currently of great interest in the community.

Keywords: Pathways, Petri nets; model validation, model checking

Joint work of: Heiner, Monika; Koch, Ina

Metabolic Pathway Prediction

Ralf Hofestädt (Universität Bielefeld, D)

Metabolic pathway alignment represents one of the most powerful tools for comparative analysis of metabolism. It involves recognition of metabolites common to a set of functionally-related metabolic pathways, interpretation of biological evolution processes and determination of alternative metabolic pathways. Moreover, it is of assistance in function prediction and metabolism modeling. Although researches on genomic sequence alignment have been intensively conducted, metabolic pathway alignment so far is less studied. We are motivated to develop an algorithm of metabolic pathway alignment to reveal the similarities between metabolic pathways.

Joint work of: Hofestädt, Ralf; Chen, Ming

Genome-wide Approach to Mammalian Cell Fate Regulation - Are Cell Fates Attractors ?

Sui Huang (Harvard Medical School, USA)

Cells in multicellular organisms have to maintain homeostatic stability, and express a stable phenotypic state, such as quiescence, proliferation, migration, or differentiation into various types despite the noisy micro-environment of the tissue.

Yet, they need to be able to undergo conditional, all-or-nothing transitions between these discrete cell fates. This behavior of uniting stability with flexibility is governed by a genome-wide regulatory network and is central to development and function of multicellular life. Since the cell fates are mutually exclusive, they cannot be viewed simply as separate "programs" associated with independent "pathway modules", but must be subjected to global coordination. In fact, gene regulatory and protein signaling pathway networks appear to form a "giant component" that covers most of the genome. This suggests that spontaneous organization of a "macroscopically" coherent, stable behavior of the genes, and hence, of the cell, can occur if the genome-wide gene regulatory interactions impose appropriate constraints onto the dynamics of the genomic regulatory network. It has been suggested that the macroscopic cell fate behavior reflects these constraints on network dynamics. Kauffman first suggested that a differentiated cell type is the manifestation of a high-dimensional attractor state of the genome-wide regulatory network. In this talk, I will present experimental data based on gene expression profiling studies of the switch from a precursor state to a terminally differentiated state in neutrophils which suggest the existence of high-dimensional attractor state in gene expression state space. Specifically, differentiating a cell in different ways reveals that a cell can visit entirely different states but then converge towards the common final neutrophil state from different directions in gene expression state space - consistent with an attractor state. This behavior defies the common notion of specific, unique instructive pathways dedicated to distinct cellular functions. The significance of this finding for our understanding of a cell's system behavior will be discussed. A new software visualization tool that allows a holistic approach to gene expression profiles, yet permits the retrieval of gene-specific information will also be presented.

Keywords: Cell fate, attractor, state space, gene expression profile

Modelling Transcriptional Activity of the Genome of Escherichia Coli

Arkady Khordusky (University of Minnesota, USA)

We used a combination of genomic and signal processing techniques to investigate the properties of transcription in the genome of Escherichia coli as a function of the position of genes on the chromosome. This approach revealed the existence of statistically significant patterns in the spatial series of transcriptional activity. These patterns could be classified on the basis of spatial ranges of correlations into three categories: i) short-range, of up to 16 kbp; ii) medium-range, over 100-125 kbp; iii) long-range, over 600-800 kbp.

Analysis of transcriptional correlations of genes within their chromosomal context demonstrated that the significant similarities in gene activities extend far beyond the length of an operon and that such local patterns of co-expression are dependent on DNA supercoiling. Unlike in the case with the short-range

patterns, the formation of the medium and long-range patterns of transcription does not strictly depend on the level of DNA supercoiling. Instead, those patterns appear to correlate with the patterns of distribution of DNA gyrase on the bacterial chromosome.

Localization of structural components in the transcriptional signal revealed an asymmetry in the distribution of spatial patterns of transcription along the bacterial chromosome. The demonstration that spatial patterns of transcription could be modulated pharmacologically and genetically along with the identification of molecular correlates of transcriptional patterns offer for the first time strong evidence of physiologically determined higher- (than operon) -order organization of transcription in the bacterial chromosome.

Keywords: Transcriptional patterns, signal processing, chromosome organization, supercoiling

Joint work of: Jeong, Kyeong Soo; Ahn, Jaeyong; Khodursky, Arkady B.

Stochastic Effects in Biochemical Reaction Networks

Andrzej Kierzek (Polish Academy of Sciences - Warsaw, PL)

The availability of a huge amount of molecular data concerning various biochemical reactions provoked numerous attempts to study the dynamics of cellular processes by means of kinetic models and computer simulations. Biochemical processes frequently involve small numbers of molecules (e.g. a few molecules of a transcriptional regulator binding to one 'molecule' of a DNA regulatory region). Such reactions are subject to significant stochastic fluctuations.

I will present some consequences of the stochastic effects in biochemical reactions on the function of complex biochemical reaction networks. First, stochastic fluctuations arising from single gene expression process will be presented. Subsequently I will show how these fluctuations propagate through the positive and negative feedback loops in the complex biochemical reaction network and cause epigenetically inherited changes in the physiology of single cells.

Keywords: Stochastic effects in biochemical processes, Monte Carlo computer simulations, negative feedback loop, positive feedback loop, overshooting effect, epigenetic inheritence

Gene Regulation in the Pi Calculus:Simulating Cooperativity at the Lambda Switch

Céline Kuttler (Université de Lille, F)

We propose to model the dynamics of gene regulatory networks as concurrent processes in the stochastic pi calculus.

As a first case study, we show how to express the control of transcription initiation at the lambda switch, a prototypical example where cooperative enhancement is crucial.

This requires concurrent programming techniques that are new to systems biology, and necessitates stochastic parameters that we derive from the literature. We test all components of our model by exhaustive stochastic simulations. A comparison with previous results reported in the literature, experimental and simulation based, confirms the appropriateness of our modeling approach.

Keywords: Gene regulation, stochastic simulation, concurrency, modelling

Joint work of: Kuttler, Celine; Niehren, Joachim

Full Paper: http://www.lifl.fr/~kuttler

Modeling biological pathways in top-down manner with hybrid functional Petri net

Hiroshi Matsuno (Yamaguchi University, J)

Several methods such as differential equation based method for modeling biological pathways have been proposed. We have modeled many biological pathways including lambda phage genetic switch control, lac operon and glycolytic pathways, and circadian rhythms, with hybrid functional Petri net (HFPN). In this talk, the characteristics of HFPN based method will be clarified with comparing other methods. Existing other methods basically construct a biological pathway model after correcting kinetics parameters of biochemical reactions (bottom up manner). In contract, our HFPN based model firstly construct a biological pathway model with reference to biological maps in the literature and then tuning parameters so that input/output behaviors of concentrations of substances such as mRNA and proteins are matched to biological facts from experiments or literature (top down manner).

Keywords: Hybrid functional Petri net, Genomic Object Net, biological pathways, systems biology

Integration of Databases at the EBI: The Integr8 Project - A Resource for Genomic and Proteomic Data

Manuela Prüss (EBI - Cambridge, GB)

Integr8 is providing an integration layer for the exploitation of genomic and proteomic data by drawing on databases maintained at major bioinformatics centres in Europe.

Main aims are to store the relationships of biological entities to each other and to entries in other databases, to provide a framework that allows for new kinds of data to be integrated, and to offer an entity-centric view of complete genomes and proteomes. Basic tools for data integration comprise the Proteome Analysis database, the International Protein Index (IPI), the Universal Protein sequence archive (UniParc) and the Genome Reviews. Entry points for the Integr8 portal depend on the user's entity of interest – from browsing the taxonomy or with a predetermined species of interest, the species page can be used, a simple search page leads to different applications when looking for certain protein sequences or genes, customisable statistics data are available from the BioMart application, and pre-prepared data can be downloaded from the FTP site.

Statistical mechanics in Transcriptional Regulatory Networks.

Osbaldo Resendis (National University Mexico, MEX)

The recents topological analysis made in biological networks has shown to be a useful tool to uncover the cell organization in life. In addition, it is well accepted that this architecture has important implication in the dynamics behavior of cell. In this talk we present a topological analysis made in the transcriptional regulatory network of E. coli. We show how the concepts of hierarchical modularity and modular biological functionality are related between them, and how this concepts could be useful to simplify the dynamical analyze the transcriptional resulatory networks.

Keywords: Networks-Topology-Hierarchical modularity-Clustering algorithm

Dynamics of segment determination in Drosophila

Maria G. Samsonova (Polytechnical University - St. Petersburg, RUS)

Morphogen gradients contribute to pattern formation by determining positional information within morphogenetic fields. Interpretation of positional information is thought to rely on direct threshold-dependent mechanisms for establishing multiple territories of differential gene expression. In Drosophila, gradients of transcription factors encoded by maternal coordinate genes establish the initial position of zygotic gap gene expression domains at the early blastoderm stage. This is the first step in a regulatory cascade leading to the expression of a segmentation prepattern by the onset of gastrulation. Using gap gene circuits, we show that a regulatory network consisting of maternal coordinate and gap genes provides a sufficient mechanism for correct gap gene expression. We report substantial shifts in the position of gap domains during cleavage cycle 14A and present a regulatory mechanism for the dynamics of gap domain boundary shifts, which relies exclusively on gap-gap cross-regulatory interactions and does not

require diffusion. Our results imply that maternal morphogen gradients are not sufficient to determine positional information in the Drosophila blastoderm and suggest interdependence between establishment and interpretation of positional information.

Keywords: Drosophila, segmentation, morphogens, system biology

Joint work of: Samsonova, Maria, Reinitz, John; Sharp, David

Mining Network Data to score Clusters from Array Experiments

Frank Schacherer (memorec biotec GmbH - Köln, D)

Clustering is a common technique to find groups of co-regulated genes in array experiments. The assumtion is that co-regulation occurs because of a shared function between those genes.

Knowledge about gene function is stored in a multitude of public and commercial databases. We can use this knowledge to examine the genes in each cluster and identify biologically meaningful clusters. This is desireable, as some clustering procedures produce a large number of clusters, making it difficult to manually find the most interesting ones.

In the special case of signaling network or interaction data, we assign a common biological theme to the genes in a cluster if they are near each other in the network. This can be achieved either by segmenting the network into local areas, or by directly measuring the distance between the genes and calculating how probable it is that they are as near to each other.

Visual Analysis of Biological Networks

Falk Schreiber (IPK Gatersleben, D)

Biological data is often structured in the form of complex networks. Dynamic visualizations may help biologists to extract information out of the data and are very useful for building sophisticated research tools. Biological networks can be modelled as graphs and represented visually, using graph drawing algorithms to obtain understandable and meaningful pictures and diagrams.

This talk considers three examples: the animated exploration of protein-protein interaction networks, the visualization of network-based phylogenetic trees and the representation and visualization of experimental data in the context of biological networks.

Keywords: Information visualization, graph drawing, networks

From 2D to 4D Bioinformatics

Christoph W. Sensen (University of Calgary, CDN)

My laboratory is involved in building Bioinformatics tools for Systems Biology. Initially most of our efforts were focused on Genome Analysis and Annotation. MAGPIE (http://magpie.ucalgary.ca/) and Bluejay (http://bluejay.ucalgary.ca/) are the results of these efforts. We now have begun to explore what is needed to fully understand Genome Expression and Proteomics Information and how to connect this information to Medical Data. We call this emerging Bioinformatics area 4D Bioinformatics, as the handling of spatial and temporal aspects is of great importance.

Keywords: Genome Analysis, Genome Annotation, 2D Bioinformatics, 4D Bioinformatics, Gene Expression, Proteomics

Ontological Analysis of Cell Signaling Mechanisms

Takako Takai-Igarashi (University of Tokyo, J)

Although databases for cell signaling system contain numbers of chemical reaction data, the data cannot be used yet to deduce biological functions from them. For the deduction, we need systematic and consistent description of the biological functions in the context of "cell signaling". To address this issue, we have ontologically analyzed the "context of cell signaling" and developed Cell Signaling Network Ontology (CSNO). CSNO defines distinction between cell signaling and metabolic system and offers how to integrate various data-models for cell signaling reactions.

We consider that CSNO explicates implicit domain knowledge so as to provide a framework that enables us to reconstruct the knowledge in computers systematically. This kind of ontologies differs in purpose of development from those mainly for control vocabulary represented by Gene Ontology.

Keywords: Ontology, cell signaling pathway, biological function

Joint work of: Takai-Igarashi, Takako; Mizoguchi, Riichiro

Metabolic Modeling with Time Series Data

Eberhard Voit (Medical University of South Carolina, USA)

Novel high-throughput techniques produce dense time series of in vivo measurements of the expression of genes at the genomic scale, of many simultaneous concentrations of metabolites, or of the prevalence and activation states of proteins at the proteomic scale.

These data will dramatically affect strategies for modeling, analyzing, and optimizing metabolic systems. At present, the analysis of metabolic systems follows one of two paths. The first focuses purely on the stoichiometry of the metabolic network and has as its goal the identification, analysis, and manipulation of flux distributions. This approach leads directly to linear node equations, which, in turn, allow a vast spectrum of algebraic and computational tools that can be applied to networks of almost arbitrarily large magnitude. Linearity is also the greatest drawback of the stoichiometric focus, because it is not possible to account for regulatory features, even though they clearly affect the functioning of metabolic networks in a fundamental fashion.

The second path toward understanding metabolic networks is based on nonlinear kinetic descriptions, which are naturally much more flexible than linear systems. This flexibility comes at a much higher cost than for the linear approach, because there are infinitely more nonlinear than linear possibilities for setting up models, which necessitates a non-trivial selection and classification task, and because nonlinear formulations usually require more parameters that must be estimated from experimental data. The design of a nonlinear kinetic model presently consists of four generic steps: (1) identifying which variables and processes are to be considered; (2) characterizing each variable and process based on information from the literature; (3) integrating this information in systems equations; and (4) adjusting parameters secondarily to minimize discrepancies between model responses and observations. Two crucial challenges of this process are that there are few objective criteria supporting the choice of a particular mathematical process description and that the data necessary for estimation had very often been obtained under different conditions or even from different strains or organisms.

The presentation projects how genomic, metabolic, and physiological time series data, combined with novel computational methods, for instance, based on Biochemical Systems Theory, will change the present strategies of nonlinear metabolic modeling so much that it seems legitimate to speak of a new era of metabolic systems analysis.

Keywords: Biochemical Systems Theory, Metabolic Modeling, Time Series

Stochastic changes in expression for a gene controlled by the PhoB/PhoR two-component regulatory system

Barry L. Wanner (Purdue University, USA)

Both experimental evidence and theoretical simulations have demonstrated stochastic nature of gene expression in diverse biological systems from bacteria to multicellular organism (1-4). Many have proposed that cells exploit stochasticity to achieve genetic and nongenetic diversity for generation of populations capable of surviving in different environments. Stochastic mechanisms are thought to be one of the driving forces in the determination of cell fate and development in

complex organisms. Thus, a better quantitative understanding of the stochastic behavior of gene expression will have broad implications in cell biology.

Here we have exploited the simple bacterium Escherichia coli as a model cell to study stochasticity of native promoters in the context of well-defined signal transduction regulatory networks. Our approach is to measure gene expression in single cells within isogenic populations carrying a promoter fusion to fluorescent reporter proteins (such as GFP) in a FACS flow cytometer.

Two-component regulatory systems are the most prevalent mechanism of transmembrane signal transduction in bacteria. One such two-component system is comprised of the sensor protein PhoR and its response regulator PhoB. This system responds to environmental phosphate concentrations; phosphate starvation causing activation. Examination of two PhoBR responsive promoters (phoAp and phoEp) phosphate starvation revealed that, as expected, both phoAp-gfp and phoEp-gfp fusions displayed graded (non-binary) changes in gene expression under conditions of starvation.

Previous studies from our laboratory showed that six other sensor proteins can also activate PhoB. We therefore examined expression of the phoAp-gfp fusion under conditions in which each of these sensor proteins activates PhoB. For two of these sensors, we found that phoAp-gfp expression displays a combination of graded and binary behaviors. Two cell populations are evident in which the phoAp-gfp fusion is expressed at different levels. The proportion of cells in these populations changes with the culture cell density. Further studies are underway to determine the biochemical, genetic, and molecular basis of this phenomenon.

References

(1) Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. Science 297:1183-1186.
(2) McAdams, H. H. and A. Arkin. 1997. Stochastic mechanisms in gene expression. Proc.Natl.Acad.Sci.U.S.A 94:814-819.
(3) Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. Nat.Genet. 31:69-73.
(4) Thattai, M. and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. Proc.Natl.Acad.Sci.U.S.A 98:8614-8619.

This work is support by a David Ross graduate fellowship to L. Z., NIH NCI-2P30CA23168 to J.P.R., and NIH GM62662 to B. L. W.

Keywords: Gene regulation, stochasticity

Joint work of: Zhou, Lu; Gregori, Gerald; Robinson, J. Paul; Wanner, Barry L.

An Integrative Approach to Modelling E. coli

David Wishart (University of Alberta, CDN)

Cellular modeling requires the integration of many sources of existing data as well as the generation of new kinds of data. These data then have to be combined into a model that can be used to explain or predict the behaviour of a cell in a comprehensive, measurable fashion. As part of an on-going effort to model a simple bacterial cell (E. coli) we have undertaken a multifaceted approach that attempts to integrate data collection and archiving (backfilling), experimental data generation (metabolomics) and spatio-temporal modeling (via cellular automata) into a coherent whole. In this presentation I will provide examples from each of these three efforts to demonstrate our progress in this project. With regard to data backfilling, I will describe the development and planned extensions of the CyberCell Database (CCDB), a comprehensive, queryable, self-updating database designed to facilitate computer modeling of E. coli. The CCDB is now the most comprehensive, quantitative database of its kind. It contains key quantitative information on must about every aspect of the genomics, proteomics and metabolomics of E. coli. With regard to our work on metabolomics, I will describe the development of a novel NMR technique and its associated software that allows rapid (<30 s) quantitative determination of hundreds of metabolites from cell extracts or media. I will further demonstrate the utility of this technique in identifying the chemical phenotypes and associated metabolic mutations of two E. coli strains. This NMR-based approach allows rapid time-series data to be acquired which could have significant implications for testing and developing of kinetic models of metabolism. Finally, I will describe the development of a unique software tool (called Cell-Sim) that can be easily used to generate accurate spatio-temporal models of E. coli transcriptional and metabolic control. The Cell-Sim software uses dynamic cellular automata to stochastically model macromolecular interactions over enormous dimensional and temporal scales. I will demonstrate its utility in modeling/visualizing general enzyme kinetics, diffusion, the trp operon and the repressilator gene circuit.

Keywords: E. coli, metabolomics, cybercell, database, cellular automata

Hierarchical organization principle and analysis of metabolic and regulatory networks

An-Ping Zeng (GBF - Braunschweig, D)

Modular structure is a common feature for robust complex systems including various biological networks. Therefore identifying the functional modules in complex cellular networks is a key issue in modelling cellular processes and in general for system biology studies. In this presentation, recent results from studies of the hierarchical organization principle and modular structure analysis of genome-wide metabolic and regulatory networks will be presented and discussed. In

particular, for the first time, we will demonstrate that a core-periphery modular organization, which exists in networks of cellular metabolism as well as in social community, also governs the transcriptional regulation network of cells. The core-periphery modular network is considered as the most advanced network type that well balances the stability, flexibility and efficiency of the system.

References

1. H.W. Ma and A.-P. Zeng (2003a) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19: 270-277.
2. H. W. Ma and A.-P. Zeng (2003b) The connectivity structure, giant strong component and centrality of metabolic networks, Bioinformatics, 19: 1423-1430.
3. H. W. Ma and A.-P. Zeng et al. (2004) Decomposition of metabolic network based on the global connectivity structure of reaction graph , Bioinformatics (in print: http://bioinformatics.oupjournals.org/cgi/reprint/bth167v1.pdf)
4. H. Ma, J. Buer and A.-P. Zeng (2004) Hierarchical structure, modules and their core-periphery organization in E. coli transcriptional regulatory network. (submitted).

Keywords: Metabolic network, regulatory network, modual structure, modeling, complex system

Joint work of: Zeng, An-Ping; Ma, Hongwu

Qualitative Modeling and Simulation of Genetic Regulatory Networks

Hidde de Jong (INRIA Rhône-Alpes, F)

In order to cope with the large amounts of data that have become available in genomics, mathematical tools for the analysis of networks of interactions between genes, proteins, and other molecules are indispensable [1]. I will present a method for the qualitative simulation of genetic regulatory networks, based on a class of piecewise-linear (PL) differential equations that has been well-studied in mathematical biology [2]. The simulation method is well-adapted to state-of-the-art measurement techniques in genomics, which often provide qualitative and coarse-grained descriptions of genetic regulatory networks. Given a qualitative model of a genetic regulatory network, consisting of a system of PL differential equations and inequality constraints on the parameter values, the method produces a graph of qualitative states and transitions between qualitative states, summarizing the qualitative dynamics of the system. The qualitative simulation method has been implemented in Java in the computer tool GNA (Genetic Network Analyzer, available at http://www-helix.inrialpes.fr/gna) [3]. I will discuss the application of the computer tool to the modeling and simulation of several bacterial regulatory systems, in particular the networks controlling the

initiation of sporulation in Bacillus subtilis and the nutritional stress response in Escherichia coli [4].

References

[1] H. de Jong (2002), Modeling and simulation of genetic regulatory systems: A literature review, Journal of Computational Biology 9(1):69-105.

[2] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselmann (2004), Qualitative simulation of genetic regulatory networks using piecewise-linear models, Bulletin of Mathematical Biology, 66(2):301-340.

[3] H. de Jong, J. Geiselmann, C. Hernandez, M. Page (2003), Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks, Bioinformatics, 19(3):336-344.

[4] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, M. Page (2004), Qualitative simulation of the initiation of sporulation in Bacillus subtilis, Bulletin of Mathematical Biology, 66(2):261-300.