

PM9: ISMB2006 Tutorial

Title: *From Pathways Databases to Network Models*

Topic Area: *Systems Biology*

Main Presenter:

Baltazar D. Aguda

Senior Research Scientist and Group Leader

Biopathways Group, Bioinformatics Institute

#07-01 Matrix, 30 Biopolis St., Singapore 138671

Adjunct Associate Professor, Dept Physiology, National University of Singapore

Email: baltazar@bii-sg.org ; *Tel* +65 6478 8298

Teaching Experience: 15 years (Universities in Canada & Boston Univ, USA)

Earlier Tutorial Representations:

[1] Invited tutorial lecturer at the annual American Physical Society Meeting (Montreal, 21 March 2004)

[2] Invited to give a series of 3 tutorial lectures at the Summer Course on Modeling Biological Systems, Humboldt University, Berlin, Germany, July 10-11, 2003.

Second Presenter:

Andrew B. Goryachev

Research Scientist and Team Leader,

Systems Biology Group, Bioinformatics Institute

#07-01 Matrix, 30 Biopolis St., Singapore 138671

Adjunct Assistant Prof., Div. of Bioengineering, National University of Singapore

Email: andrewg@bii-sg.org ; *Tel:* +65 6478 8269

<http://web.bii.a-star.edu.sg/~andrewg/>

Teaching Experience: 3 years, taught graduate and senior undergraduate courses in physical chemistry, bioengineering and systems biology.

Abstract (50 words):

This tutorial will first provide a primer on online pathway resources and ontologies. The focus will then shift to the topics of extracting network models from pathways databases, modeling at different levels of resolution, the methods and tools of network analysis and simulation, and on the qualitative analysis of networks with incomplete or uncertain information. Lastly, a specific biological network involved in the mammalian cell cycle will be used to illustrate the methods discussed.

Tutorial Level: *Introductory to Intermediate*

Prior Knowledge Required:

Attendees are expected to have the basic knowledge of general molecular biology and, in particular, some previous encounters with gene or protein regulatory and signaling networks. It is not required to have advanced mathematical training beyond an undergraduate first-year level calculus. Basic bioinformatics skills, such as searching online databases, are assumed.

Suitability of This Tutorial for ISMB:

The excitement in today's biology is driven by the huge amounts of data generated by high-throughput technologies, and by the expectation that these datasets will soon provide detailed understanding of life's processes. Ultimately, these datasets have to be integrated into a system-theoretic framework that should allow the study of the dynamics arising from networks of physico-chemical interactions orchestrating the physiology of a biological cell. The bioinformatics community is actively responding to this call for integration with the creation of a wide array of pathways databases. The proposed tutorial will give an overview of these databases, a classification scheme to guide users according to their needs, and a discussion of the current problems of pathway representation. The underlying objective of the tutorial is to demonstrate possible methods of extracting network models from databases. Models come at different resolutions, and pathways databases often provide only information on the connectivity (topology) of the interactions involved in a biological process. Thus, a unique feature of this tutorial is a discussion of a method of qualitative network analysis that the presenters think are most appropriate in the treatment of uncertain or incomplete pathway datasets. Also summarized in the tutorial are existing methods and tools for network visualization, analysis, and simulation. The mammalian cell cycle will be used to illustrate the methods discussed.

Profile of Presenter 1:

Baltazar Aguda obtained his PhD (Chemical Physics program) from the University of Alberta (Canada) under the tutelage of Prof. Bruce Clarke, the inventor of stoichiometric network analysis. After having worked for many years on the nonlinear dynamics and network analysis of enzyme reaction systems, he turned his attention (in 1997) to biological networks as complex as the mammalian cell cycle. In parallel with the refocusing of his research, he moved from Canada (Laurentian University, Chemistry & Biochemistry Dept) to take an Associate Professor position in the USA (Boston University, School of Medicine, Dept of Genetics & Genomics; with joint appointments in the Biomedical Engineering Dept, and in the Bioinformatics Program) to further his research on cell cycle checkpoints, a major topic in cancer research. Currently, Baltz Aguda is a Senior Research Scientist & Group Leader in Singapore's Bioinformatics Institute where he heads a biopathways team that utilizes (extracts models) and creates

pathways databases relevant to regulatory networks of the cell cycle, apoptosis, and associated intracellular signaling. Baltz and Avner Friedman (USA) are currently writing a graduate textbook (invited by Oxford Univ Press) on mathematical models of cellular regulation. This book will feature the use of genomic and other bioinformatic data to generate models. A key reference and resource for the proposed ISMB tutorial is a recent book chapter (on pathways databases and modeling gene regulatory networks) published recently by Dr. Aguda and his co-workers (cited after the *Tutorial Outline*: Aguda, Craciun & Cetin-Atalay, 2005).

Profile of Presenter 2:

Andrew Goryachev has at least seven years of experience in computational biology in both academic and industrial settings. His PhD work at the University of Toronto involved computational modeling of complex reaction networks and formation of spatial patterns in chemical and biological systems. As a postdoctoral fellow and later as a Senior Scientist at a leading bioinformatics software provider (GeneData AG), he worked on the development of algorithms and software for analysis of functional genomics data and integration of heterogeneous high-throughput omics data with pathway information. Presently he is working on modeling spatio-temporal dynamics of complex molecular networks in prokaryotic and eukaryotic systems. The main research focus of his group is on the emergence of cooperative self-organized behavior in the complex networks of interacting molecules and cells. Specifically, he performed a detailed analysis of the bacterial quorum sensing phenomenon from the level of intracellular signaling network to the population-wide scale. He is also currently working on the regulatory networks that control cycling of small GTPases and developing methods for the representation, database storage and analysis of the cell interaction networks in the immune system. Concurrently, he is actively involved in graduate teaching in bioengineering and systems biology. His expertise in qualitative and quantitative methods of network analysis as well as a substantial experience of work with real biological systems should enable him to present the tutorial material authoritatively and effectively.

Tutorial Outline

I. *Pathways databases and ontologies: Where to find and how to extract network information [1 hr]*

- A. Survey of existing databases and pathway resource compendia [40 min]
 - 1. Binary interaction resources (BIND, MINT, InterDom, etc.)
 - 2. Resources for metabolic systems (KEGG, BioCyc, Brenda, WIT)
 - 3. Databases of networks and pathways (TRANSPATH, aMAZE, GeneNet, PATIKA, etc.)
 - 4. Commercial pathway software (e.g. PathwayAssist and Ingenuity Pathway Analysis)

- B. Pathway ontologies and data exchange formats [20 min]
 - 1. BIOPAX and SBML
 - 2. Ontologies for incomplete pathway information (e.g. PATIKA)

15 min break

II. **Methods and tools for network analysis and visualization: Modeling at different levels of resolution** [1 hr]

- A. Representation of pathways and networks [15 min]
 - 1. Mechanistic and qualitative interactions
 - 2. Network graphs & visualization tools (CellDesigner, Cytoscape, PATIKA)

- B. Survey of methods and tools for network analysis and modeling [45min]
 - 1. Network dynamics simulation tools (E-cell, SBW, Gepasi, etc.)
 - 2. Stoichiometric network analysis in a nutshell (SNA toolbox)
 - 3. Short introduction into practical metabolic control analysis and tools (SBToolbox, SBW, Gepasi)
 - 4. Qualitative network analysis: exploiting topologies of ill-defined networks

15 min break

III. **Extracting a cell cycle model from pathways databases: network topology and stability** [1 hr]

- A. A brief introduction to the mammalian cell cycle
- B. From pathways databases, to qualitative network, and to a kinetic model
- C. Analysis and computer simulation of the model

30 min discussion and Q&A session

Total Time: 4 hours

Key references are from the presenters' original published work:

- [1] B.D. Aguda, G. Craciun, R. Cetin-Atalay, "Data Sources and Computational Approaches for Generating Models of Gene Regulatory Networks," in *Reviews in Computational Chemistry*, Volume 21, p.381-411. (edited by Kenny B. Lipkowitz, Raima Larter, and Thomas R. Cundari) Wiley-VCH, John Wiley & Sons, Inc. (2005).
- [2] B.D. Aguda, "Modeling the Cell Division Cycle," in *Lectures Notes in Mathematics*, Vol. 1872 : Tutorials in Mathematical Biosciences III: Cell Cycle, Proliferation, and Cancer (Editor: A. Friedman, Ed.), Springer-Verlag GmbH, pp. 1-22 (2005).
- [3] A. Goryachev, D.-J. Toh, T. Lee, Systems analysis of a quorum sensing network: design constraints imposed by the functional requirements, network topology and kinetic constants. *BioSystems* article in press (2006).
- [4] A. Goryachev, D.-J. Toh, K. B. Wee, T. Lee, H. B. Zhang, and L. H. Zhang, Transition to quorum sensing in an *Agrobacterium* population: A stochastic model. *PLoS Computational Biology* 1(4), 265 – 275 (2005). *Reviewed by Faculty of 1000.*