Stochasticity in Biochemical Reaction Networks

David Thorsley (University of Washington),
Brian Munsky (Los Alamos National Laboratory)

09/25/09 - 09/27/09

1 Overview of the Field

Cellular processes are subject to vast amounts of random variation, which can cause isogenic cells to respond differently, despite identical environmental conditions. Recent experimental techniques make it possible to measure this variation in gene expression, protein abundance, and cellular behavior. Combined with computational modeling, these techniques enable us to uncover the causes and effects of stochastic cellular dynamics. Depending on cellular function, biochemical processes may act to minimize stochastic variations or exploit them to the cell's advantage; in both cases, cellular processes have evolved to be remarkably robust to both intrinsic and extrinsic noise. By exploring this robustness in naturally occurring biological systems, we hope not only to improve our understanding of cellular biology, but also to formulate the "design principles" necessary to build similarly robust biochemical circuits and nanoscale devices. The second workshop on **Stochasticity in Biochemical Reaction Networks** held at Banff International Research Station, 25-27, September 2009 served as an excellent venue to discuss the multifaceted progress in this field.

The exciting research topic of stochasticity in biochemical networks combines many different aspects from multiple disciplines. First, experimental molecular biologists have begun to develop and perfect new quantitative techniques to observe single cell and single molecule dynamics. Tools such as flow cytometry and fluorescence activated cell sorting (FACS) enable researchers to measure the protein levels for millions of individual living cells in the time span of a single minute–thus conducting millions of simultaneous experiments. Time-lapse fluorescence microscopy and microfluidics have made it possible for researchers to measure, track and manipulate the behavior of single cells in carefully controlled micro-environments. Single molecule fluorescence *in situ* hybridization (FISH) techniques enable researchers to explore the spatial distributions of specific RNA molecule within a cell.

Next, theorists and mathematicians have derived new quantitative methods to analyze and explain the vast amounts of statistical data gathered from such experiments. It is known that stochasticity in cells is caused in part by what is referred to as "intrinsic noise" - the variability caused by the statistical dynamics of a chemical reaction with a small number of reactants - and in part by "extrinsic noise" - the variability caused by random fluctuations in a cell's environment. The participants in this workshop have developed many methods to understand and differentiate between these types of variability in experimental data.

Finally, collaborations between theorists and experimentalists can enable the multidisciplinary community to understand how, why and when different cellular mechanisms transmit variability in different ways, i.e. some suppress it while others amplify or exploit it. For example, control theory can help us understand feedback and feedforward regulatory motifs in cellular architectures, while an information theoretic perspective can help us to understand how cells in a developing multicellular organism can determine their exact spatial

location. These analyses suggest new methods and appropriate models for mathematically demonstrating how certain motifs are useful for dealing with cellular uncertainties. Such analyses are then directly applicable to the work of more applied researchers, who can use these theories to better constructing synthetic biological circuits and devices at the nanoscale level, including biomolecular motors and DNA molecular machines.

This workshop highlighted many of these recent improvements to measure, analyze, understanding and/or implement stochasticity in biological systems and has served as a starting point to devise the next crucial steps in this progress. A brief summary of some of the specific topics in these categories are discussed in more detail in the following section.

2 Presentation Highlights

The participants of this workshop form an intellectually diverse group of researchers united by their interest in the subject of stochasticity in biochemical reaction networks; they represent the fields of biology, biophysics, engineering, chemistry, mathematics, and computer science. Each has contributed to the field of biochemical networks in either the theoretical or experimental sphere and many have contributed in both areas.

Kyung-Hyuk Kim, University of Washington derived sensitivity analyses to better understand the effects of cellular variability as it passes through biochemical networks. In theory, these sensitivities could later be used in synthetic biological design to control cellular fluctuations while minimizing changes in mean concentration levels.

Mary Dunlop showed how temporal measurements of gene expression fluctuations could help scientists to determine the existence and form of regulatory links. She showed that natural stochastic noise aids in this process by exciting the systems dynamics. Then as these stochastic dynamics pass through the network it is possible to follow the signature of that noise and determine the underlying sequence of protein (in)activation. Using single cell microscopy and a well characterized three color synthetic gene regulatory construct, she validated the usefulness of these approaches with *in vivo* experiments. She then applied these approaches to discover the regulatory mechanisms in a natural galactose metabolism network.[3]

Aleksandra Walczak used analytical tools from the theory of information processing to understand how cell regulatory networks transmit information in order to process external stimuli and initiate cellular responses. Certain biological tasks require more precision and thus more information than others. However, small concentrations of regulatory components at the cellular level (and the resulting intrinsic noise) place strong limitations on cells' abilities to conduct this information transmission. Thus, by understanding how much information is necessary to complete a given biological function, one may be able to predict the qualitative and even quantitative form of the network necessary to complete that function [13]. By generalizing the theoretical considerations of information flow, Walczak also formulated spectral method to compute the joint stationary probability distribution of gene regulatory cascades [14].

James Faeder illustrated the vast complexity that can arise in signaling networks involving myriad protein-protein interactions. Through combinatorial complexity, the number of distinct chemical species in a given biochemical reaction may exceed any reasonable number, while the mechanics of these reactions and species can be understood with a handful of reaction "rules." In turn these rule-based models can be efficiently simulated with on-the-fly generation of the chemical species as they become important [4]. Faeder has successfully applied these rules to develop models to cell-surface receptor aggregation under typical experimental conditions [7]. Although biochemical networks like the ones discussed by Faeder can be incredibly complicated, they can often be reduced to much simpler systems as Ilya Nemenman illustrated in his presentation. The key component of this work was to eliminate the many fast chemical species without losing information regarding the slow, non-Poissonian fluctuations of the slow chemical species [11]. In related work, Nemenman and collaborators have shown the even very complicated multi-step processes can reduce down to much simpler dynamical systems [2, 8].

Arjun Raj presented a novel approach that he has developed to detect and count individual mRNA molecules inside a single living cell. This process known as single molecule fluorescence *in situ* hybridization promises to revolutionize the study of stochasticity gene regulatory networks [10]. Raj then uses this

approach to study the gene regulatory network responsible for the robust gut formation during early embryonic development in *C. elegans*.

Also using the FISH approach to detect single mRNA molecules, Gregor Neuert has studied the high-osmolarity glycerol (HOG) pathway, which is one of the mitogen-activated protein kinase (MAPK) pathways in Saccharomyces cerevisiae yeast cells. While the components of this regulatory network are known from many years of previous work, the dynamics of stochastic gene expression in single cells were previously unknown. With the precise single-cell experimental procedures offered by FISH, and careful modeling, a simple intuitive model has been formulated to capture and predict the all observable aspect of the single cell dynamics.

Inspired by the new wealth of quantitative single cell experiment data offered by flow cytometry and single cell microscopy experiments, Brian Munsky showed how one could use this data to better identify the parameters of gene regulatory systems. With theoretical studies, Munsky showed how the distributions of single cell population responses at a few transient time points could provide a lot of information about the underlying system's dynamics, much more information than is obtainable from just the mean behavior or even distributions taken at a stationary time point. These theoretical studies help to establish experimental guidelines that have been used help to identify and test predictive models for (*i*) *lac* regulation in E. coli using flow cytometry experiments [9] and (*ii*) the HOG pathway in yeast using Neuert's single molecule mRNA measurements.

Narendra Maheshri investigated the role that stochasticity plays in the positive feedback loop motif, that is prevalent in genetic regulatory networks. Maheshri demonstrated experimentally and in simulation that network with positive feedback can exhibit a bimodal distribution when noise is present in the feedback loop, even if the corresponding deterministic system does not exhibit bistability. Theoretical studies indicate that in order for the bimodal behavior to occur, the promoter in the feedback loop should be expressed in infrequent, large bursts and decay rapidly.

On the topic of molecular computation, David Soloveichik reported on the computational properties of stochastic chemical reaction networks and highlighted the connections between standard models of stochastic chemical kinetics and well-known computation models such as Boolean Logic Circuits, Vector Addition Systems, Petri nets, Primitive Recursive Functions, Register Machines, and Turing Machines [12]. Marc Riedel elaborated on this issue of molecular computation from the point of view of circuit design, proposing methods for automated synthesis of stochastic biochemical networks that perform mathematical computations with a high degree of accuracy.

The next pair of talks considered the role of stochasticity in molecular engineering and, specifically, the design of nanostructures and nanotransporters. William Shih presented novel results in the self-assembly of DNA structures. Building on previous results on programmable self-assembly of two-dimensional structures, Shih demonstrated how, by using stacks of flat layers of DNA, custom-designed three-dimensional structures can be made to self-assemble and explained how to control the curvature of the DNA strands in order to design complex shapes [5]. Henry Hess discussed the construction and control of molecular shuttles, consisting of cargo-binding microtubules that are propelled by surface-immobilized kinesin motor proteins. Ideally such nanoscale system can be selectively activated at programmable locations and times [1]. By controlling the sequestration of the activator compound using an enzymatic network, Hess develops a scheme for sharpening the concentration profile of the diffusing activator at the cost of decreasing activator utilization.

Michael Samoilov discussed the connections between classical, deterministic modeling of "large molecular systems," i.e., chemical reactions in which all of the reacting species are abundant, and stochastic modeling of these systems. Samoilov demonstrated that stochastic effects in large molecular systems are not, as commonly assumed, the results of low molecular counts of some species or of transient effects, but can occur in stationarity even for large systems.

David Thorsley investigated the problem of determining the state of a stochastic chemical kinetic system using time-lapse microscopy data. Because most chemical species in a single-cell experiment cannot be directly observed, Thorsley developed the concept of an observer for a stochastic chemical kinetic system and demonstrated how it could be used for state estimation, parameter estimation, and hypothesis testing.

The last two talks of the workshop focussed on approaches for simulating stochastic chemical reactions. In the basic stochastic simulation algorithms, the chamber in which the reactions occur is assumed to be well-mixed. Sotiria Lampoudi presented a spatio-temporal variant of the stochastic simulation algorithm [6]. Michael Chevalier discussed the issue of time-scale separation in stochastic biological systems. The existence of reactions on different time scales results adversely affects the computation time needed for basic stochastic simulations, and Chevalier proposed a new decomposition technique that allows for approximate solutions that trade off between computation time and guarantees of accuracy.

3 Outcome of the Meeting

The workshop emphasized recent improvements in the theoretical, computational, and experimental investigation of stochasticity at the cellular and nanoscale levels. Each of the participants at the meeting contributed to this progress in at least one, and in many cases two or three, of these advances. The workshop promoted cross-disciplinary communication and collaboration between researchers in mathematical fields such as stochastic processes, Markov models, stochastic simulation and information theory, engineering fields such as control theory, computer science, and circuit design, and scientific fields such as computational biology, nucleic acid research, biophysics, biochemistry, and nanotechnology. The event was highly successful in encouraging the development of a research community uniquely qualified to investigate the phenomenon of stochasticity in biochemical reaction networks.

In addition to presenting significant progress on the topics of stochasticity in biochemical reactions, the workshop also highlighted the persisting need for continued improvements in the analysis of such reactions. For example, combining new techniques for measuring spatial variability in cellular components with spatially non-homogenous analyses may yield new insights into cell regulatory behaviors. Similarly, the expanding usage of experimental techniques such as flow cytometry, time-lapse fluorescence microscopy, and other techniques involving the use of fluorescent proteins leads to a demand of a much more quantitative characterization of these important proteins. Finally, with researchers from many diverse disciplines exploring stochasticity in the fields of synthetic and computational biology, a real need is arising for an improved and standardized toolkit for researchers to describe and computationally analyze cellular variability. These and other discussion topics that arose during the meeting will be revisited in the next workshop on stochasticity in biochemical reaction networks.

References

[1] G. Bachand, H. Hess, B. Ratna, P. Satird, and V. Vogel. smart dust biosensors powered by biomolecular motors. *Lab on a Chip*, 9:1661–1666, 2009.

[2] G. Bel, B. Munsky, and I. Nemenman. Simplicity of completion time distributions for common complex biochemical processes. *arXiv:0904.1587*, 2009.

[3] M. Dunlop, R.S. Cox III, J. Levine, R. Murray, and M. Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics*, 40:1493–1498, 2008.

[4] J. Faeder, M. Blinov, and W. Hlavacek. Rule-based modeling of biochemical systems with bionetgen. *Methods Mol. Biol.*, 500:113–167, 2009.

[5] Y. Ke, S. M. Douglas, M. Liu, J. Sharma, A. Cheng, A. Leung, Y. Liu, W. M. Shih, and H. Yan. Multilayer dna origami packed on a square lattice. *J Am Chem Soc.*, 131(43):15903–8, 2009.

[6] Sotiria Lampoudi, Dan T. Gillespie, and Linda R. Petzold. The multinomial simulation algorithm for discrete stochastic simulation of reaction-diffusion systems. *The Journal of Chemical Physics*, 130(9):094104, 2009.

[7] M. Monine, R. Posner, P. Savage, J. Faeder, and W. Hlavacek. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *In revision*, 2009.

[8] B. Munsky, I. Nemenman, and G. Bel. Specificity and completion time distributions of biochemical processes. *arXiv:0909.2631*, 2009.

[9] B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5(318), 2009.

[10] A. Raj, P. van den Boogard, S. Rifkin, A. van Oudenaarden, and S. Tyagi. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 5(10):877–879, 2008.

[11] N. Sinitsyn, N. Hengartner, and I. Nemenman. Adiabatic coarse-graining and simulations of stochastic biochemical networks. *Proc. Nat. Acad. Sci.*, 106(26):10546–10551, 2009.

[12] D. Soloveichik, M. Cook, E. Winfree, and J. Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7:615–633, 2008.

[13] G. Tkacik, A. Walczak, and W. Bialek. Optimizing information flow in small genetic networks. *Phys. Rev. E*, 80(031920), 2009.

[14] A. Walczak, A. Mugler, and C. Wiggins. A stochastic spectral analysis of transcriptional regulatory cascades. *Proc. Nat. Acad. Sci.*, 106(16):6529–6534, 2009.