Complex systems: From chemistry to systems biology

John Ross[a,1] and Adam P. Arkin[b]

[a]*Department of Chemistry, Stanford University, Stanford CA 94305; and* [b]*Department of Bioengineering, University of California, Berkeley CA 94710*

Edited by Jack Halpern, University of Chicago, Chicago, IL, and approved March 27, 2009 (received for review March 27, 2009)

There is great interest in complex systems in chemistry, biology, engineering, physics, and gene networks, among others. The complexity comes from the fact that in many systems there are a large number of variables, many connections among the variables including feedback loops, and many, usually nonlinear, equations of motion, or kinetic and transport equations. "Many" is a relative term; a properly interacting system of just three variables can show deterministic chaos, a complex behavior indeed. For the natural scientist and the engineer, nearly all their systems are complex. Many problems still resist the arguments of symmetry, averaging, timescale separation, and covariation that often underlie complexity reductions. New tools have allowed us to peer at even the nanometer scale of the structure of materials, explore the dynamic chemical composition of the explosions in our increasingly efficient engines, and determine the organization of the genome and architecture of the molecules and molecular networks it implicitly encodes. All these are revealing extraordinary arrangements of kinetic processes, feedback loops, and spatial organization that together create complex behaviors. The recent interest is due, in part, to the substantial advances in measurement techniques of chemical and biological species and experiments on complex systems, concurrent substantial advances in theory, and the increased urgency of analyzing and understanding complex systems of fundamental importance.

Nowhere is the importance of complex dynamics and architectures clearer than in biological systems. In this issue, all of the articles address problems of complexity in organisms. Topics range from information processing in their signaling network and the organization of their metabolism, to how populations of differentiated cells communicate with one another to coordinate behavior, and to how evolution has arrived at different recurrent motifs of control and linked together different physiological functions. These studies are enabled by the rapid progress in our ability to sequence genomes, measure molecular species and their interactions at genome scale, image their spatial distribution and dynamics during perturbation (at extraordinary resolution), and genetically change the structure of these systems to test theories of function. Together these methods are allowing the unprecedented mapping of entire cellular systems when, not long ago, following only a few chemical or biochemical species was considered state of the art.

Such datasets themselves are only substrates for theories of complex function and behavior. Methods of multifactorial clustering and dimensionality reduction of data have been progressing at a good rate. The tools for inferring networks and parameterizing models from different sorts of direct and indirect measurements have also been arising at increasing rates and methods for assessing and comparing these tools have also arisen. Computationally more efficient methods for multiscale simulation of the dynamical chemical and mechanochemical representations of these systems are in continual development and have begun to move from the hands of applied mathematicians and physical chemists to the broader biological community. These methods have driven the development of significant new theory, for example, in the determination, not guessing, of complex reaction mechanisms, and in the representation and analysis of stochastic chemical systems, which impacts fields far beyond biology. Finally, approaches for extracting principles of operation of these networks, their recurrent control motifs, and their optimality with respect to different performance metrics have begun to arise as we begin to ask the dangerous question of the "purpose" of a particular biological network architecture.

Here, we focus on recent advances in complex biology. These articles are at a level that complexity theorists can appreciate and yet communicates to the biological and biophysical communities as well. Even in this subsector of the complexity field the articles appearing in this Special Feature are indicative of only some current interests but are far from inclusive; the field has grown too quickly for that. But they give a spectrum of approaches that are representative of the challenges that have been overcome and make clear a number of the challenges that remain.

Two articles address the difficult issue of the relation of the genome of an organism and the possible phenotypes of that organism. In one of these, "Phenotypes and tolerance in the design space of biochemical systems" Savageau et al. (1) give a definition of phenotype at the molecular level in terms of the dynamic properties of a deterministic kinetic description and the boundaries of a given phenotype in the design space, that is, the space of concentrations of pertinent biochemical species. Three examples of simple kinetic mechanisms are discussed in some detail: pathways, cycles, and branch points. The boundaries between phenotypic regions yield a method for discussion of the tolerance of a system to large changes of its parameters and the identification of design principles.

The other article devoted to the subject of the effect of genetic variation on phenotype is by Pe'er et al., "Modularity and interactions in the genetics of gene expression" (2). The authors develop a statistical method to identify a large number of linked regions for each gene. Genetic polymorphism can yield distinct cellular states (phenotypes) in which metabolic pathways and biological processes are activated to different extents. They find the interesting phenomenon of allele-specific interactions, that is, a gene has an influence on a phenotype only in the presence of a given allele at the primary locus. The authors believe that "state changes driven by intrinsic genetic variation and the resulting allele-specific interactions are likely common in human and disease associated genetics".

The complexity of biological processes, such as molecular signaling, frequently requires simplification procedures of various types, such as reduction of the number of variables. Spang et al., "Modeling the temporal interplay of molecular signaling and gene expression using dynamical nested effects models" (3), introduce such a method. They present a statistical method based on Bayesian models, which they call the "dynamical nested effect model" for investigating the interplay of cell signaling and gene expression. They contract ob-

Author contributions: J.R. and A.P.A. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: john.ross@stanford.edu.

served time delays of multiple step signaling processes into single steps. Their method allows the separation of biological processes into signaling and expression events and they apply it to murine stem cell development.

In another approach to deal with the immense complexity of many biological systems, Fontana et al. "Internal coarse-graining of molecular systems" (4), and others, have constructed rule-based models that consist of formal rules which describe specific interactions, for example, protein–protein interactions. The authors provide an automatic procedure for converting a rule-based model into a set of differential equations of much reduced dimension. The procedure involves the construction of course-grained variables that are determined by the dynamics of the system according to the rules.

In an article that spans analysis of molecular networks and cellular populations, Bischofs et al. (5) explore how the architecture of a central signal-transduction phosphorelay in *Bacillus subtilis* affects decision making during starvation. This pathway can integrate signals of nutrient deprivation and cell density (quorum), the latter through secretion and reuptake of small peptides. They show that the network can support a robust ratiometric calculation of food/cell. Experimentally they show that a key quorum system is heterogeneously expressed in a subpopulation during starvation. Cells not expressing this system commit early to sporulation and cells that do express it continue to grow. The restriction of the communication system to only the growing subpopulation leads to the interpretation that the phosphorelay is making a calculation of food per *growing* cell which is a better metric of the available resources. This proposed new role for quorum signal underscores that even isogenic bacterial populations can employ complex signaling among differentiated subpopulations to coordinate their behaviors.

The article by Ross et al., "Kinetic laws, phase-phase expansion, renormalization group, blood coagulation, and INR calibration" (6), presents first a systematic approach to deterministic chemical kinetics based on a phase-phase, or log-log, expansion. The first order in this expansion is the mass action law of kinetics. Higher-order terms lead to corrections of this law. If recycling occurs in the reaction mechanism, as for example in enzyme catalysis, then a generalized mass action law can be derived as a result of the recycling. These approaches are applied to the biological case of blood coagulation and the recycling model yields the empirical equations for the International Normalized Ratio (INR), and the dependence of the INR on the concentration of coagulation factors.

In "Exploring the role of noise in the eukaryotic cell cycle," Tyson et al. (7) investigate the relative contributions of intrinsic and extrinsic noise to the variability in the cell cycles of yeast. They work on a model of the system and carry out calculations of the deterministic kinetics. Then on a smaller, reduced model they do a fully stochastic calculation of the kinetics. Both intrinsic and extrinsic sources of noise contribute to observed variations of the cell cycle, but the contributions of the intrinsic molecular fluctuations are substantially larger. They conclude that "accurate stochastic models of cell cycle regulation are needed to confront quantitative measurements of specific regulatory proteins and mRNAs in single cells."

Stephanopolous et al. (8) address the difficult problem of integrating metabolic flux measurements across the functional levels of the cell: they develop a model-based approach to correlate mRNA data with metabolic flux measurements in the absence and presence of global regulators. The work yields evidence of rewiring of metabolic flux by transcriptional regulation, and of metabolic interaction density being a key biosynthetic control factor. By linking metabolic control and genetic regulatory networks, the authors emphasize the importance of integrating diverse types of data in the investigation of large-scale cellular models.

1. Savageau M, Coelho P, Fasani R, Tolla D, Salvador A (2009) Phenotypes and tolerance in the design space of biochemical systems. *Proc Natl Acad Sci USA* 106:6435–6440.
2. Litvin O, Causton H, Chen B, Pe'er D (2009) Modularity and interactions in the genetics of gene expression. *Proc Natl Acad Sci USA* 106:6441–6446.
3. Anchang B, et al. (2009) Modeling the temporal interplay of molecular signaling and gene expression using dynamical nested effects models. *Proc Natl Acad Sci USA* 106:6447–6452.
4. Feret J, Danos V, Krivine J, Harmer R, Fontana W (2009) Internal coarse-graining of molecular systems. *Proc Natl Acad Sci USA* 106:6453–6458.
5. Bischofs I, Hug J, Liu A, Wolf D, Arkin A (2009) Complexity in bacterial cell-cell communication: Quorum signal integration and subpopulation signaling in the B. subtilis phosphorelay. *Proc Natl Acad Sci USA* 106:6459–6464.
6. Vlad M, et al. (2009) Kinetic laws, phase-phase expansion, renormalization group, blood coagulation, and INR calibration. *Proc Natl Acad Sci USA* 106:6465–6470.
7. Kar S, Baumann W, Paul M, Tyson J (2009) Exploring the role of noise in the eukaryotic cell cycle. *Proc Natl Acad Sci USA* 106:6471–6476.
8. Moxley J, et al. (2009) Linking high resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc Natl Acad Sci USA* 106:6477–6482.