

Systems Biology: A Brief Overview

Hiroaki Kitano

To understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism. Properties of systems, such as robustness, emerge as central issues, and understanding these properties may have an impact on the future of medicine. However, many breakthroughs in experimental devices, advanced software, and analytical methods are required before the achievements of systems biology can live up to their much-touted potential.

Since the days of Norbert Wiener, system-level understanding has been a recurrent theme in biological science (1). The major reason it is gaining renewed interest today is that progress in molecular biology, particularly in genome sequencing and high-throughput measurements, enables us to collect comprehensive data sets on system performance and gain information on the underlying molecules. This was not possible in the days of Wiener, when molecular biology was still an emerging discipline. There is now a golden opportunity for system-level analysis to be grounded in molecular-level understanding, resulting in a continuous spectrum of knowledge.

System-level understanding, the approach advocated in systems biology (2), requires a shift in our notion of “what to look for” in biology. While an understanding of genes and proteins continues to be important, the focus is on understanding a system’s structure and dynamics. Because a system is not just an assembly of genes and proteins, its properties cannot be fully understood merely by drawing diagrams of their interconnections. Although such a diagram represents an important first step, it is analogous to a static roadmap, whereas what we really seek to know are the traffic patterns, why such traffic patterns emerge, and how we can control them.

Identifying all the genes and proteins in an organism is like listing all the parts of an airplane. While such a list provides a catalog of the individual components, by itself it is not sufficient to understand the complexity underlying the engineered object. We need to know how these parts are assembled to form the structure of the airplane. This is analogous to drawing an exhaustive diagram of gene-regulatory networks and their biochemical interactions. Such diagrams provide limited knowledge of how changes to one part of a system may affect other parts, but to understand how a particular system functions, we

must first examine how the individual components dynamically interact during operation. We must seek answers to questions such as: What is the voltage on each signal line? How are the signals encoded? How can we stabilize the voltage against noise and external fluctuations? And how do the circuits react when a malfunction occurs in the system? What are the design principles and possible circuit patterns, and how can we modify them to improve system performance?

A system-level understanding of a biological system can be derived from insight into four key properties:

1) System structures. These include the network of gene interactions and biochemical pathways, as well as the mechanisms by which such interactions modulate the physical properties of intracellular and multicellular structures.

2) System dynamics. How a system behaves over time under various conditions can be understood through metabolic analysis, sensitivity analysis, dynamic analysis methods such as phase portrait and bifurcation analysis, and by identifying essential mechanisms underlying specific behaviors. Bifurcation analysis traces time-varying change(s) in the state of the system in a multidimensional space where each dimension represents a particular concentration of the biochemical factor involved.

3) The control method. Mechanisms that systematically control the state of the cell can be modulated to minimize malfunctions and provide potential therapeutic targets for treatment of disease.

4) The design method. Strategies to modify and construct biological systems having desired properties can be devised based on definite design principles and simulations, instead of blind trial-and-error.

Progress in any of the above areas requires breakthroughs in our understanding of computational sciences, genomics, and measurement technologies, and integration of such discoveries with existing knowledge.

Identification of gene-regulatory logic (3) and biochemical networks is a major challenge. The conventional methods for creating a network model include performing a series of ex-

periments to identify specific interactions and conducting extensive literature surveys. Several attempts are under way to create a large-scale, comprehensive database on gene-regulatory and biochemical networks (4). Although such databases are useful sources of knowledge, many network structures remain to be identified. Substantial research has been done on expression profiling, in which clustering analysis is used to identify genes that are coexpressed with genes of known function (5, 6). Although clustering analysis provides insight into the “correlation” among genes and biological phenomena, it does not reveal the “causality” of regulatory relationships. Several methods have been proposed to automatically discover regulatory relationships solely on the basis of microarray data (7–9). At present, such methods use information derived from mRNA abundance, so there is limited scope to infer causality based on transcriptional regulation. Posttranscriptional and posttranslational mechanisms of regulation must be incorporated as large-scale data become available, but many properties have yet to be measured with sufficient accuracy or in high throughput. Although it is not possible to incorporate all the desired data into the automated discovery system, analysis of transcriptional regulation may provide very useful information because of the possible hypotheses it generates to allow us to infer the network structure. In general, when multiple hypotheses are generated by automated discovery analysis, it reflects a lack of information. This type of analysis can be combined with entropy-based decision-making algorithms to theoretically suggest an experiment that most reduces the number of ambiguous network hypotheses. Although such algorithms have yet to reach a level of practical application, they may prove useful for determining the optimal order of experiments needed to resolve ambiguous hypotheses (10). Progress in this area would lead to an increased emphasis on hypothesis-driven research in biology (Fig. 1).

Once we have attained an understanding of network structure, we will be able to investigate network dynamics. In reality, analysis of dynamics and structure on the basis of network dynamics are overlapping processes, because dynamic analysis may yield useful predictions of unknown interactions. For dynamic analysis of a cellular system, we need to create a model. But first it is important to carefully consider the purpose of model building: Whether it is to obtain an in-depth understanding of system behavior or to predict complex behaviors in response to complex stimuli, we must first define the scope and abstraction level of the model.

Sony Computer Science Laboratories, Inc., 3-14-13 Higashi-Gotanda, Shinagawa, Tokyo 141-0022, Japan, and Kitano Symbiotic Systems Project, ERATO, JST, and the Systems Biology Institute, Suite 6A, M31, 6-31-15 Jingumae, Shibuya, Tokyo 150-0001, Japan. E-mail: kitano@csl.sony.co.jp

The choice of analytical method used depends on the availability of biological knowledge to incorporate into the model. A steady-state analysis can be done using only the network structure, without knowing the rate constants for a particular reaction. For example, flux balance analysis (FBA) was used to predict switching of the metabolic pathway in *Escherichia coli* under different nutritional conditions based on knowledge of only the metabolic network structure; this was experimentally confirmed (11). With some knowledge of steady-state rate constants, traditional stability analysis and sensitivity analysis provide insights into how systems behavior changes when stimuli and rate constants are modified to reflect dynamic behavior. Bifurcation analysis, in which a dynamic simulator is coupled with analysis tools, can provide a detailed illustration of dynamic behavior (12, 13). This type of analysis has become conventional in dynamic systems and is already used in many studies on biological simulation.

Once both the network structure and its functional properties are understood for a large number of regulatory circuits, studies on classifications and comparison of circuits will provide further insights into the richness of design patterns used and how design patterns of regulatory circuits have been modified or conserved through evolution. The hope is that intensive investigation will reveal a possible evolutionary family of circuits as well as a "periodic table" for functional regulatory circuits.

Robustness is an essential property of biological systems (14). Understanding the mechanisms and principles underlying biological robustness is necessary for an in-depth understanding of biology at the system level. The phenomenological properties exhibited by robust systems can be classified into three areas: (i) adaptation, which denotes the ability to cope with environmental changes; (ii) parameter insensitivity, which indicates a system's relative insensitivity to specific kinetic parameters; and (iii) graceful degradation, which reflects the characteristic slow degradation of a system's functions after damage, rather than catastrophic failure. In engineering systems, robustness is attained by using (i) a form of system control such as negative-feedback and feed-forward control; (ii) redundancy, whereby multiple components with equivalent functions are introduced for backup; (iii) structural stability, where intrinsic mechanisms are built to promote stability; and (iv) modularity, where subsystems are physically or functionally insulated so that failure in one module does not spread to other parts and lead to system-wide catastrophe. Not surprisingly, these approaches used in engineering systems are also found in biological systems. Bacterial chemotaxis is an example of negative feedback that attains all three aspects of robustness (15–17). Redundancy is seen at the gene level, where it functions in control of

the cell cycle and circadian rhythms, and at the circuit level, where it operates in alternative metabolic pathways in *E. coli*. Structural stability provides insensitivity to parameter changes in the network responsible for segment formation in *Drosophila* (18). And modularity is exploited at various scales, from the cell itself to compartmentalized yet interacting signal-transduction cascades (19).

To conduct a systems-level analysis, a comprehensive set of quantitative data is required. Projects already under way, such as the Alliance for Cellular Signaling (AfCS) (20), are making large-scale measurements with the ultimate goal of creating an in-depth simulation model of cells. Exploratory studies on modeling should be done at the earliest stage of such a project to identify where measurement bottlenecks exist in building the final model and to avoid acquiring data with little value for model building, such as measurements of insufficient coverage and accuracy.

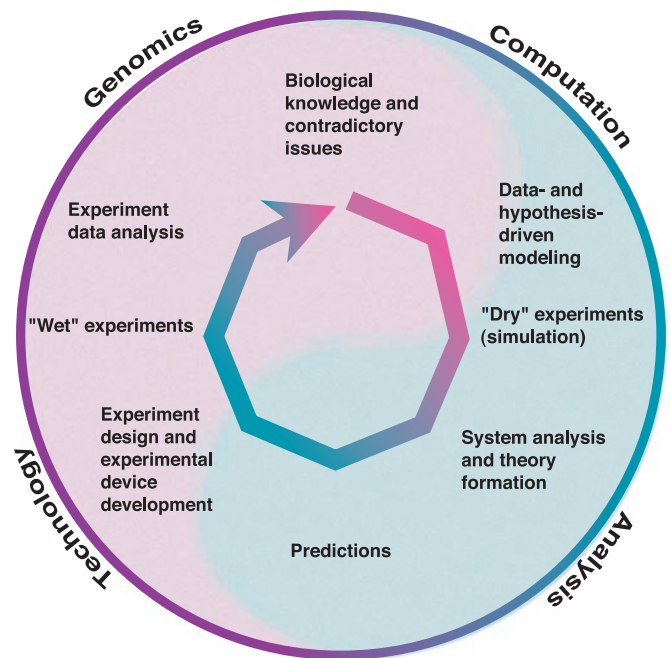
Comprehensiveness in measurements requires consideration of three aspects: (i) factor comprehensiveness, which reflects the numbers of mRNA transcripts and proteins that can be measured at once; (ii) time-line comprehensiveness, which represents the time frame within which measurements are made; and (iii) item comprehensiveness, which refers to the simultaneous measurement of multiple items, such as mRNA and protein concentrations, phosphorylation, lo-

calization, and so forth. Model-based experiment planning dictates where accuracy is critical and where it is not, so that resources can be optimally allocated.

Complete system-level analysis of biological regulation requires high throughput and accurate measurements, goals that are perhaps beyond the scope of current experimental practices. Technical innovations in experimental devices, single-molecule measurements, femto-lasers that permit visualization of molecular interactions, and nano-technologies are critical aspects of systems biology research. For example, microfluidic systems, also known as micro-TAS (total analysis system), enable minute quantities (picoliters) of samples to be measured more rapidly and more precisely. Various prototypes for polymerase chain reaction and electrophoresis have been developed (21–24). Such methods not only speed up measurements, but also encourage automation.

Software infrastructure is another critical component of systems biology research. Although attempts have been made to build simulation software and to make use of the many analysis and computing packages originally designed for general engineering purposes, there is no common infrastructure or standard to enable integration of these resources. The Systems Biology Mark-up Language (SBML), along with CellML, represent attempts to define a standard for an XML-based computer-

Fig. 1. Hypothesis-driven research in systems biology. A cycle of research begins with the selection of contradictory issues of biological significance and the creation of a model representing the phenomenon. Models can be created either automatically or manually. The model represents a computable set of assumptions and hypotheses that need to be tested or supported experimentally. Computational "dry" experiments, such as simulation, on models reveal computational adequacy of the assumptions and hypotheses embedded in each model. Inadequate models would expose inconsistencies with established experimental facts, and thus need to be rejected or modified. Models that pass this test become subjects of a thorough system analysis where a number of predictions may be made. A set of predictions that can distinguish a correct model among competing models is selected for "wet" experiments. Successful experiments are those that eliminate inadequate models. Models that survive this cycle are deemed to be consistent with existing experimental evidence. While this is an idealized process of systems biology research, the hope is that advancement of research in computational science, analytical methods, technologies for measurements, and genomics will gradually transform biological research to fit this cycle for a more systematic and hypothesis-driven science.



readable model definition that enables models to be exchanged between software tools. Systems Biology Workbench (SBW) is built on SBML and provides a framework of modular open-source software for systems biology research. Both SBML and SBW are collective efforts of a number of research institutions sharing the same vision (25).

How does the idea of systems biology impact pharmaceutical industries and medical practice? The most feasible application of systems biology research is to create a detailed model of cell regulation, focused on particular signal-transduction cascades and molecules to provide system-level insights into mechanism-based drug discovery (26–28). Such models may help to identify feedback mechanisms that offset the effects of drugs and predict systemic side effects. It may even be possible to use a multiple drug system to guide the state of malfunctioning cells to the desired state with minimal side effects. Such a systemic response cannot be rationally predicted without a model of intracellular biochemical and genetic interactions. It is not inconceivable that the U.S. Food and Drug Administration may one day mandate simulation-based screening of thera-

peutic agents, just as plans for all high-rise building are required to undergo structural dynamics analysis to confirm earthquake resistance.

Although systems biology is in its infancy, its potential benefits are enormous in both scientific and practical terms. A transition is occurring in biology from the molecular level to the system level that promises to revolutionize our understanding of complex biological regulatory systems and to provide major new opportunities for practical application of such knowledge.

References and Notes

1. N. Weiner, *Cybernetics or Control and Communication in the Animal and the Machine* (MIT Press, Cambridge, MA, 1948).
2. H. Kitano, *Foundations of Systems Biology* (MIT Press, Cambridge, MA, 2001).
3. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
4. Examples of such databases are: Signal Transduction Knowledge Environment (STKE <http://www.stke.org/>); KEGG (<http://www.genome.ad.jp/>); EcoCyc (<http://ecocyc.org/>).
5. M. Eisen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
6. S. Chu *et al.*, *Science* **282**, 699 (2000).
7. S. Onami *et al.*, in *Foundations of Systems Biology*, H. Kitano, Ed. (MIT Press, Cambridge, MA, 2001), pp. 59–75.

8. S. Imoto *et al.*, *Pacific Symposium on Biocomputing 2002* (World Scientific, Singapore, 2002), pp. 175–186.
9. C. Yoo *et al.*, *Pacific Symposium on Biocomputing 2002* (World Scientific, Singapore, 2002), pp. 498–509.
10. T. Ideker *et al.*, *Pacific Symposium on Biocomputing* (World Scientific, Singapore, 2000), pp. 302–313.
11. J. Edwards *et al.*, *Nature Biotechnol.* **19**, 125 (2001).
12. M. Borisuk *et al.*, *J. Theor. Biol.* **195**, 69 (1998).
13. K. Chen *et al.*, *Mol. Biol. Cell* **11**, 369 (2000).
14. M. E. Csete, J. C. Doyle, *Science* **295**, 1664 (2002).
15. N. Barkai *et al.*, *Nature* **387**, 913 (1997).
16. U. Alon *et al.*, *Nature* **397**, 168 (1999).
17. T.-M. Yi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4649 (2000).
18. V. Dassaw *et al.*, *Nature* **406**, 188 (2000).
19. G. Weng *et al.*, *Science* **284**, 92 (1999).
20. Alliance for Cellular Signaling (<http://www.cellular-signaling.org/>).
21. M. Burns *et al.*, *Science* **282**, 484 (1998).
22. R. Anderson *et al.*, *Nucleic Acids Res.* **28**, e60 (2000).
23. P. Simpson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2256 (1998).
24. P. Gilles *et al.*, *Nature Biotechnol.* **17**, 365 (1999).
25. Additional information can be obtained at <http://www.cds.caltech.edu/erato/> or <http://www.sbml.org>.
26. J. Gibbs, *Science* **287**, 1969 (2000).
27. C. Sander, *Science* **287**, 1977 (2000).
28. D. Noble, *Science* **295**, 1678 (2002).
29. I thank J. Doyle, M. Simon, and members of ERATO Kitano project for fruitful discussions. Supported by the ERATO and BIRD program of the Japan Science and Technology Corporation, and the Rice Genome and Simulation Project of the Ministry of Agriculture, Japan.

REVIEW

Reverse Engineering of Biological Complexity

Marie E. Csete¹ and John C. Doyle^{2*}

Advanced technologies and biology have extremely different physical implementations, but they are far more alike in systems-level organization than is widely appreciated. Convergent evolution in both domains produces modular architectures that are composed of elaborate hierarchies of protocols and layers of feedback regulation, are driven by demand for robustness to uncertain environments, and use often imprecise components. This complexity may be largely hidden in idealized laboratory settings and in normal operation, becoming conspicuous only when contributing to rare cascading failures. These puzzling and paradoxical features are neither accidental nor artificial, but derive from a deep and necessary interplay between complexity and robustness, modularity, feedback, and fragility. This review describes insights from engineering theory and practice that can shed some light on biological complexity.

The theory and practice of complex engineering systems have progressed so radically that they often embody Arthur C. Clarke's dictum, "Any sufficiently advanced technology is indistinguishable from magic." Systems-level

¹Departments of Anesthesiology and Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA. ²Control and Dynamical Systems, Electrical Engineering, and Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: doyle@cds.caltech.edu

approaches in biology have a long history (1, 2) but are just now receiving renewed mainstream attention (3–13), whereas systems-level design has consistently been at the core of modern engineering, motivating its most sophisticated theories in controls, information, and computation. The hidden nature of complexity ("magic") and discipline fragmentation within engineering have been barriers to a dialog with biology. A key starting point in developing a conceptual and theoretical bridge to biology is robustness, the preservation of particular characteristics despite uncertain-

ty in components or the environment (14).

Biologists and biophysicists new to studying complex networks often express surprise at a biological network's apparent robustness (15). They find that "perfect adaptation" and homeostatic regulation are robust properties of networks (16, 17), despite "exploratory mechanisms" that can seem gratuitously uncertain (18–20). Some even conclude that these mechanisms and their resulting features seem absent in engineering (20, 21). However, ironically, it is in the nature of their robustness and complexity that biology and advanced engineering are most alike (22). Good design in both cases (e.g., cells and bodies, cars and airplanes) means that users are largely unaware of hidden complexities, except through system failures. Furthermore, the robustness and fragility features of complex systems are both shared and necessary. Although the need for universal principles of complexity and corresponding mathematical tools is widely recognized (23), sharp differences arise as to what is fundamental about complexity and what mathematics is needed (24). This article sketches one possible view, using experience and theoretical insights from engineering complexity that are relevant to biology.