Curr Genet (2002) 41: 1–10
DOI 10.1007/s00294-002-0285-z

REVIEW ARTICLE

Hiroaki Kitano

Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology

Received: 12 December 2001 / Accepted: 13 February 2002 / Published online: 4 April 2002

Abstract With the ever-increasing flow of high-throughput gene expression, protein interaction and genome sequence data, researchers gradually approach a system-level understanding of cells and even multi-cellular organisms. Systems biology is an emerging field that enables us to achieve in-depth understanding at the system level. For this, we need to establish methodologies and techniques that enable us to understand biological systems as systems, which means to understand: (1) the structure of the system, such as gene/metabolic/signal transduction networks and physical structures, (2) the dynamics of such systems, (3) methods to control systems, and (4) methods to design and modify systems to generate desired properties. However, the meaning of ''system-level understanding'' is still ambiguous. This paper reviews the current status of the field and outlines future research directions and issues that need to be addressed.

Keywords Systems biology · Gene network · High-throughput measurement · Robustness

Communicated by S. Hohmann

H. Kitano
Sony Computer Science Laboratories,
Inc., 3-14-13 Higashi-gotanda, Shinagawa,
Tokyo 141-0022, Japan
E-mail: kitano@csl.sony.co.jp

H. Kitano
ERATO Kitano Symbiotic Systems Project,
Japan Science and Technology Corporation,
Suite 6A, M31 6-31-15 Jinguu-mae, Shibuya,
Tokyo 150-0001 Japan

H. Kitano
The Systems Biology Institute, Suite 6A,
M31 6-31-15 Jinguu-mae, Shibuya,
Tokyo 150-0001 Japan

Introduction

This paper elucidates the scope of and issues in systems biology (Kitano 2000, 2001, 2002), an emerging discipline that attempts to understand organisms at the system level. Systems biology is both an old and new field in biology. It is an old field because system-level understanding has been proposed and tried in the past. It was perhaps originated by Norbert Wiener who proposed the concept of cybernetics and devised mathematical formulae for physiological systems nearly 40 years ago (Wiener 1965). A precursor to Wiener was the concept of homeostasis by Cannon (1933). The philosopher Von Bertallanffy attempted to establish a general systems theory (Von Bertallanffy 1969), but it was too abstract to be a serious scientific discipline. Concepts such as robustness and feedback control were already discussed at that time and extensively investigated.

In the late 1980s, a group of scientists including Christopher Langton tried to develop a theoretical basis for living systems, using a computational approach that has progressed far since the days of Wiener. They claimed artificial life (Langton 1988) and complex systems research to be the central thrust for understanding life through theoretical methods. However, these attempts had minimal impact, if any, on the biological community because their theories were not able to provide useful predictions, guiding principles, or verifications of actual biological issues. In addition, with only a few exceptions, the community is dominated by computer scientists and physicists, who are not much interested in tangible biological phenomena. While artificial life has been an interesting source of novel ideas for engineering transformation of some biological knowledge, such as genetic algorithms and evolvable hardware, it never grew into the biological sciences.

At the same time, molecular biology and genetic analyses made remarkable progress. The main focus of those disciplines has been (and is) to identify essential elements and fundamental mechanisms that enable cell

functions, as we perceive them. The approach has been deductive, in the sense that major efforts have been made on the identification of genes and proteins responsible for specific phenomena, to create an encyclopedic knowledge of the genes and proteins involved. Although interactions have been investigated to understand how particular genes and proteins function, studies have only been made in an ad hoc manner and on a small scale.

The situation, however, is now changing, or being forced to change, due to the availability of complete genome sequences and the emergence of high-throughput measurement systems for gene expressions, protein interactions, and other essential factors. A group of new experimental methods enables us to measure large numbers of components simultaneously, thus opening up the possibility of system-level studies. Already, there have been various early attempts to make use of the massive amount of expression array and protein interaction data.

Systems biology is built on progress in related fields and past attempts. Thus, attempts to achieve system-level understanding are not new, but for the first time we can base such understanding on a molecular-level understanding. This was not possible in the days of Norbert Wiener, because molecular biology was only in its infancy and control theory and computer science had not matured enough to handle the immense complexity inherent in biological systems.

The goal of systems biology is to offer a comprehensive and consistent body of knowledge of biological systems tightly grounded on the molecular level, thus enabling us to fully integrate biological systems into more fundamental principles. It should be noted, however, that this does not imply that all biological phenomena can be explained directly by a set of basic principles, such as the basic laws of physics. System-level knowledge should be grounded in such a way that the system is composed of molecules; and molecules follow the laws of physics. However, how a system operates can only be described by a set of theories that focuses on system-level behaviors. The point is that such theories must reflect the realities of biological systems and molecules, without abstracting the essential aspects of biology.

It is important to distinguish between the types of theories. There are fundamental theories, as often seen in physics, that can be a basis for all phenomena, such as quantum electro-dynamics and super-string theory. However, there are theories that I would call "structural theories" that define structural constraints for specific phenomena. An example of this is the Bardeen Cooper Schrieffer theory (the BCS theory) that defines specific constraints causing the cooper pair, a state that is responsible for super-conductivity. The value of the BCS theory is that it explains super-conductivity, while it is grounded on the quantum theory. On top of these theories, there are designs that define specific realization of theoretical constraints, such as specific composition, to comply with the BCS theory.

In systems biology, research on all three categories of theories and design must be conducted, but more emphasis will be placed on identifying structural theories and using such theories to quickly identify specific designs and vice versa. In the past, molecular biologists focused almost exclusively on finding the design of life, by isolating specific genes and proteins, but paid little attention to structural theory. Researchers who have a physics background mislead themselves by trying to find fundamental theories where they should instead seek structural theories that can sufficiently constrain design spaces. This situation needs to be rectified.

The scope of systems biology is potentially very broad and different sets of techniques may be deployed for each specific research target. It requires collective efforts from multiple research areas, such as genetics and molecular biology, high-precision measurement, computer science, control theory, and other scientific and engineering fields. There are four areas where research needs to be carried out: (1) genomics and other molecular biology research, (2) computational studies, such as simulation, bioinformatics, and software tools, (3) analysis of dynamics of the system, and (4) technologies for high-precision and comprehensive measurements.

One issue that may puzzle traditional biologists when they encounter systems biology research is that they have to deal with the state of the system, rather than components of the system. Molecular biologists and geneticists have been working on understanding genes and proteins that are tangible objects. They are physical substances that we can point to and say "this is a protein". In contrast, systems biology has to deal with not only genes and proteins, but also the states of systems that are not tangible. Of course, components of the systems are tangible, but a system is not merely an assembly of components. What matters in the system is its dynamics, rather than a mere list of components or static structures. This raises an important question "what does it mean to understand a system?" A system is an abstract concept by itself. It is basically an assembly of components in a particular way that exhibits certain behaviors. Obviously, it is not sufficient to draw a diagram of all gene and protein interactions. Such a diagram is, of course, necessary and provides insights on how the system works. But, it is like a road map where we wish to understand traffic patterns and their dynamics. The focus of the research shifts from elements to networks, from matters to states, and from structures to dynamics.

Research in systems biology can be divided into two major groups that are complementary. One is a research on tools and algorithms for system-level studies. The other is research on system properties of specific biology, using the tools and algorithms developed.

Four phases of systems biology research

System-level understanding involves four phases depending on how much we understand the system: (1)

structure identification, (2) behavior analysis, (3) system control, and (4) system design.

System structure identification

The most basic level is to understand the structure of the system, such as the network of gene regulation, metabolism, and signal transduction. This level of understanding provides the basic organization of the system. Of course, not only the network of the system, but also physical structures need to be identified and included in the model. For the sake of explanation, this review article focuses on the network of gene regulation, metabolism, and signal transduction. The network structure consists of: (1) elements of the network, (2) interaction between elements, and (3) associated parameters.

Identification of elements of networks, i.e. genes, mRNA, and proteins, has been the main focus of research in molecular biology; and the pace of research will accelerate. Interactions have been investigated, but only on a small scale. Some efforts has been made to create a map of interactions, as seen in KEGG (Kanehisa and Goto 2000) and EcoCyc (Karp 2001). However, these efforts are basically assembling literature reports of independent experiments and are thus dependent on individual discoveries. Various early attempts at modeling and simulation of biological systems took a similar approach in building up their models. Examples of such approaches are seen in studies on the lambda phage decision circuit (McAdams and Shapiro 1995), early embryogenesis and morphogenesis of *Drosophila* (Reinitz et al. 1995; Kitano et al. 1997; Hamahashi and Kitano 1998; Kyoda and Kitano 1999), cell cycle modeling (Novak and Tyson 1997; Borisuk and Tyson 1998), cellular aging (Kitano and Imai 1998), circadian rhythms (Leloup and Goldbeter 1999; Ueda et al. 2001), Calcium oscillation (De Young and Keizer 1992; DuPoint et al. 1996), IP3 receptor-based calcium oscillation (LeBeau et al. 1999), bacterial chemotaxis (Barkai and Leibler 1997; Bray et al. 1998), and the MAPK cascade (Ferrell and Machleder 1998). Extensive research has been done in the past and this list represents only a part of the effort.

Despite intensive investigations on simulation-based studies, there has not been an increased recognition on the usefulness of such an approach. One bottleneck, which is still the case even today, is a lack of precise data and knowledge that can provide a basis for precision simulation. For example, simulation of the G protein cascade requires kinetic constants, such as the rate of Gα activating PLCβ and the rate of active PLCβ inducing IP3 production, which are not precisely measured. In addition, very few simulations take into account spatial distribution and movement of proteins within a cell and assume all ingredients are in an homogeneous mixture. Although there are cases where such approximations do not substantially alter the nature of the systems, many cellular processes are influenced by sub-cellular localization. This drawback is readily recognized by many researchers involved in simulations and is expected to be resolved in the future.

Despite these problems, simulation-based study is expected to be very effective for resolving conflicts in hypotheses or to find a hypothesis to explain counter-intuitive and contradictory data, because it enables us to make implicit assumptions explicit and to carry out complex calculations underlying counter-intuitive behavior. However, it is not suitable when so many factors are left unknown and only fragmented data are available. This situation is gradually improving due to the emergence of various high-throughout measurement techniques. With an increase in data precision, the accuracy of simulation and hypotheses will be enhanced, so that simulation-based studies will enjoy more opportunity to make predictions which can be experimentally confirmed and to resolve conflicting ideas. It may takes years for this approach to be used ubiquitously in experimental molecular biology, but it will certainly impact on how biology is pursued.

A high-throughput approach has been designed and some pioneering studies on the reverse engineering of gene regulatory networks from expression profile data have been reported. Clustering of expression profiles has been used to identify genes that are involved in a certain phase of biological processes. Such an approach has been applied for instance to the yeast cell cycle (DeRisi et al. 1997; Brown and Botstein 1999; Spellman et al. 1998) and the development of mouse central neural systems (Michaels et al. 1998). While clustering provides some insights as to which genes may be involved in the process, it does not identify the causal relationship between genes and thus is not able to recover network structures.

Alternative methods are now being developed to directly infer the network structure from expression profiles (Liang et al. 1999) and extensive gene-disruption data (Akutsu et al. 1999; Ideker et al. 2000). Most of the methods developed in the past translate expression data into binary values, so that algorithms are made tractable and computing cost can be reduced. But, such methods seriously suffer from information loss in the binary translation process and cannot accurately obtain network structure. A method that can directly handle continuous-value expression data was proposed (Kyoda et al. 2000) and reported highly accurate performance. However, only a limited relationship can be inferred, because it uses only the mRNA-level. Protein–protein interaction data, such as yeast two-hybrid data are, at this moment, too noisy to be used for pathway inference. With improvements in measurement accuracy, these data will be integrated and provide us with an overall picture of complex networks. To accomplish this, there are several issues that have to be overcome, including: (1) the fusion of various data, such as protein interaction, sequence, and modification, (2) extension to multi-cellular organisms, and (3) improvements in measurement accuracy.

In addition to the reverse engineering of network structures, various parameters need to be measured to investigate the dynamic properties of the system. While computational estimation of such parameters is possible to a certain extent, parameters should be measured and verified through high-precision experiments.

System behavior analysis

Once a system structure has been identified to a certain degree, its behavior needs to be understood. Various analytical methods can be used for this purpose. System behavior can be analyzed at several levels, depending upon how much we know about parameters associated with the network.

First, steady-state analysis can be performed even without knowing various parameters. Methods such as flux balance analysis (FBA) and metabolic control analysis can provide theoretical upper-bound, lower-bound, and optimal operation points of the circuit in the steady-state condition, using only the structure of the network (Fell 1997; Edward and Palsson 1999). FBA is a combination of linear algebra and linear programming applied to biological systems. Given the structure of a metabolic network, it generates a high-dimension space, where the basis vector spans the possible subspace in which this network operates, constrained by the law of mass action, etc. The basis is transformed to correspond to a biologically meaningful vector space, where each dimension roughly corresponds to extreme cases of metabolic balance, so that the combination of such balances covers all possible metabolic states of the system. Finally, the object function is defined, such as growth rate, so that optimal operating points can be calculated using linear programming, a conventional method in operations research. Palsson demonstrated the utility of this approach through the analysis of controlled nutrition intake in *Escherichia coli* and succeeded in predicting shifts in metabolic pathways due to external conditions (Edward et al. 2001). Although this method is limited to steady-state analysis, it is one of the most powerful methods for system analysis, because of its mathematical transparency and no need for parameters.

Some of the other approaches used in operations research, such as sensitivity analysis, can be used for understanding inter-relationships, both in the steady state and in the quasi-dynamic state.

Bifurcation analysis and other dynamic analytical methods provide means to understand the dynamic properties of biological systems, but require the identification of associated parameters and massive computing power. Some pioneering studies were done by John Tyson and his colleagues for the cell cycle network (Novak and Tyson 1997; Borisuk and Tyson 1998), where complex combinations of attractors are identified. Bifurcation analysis enables us to draw a phase portrait in which bifurcation points of the system can be mapped out, so that we can understand the dynamics of the system, potential operating points, and phase transition boundaries. Such an analysis not only reveals system-level characteristics, but also provides important insights for medical treatments, by discovering the cell response to certain chemicals, so that effects can be maximized while lowering possible side-effects.

System control

In order to apply insights obtained by system structure and behavior understanding, research on establishing a method to control the state of biological systems would be required. Control may be done through controlled environmental stimuli, chemical injection, drug absorption, and physical intervention. The current usage of drugs is a crude means to control cellular states, but it has not been precisely arranged, due to the limitations of our knowledge and a lack of simulation capabilities. Although research in this area has not taken off yet, this is one of the most important areas in systems biology, because major drugs and new treatment methods may be discovered. Central questions at this stage are: How can we transform malfunctioning cells into healthy cells? How can we control cancer cells, to turn them into normal cells or cause apoptosis? Can we control the differentiation of a specific cell into a stem cell and control its subsequent differentiation into the desired cell type? Technologies to accomplish such control would enormously benefit human health.

Obviously, the first precise treatment may be enabled by using various individual genetic variations, including single nucleotide polymorphisms (SNPs). It is likely that some of the effects of SNPs are masked by a mechanism that compensates for such variations. In this case, corresponding SNPs do not seem to affect the behavior of the cell. However, if such a compensation mechanism is disrupted by a SNP in a locus that constitutes the compensation mechanism, the effects of SNPs will directly show up in the cell's behavior. In such a case, it will be observed that, for a certain group of cells, SNPs affect phenotype, but for other groups SNPs do not seem to affect phenotype. By the same token, the effects of drug administration may be balanced out by a certain feedback system controlling metabolism, but effects may be gained if such a feedback system is temporarily suspended by using another drug prior to the effective drug. Understanding and exploiting such complexity cannot be captured just by looking at single gene mutations: it needs system-level research.

System design

Ultimately, we would like to establish technologies that allow us to design biological systems, for instance with the aim of curing diseases. A futuristic example would be to actually design and grow organs from the tissue of the patient him/herself. There may be some engineering

applications by using biological materials for robotics or computation. By using materials that have self-repair and self-sustaining capabilities, industrial systems will undergo revolutionary progress.

Measurement issues

Computational efforts alone will never solve the problem of structure identification of gene and metabolic networks. The popular notion that ''biological science will turn into information science'' overestimates the power of the computational approach and ignores the pressing need for high quality data for modeling and analysis. Any knowledgeable computer scientist knows the ''garbage-in garbage-out'' golden rule in computing. There are various issues that need to be improved in the measurement technique.

Comprehensiveness of the measurement is a critical issue. Although the word ''comprehensive'' is used in various contexts, there are three main types of comprehensiveness.

Factor comprehensiveness: this refers to the percentage of gene or protein covered in the experiment. This includes a measurement expression level of large numbers of genes, using DNA micro-arrays and a large-scale protein interaction using the yeast two-hybrid method (Ito et al. 2000; Schwikowski et al. 2000).

Time comprehensiveness: numbers of time points measured may play a critical role in understanding dynamic phenomena that are time-dependent.

Item comprehensiveness: multiple items may need to be measured, such as gene expression, protein concentration, and localization. Comprehensiveness in terms of numbers of items measured simultaneously is an important index.

Although there have been many projects that have performed comprehensive measurements of the various aspects of biological systems, most notably for yeast and *Caenorhabditis elegans*, such as cell lineage identification (Sulston and Horvitz 1977; Sulston et al. 1983), in situ hybridization (Tabara et al. 1996), and other experiments (White et al. 1986), they are not at the level of accuracy and comprehensiveness required for computer simulation modeling and systems analysis. The use of new technologies and automation may further improve the accuracy and throughput of such measurements.

One of the interesting examples of automated and computerized measurement is a new microscopy system that can automatically trace the cell lineage of *C. elegans* (Onami et al. 2001). It can identify the nucleus position of each cell from 4-D differential interference confocal images, using a special image analysis software running on a massively parallel PC cluster of 32 CPUs. With such a system, we can identify a very large number of cell lineages produced from an exhaustive RNAi knockout project and obtain highly reliable 3-D position data throughout the time course. The automatic nature of the system enables us to repeat the same experiment much more easily than before. For example, we can identify the 4-D lineage of the same mutant or wild-type strain for 100 samples and obtain statistical properties. Quantity changes the quality of research, because the availability of a large amount of 4-D lineage data enables us to define wild-type and each mutant lineage at the desired statistical significance, whereas we currently depend on individual judgement of mutant isolation and no quantitative measure has been established. With such a system, we can map all mutants in the phenotype space, statistically defined and grounded on massive experimental data. Similar efforts can be made on various aspects of experiments.

Aside from comprehensiveness, there are serious needs to improve the accuracy of measurement and the capability of single-cell measurement, instead of cell culture.

The needs for high quality and comprehensive measurement demand a radical change in experimental systems for further automation and the introduction of micro-fluidic systems. Some pioneering work on micro-fluid measurement systems has already been done (Anderson et al. 2000; Ikuta et al. 2001). The power of the micro-fluidic system is that it enables a higher level of automation and reduces the need for sample quantities. A series of research projects in so-called nano-biology follow another approach that may enable highly precise measurement. As represented by the work of Yanagida and others, nano-biology enables single molecular measurements (Ishijima and Yanagida 2001).

The way we design and plan experiments may be seriously affected. The current method is generally of an exploratory nature. In future, experiment planning will be more model- and hypothesis-driven. Models and analytical results dictate what is to be measured, with what level of accuracy. This resembles high-energy physics experiments carried out to verify theoretical predictions, so that specific energy ranges, properties measured, and minimum number of experiments are well defined. We cannot expect the same to apply to biology, where systems are so complex. However, it is possible to identify what needs to be measured and with what accuracy from models and analysis. We may just need to measure more items.

Software platform

A comprehensive body of software systems is necessary to make best use of massive data sets. Such a body of software includes:

1. Database for storage of experimental data
2. Cell and tissue simulator
3. Parameter optimization software
4. Flux balance analysis software
5. Bifurcation and systems analysis software
6. Hypothesis generator and experiment-planning advisor software
7. Data visualization software

It is neither feasible nor practical to assume that all modules will be developed to the highest quality by a single development group. Currently, a collective effort is underway to create a standard for models and data exchange, together with a plug-in modular software environment, so that users can create their own environment by combining the best modules available. Systems biology mark-up language (SBML) is an effort to define consistent data and model exchange formats among simulator and analysis tools (Hucka et al. 2000, Kitano 2002). It is an extension of XML and is expected to be the industrial and academic standard of the data and model exchange format. Already SBML level-1 has been defined and was announced in March 2001; and SBML level-2 is now being defined. The systems biology workbench project is a joint project that enables versatile modular software to be used to create a software environment for systems biology (Hucka et al. 2001). The first version of this software was recently released (http://www.cds.caltech.edu/erato/).

Among various components in the software environment, simulation of gene and metabolism network behavior plays an important role. There are several ongoing efforts for simulator development (Mendes and Kell 1998; Tomita et al. 1999; Kyoda et al. 2000). There are several issues that need to be resolved.

First, a method to efficiently compute stochastic processes needs to be developed. Many existing simulation software packages use differential equations to describe biochemical processes, but the stochastic nature of biochemical processes will be explicit when only a small number of molecules is involved in the reaction. Simulation using differential equations only provides average behaviors, instead of actual behaviors for each case. Such a problem has been pointed out in the simulation of phage decision circuits where a stochastic process is dominant (Arkin et al. 1998).

Second, there are many features of biological systems that we do not know how to simulate at a reasonable level of accuracy. Simulation can be an extremely powerful tool when it is applied to issues that are sufficiently understood to make the simulation reasonably accurate, so that predictions derived from the simulation can effectively constrain possible hypotheses. However, simulation is powerless when there is too little knowledge to provide a solid basis. Decisions on which state the assumed problems fall into are left to the intuition and experience of the researcher.

Third, biological systems are complex and highly nonlinear. This is exactly where we need simulation, instead of human intuition. But, this is precisely where simulation is of limited usefulness. Although various research has been done, simulation and equation solver technologies have certain limitations for high-order nonlinearity.

There are additional problems that will be discussed in detail here. However, these problems can be overcome and simulation will be an indispensable means of biological research. It is not foolish to imagine that some day the Food and Drug Administration may require simulation data to be attached to drug approval applications, just like construction approval applications for high-rise buildings must be accompanied by structural dynamics simulations.

Scientific issues

Robust systems

One of the essential traits of biological systems is robustness. Understanding how robustness is established and how it will break down is one of the most important aspects of understanding a system. The term ''robustness'' is used in various contexts. It often refers to the system's ability to adapt to the environment (adaptation). In a different context, it means that the system can tolerate damage by preventing catastrophic degradation of its functions (graceful degradation). It is also used to capture the system's ability to operate normally under a range of parameters and constants (parameter insensitivity).

In engineering systems, robustness is attained by various methods. First, extensive use of feedback and feedforward control significantly improves robustness by monitoring and adjusting the difference between the desired state and the current state of the system. Second, built-in redundancy makes the system more tolerant against disruption of the system components. Third, the use of a structurally stable network ensures that the state of the system converges to the desired state, despite various noises. Finally, modular design isolates subsystems from possible disruptions.

The exciting fact is that a number of cases can be identified in biological systems that exploit these engineering methods for robust system design. Some examples will illustrate how these mechanisms are used in biological systems.

Feedback control

Bacterial chemotaxis is one of the most well studied phenomena in biological systems. It demonstrates robust adaptation against a broad range of chemical attractant concentrations, so that it can always sense changes in chemical concentration and adjust its behavior accordingly. The mechanism behind this ability is a closed-loop feedback circuit (Barkai and Leibler 1997; Alon et al. 1999) that enables the bacteria to sense only acute changes in chemical concentration. In addition, it was shown that this feedback loop consists of integral feedback, so that it can perfectly adapt under a broad range of internal parameters (Yi et al. 2000). In this case, ligands that are involved in chemotaxis bind to a specific receptor (MCP) that forms a stable complex with CheA and CheW. CheA phosphorylates CheB and

CheY. Phosphorylated CheB de-methylates the MCP complex and phosphorylated CheY triggers tumbling behavior. It was shown through experiments and simulation studies that this forms feedback circuits which enable adaptation to changes in ligand concentration. Specifically, for any sudden change in the ligand concentration, the average activity level that is characterized by the tumbling frequency quickly converges to the steady-state value. This means that the system only detects acute changes in ligand concentration that can be exploited to determine tumbling frequency, but is insensitive to the absolute value of ligand concentration. Therefore, the system can detect and control its behavior to move to an area of highly attractant concentration in the field regardless of the absolute concentration level, without saturating its sensory system. Detailed analysis revealed that this circuit functions via integral feedback (Yi et al. 2000), the most typical automatic control strategy. System-level analysis revealed that this sub-system is relatively parameter-insensitive, so that adaptation behaviors can be sustained even under varying environments and internal disruption.

Structural stability

The simplest example of how a biological system exploits a structurally stable network can be seen in the lambda phage fate decision circuit (McAdams and Shapiro 1995). Lambda phage exploits multiple feedback mechanisms to stabilize the committed state and to enable switching of its pathways. When lambda phage is infected to *E. coli*, it chooses one of two pathways: lysogeny or lysogen. While a stochastic process is involved in the early stage of commitment, two positive and negative feedback loops involving CI and Cro play a critical role in stable maintenance of the committed decision. In this case, whether to maintain feedback or not is determined by the dosage of activator binding to the Or region; and the activator itself cuts off feedback when the dosage exceeds a certain level. Overall, the concentration of Cro is maintained at a certain level using positive feedback and negative feedback.

In the *Drosophila* development process, circuits are shown to be structurally stable, so that it exhibits converging behaviors over a broad range of parameters (Von Dassaw et al. 2000). This is an important finding because it demonstrates that structure is an essential part of a biological network, instead of specific parameters. Some feedback is used in the circuit to attain this parameter insensitivity. This is not used for adaptation, but is used to attain structural stability.

Feedback circuits also play important roles in development. A recent review article (Freeman 2000) elucidated some interesting cases in which feedback circuits play a dominant role in the development process. Such cases include: temporal control of signaling in JAK/STAT-signaling pathways, spatial control in pattern formation in *Drosophila* involving Ubx and Dpp, maintenance of patterns of expression for sonic hedgehog (Shh) and ZPA in limb development, etc. Detailed analysis has yet to be performed on specifically what kind of control is imposed on these feedback systems. Analysis of feedback control in development is highly complicated, because it involves both spatio-temporal dynamics and changes in morphology of the organism.

Redundancy

Redundancy also plays an important role in attaining robustness of the system. It is critical to cope with accidental damage to components of the system. If there are multiple independent signal transmission sub-systems, the system functions normally even when one or two of them are damaged. Redundancy design is generally used to protect the essential parts of the system against accidental disruption. Therefore, in an aircraft, control systems and engines are designed to have a high level of redundancy. In a cellular system, signal transduction and cell cycle are equivalent to control systems and engines.

The cell cycle is the essential process of cellular activity. For example, in the yeast cell cycle, the Cln and Clb families play a dominant role in the progress of the cell cycle. They bind with Cdc28 kinase to form Cdk complex. Cln is redundant because knock-out of up to two of three Cln (Cln1, Cln2, Cln3) does not affect the cell cycle. All three Cln have to be knocked out to stop the cell cycle. Six Clb have very similar features and part their origin may be gene duplication. No single loss-of-function mutant of any of six Clb affects growth of the yeast cell. The double mutants CLB1 and CLB2, or CLB2 and CLB3 are lethal, but other double-mutant combinations do not affect phenotype. It is reasonable that the basic mechanism of the cell cycle has evolved to be redundant and is thus robust against various perturbations.

Modular design

Modular design is a critical aspect of robustness. It ensures that damage in one part of the system does not spread to the entire system. Modular design can be implemented by three means: physical isolation, spatial isolation, and structural isolation.

The cellular structure of a multi-cellular organism is a clear example of physical isolation. It physically partitions the structure so that it prevents the entire system from collapse due to local incidents.

Spatial isolation may take place within the cell, where specific chemicals are localized and isolated from other parts of the cell.

Gene regulatory circuits may embed structural isolation by shunting effects in certain parts of the network from other parts of the network, using feedback loops and other methods. Even if a certain part of the circuit is disrupted, due to mutation or injection of chemicals, it does not necessarily affect other parts of the circuit. For

example, mutation in p53 may destroy the cell-cycle check-point system that leads to cancer. But, it does not destroy metabolic pathways, so that the cell continues to proliferate. How and why such modularity is maintained is not well understood at present.

Design patterns

Considering how biological systems are formed, it is hard to imagine that we can understand every detail from first principles. Thus, instead of trying to identify design principles, we should try to find design patterns that are widely used in biological system with certain variations.

Although there are very large numbers of gene network topologies and associated parameters, it is certainly not infinite and the number of useful patterns should be countable. With careful analysis and categorization, the author expects that something like a periodic table of biological networks can be created.

For example, many of the circuits that are involved in oscillatory behaviors, either temporal or spatial, have common network structures and are classified into a few typical patterns. Such auto-regulatory loops can be seen in various other places in biological systems. Calcium oscillation based on the activation of the G protein-coupled receptor (GPCR) complex is caused by an elevated IP3 level, induced by PLCβ activation, triggering the release of calcium pooled in the endoplasmic reticulum (ER) into the cytoplasm, binding with CaM to form Ca^{2+}-CaM. Ca^{2+}-CaM activates the hydrolytic activity of the RGS protein, which shunts Gα activity. Through this loop, the initial activation of GPCR is reduced and thus calcium release is stopped. Calcium in the cytoplasm is pumped out or back into the ER. While this is a greatly simplified picture of this process, the point is that calcium oscillation may be an attribute of a feed-back loop (DuPoint and Goldbeter 1996). There is a potentially conflicting hypothesis that claims that prime cause of calcium oscillation is the properties of ER surface IP3 receptor dynamics (LeBeau et al. 1999). In a totally different system, p53 activates the transcription of mdm2 that binds to p53. This creates a negative feedback loop, whereas expression of p53 invokes mdm2 that effectively causes proteolytic degradation of p53. Oscillatory behavior is observed when a high dose of ionized radiation is imposed on cells in culture (Lev Bar-Or et al. 2000). This is a simple auto-regulatory system that causes oscillation. Both use completely different molecules, but attain induced-oscillation because of the structural properties of interaction, i.e. an auto-regulatory loop, or negative feedback. Obviously, the proteins involved are all different, but the point is that the structure of the network at the abstract level remains identical, thus both share similar, but not identical, dynamic properties.

Circadian oscillation is a famous oscillatory behavior that has been extensively investigated. Circadian oscillation is evoked and maintained by the proteins PER and TIM, that create a dimer on the cytosol and inhibit their own transcription when transported back to the nuclei. On the surface, circadian oscillation may strike one as being similar to oscillatory systems, such as calcium oscillation and p53-mdm2 oscillation. However, unlike calcium and p53-mdm2 oscillation, which oscillate in the presence of inducing stimuli, such as ligand-induced GPCR activity and ionized radiation, circadian oscillation is more sustained and can oscillate for very long periods of time without any inducing stimuli; and it is very robust against environmental changes. It was discovered that the circadian system actually encompasses one more feedback loop involving CYC and CLK interlocked with the PER-TIM loop. A salient feature is that PER-TIM activates transcription of CYC-CLK, which are auto-regulatory; and CYC-CLK activates transcription of PER and TIM, which are also auto-regulatory. The two loops are inter-locked, so that either is always stimulated from the complementary loop (Glosopp et al. 1999). While the three systems briefly described here attain oscillation, the circadian system uses different structural dynamics from other two systems; and such structural differences affect the robustness of oscillatory behavior itself. Thus, careful classification at structural level may provide significant insights on the dynamics of biological systems. Circadian systems themselves are known to exist in different species, but using slightly different proteins; and it may be considered to be an evolutionary well conserved circuit. Similarly, there may be circuits that are almost identical, but with some elements substituted. These are the subject of investigation for evolutionary conserved circuits. Calcium oscillation and p53-mdm2 share only a very abstract circuit structure, but cell cycle circuits in different organisms share details of circuits and molecules involved which should share details of dynamics, too.

Orthologous and homologous circuits need to be identified to gain a more detailed picture of the evolutionary change in genetic information. Circuits that may be found in yeast and *C. elegans* may also be present in mice and humans, similar to the idea of homologous genes. Some of the feedback circuits, for example, may be so essential that they have been conserved through the course of evolution. Also, some circuits may be duplicated and revised versions applied in other parts of the system. With the progress of the systeome project in various model systems, such comparative studies and homology searches at the circuit level will became possible.

Future directions

Systems biology will ultimately change biological research and medical practice, because it offers a more precise understanding of the state of the system and a prediction of the effects of genetic alterations or treatments of patients. To accelerate progress in the field, a

project comparable to the human genome project needs to be established, but combining both centralized task-oriented research and distributed exploratory research. Such a project may be called "The human systeome project".

The goal of the human systeome project, if it is realized at all, shall be defined as "to complete a detailed and comprehensive simulation model of human cells at an estimated error margin of 20% by the year 2020 and to finish the identification of the system profile for all genetic variations, drug responses, and environmental stimuli by the year 2030."

Undoubtedly, this is an ambitious project and needs several milestones and pilot projects, such as in yeast, leading to the final goal. Already, there are projects having similar targets. The Alliance for Cellular Signaling, headed by Alfred Gilman, aims at building models of cells for cardiomyocytes and B cells, focusing on the G protein cascade. The initial part of the project is focused on the development of measurement systems to generate quantitative data to form the basis of high quality simulation models.

While these projects focus on modeling at the moment, the final target will be restated to identify the systeome. The systeome is an assembly of system profiles for all genetic variations and environmental stimuli responses. A system profile means a set of information on the properties of the system that includes the structure of the system and its behavior, analysis results such as phase portfolio, and bifurcation diagrams. The structure of the system means the structure of the gene and metabolic network, its associated constants, physical structures, and their properties. The systeome is different from a simple cascade map, because it assumes active and dynamic simulations and profiling of various system statuses, not static entities.

The impact of this project will be far-reaching. It will be a standard asset for biological research and a fundamental basis for diagnostics and prediction for a wide range of medical practices.

The systeome project will be a major commitment. However, it is indispensable for accelerating research on systems biology and contributing to a better understanding of living systems and medical practice. The systeome project includes a major engineering project for measurement and software platform development, followed by a wide range of scientific projects. Ideally, this project should be initiated as an international joint project on a scale comparable to the human genome project.

Conclusion

Systems biology is a new and emerging field in biology, that aims at system-level understanding of biological systems. System-level understanding requires a range of new analytical techniques, measurement technologies, experimental methods, software tools, and new concepts for looking at biological systems. The work has just begun and much remains to be done to develop a deep understanding of biological systems. Nevertheless, the author believes that systems biology will be the dominant paradigm in biology and that many medical applications as well as scientific discoveries can be expected.

References

Akutsu T, Miyano S, Kuhara S (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. Pac Symp Biocomput 4:17–28

Alon U, Surette MG, Barkai N, Leibler S (1999) Robustness in bacterial chemotaxis. Nature 397:68–171

Anderson R, Su X, Bogdan G, Fenton J (2000) A miniature integrated device for automated multistep genetic assays. Nucleic Acids Res:28–60

Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. Genetics 149:1633–1648

Barkai N, Leibler S (1997) Robustness in simple biochemical networks. Nature 387:913–917

Borisuk M, Tyson J (1998) Bifurcation analysis of a model of mitotic control in frog eggs. J Theor Biol 195:69–85

Bray D, Levin MD, Morton-Firth CJ (1998) Receptor clustering as a mechanism to control sensitivity. Nature 393:85–88

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37

Cannon WB (1933) The wisdom of the body. Norton, New York

DeRisi JL, Lyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

De Young GW, Keizer J (1992) A single-pool inositol 1,4,5-trisphosphate-receptor-based model for agonist-stimulated oscillations in Ca^{2+} concentration. Proc Natl Acad Sci USA 9:9895–9899

DuPoint G, Pontes J, Goldbeter A (1996) Modeling spiral Ca^{2+} waves in single cardiac cells: role of the spatial heterogeneity created by the nucleus. Am J Physiol 271:C1390–C1399

Edward JS, Palsson BO (1999) Systems properties of the *Haemophilus influenzae Rd* metabolic genotype. J Biol Chem 274:17410–17416

Edward JS, Ibarra R, Palsson BO (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130

Fell D (1997) Understanding the control of metabolism. Portland Press, London

Ferrell J, Machleder E (1998) The biochemical basis of an all-or-none cell fate switch in *Xenopus* ooctytes. Science 280:895–898

Freeman M (2000) Feedback control in intercellular signalling in development. Nature 408:313–319

Glossop NR, Lyons LC, Hardin PE (1999) Interlocked feedback loops within the *Drosophila* circadian oscillator. Science 286:766–768

Hamahashi S, Kitano H (1998) Simulation of fly embryogenesis. In: Adami C, Belew RK, Kitano H, Taylor CE (eds) Proceedings of the international conference on artificial life. MIT Press, Cambridge, Mass., pp 151–160

Hucka M, Sauro H, Finney A, Bolouri H (2000) An XML-based model description language for systems biology simulations. (ERATO Kitano Project) CALTECH Group, Tokyo

Hucka M, Finney A, Sauro H, Bolouri H, Doyle J, Kitano H (2001) The ERATO systems biology workbench: an integrated environment for multiscale and multitheoretic simulation in systems biology. In: Kitano H (ed) Foundations of systems biology. MIT Press, Cambridge, Mass., pp 125–143

Ideker T, Thorsson V, Karp R (2000) Discovery of regulatory interactions through perturbation: inference and experimental design. Pac Symp Biocomput 5:302–313

Ikuta K, Takahashi A, Maruo S (2001) In-chip cell-free protein synthesis from DNA by using biochemical IC chips. Proc IEEE Int Conf Micro Electro Mech Syst 14:455–458

Ishijima A, Yanagida T (2001) Single molecule nanobioscience. Trends Biochem Sci 26:438–444

Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y (2000) Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci USA 97:1143–1147

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Karp P (2001) Pathway database: a case study in computational symbolic theories. Science 293:2040–2044

Kitano H (2000) Perspectives on systems biology. New Generation Comput 18:199–216

Kitano H (2001) Systems biology: toward system-level understanding of biological systems. In: Kitano H (ed) Foundations of systems biology. MIT Press, Cambridge, Mass., pp 1–36

Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664

Kitano H (2002) Standards for Modeling. Nature Biotechnology, April, Vol. 20, page 337

Kitano H, Imai S (1998) The two-process model of cellular aging. Exp Gerontol 33:393–419

Kitano H, Hamahashi S, Takao K, Imai S (1997) Virtual biology laboratory: a new approach of computational biology. In: Husbands P, Harvey I (eds) European conference on artificial life. MIT Press, Cambridge, Mass., pp 274–283

Kyoda K, Kitano H (1999) Simulation of genetic interaction for *Drosophila* leg formation. Pac Symp Biocomput 4:77–89

Kyoda K, Muraki M, Kitano H (2000) Construction of a generalized simulator for multi-cellular organisms and its application to SMAD signal transduction. Pac Symp Biocomput 5:317–328

Langton CG (1988) Artificial life. Addisson-Wesley, Reading, Mass.

LeBeau A, Yule D, Groblewski G, Sneyd J (1999) Agonist-dependent phosphorylation of the inositol 1,4,5-trisphosphate receptor: a possible mechanism for agonist-specific calcium oscillations in pancreatic acinar cells. J Gen Physiol 113:851–872

Leloup J, Goldbeter A (1999) Chaos and birhythmicity in a model for circadian oscillations of the PER and TIM proteins in *Drosophila*. J Theor Biol 198:445–459

Lev Bar-Or R, Maya R, Segel L, Alon U, Levine A, Oren M (2000) Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. Proc Natl Acad Sci USA 97:11250–11255

Liang S, Fuhrman S, Somogyi R (1999) REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. Pac Symp Biocomput 4:18–29

McAdams H, Shapiro L (1995) Circuit simulation of genetic networks. Science 269:650–656

Mendes P, Kell DB (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics: 14:869–883

Michaels GS, Carr DB, Askenazi M, Fuhrman S, Wen X, Somogyi R (1998) Cluster analysis and data visualization of large-scale gene expression data. Pac Symp Biocomput 3:42–53

Novak B, Tyson J (1997) Modeling the control of DNA replication in fission yeast. Proc Natl Acad Sci USA 94:9147–9152

Onami S, Hamahashi S, Kitano H (2001) Automatic construction of cell lineage of *C. elegans*. In: Kitano H (ed) Foundations of systems biology. MIT Press, Cambridge, Mass., pp 39–55

Reinitz J, Mjolsness E, Sharp DH (1995) Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. J Exp Zool 271:47–56

Schwikowski B, Uetz P, Fields S (2000) A network of protein–protein interactions in yeast. Nat Biotechnol 18:1257–1261

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen M, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 9:3273–3297

Sulston J, Horvitz HR (1977) Post-embryonic cell lineage of the nematode, *Caenorhabditis elegans*. Dev Biol 56:110–156

Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *Caenohabditis elegans*. Dev Biol 100:64–119

Tabara H, Motohashi T, Kohara Y (1996) A multi-well version of in situ hybridization on whole mount embryos of *Caenorhabditis elegans*. Nucleic Acids Res 24:2119–2124

Tomita M, Shimizu K, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter C, Hutchison C (1999) E-Cell: software environment for whole cell simulation. Bioinformatics 15:72–84

Ueda H, Hagiwara M, Kitano H (2001) Robust oscillation within the interlocked feedback model of *Drosophila* circadian rhythm. J Theor Biol 210:401–406

Von Bertalanffy L (1969) General system theory. Braziller, New York

Von Dassaw G, Meir E, Munro EM, Odell G (2000) The segment polarity network is a robust developmental modules. Nature 406:188–192

White JG, Southgate E, Thomson JN, Brenner S (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans*. Philos Trans R Soc Lond B Biol Sci 314:1–340

Wiener N (1965) Cybernetics or control and communication in the animal and the machine. MIT Press, Cambridge, Mass.

Yi T-M, Huang Y, Simon M, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc Natl Acad Sci USA 97:4649–4653