Perspectives on Systems Biology

Attacking Complex Problems with the Power of Systems Biology

The ultimate goal of biology is to understand biological systems in sufficient detail to enable accurate, quantitative predictions about the behaviors of biological systems, including predictions of the effects of modifications of the systems. In other words, what we want to precisely simulate the behavior of the system using a computer model—we are nowhere close to this goal yet. In this short essay, I will discuss the problems associated with conventional approaches, describe how systems biology will help us get closer to this goal, and present some concepts that we biologists need to adopt to take advantage of the power of systems biology.

THE REDUCTIONIST APPROACH TO THE STUDY OF A BIOLOGICAL SYSTEM

The last revolutionary advance in biological research was driven by the concepts and technology of molecular biology, which links information about genetic traits to physical entities such as DNA or proteins. It is important to point out that this advance demanded that biologists adopt a new approach—molecular thinking. Molecular biology strongly promoted the reductionist approach, resulting in attribution of biological phenomena to the actions of one or a few genes. Although the reductionist approach is powerful in building logically simple hypotheses and devising ways to test them, it is very difficult to reconstitute a model for a whole biological system by combining the pieces of information it generates. First, using a reductionist approach, the entire system model must be reconstituted by combining information about every molecular step in the system (which I call a "bottom-up" approach). Any missing pieces of information block reconstitution of the system. Therefore, the bottom-up approach requires essentially complete information, including the dynamics of each step, to build a system model. Study of each molecular step requires intensive research. It is not realistic to expect that we will obtain nearly complete information about any biological system any time soon. Second, reductionism by definition focuses on information essential to a simplified question and intentionally discards extra information. Such extra information could be important for reconstitution. For example, relatively weak interactions among molecular steps may be important for explaining the dynamics of the whole system. Third, the behavior of the system may depend heavily on complex dynamics of the system. In such cases, if some parts are missing, the reconstituted system model may not behave like the real system at all.

SYSTEMS BIOLOGY ENABLES A "TOP-DOWN" APPROACH

It is fortunate for us to have an opportunity to witness and participate in another revolutionary advance. The advance is driven by systems biology. One aspect of systems biology is high-throughput reductionism (e.g. assigning biological functions to [ideally] all the genes of an organism). I suspect that other authors contributing to this issue emphasize this aspect, so I will not discuss it much. Another aspect of systems biology is characterized by exhaustive, simultaneous descriptions of a biological system, such as quantitative global profiling. Such broad and detailed information about a biological system provides us with a view different from reductionism—a view of how the system model should behave as a whole. For example, if we have mRNA expression profiles for 25,000 Arabidopsis genes under 500 different experimental conditions, our initial objective is to build a model to explain the expression levels of each of the 25,000 genes in each of the 500 experimental conditions. If the model building is successful, the model should be able to predict expression profiles under experimental conditions other than the initial 500 conditions. It will not be totally deterministic. Stochastic events, measurement errors in data, unknown environmental factors, etc. can affect the outputs of the system. The model must contain these factors within it and, therefore, will be probabilistic. The point is that broad and detailed information about a biological system obtained by systems biology enables us to start with a view of how the entire system model should behave. Then, we will define behaviors of parts of the system so that they are consistent with the expected behavior of the entire system.

HOW CAN WE TAKE A TOP-DOWN APPROACH IN PRACTICE?

If we try to build a system model by collecting a large amount of data describing inputs and outputs of the system while we treat the entire system as a black box, we will fail. It is mathematically proven that regardless of the size of data sets, descriptive data of a black box cannot specify a unique model—or even a small number of models—for the black
box. Therefore, for the top-down approach to work, we need to add interpretations based on biological assumptions and constraints. Generally, we assume that the degree of similarity between profiles reflects the closeness of relationships in the system. We can use this assumption to classify inputs and outputs of a system and to map relative positions of components in the system. For example, if we perform mRNA expression profiling of many mutant plants undergoing a particular response, the similarity relationships among the profiles for the mutants can be used for classification of inputs (perturbations of a system, such as mutations, can be considered as inputs to the system). The relationships among the inputs can be used to determine relative positions of mutations in the system. Similarly, the relationships among the profiles for genes whose expression is measured can be used for classification of outputs (any responses of a system to inputs) and defining relationships among the output classes. By identification of input and output classes and definition of their relationships, what we are effectively doing is decomposing the system into smaller parts. Note that to accomplish this task, we need a powerful similarity analysis method. If you have performed clustering analysis of expression profiles using commonly available methods, such as hierarchical clustering or self-organizing maps, you probably obtained results like: “The expression pattern of gene A is very similar to those of genes in cluster 8 and that of gene B is not. However, gene B included in cluster 8, and gene A is not.” This is because these methods are not sensitive to the local and multidimensional context of the similarity. We need a more sophisticated analytical method for recognition of patterns in expression profiling data.

Another important requirement of the top-down approach is that experiments must be run recursively. One round of profiling and data analysis leads to certain interpretations about the system. The next round of profiling and data analysis should be designed to investigate each interpretation in more detail. In this way, the system can be decomposed into smaller parts in every round. Note that the levels of time and space resolution and the accuracy required can change round-by-round and according to which part of the system is under investigation. Analytical methods may need to be adjusted for analysis of robust and non-robust sections (see below for a discussion of robustness).

CONNECTING TOP-DOWN AND BOTTOM-UP

The top-down approach discussed above is based on the assumption that similar profiles are correlated with close positions in the system. Although this is generally a good assumption in biological systems, it is not a mathematical necessity. How can we be sure that we are on the right track in a top-down approach? This is the reason we need to take both top-down and bottom-up approaches. The top-down approach is to decompose the system to smaller parts. The bottom-up approach is to reconstitute elemental steps into larger parts. If the results of these approaches meet in the middle, and if they are consistent, we can be confident that we are on the right track. In other words, we can use information from the reductionist approach as constraints in model building. Although high-throughput reductionism driven by systems biology will produce a large amount of information, it is unlikely that it will be sufficient to reconstitute entire systems using only the bottom-up approach. Using the top-down approach, some parts can be tentatively left as black boxes defined by mathematical functions. These mathematical functions would define relationships between the inputs and outputs of a box; in practical terms, this would require extrapolation and extrapolation of the relationships determined by the data. In this way, we could build approximate system models even when our knowledge about the system is not complete.

CONSTRAINTS ON SYSTEMS BEHAVIOR—ROBUSTNESS

Intrinsic features of systems behavior can be used as constraints in model building. Many biological systems are intrinsically robust. For example, under many different environmental conditions, a catfish develops as a catfish. The developmental program of the fish is robust. If the system were not robust, small fluctuations in things such as the gradient of a morphogen or the number of morphogen receptors in certain cells could change the fish shape substantially. The model for development must include intrinsic mechanisms such as topology and dynamics of the system that explain the robustness. This is a strong constraint. Therefore, it is important to identify robust components of the behavior of the system of interest. When we decompose the system by the top-down approach, we will be able to observe which parts of the system are robust and the points at which robust parts start to lose robustness or lose robustness due to another decomposition. In this way, we can define minimum sizes of subsystems that are responsible for the robust system behavior. We need to build models of the subsystems that satisfy the robustness constraints of the whole system.

MAKING USE OF UNKNOWN FACTORS

We should be aware that there are always experimental conditions that we cannot control. This situation is much worse with plants than with yeast, which grows in a synthetic medium in the same tubes at the same temperature. In contrast, it would be very costly to make the light intensity/quality and airflow...
very homogeneous for every single plant. Individual differences in plant body shape and size makes the microenvironment for each plant different. In addition, some biological mechanisms may involve stochastic events. We often observe that some quantitative aspects of phenotype fluctuate significantly. For example, in planta growth of phytopathogenic bacteria fluctuates significantly (but generally within a definable range) day-to-day although the relative bacterial growth between plants different genotypes is fairly stable. This type of day-to-day fluctuation may be caused by uncontrolled environmental factors or stochastic events. In either case, if the phenotype variation is correlated with certain conditions of the cell, which can be described by global profiling, we can use this correlation information as a constraint for the model building. The power of exhaustive, simultaneous descriptions provides a possibility of extracting information about less robust aspects of system behavior.

IMPORTANT FACTORS FOR SUCCESS

Whether we take top-down or bottom-up approaches, the quality of data is always crucial. If our data are not of sufficient quality, we will experience garbage-in-garbage-out. Of course, it is best to have high-quality data all the time. However, in general, the money and time required to obtain higher quality data increase exponentially, and in many cases the current technological capability is the limiting factor. Practically, we need to settle at a sufficient level. Criteria, such as what error level is tolerable and what resolution in space and time is sufficient, depend on what sort of analysis we perform. It also depends on characteristics of the part of the system that is analyzed. Therefore, experimenters must be familiar with the factors that are important for the analysis of the part of the system they are studying. As illustrated above, in contrast to the reductionist approach in which simple, intuitive logic is applied, the top-down approach demands more mathematical skills because complex relationships and dynamics must be derived from large data sets. We need to use mathematical methods to make large data sets interpretable to us. Furthermore, we will need to push the frontier of mathematics to enable model building for complex biological systems. These activities will require close collaborations between biologists and mathematicians. General understanding of what collaborators do is crucial for an efficient collaboration. These considerations lead to one conclusion: We need more scientists who are truly interdisciplinary, including biologists with broad mathematical backgrounds and mathematicians with broad biological backgrounds. As the advance of molecular biology demanded that biologists learn molecular thinking, the advance of systems biology demands that biologists learn mathematical thinking.

CONCLUDING REMARKS

The top-down approach that has been made feasible by advances in systems biology, in combination with conventional and high-throughput reductionist approaches, has great potential to enable us to build quantitative models of complex biological systems. We are at the verge of the next revolution in biology research, which is driven by systems biology. As we witnessed before, a scientific revolution demands that scientists obtain new skill sets. This time the skill set required for biologists is mathematics.

ACKNOWLEDGMENT

I thank Jane Glazebrook for critical reading of the manuscript.

Fumiaki Katagiri*  
Torrey Mesa Research Institute  
Syngenta Research and Technology  
San Diego, California 92121