

CHAPTER 1

EXECUTIVE SUMMARY AND INTRODUCTION

Marvin Cassman

BACKGROUND

Systems biology has become a major force in the past five to seven years. As with all new developments in science, the emergence of new approaches is a result of limitations in the existing model, in this case the limitations of molecular biology. For the past 40 years the paradigm for predicting phenotype has focused on single gene defects. This extraordinarily powerful approach has been the major contributor to an understanding of the function of individual genes and proteins. It seems less likely that it will yield an understanding of complex biological behavior, from individual cellular activities such as motility to the operation and integration of organ systems.

The current textbook image of biological processes is that of a static model of loosely linked, highly detailed, molecular devices. However, every biologist knows that dynamic processes drive biology. The physiologist Walter Cannon provided a clear statement of this concept in 1932 when he coined the term “homeostasis.” One contemporary definition of a homeostatic system is “... an open system that maintains its structure and functions by means of a multiplicity of dynamic equilibriums rigorously controlled by interdependent regulation mechanisms.” This description neatly encapsulates several of the key issues in systems biology—dynamic processes, interdependent regulatory controls, and the operation of multiple interacting components. The inability of most of modern biology to take these behaviors into account is certainly a major reason why the function and malfunction of complex biological processes is still poorly understood, despite biologists’ increasingly detailed knowledge of the components of these processes. Systems biology brings the promise, if not yet the reality, of offering a more complete understanding of health and disease. (It is striking that at about the same time as Cannon was developing the concept of homeostasis, H. S. Black was refining the concept of negative feedback control at Bell Labs. The current active involvement of engineers in systems biology can be considered to be a convergence of these two threads of thought.)

All this interest has emerged despite the inability to arrive at a consensus definition of systems biology. Elements that appear in virtually all definitions are “networks,” “computation,” “modeling” and often, “dynamic properties.” For the purposes of this study the objective of systems biology has been defined as the understanding of network behavior, and in particular their dynamic aspects, which requires the utilization of mathematical modeling tightly linked to experiment. This involves a variety of approaches, such as the identification and validation of networks, the creation of appropriate datasets, the development of tools for data acquisition and software development, and the use of modeling and simulation software in close linkage with experiment, often done to understand dynamic processes. Of course, the definition becomes ambiguous at the margins. But at the core is the focus on networks, which makes it clear that the goal is to understand the operation of the systems, rather than the component parts. It also tries to distinguish between systems biology and what the WTEC study panel defined as systematic biology. Systematic biology can be considered the large-scale, high-throughput collection of specific data sets and their organization and interpretation, usually through the application of advanced bioinformatics tools. Clearly, this is at one boundary of systems biology, since the data sets may be used for systems analysis. Although these two very recent approaches to understanding biology are closely linked and have developed over more or less the same

time frame, they are quite distinct. Systematic biology is a consequence of the enormous success of the genome program, together with the development of new technologies for the high-throughput collection and analysis of data on individual molecular events. In this respect it is a lineal descendent of the primary driver for biology in the second half of the 20th century, molecular biology/genetics. In contrast, systems biology has a much older lineage, and can be considered a variant of computational physiology, one example of which is the Hodgkins-Huxley equation of 1952. More recent approaches include the development of metabolic control analysis in the 1970s (Kacser, Burns, 1973; Heinrich, Rapoport, Rapoport, 1977) and detailed anatomic- and molecularly based models of cardiac function (Kohl, Noble, Winslow, Hunter, 2000). These approaches, as well as systems biology, are distinguished from earlier attempts at systems modeling in that they incorporate the detailed molecular knowledge that has been and continues to be generated. However, they all assume that an understanding of physiological functions (phenotype) requires knowledge of the behavior of systems or networks.

At the heart of systems biology is the need to couple advanced modeling and simulation with experiment. This thrusts biology into a new era. For the past century the tools and concepts of chemistry have driven biology. This is so much a part of the landscape of modern biology that it will be shocking to many when, in the next century, biology becomes largely driven by engineering and physics. This is a consequence of the fact that understanding the dynamics of even the simplest biological networks requires the application of mathematical approaches and the generation of models and simulations. These mathematical tools are not now part of the average biologists' training. Indeed, biology has almost become the province of those who want to do science without learning mathematics. Consequently, two major issues in the evolving field of systems biology are needed to create functional collaborations between engineers, physicists, and biologists, and to produce a new generation of scientists that will be conversant with both the mathematical tools and the biological systems. Both of these issues were a major concern of the study.

Having said all this, it must be stressed that systems biology is still a very new field. Although individual investigators have for some years been studying the properties of biological networks using quantitative approaches, until recently they were few, relatively isolated and, to a significant extent, ignored. In part, this was due to a long-standing bias against modeling and simulation, particularly in cell biology. This bias was not completely irrational, since many modeling approaches had very little connection to experiment and consequently rarely told biologists anything they wanted to know. In contrast, the close ties of computation and modeling with experiment distinguish modern systems biology. This will be repeated often in the following chapters. It is also worth repeating Michael Faraday's comment, "All this is a dream. Still, examine it by a few experiments. Nothing is too wonderful to be true, if it be consistent with the laws of nature, and in such things as these, experiment is the best test of such consistency" (Hamilton, 2005).

The investment by government agencies of millions of dollars of federal funding since 1998 demonstrates the increasing interest in systems biology. In addition, the major journals have devoted special issues to the subject. A new journal of systems biology, called *IEE Systems Biology*, and most recently an online journal entitled "Molecular Systems Biology" published jointly by EMBO and Nature Publishing Group, have recently been established. Institutes have sprung up, numerous meetings are held, and increased funding can be observed both in the U.S. and abroad. Yet core issues affecting the progress of the field remain to be resolved, and some of these will be addressed in this volume. If systems biology is not quite yet a discipline, it is clearly more than a fad.

A final caveat is that limitations of time and resources prevented us from visiting many important research sites around the world. For example, we regret that we were not able to see the work underway in the Scandinavian countries and Israel, as well as Asian nations such as Singapore and Korea. Additionally, we missed numerous laboratories even in the countries we visited, and consequently some important research areas were surely neglected.

OBJECTIVES OF THE WTEC STUDY

The recent growth in interest in systems biology has not been accompanied by a systematic evaluation of activities in the U.S. and abroad. Led by the National Science Foundation, a number of U.S. governmental agencies involved in the support of research asked the World Technology Evaluation Center (WTEC) to

conduct a study of systems biology activities in the U.S. and abroad to support important policy and funding initiatives. The goals are:

- To understand the state of current research
- To determine what is needed to support future research
- To understand the opportunities for international collaboration

The United States needs knowledge of and access to the latest international developments in this field in order to proceed expeditiously with promising applications in this rapidly developing field. The number and diversity of its sponsors reflect the breadth of interest in this study. These include the National Science Foundation, the Department of Energy (DOE); the Defense Advanced Research Projects Agency (DARPA) of the Department of Defense; the National Aeronautics and Space Administration (NASA); the National Cancer Institute (NCI) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health; the National Institute of Standards and Technology (NIST); and the Environmental Protection Agency (EPA).

PANEL MEMBERS

- Marvin Cassman, San Francisco, CA (Chair)
- Adam Arkin, University of California, Berkeley
- Frank Doyle, University of California, Santa Barbara
- Fumiaki Katagiri, University of Minnesota
- Douglas Lauffenburger, MIT
- Cynthia Stokes, Entelos Corp.

STUDY SCOPE

Broadly, the scope of the study included:

- Organization and regulation of biological networks
- Tools for analyzing the spatial and temporal behavior of networks
- Approaches to the education of graduates and undergraduates in systems and computational biology
- Trends in government interest and support of systems biology programs

The report follows an outline that was first presented at a meeting with sponsors on February 23, 2004, further defined at a workshop with prominent U.S. researchers on June 4, 2004, and refined over the months of discussion and visits carried out by the panel. It reflects not only activities observed during the site visits but also research directions which, in the opinion of the panelists, were often underrepresented but require more emphasis to ensure the progress of the field. The format for this volume is:

1. Introduction and Executive Summary (Marvin Cassman)
2. Data Generation and Analysis (Fumiaki Katagiri and Adam Arkin)
3. Systems Inference (Frank Doyle and Douglas Lauffenburger)
4. Network Organization and Modeling (Cynthia Stokes and Adam Arkin)
5. Education, Infrastructure, and National Programs (Marvin Cassman, Frank Doyle, Douglas Lauffenburger)
6. Plant Science (Fumiaki Katagiri)

PLAN OF THE STUDY

The first formal discussion of the study occurred at a workshop at NSF on February 23, 2004. In addition to most of the panelists and WTEC staff, there were representatives from all of the organizations sponsoring the study. On top of a broad discussion of the goals of the study, an outline was generated for a workshop to provide baseline information on U.S. activities in systems biology. This workshop was held on June 4, 2004, at NSF.

Site visits were made to Europe and Great Britain on July 5–9, 2004, and to Japan on December 13–17, 2004. Team members visited 16 sites in the EU and Switzerland and 12 in Japan. The site reports are appended to this report.

At all the sites that were visited (see Table 1.1), the hosts treated us with the utmost consideration. The study sponsors (in particular Fred Heineken and Semahat Demir of NSF who traveled with the panel) and participants thank them for their hospitality with the hope that this volume and anything that emerges from it will prove of value to them.

Table 1.1
Sites visited in Europe and Japan

Europe		
Site	Panelists	Date
Cambridge University, Department of Anatomy	Katagiri, Lauffenburger, Stokes	9 July 2004
European Bioinformatics Institute	Katagiri, Lauffenburger, Stokes	9 July 2004
European Commission Office	Ali, Arkin, Cassman, Doyle, Heineken	8 July 2004
German Cancer Research Center (DKFZ) Heidelberg	Arkin, Cassman, Doyle, Heineken	7 July 2004
Humboldt University	Arkin, Cassman, Doyle, Heineken	6 July 2004
Max Planck Institute for Molecular Genetics	Cassman, Doyle, Heineken, Katagiri	5 July 2004
Max Planck Institute for Molecular Plant Physiology	Arkin, Cassman, Heineken, Katagiri	5 July 2004
Oxford Brookes University, School of Biological and Molecular Sciences	Lauffenburger, Stokes	6 July 2004
Oxford University, Centre for Mathematical Biology/Mathematical Institute	Ali, Lauffenburger, Stokes	5 July 2004
Oxford University, Department of Physiology	Ali, Lauffenburger, Stokes	6 July 2004
Sheffield University, Computational Biology Research Group	Katagiri, Lauffenburger, Stokes	8 July 2004
SystemsX	Cassman	29–30 June 2004
University College London	Ali, Lauffenburger, Stokes	5 July 2004
Université Libre (Free University) De Bruxelles	Ali, Arkin, Cassman, Doyle, Heineken	8 July 2004
University of Warwick, Mathematics Institute	Katagiri, Lauffenburger, Stokes	7 July 2004
Vrije Universiteit (Free University) Amsterdam	Ali, Arkin, Cassman, Doyle, Heineken	9 July 2004
Japan		
Computational Biology Research Center (CBRC)	Arkin, Cassman, Demir, Doyle, Horning, Katagiri, Stokes	14 Dec 2004
Japan Biological Information Research Center (JBIRC)	Arkin, Cassman, Demir, Doyle, Horning, Katagiri, Stokes	14 Dec 2004
Kazusa DNA Research Institute	Cassman, Horning, Katagiri	15 Dec 2004
Keio University, Institute for Advanced Biosciences (IAB)	Arkin, Cassman, Horning, Katagiri	17 Dec 2004
Keio University, Symbiotic Systems Project	Arkin, Cassman, Demir, Doyle, Horning, Katagiri, Stokes	13 Dec 2004
Kyoto University, Bioinformatics Center	Demir, Doyle, Stokes	17 Dec 2004
Kyoto University, Cell/Biodynamics Simulation Project	Demir, Doyle, Hane, Stokes	17 Dec 2004
RIKEN Yokohama Institute	Arkin, Cassman, Demir, Doyle, Hane, Katagiri, Stokes	16 Dec 2004
Tokyo Medical and Dental University	Arkin, Cassman, Demir, Doyle, Horning, Katagiri, Stokes	15 Dec 2004
University of Tokyo, Department of Computational Biology	Arkin, Demir, Doyle, Stokes	15 Dec 2004
University of Tokyo, Institute of Medical Science	Arkin, Cassman, Demir, Doyle, Horning, Katagiri, Stokes	13 Dec 2004
University of Tokyo, Laboratory of Systems Biology and Medicine (LSBM)	Arkin, Cassman, Demir, Doyle, Horning, Katagiri, Stokes	13 Dec 2004

PRINCIPAL FINDINGS

General Conclusions

Over the past decade numerous individual laboratories around the world have been engaged in systems biology. Some of these investigators include Lauffenburger (Lauffenburger, Forsten, Wiley HS (1995), Arkin (Arkin, Ross, McAdams (1998), McAdams (McAdams, Shapiro, 1995), Leibler (Barkai, Leibler, 1997), and Savageau (Hlavecek and Savageau, 1997) in the U.S.; Bray (Levin, Morton-Firth, Abouhamad, Bourret, Bray, 1998) and Noble (McCulloch, Bassingthwaighe, Hunter, Noble, 1998) in the U.K.; Heinrich (Wolf, Heinrich, 1997), Westerhoff (Westerhof, 1995), and Goldbeter (Goldbeter, 2002) in Europe; and Kitano (Kitano, 2002), Tomita (Tomita, Hashimoto, Takahashi, Shimizu, Matsuzaki, Miyoshi, Saito, Tanida, Yugi, Venter, Hutchison, 1999), and Kanehisa (Kanehisa, 2000) in Japan. However, few of these efforts were matched by either significant national funding or institutional interest. This changed as the profusion of data and the complexity of regulatory processes mounted. The interest generated by the need to integrate molecular data into a systems approach in turn stimulated events over the last five to seven years in the U.S., and more recently elsewhere, when large investments in systems biology began to be made by national entities and research institutions.

Largely because of its head start, the WTEC panel rates the U.S. as currently ahead of the rest of the world in systems biology. The lead is reflected in the larger number of active groups, greater number of educational programs underway, and the more diverse and growing funding base. However, there is evidence of rapid development outside the U.S., much of it begun in the last two to three years. It must be stressed that the attempt to incorporate the details of molecular events obtained over the past half-century into a dynamic picture of network behavior in biological systems is only just beginning, in the U.S. and elsewhere. In particular, progress in the core activity of systems biology—modeling tied to experiment—is still limited. Successes, however defined, remain few and controversies abound. Training, research, and infrastructure all would benefit from strong international collaborations that could provide examples of novel approaches. For example, Japan and Germany have developed large-scale organizations that can address specific research issues, e.g. the Max Planck Institutes in Germany and RIKEN in Japan. The U.S. has a limited capability to create such structures and needs to develop inter-agency collaborations that will identify and support activities of this kind. Overall, the picture is of an active field in the early stages of explosive growth.

Databases and Data

The production of data and the construction of databases are visible at a roughly similar scale in the U.K., EU, and Japan, although the development of large databases in Japan was particularly striking. However, the databases examined, in the U.S. and abroad, were not always valuable to investigators developing and testing models of biological processes.

Two opposite trends exist in database organizations: large inclusive databases and small specialty databases. Both approaches have advantages and disadvantages. Large-scale databases, which primarily collect information that is not closely tied to the state of a cell, have become quite common, and their value is well understood. (How useful these data are for systems biology is less clear. In general, the degree of quantitation is too limited to be used by investigators developing and testing models of biological processes.) Standardization, although not complete, is progressing. This is not the case for many other kinds of data, particularly those tied to biological processes that are strongly conditioned by the state of the cell. It is not even clear how much “meta-data” is needed. Gene expression, protein expression, molecular localization, interactions, and post-translational modification are highly conditional. Indeed, the strain of cell used, the media, and other measurement conditions can appreciably affect the measured outcomes. There are a number of related issues, such as the amount of raw data needed, and the availability of statistical analysis and software packages used. These issues are not unique to data used for systems biology, but their absence is even more critical than in the analysis of state-independent data.

Models for data production and data storage in systems biology are highly variable, ranging from large centers with massive accumulations of high-throughput data, to small, manually curated databases. Whichever model is used, the absence of data standards that permit groups other than the producer to use, analyze, and evaluate the results is clearly a significant barrier to progress. This is an international issue, and must be solved by broad collaborative interactions.

System Inference

One finds numerous network inference studies in all of the regions described with the U.S. and Japanese (and Israeli) groups leading in the development of methodologies. All regions showed exciting application studies, with significant potential for “success stories” to emerge in the coming years.

The encouraging trends that were observed included: (i) multiple, complementary approaches to the regression of models for network inference, (ii) the incorporation of motifs and modules into network inference methods, (iii) the emergence of a nice interplay between the classical static network databases and the formats for dynamic systems biology models (*e.g.*, SBML), and (iv) the initiation of a considerable amount of curricular development in this area (notably in bioinformatics).

Of concern was the fact that the issues of: (i) explicit incorporation of dynamics, (ii) identifiability and (in)validation of models, and (iii) model iterations with design of experiment, were receiving only modest attention in the regions, with noteworthy efforts in the U.S. and Europe (particularly Germany). There were many reported examples of researchers identifying large numbers of parameters from relatively small data sets. However, there appear to be a number of groups working towards solutions to these challenges, and considerable progress can be expected in the next two to three years.

Modeling and Network Organization

Modeling and network organization analysis efforts are utilized in many areas of biological study and in all countries visited, but are definitely not ubiquitous throughout biological and biomedical research. The panel found that research efforts that closely integrated modeling with experimental work were the most productive in terms of driving new understanding of a biological system. Related to this, substantially more effort using model-based experimental design is needed to attain the data that most efficiently leads to maximally useful models. In addition, better tools for model-experiment comparison would be helpful. Significant resources are being invested in the development of modeling and simulation software worldwide, and at least some duplication of effort is apparent. Sharing of models between researchers remains a challenge but is being addressed by the development of several markup languages. Finally, the involvement and interest of industry in use of modeling in biology is significant although, again, not ubiquitous.

Plant Systems Biology

Progress of systems biology research in the plant field has been slow. However, some advanced studies shed light on unique aspects of plants. Several actions are needed to promote systems biology of plants.

To make the most out of limited funding:

- Focus on model plant species. It is clear that the majority of advanced studies have been performed with model plant species, such as *Arabidopsis*.
- Cooperate rather than compete at the global level.

To compensate for the bias against promotion of systems biology in the research community:

- Implement a sustaining, targeted funding program in plant systems biology.

To raise the next generation of researchers:

- Train biology-major students in quantitative science.
- Recruit students oriented to mathematics, engineering, physics, and chemistry into plant biology.

Education, National Programs, and Infrastructure

The future of systems biology will depend on three critical elements: education of a new generation of scientists who have both biological and mathematical training; the availability of funding that operates outside of disciplinary boundaries; and the availability of a supportive infrastructure that can accommodate the needs of an intrinsically interdisciplinary research area.

Education

The general impression is that most of the formal teaching programs, in the U.S. and abroad, are in bioinformatics rather than systems biology. Relatively few examples exist of training in modeling focused on biological systems, and where they do exist they tend to be isolated courses rather than fully integrated

programs in systems biology. Most of the programs are somewhat *ad hoc* “menu selection” curricula. The difficulty of training quantitative students in biology and *vice versa* is clearly well understood and no real solution has yet been provided, although a number of experiments are underway. It is much too early to tell which, if any, of these are successful in producing qualified researchers in systems biology. Given the importance of this issue and its embryonic state, some mechanisms for exchanging information internationally and locally on best practices is essential.

National Programs

The U.S. remains one of the few countries that offers a significant targeted investment in systems biology. A clear exception is Germany, which has developed a new initiative in the systems biology of hepatocytes, beginning in January 2004. In the last few years, national programs have also been initiated in Switzerland and the U.K., and international programs at the EU level. Additionally, activities in systems biology are underway in many locations, as part of ongoing “traditional” governmental support programs. This is particularly noticeable in Japan. However, it is hard to avoid the conclusion that both the breadth and the scale of systems biology support from governmental entities are significantly greater in the U.S. than elsewhere in the world.

A possible caveat to this conclusion depends on the definition of systems biology. As noted earlier, there is a distinction between “systems biology” and “systematic biology.” Systematic biology, the high-throughput collection of targeted data sets, is a booming business everywhere, fuelled by the success of the genome project. Systems biology, the computational analysis of biological networks, is much more sparsely represented. Although this is also true in the U.S., encouragement of these activities through federal funding programs is significant and growing. It was slightly discouraging to see how frequently systems and systematic biology were conflated. Although data collection is clearly critical, it was not often the case that there was a connection between the data collected and its potential use in modeling and simulation of biological systems. In general, the future of systems biology worldwide depends on the support of programs that consider experimental and data-driven approaches together with the computational methods needed to model specific biological problems. Relatively few funding programs focus explicitly on this.

Infrastructure

The infrastructure to be discussed in this study is limited to large-scale resources, specifically databases, software repositories, and centers. In order to ensure both standardization and access, it is strongly recommended that centralized resources be developed for both software and data. The third issue is the value of centers for systems biology. The creation of specialized centers is much more common in Europe and Japan than in the U.S., although the development of high-throughput centers for DNA sequencing and structural biology have proven their value. It is suggested that centers targeted to specific research problems, and specific experimental systems, could benefit systems biology in the U.S. The need for consistent and reproducible data and the need for close collaboration between theorists and experimentalists are both arguments for co-located groups that can interact easily and often. It is also far easier to enforce standards at such centers. At this point in time, systems biology can benefit from stronger centralized approaches that will allow the testing of model systems in an optimum environment.

REFERENCES

- Arkin, A., J. Ross, H. H. McAdams. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149: 1633–48.
- Barkai, N., S. Leibler. 1997. Robustness in simple biochemical networks. *Nature* 387: 913–7.
- Goldbeter, A. 2002. Computational approaches to cellular rhythms. *Nature* 420: 238–45.
- Hamilton, J. 2005. “A Life of Discovery,” cited in NYT Book Reviews, p. 16, March 13, 2005.
- Heinrich, R., S. M. Rapoport, and T. A. Rapoport. 1977. Metabolic regulation and Mathematical Models. *Prog Biophys Mol Biol* 32: 1–82.
- Hlavacek W. S., M. A. Savageau. 1996. Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol* 255: 121–39.
- Kacser, H. and J. A. Burns. 1973. The Control of Flux. *Symp Soc Exp Biol* 27: 65–104.

- Kanehisa, M. 2000. Pathway databases and higher order function. *Adv Protein Chem* 54: 381–408.
- Kitano, H. 2002. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr Genet* 41: 1–10.
- Levin, M. D., C. J. Morton-Firth, W. N. Abouhamad, R. B. Bourret, D. Bray. 1998. Origins of individual swimming behavior in bacteria. *Biophys J* 74: 175–81.
- Kohl, P. D., R. L. Noble, R. L. Winslow, and P. Hunter. 2000. Computational modeling of biological systems: tools and visions. *Phil Trans R.Soc Lond A* 358: 576–610.
- Lauffenburger, D. A., K. E. Forsten, B. Will, H. S. Wiley. 1995. Molecular/cell engineering approach to autocrine ligand control of cell function. *Ann Biomed Eng* 23: 208–15.
- McAdams, H. H., L. Shapiro. 1995. Circuit simulation of genetic networks. *Science* 269: 650–6.
- McCulloch, A., J. Bassingthwaight, P. Hunter, D. Noble. 1998. Computational biology of the heart: from structure to function. *Prog Biophys Mol Biol* 69: 153–5.
- Tomita, M., K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter, C. A. Hutchison 3rd. 1999. E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15: 72–84.
- Westerhoff, H. V. 1995. Subtlety in control—metabolic pathway engineering. *Trends Biotechnol* 13: 242–4.
- Wolf, J., R. Heinrich. 1997. Dynamics of two-component biochemical systems in interacting cells; synchronization and desynchronization of oscillations and multiple steady states. *Biosystems* 43: 1–24.