Making Sense of Complexity

Summary of the Workshop on Dynamical Modeling of Complex Biomedical Systems

George Casella, Rongling Wu, and Sam S. Wu
University of Florida

Scott T. Weidman
National Research Council

Board on Mathematical Sciences and Their Applications
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This summary is based on work supported by the Burroughs Wellcome Fund, Department of Energy, Microsoft Corporation, National Science Foundation (under Grant No. DMS-0109132), and the Sloan Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors.

International Standard Book Number 0-309-08423-7

Additional copies of this report are available from:

Board on Mathematical Sciences and Their Applications
National Research Council
2101 Constitution Avenue, N.W.
Washington, DC 20418

Printed in the United States of America

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

BOARD ON MATHEMATICAL SCIENCES AND THEIR APPLICATIONS

Staff

Preface

On April 26-28, 2001, the Board on Mathematical Sciences and Their Applications (BMSA) and the Board on Life Sciences of the National Research Council cosponsored a workshop on the dynamical modeling of complex biomedical systems. The workshop's goal was to identify some open research questions in the mathematical sciences whose solution would contribute to important unsolved problems in three general areas of the biomedical sciences: disease states, cellular processes, and neuroscience. The workshop drew a diverse group of over 80 researchers, who engaged in lively discussions.

To convey the workshop's excitement more broadly, and to help more mathematical scientists become familiar with these very fertile interface areas, the BMSA appointed one of its members, George Casella, of the University of Florida, as rapporteur. He developed this summary with the help of two colleagues from his university, Rongling Wu and Sam S. Wu, assisted by Scott Weidman, BMSA director.

This summary represents the viewpoint of its authors only and should not be taken as a consensus report of the BMSA or of the National Research Council. We are grateful to the following individuals who reviewed this summary: Peter J. Bickel, University of California at Berkeley; Ronald Douglas, Texas A&M University; Nina Fedoroff, Pennsylvania State University; and Keith Worsley, McGill University.

Funding for the workshop was provided by the Burroughs Wellcome Fund, the Department of Energy, Microsoft Corporation, the National Science Foundation, and the Sloan Foundation. The workshop organizers were Peter J. Bickel, University of California at Berkeley; David Galas, Keck Graduate Institute; David Hoel, Medical University of South Carolina; Iain Johnstone, Stanford University; Alan Perelson, Los Alamos National Laboratory; De Witt Sumners, Florida State University; and James Weiss, University of California at Los Angeles

Videotapes of the workshop's presentations are available online at <http://www.msri.org/publications/video/index6.html/> and also through a link at <http://www.nas.edu/bms>.

Contents

1

Introduction

This report documents a recent workshop[1] at which approximately 85 biomedical scientists, mathematicians, and statisticians shared their experiences in modeling aspects of cellular function, disease states, and neuroscience. The topics were chosen to provide a sampling of the rapidly emerging research at the interface of the mathematical and biomedical sciences, and this summary has been prepared as an introduction to those topics for mathematical scientists who are exploring the opportunities from biomedical science. While a range of challenges and approaches was discussed at the workshop, its overall theme was perhaps best summarized by discussant Jim Keener, of the University of Utah, who noted that what researchers in these areas are really trying to do is "make sense of complexity." The mathematical topics that play important roles in the quest include numerical analysis, scientific computing, statistics, optimization, and dynamical systems theory.

Many biological systems are the result of interwoven interactions of simpler behaviors, with the result being a complex system that defies understanding through intuition or other simple means. In such a situation, it is critical to have a model that helps us understand the structure of the phenomenon, and we look to the mathematical sciences for the tools with which to construct and investigate such models. Although the experimental data from biological systems and the resulting models can be bewildering in their complexity, a minimal model can sometimes expose essential structure. An example is given in Figure 1-1, which shows the simple (and pleasing) linear relationship between the level of DNA synthesis in a cell and the integrated activity of the ERK2 enzyme.[2] After understanding such basic elements of cell signaling and control, one may then be able to construct a more complex model that better explains observed biomedical phenomena. This evolution from basic to more complex was illustrated by several workshop talks, such as that of Garrett Odell, of the University of Washington,

[1] "Dynamical Modeling of Complex Biomedical Systems," sponsored by the Board on Mathematical Sciences and Their Applications and the Board on Life Sciences of the National Research Council, held in Washington, D.C., April 26-28, 2001.

[2] ERK2, the extracellular-signal-regulated kinase 2, is a well-studied human enzyme. In response to extracellular stimuli, such as insulin, it triggers certain cellular activity, including, as suggested by Figure 1-1, DNA synthesis.

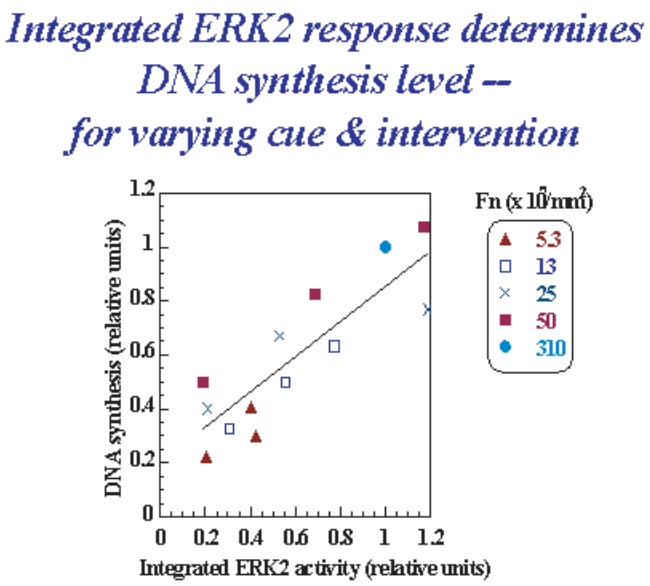

FIGURE 1-1 DNA synthesis dependence on ERK signal for varying cue and intervention. Fn is fibronectin. Figure courtesy of Douglas Lauffenburger.

which presented a model that grew from 48 to 88 parameters, and that of Douglas Lauffenburger, of the Massachusetts Institute of Technology, which described how a model grew in complexity as his group worked to capture the relationship between insulin response and ERK2. Because the phenomenology of most biomedical processes is so complex, a typical development path for biomedical modeling is to start with a model that is clearly too simple and then evolve it to capture more of nature's complexity, always avoiding any detail whose effect on the phenomenology is below some threshold of concern.

The workshop opened with a welcome from Peter Bickel, the chair of the Board on Mathematical Sciences and Their Applications (BMSA). Bickel remarked that one mission of the BMSA is to showcase the role that the mathematical sciences play in other disciplines, and this workshop was planned to do that. The 16 talks, given by researchers at the interface between the mathematical and biomedical sciences, all illustrate how the mathematical and biological sciences can interact for the benefit of both. The presentations were videotaped and subsequently made available at <www.msri.org/publications/video/index6.html/>, with a link from <www.nas.edu/bms>.

Two important principles emerged from the workshop:

1. Successful modeling starts with simple models to gain understanding. If the simple model succeeds somewhat in capturing the known or anticipated behavior, then work to refine it.

2. When biomedical processes are modeled with mathematical and statistical concepts, the underlying structure of the biological processes can become clearer. Knowledge of that structure, and of the way its mathematical representation responds to change, allows one to formulate hypotheses that might not be apparent from the phenomenological descriptions.

While these principles are not new or unique to modeling in the biomedical sciences, they may not be obvious to mathematical scientists whose previous experience is with models that are based on well-established laws (e.g., mechanical or electromagnetic modeling) or who have not worked in data-intensive fields. In modeling very complex behaviors such as biomedical phenomena, these principles are the hallmark of good research.

2

Modeling Processes Within the Cell

In the 20th century our ability to describe and categorize biological phenomena developed from the organismal level down to the gene level. The 21st century will see researchers working back up that scale, composing genetic information to eventually build up a first-principles understanding of physiology all the way to the level of the complex organism. Figure 2-1 shows this challenge schematically. The 21st century advances in bioinformatics, structural biology, and dynamical systems modeling will rely on computational biology, with its attendant mathematical sciences and information sciences research. As a first step, the huge amount of information coming from recent advances in genomics (e.g., microarray data and genetic engineering experiments) represents an opportunity to connect genotype and phenotype[1] in a way that goes beyond the purely descriptive.

Workshop speaker James Weiss, of the University of California at Los Angeles, outlined a strategy for going in that direction by first considering a simple model that might relate the simple gene to the complex organism. His strategy begins by asking what the most generic features of a particular physiological process are and then goes on to build a simple model that could, in principle, relate the genomic input to those features. Through analysis, one identifies emergent properties implicit in the model and the global parameters that identify the model's features. Physiological details are added later, as needed, to test experimental predictions. This strategy is counter to a more traditional approach in which all known biological components would be included in the model. Weiss's strategy is necessary at this point in the field's development because we do not know all the components and their functions, nor would we have the computational ability to model everything at once even if that information were available.

This principle of searching for a simple model was apparent throughout Weiss's presentation, which showed how a combination of theoretical and experimental biology could be used to study a complex problem. He described research that modeled the causes of ventricular fibrillation. The first attempts at

[1] A genotype is a description or listing of a cell or organism's genetic information, while the cell or organism's phenotype is a description of its resulting features and/or functions.

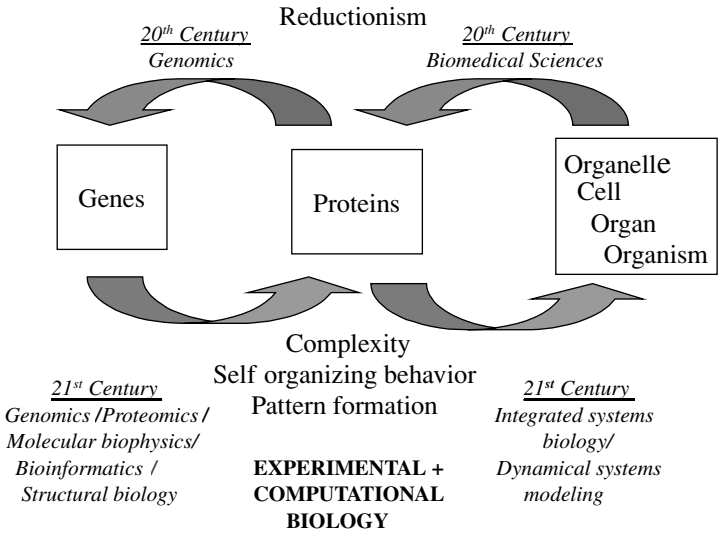

FIGURE 2-1 Directions of scientific investigation. Figure courtesy of James Weiss.

controlling fibrillation focused on controlling the triggering event, an initial phase of ventricular irregularity. However, it was found that a drug therapy that controlled this event did not decrease mortality from ventricular fibrillations. Thus, there was a need to understand better the chain of causality behind ventricular fibrillation.

Using the basic premise that cardiac tissue is an excitable medium, Weiss proposed a wave model. In his model, fibrillation is the result of a breaking wave, and the onset of fibrillation occurs when the wave first breaks; it escalates into full fibrillation as the wave oscillation increases. The cause of the wave breakage was thought to be connected to the occurrence of a premature beat. If the wave could not recover from this premature impulse (recovery is called "electric restitution"), oscillation would develop. This basic concept was modeled through the following simple equation:

$$\text{Wavelength} = \text{APD} \times \text{Conduction velocity}$$

where APD is the action potential duration. Supercomputer simulations of wave patterns in two- and three-dimensional cardiac tissue, based on this simple equation, showed that the wave patterns undergo a qualitative shift in their characteristics (being either spiral or scroll waves) depending on whether the parameter APD is less than or greater than unity. When APD > 1, the impulses come too rapidly for the wave to recover (i.e., for electric restitution to take place), and fibrillation results. Thus the simulations suggested that holding APD below unity might result in tissue that can recover rather than fall into fibrillation mode. Because drugs are available that can lower APD, it was possible to verify the simulated results in real tissue (a pig ventricle). This suggests the possibility of an important drug intervention that was not of obvious importance before Weiss carried out his simulations. See Garfinkel et al. (2000) for more details.

The graph of DNA synthesis as a function of integrated ERK2 activity shown in Figure 1-1 is another example of how a simple model can sometimes capture the effective behavior of a complex process. The complex process here is one case of how a molecular regulating network governs cell functions. In general, protein signaling causes interconnected, complicated networks to form (see, e.g., Hanahan and Weinberg, 2000, or Figure 2-2 below). The protein signaling pathways include membrane receptors (sensors), intracellular signal cascades (actuators), and cell functional responses (outputs), and one obvious approach to modeling this network would view it as consisting of three parts:

Cues → Intracellular signals → Cell function

Douglas Lauffenburger of MIT adopted this approach to model the quantitative dynamics of the ERK2 signal as it responds to an external cue (fibronectin, a protein involved in many important cellular processes) and helps lead to the cell function of synthesizing DNA. After introduction of the external cue, ERK2 activity increases and peaks at 15 minutes, and then it drops. Amazingly, the DNA synthesis level appears to be linearly dependent on the integrated ERK2 activity, as shown in Figure 1-1. This striking result suggests that there is no need, at this level, to model the network of intracellular signals in detail. Instead, they can be replaced by the de facto linear relationship.

However, simple models are not always sufficient, and in the case of multiple external cues—e.g., insulin's synergy with fibronectin (Fn) in the regulation of DNA synthesis—the insulin/Fn cue-response synergy is not explained by an integrated ERK2 signal. The more complex behavior in this case is shown in Figure 2-2. Multidimensional signal analysis is probably required for this scenario. More detail about this research may be found in Asthagiri et al. (2000) or at <http://web.mit.edu/cbe/dallab/research.html>.

The reason we seek the simplest models with the right functionality is, of course, that science needs to understand the biological process (ultimately to influence it in a positive way) in terms that are simple enough to develop a conceptual understanding, even an intuition, about the processes. Thus, there is a

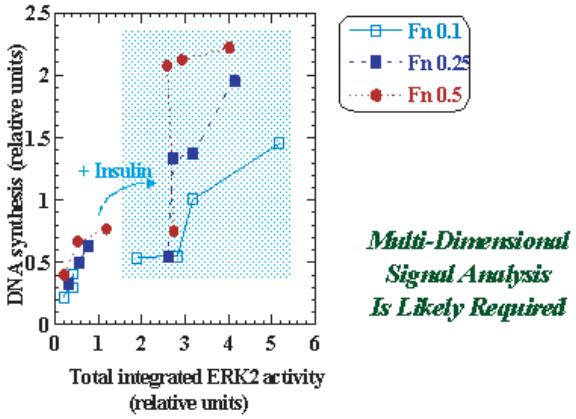

FIGURE 2-2 The insulin/fibronectin (Fn) cue-response synergy is not explained by the integrated ERK2 signal. Multidimensional signal analysis is probably required. Figure courtesy of Douglas Lauffenburger.

balance between simplicity and capturing the essentials of the underlying process. The definition of "essential" will vary according to the investigator's needs.

Other workshop presentations, by John Tyson, of the Virginia Polytechnic Institute and State University, and Garrett Odell, also delved into the modeling of cellular networks. Tyson investigated the cell cycle, the sequence of events in which a growing cell replicates its components. The network (molecular interactions) of the cell cycle is very complex (see, e.g., Kohn, 1999), as shown in Figure 2-3. Using a compartment model approach, Tyson models the cell cycle with a system of differential equations that represent the molecular interactions. His goal is to produce a model that is tailored to the properties of yeast: that is, having parameter values for which the output of the model agrees with representative experimental data for yeast.

The network diagram shown in Figure 2-3 leads to a system of differential equations with more than 50 rate constants. This mathematical model was fit to data and then tested by looking at its predictions in approximately 100 mutant strains of yeast. The agreement was very good.

Figure 2-4 shows the modeling process that Tyson went through. Neither intuition nor direct experimental data could explain some aspects of the yeast cell's physiology, but there was enough understanding to hypothesize a molecular signaling network. That network could be described by a system of differential equations, and the output of that system (seen through tools of dynamical system theory) sheds light on the physiology of the cells. Finally, the proposed physiology was verified experimentally.

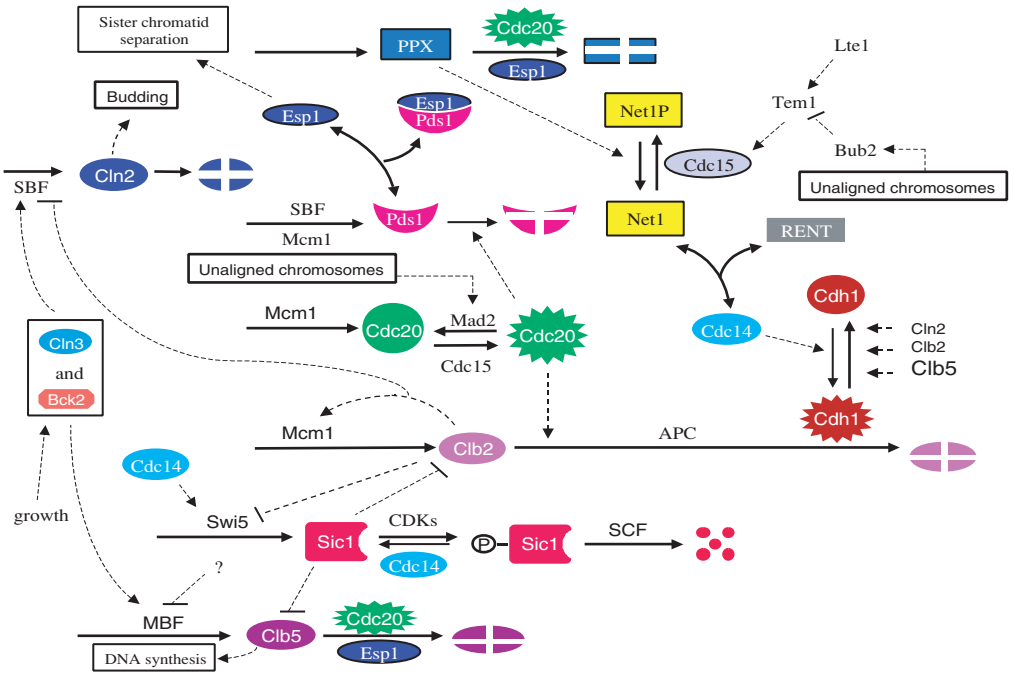

FIGURE 2-3 Network diagram of the yeast cell cycle. Figure courtesy of John Tyson.

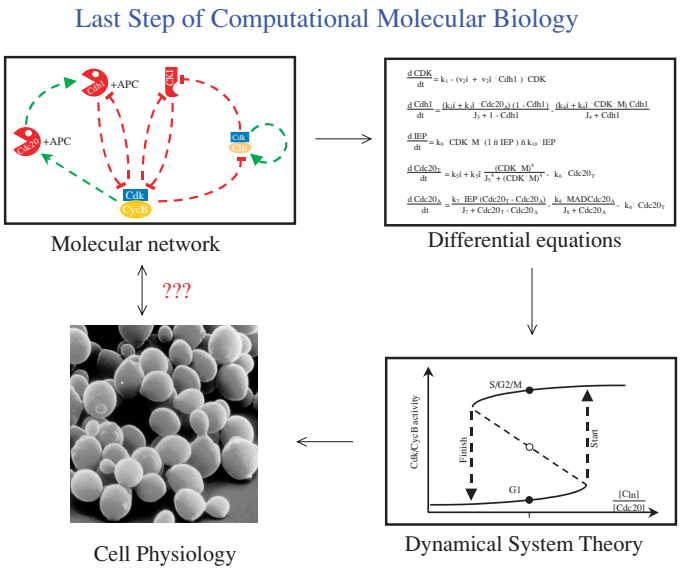

FIGURE 2-4 The modeling process. Figure courtesy of John Tyson.

To construct his very complex model, Tyson did the work in segments. The model was split into simple pieces, and each piece was provisionally fit to data. Then the pieces were joined together and refit as a complete unit. As was the case with the other modeling efforts described in this summary, Tyson's process began with simple models that didn't necessarily emulate every known aspect of cellular physiology or biochemistry, and additional complexity was added only as needed to produce output that captures important features observed experimentally.

Garrett Odell used a similar approach to uncover what cellular mechanism controls the formation of stripes in arthropods (see Nagy, 1998, and von Dassow et al., 2000). To model the cell-signaling network, Odell needed 33 differential equations with 48 free parameters. The model was fit using nonlinear optimization with an objective function that was "crafted" so that, at its minimum, the desired genetic pattern would be observed.

Figure 2-5 shows the connection between the network diagram and the mathematical model, where the model parameters ν_{ENhh} and κ_{ENhh} need to be estimated. A parametric form is specified for the rate of exchange between the components of a network diagram such as that in Figure 2-3, and the resulting model equations, the solutions to the differential equations, are then estimated. The model was fit using nonlinear optimization with an objective function that was "crafted" so that, at its minimum, the desired genetic pattern would be observed.

A quote that resurfaced at times throughout the conference was the following one, attributed to Stanislaw Ulam: "Give me 15 parameters and I can make an elephant; give me 16 and I can make it dance." Odell noted, "I cannot make four lousy stripes with 48 parameters"—his first model did not work, and it was found later that the network it was modeling was not correct. (The evidence in the literature was ambiguous about the exact details of the network.)

In fact, this failure demonstrated that the network, as originally conceived, lacked some necessary connections. The process of representing the network with a differential equations model made the

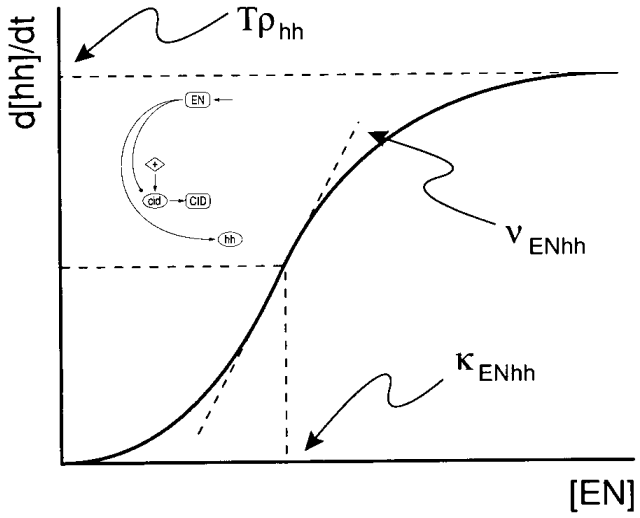

FIGURE 2-5 Parameters control the shape of the typical connection in the network. Figure courtesy of Garrett Odell.

absence of these connections more apparent because the erroneous set of equations did not have the mathematical capacity to create the stripes that are known to occur in nature. After recognizing the missing network links and representing them in the differential equations, the resulting set of equations not only produced the proper pattern, but the choice of parameters also turned out to be extremely robust. That is, the same pattern of stripes occurs over a wide range of parameter values, and it was no longer necessary to use optimization to tune the parameter set. In what was now a 50-dimensional parameter space, choosing the parameters at random (within reasonable bounds) still gave a 1/200 chance of achieving the desired pattern. Further study of the robustness confirmed that the function represented by the differential equations—and, accordingly, the molecular network implied—was extremely stable. Compare this to a radio wiring-diagram, where a change in one connection will render the network inoperable. Here, the robustness of the network is similar to replacing a blown capacitor with whatever is handy and still having an operable radio.

The search for a simple model, indeed for any model, is the search for an underlying structure that will help us to understand the mechanism of the biological process, and—if we are successful—to lead us to new science. The solutions to the yeast differential equations led to understanding a bifurcation phenomenon, and the model also predicts an observed steady-state oscillation. So the mathematical model not only shed new understanding on a previously observed phenomenon, but also opened the door to seeing behavior that had not been explained by biology.

3

Probabilistic Models That Represent Biological Observations

One of the common major goals of the work described in Chapter 2 is the derivation of simple models to help understand complex biological processes. As these models evolve, they not only can help improve understanding but also can suggest aspects that experimental methods alone may not. In part, this is because the mathematical model allows for greater control of the (simulated) environmental conditions. This control allows the researcher to, for example, identify stimulus-response patterns in the mathematical model whose presence, if verified experimentally, can reveal important insights into the intracellular mechanisms.

At the workshop, John Rinzel, of New York University, explained how he had used a system of differential equations and dynamical systems theory to model the neural signaling network that seems to control the onset of sleep. Rinzel's formulation sheds light on the intrinsic mechanisms of nerve cells, such as repetitive firing and bursting oscillations of individual cells, and the models were able to successfully mimic the patterns exhibited experimentally. More detail may be accessed through his Web page, at <http://www.cns.nyu.edu/corefaculty/Rinzel.html>.

In another approach, based on point processes and signal analysis techniques, Don Johnson, of Rice University, formulated a model for the neural processing of information. When a neuron receives an input (an increase in voltage) on one of its dendrites, a spike wave—a brief, isolated pulse having a characteristic waveform—is produced and travels down the axons to the presynaptic terminals (see Figure 3-1). The sensory information in the nervous system is embedded in the timing of the spike waves. These spikes are usually modeled as point processes; however, these point processes have a dependence structure and, because of the presence of a stimulus, are nonstationary. Thus, non-Gaussian signal processing techniques are needed to analyze data recorded from sensory neurons to determine which aspects of the stimulus correlate with the neurons' output and the strength of the correlation.

Johnson developed the necessary signal processing techniques and applied them to the neuron spike train (see details in Johnson et al. (2000) and also at <http://www.ece.rice.edu/~dhj/#auditory>). This theory can be extended to an ensemble of neurons receiving the same input, and under some mild assumptions the information can be measured with increasing precision as the ensemble size increases.

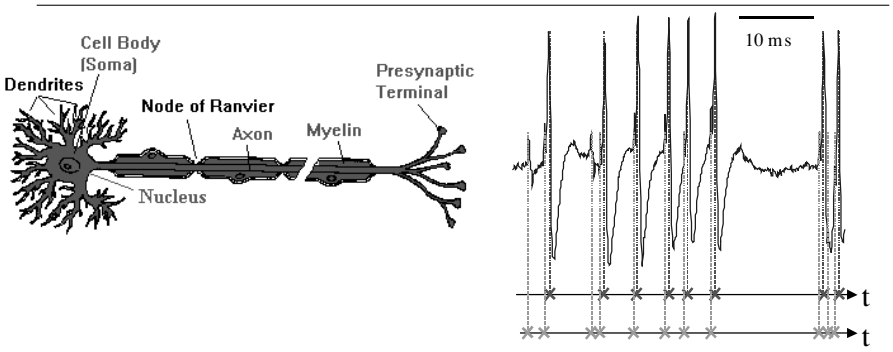

FIGURE 3-1 Neural representation of information. Information is represented by *when* spikes occur, either in single-neuron responses or, more importantly, jointly, in population (ensemble) neural responses. A theoretical framework is needed for analyzing and predicting how well neurons convey information. Figure courtesy of Don Johnson.

Larry Abbott, of Brandeis University, also explored the characteristics of neuron signals. He presented research on the effect of noise as an excitatory input to obtain a neural response, and his methods took advantage of the difference between in vivo measurements and in vitro measurements. His work counters one of the most widespread misconceptions, that conductance alone changes the neural firing rate. Instead, a combination of conductance and noise controls the rate. As Figure 3-2 shows, although a constant current produces a regular spike train in vitro, this does not happen in vivo, where there is variance in the response, and thus more noise in the signal.

It is of great interest to study the input and output relations in a single neuron, which has more than 10,000 excitatory and inhibitory inputs. Let I denote the mean input current, which measures the difference between activation and inhibitory status, and let σ_I^2 be the input variance. For an output with a mean firing rate of r hertz, neuroscientists typically study the output's variance σ_v^2 and coefficient of variation CV. Abbott also studies how the mean firing rate changes as the mean input current varies; this is labeled as the "gain," dr/dI, in Figure 3-3. The standard view is as follows:

- The mean input current I controls the mean firing rate r of the output.
- The variance of the input current affects σ_v^2 and CV.

Abbott disputes the second statement and concludes that the noise channel also carries information about the firing rate r. To examine this dispute, Abbott carried out in vitro and in vivo current injection experiments.

In the first experiment, an RC circuit receiving constant current was studied. Such a circuit can be represented with a set of linear equations that can be solved analytically. The result from this experiment showed that the output variance increases as input variance increases, and that it reaches an asymptote at large σ_I^2. The firing rate r increases as the input I increases, and the CV decreases as r increases.

Abbott's second experiment studied real neurons in an artificial environment: Laboratory-generated signals were used as the input to actual neurons in vivo (see Figure 3-4). Both excitatory and

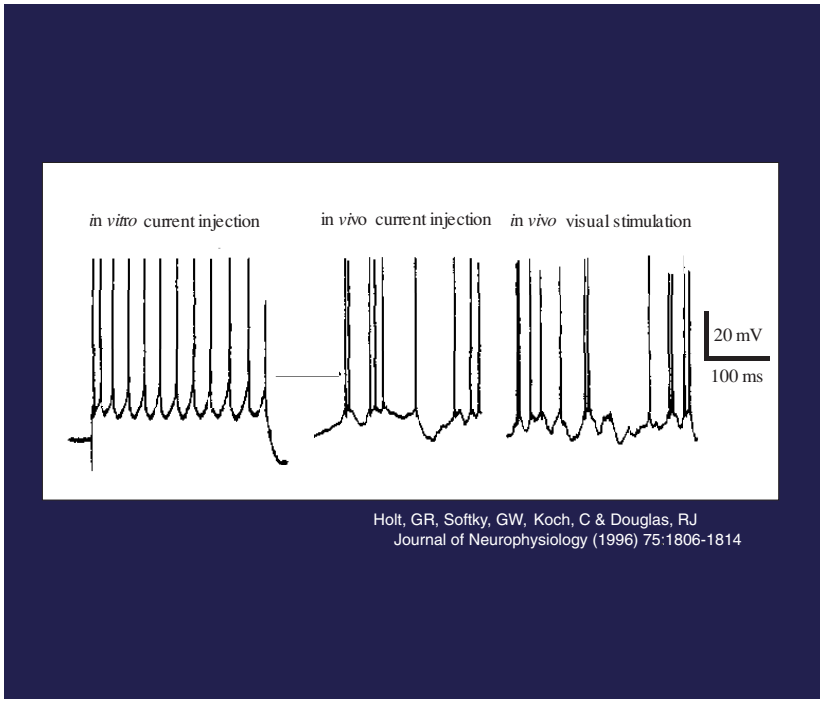

FIGURE 3-2 Neural responses. SOURCE: Holt et al. (1996).

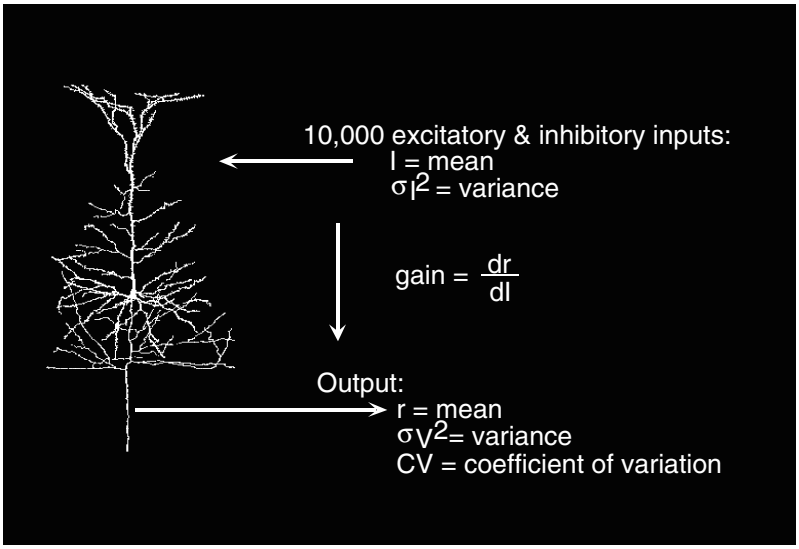

FIGURE 3-3 Neural input and output. Figure courtesy of Larry Abbott.

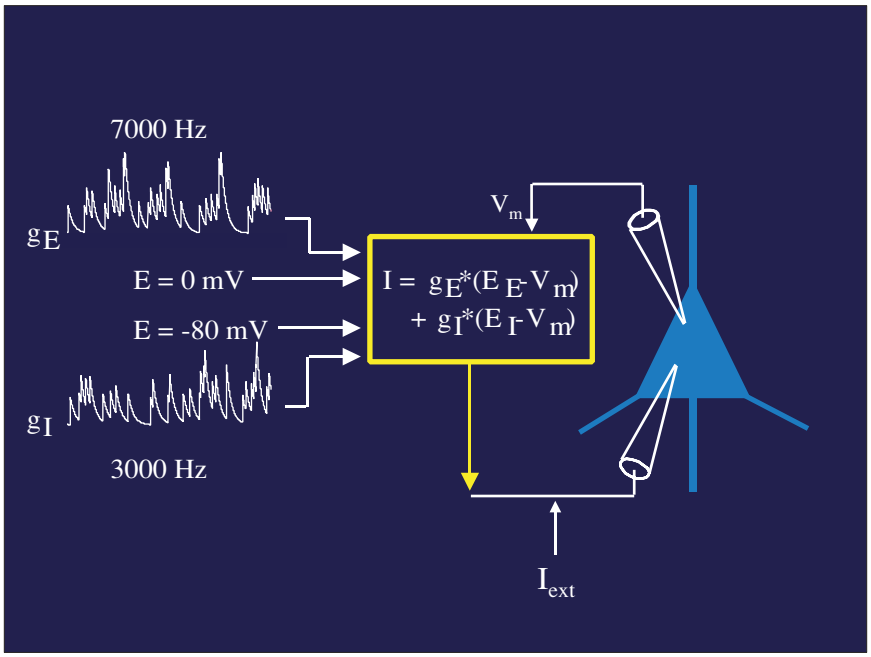

FIGURE 3-4 Neural stimuli. Figure courtesy of Larry Abbott.

inhibitory inputs (g_E and g_I), at different voltages, combine to create the input I that is fed into the neuron (triangle in Figure 3-4). Through this experiment it was shown that the mean of the input affects the rate but that the variance of the input is not correlated with the variance of the output. Instead, the input variance acts more like a volume control for the output, affecting the gain of the response. Dayan and Abbott (2001) contains more detail on this subject.

The workshop's last foray into neuroscience was through the work of Emery Brown, of the Harvard Medical School, whose goal was to answer two questions:

- Do ensembles of neurons in the rat hippocampus maintain a dynamic representation of the animal's location in space?
- How can we characterize the dynamics of the spatial receptive fields of neurons in the rat hippocampus?

The hippocampus is the area in the brain that is responsible for short-term memory, so it is reasonable to assume that it would be active when the rat is in a foraging and exploring mode. For a given location in the rat's brain, Brown postulated that the probability function describing the number of neural spikes would follow an inhomogeneous Poisson process:

$$\text{Prob}(k \text{ spikes}) = e^{-\lambda(t)}\lambda(t)\ k/k!$$

where $\lambda(t)$ is a function of the spike train and location over the time interval $(0, t)$. (Brown later generalized this to an inhomogeneous gamma distribution.) Given this probability density of the

number of spikes at a given location, we next assume that the locations $x(t)$ vary according to a Gaussian spatial intensity function given by

$$f(x(t)) = \exp\{\alpha - {}^1/_2[x(t) - \mu]^T W^{-1}[x(t) - \mu]\}$$

where μ is the center, W is the variance matrix, and $\exp\{\alpha\}$ is a scaling constant.

This model was fit to data, and an experiment was run to see how it performed. In the experiment, a rat that had been trained to forage for chocolate pellets scattered randomly in a small area was allowed to do so while data on spike and location were recorded. The model was then used to predict the location of brain activity and validated against the actual location. The agreement was reasonable, with the Poisson prediction interval covering the actual rate of activation 37 percent of the time and the inhomogeneous gamma distribution covering it 62 percent of the time. Brown concluded that the receptive fields of the hippocampus do indeed maintain a dynamic representation of the mouse's location, even when the mouse is performing well-learned tasks in a familiar environment, and that the model, using recursive state-space estimation and filtering, can be used to analyze the dynamic properties of this neural system. More information about Brown's work may be found at <http://neurostat.mgh.harvard.edu/brown/emeryhomepage.htm>.

4

Modeling with Compartments

Turning to other modeling domains, Lauffenburger proposed to the workshop participants a simple taxonomy of modeling according to what discipline and what goal are uppermost in the researcher's mind:

- *Computer simulation*. Used primarily to mimic behavior so as to allow the manipulation of a system that is suggestive of real biomedical processes;
- *Mathematical metaphor*. Used to suggest conceptual principles by approximating biomedical processes with mathematical entities that are amenable to analysis, computation, and extrapolation; and
- *Engineering design*. Used to emulate reality to a degree that provides real understanding that might guide bioengineering design.

Byron Goldstein, of Los Alamos National Laboratory, presented work that he thought fell under the first and third of these classifications. He described mathematical models used for studying immunoreceptor signaling that is initiated by different receptors in general organisms. He argued that general models could be effectively used to address detailed features in specific organisms.

Many important receptors—including growth factor, cytokine (which promotes cell division), immune response, and killer cell inhibitory receptors—initiate signaling through a series of four biological steps, each having a unique biological function. Building on work of McKeithan (1995) that proposed a generic model of cell signaling, Goldstein developed a mathematical model for T-cell receptor (TCR) internalization in the immunological synapse. Goldstein's model takes different contact areas into account and was used to predict TCR internalization at 1 hour for the experiments in Grakoui et al. (1999).

To date, the major effort in cell signaling has been to identify the molecules (e.g., ligands, receptors, enzymes, and adapter proteins) that participate in various signaling pathways and, for each molecule in the pathway, determine which other molecules it interacts with. With an ever-increasing number of participating molecules being identified and new regulation mechanisms being discovered, it has become clear that a major problem will be how to incorporate this information into a useful predictive model.

To have any hope of success, such a model must constantly be tested against experiments. What makes this possible is the ability of molecular biologists to create experimental systems containing only small numbers of signaling molecules. Thus, separate parts of the model can be tested directly.

Where are we at the moment in our attempt to build a detailed model of cell signaling? Goldstein has used deterministic and stochastic approaches to create the following detailed models of cell signaling:

- An algorithm has been created to generate the chemical rate equations that describe the dynamics of the average concentrations of chemical species involved in a generic signaling cascade.
- A stochastic model for the time dependence of the state concentrations has been developed, and it has been shown that the stochastic and deterministic formulations agree in the cases studied to date.
- A model has been created for the signaling cascade that is mediated by the immunoreceptor that plays a central role in allergic reactions. This model includes a bivalent ligand, a monovalent receptor, and the first two enzymes in the cascades, Lyn and Syk.

Additional information on Goldstein's modeling may be found at <http://www.t10.lanl.gov/profiles/Goldstein.html>.

Moving from intracellular processes, Bruce Levin, of Emory University, presented some research that uses mathematical models to understand trends in antibiotic resistance, a serious public health concern worldwide. Levin is addressing the need to know what trends are and are not of serious importance. As an example, he noted that resistance to vancomycin (an antibiotic) increased from approximately 1 percent in 1989 to 16 percent in 1997. It does not necessarily follow, however, that this is a serious problem. Lipsitch et al. (2000) state as follows:

> Although it generally is assumed that use of a particular antibiotic will be positively related to the level of resistance to that drug . . . it is difficult to judge whether an intervention has been successful . . . . Mathematical models can provide such quantitative predictions, which naturally give rise to criteria for evaluating the interventions.

Population dynamics can be examined with a compartment model, as shown in Figure 4-1. The compartments represent the disease state of the individual (S, susceptible; IS, immune/susceptible; IR, immune/resistant). The proportion *p* stands for those under treatment, and the parameters represent the rate of movement from one compartment to another. Based on such a model, one can calculate parameters such as basic reproductive numbers and then establish rates and conditions under which the percent of resistance will increase when a proposed treatment is applied. What is often observed in public health is that the rate of resistance changes as the efficacy of the treatment changes, with high efficacy corresponding to high resistance, and the rate of resistance increases more rapidly than it decreases.

To further investigate how a host controls infection, Levin examined *E. coli* infection in mice, where the following threshold effect has been observed experimentally: While high doses of *E. coli* kill mice, lower doses can be brought under control. A differential equations model was developed that includes this threshold effect, and it was found to fit the data quite well. Levin's results again illustrate one of the common themes of the workshop, that a mathematical model—built on a functional premise, even if simple, and verified with data—allows us to quantify biophysical processes in a way that can lead to valuable insight about the underlying structure of the processes.

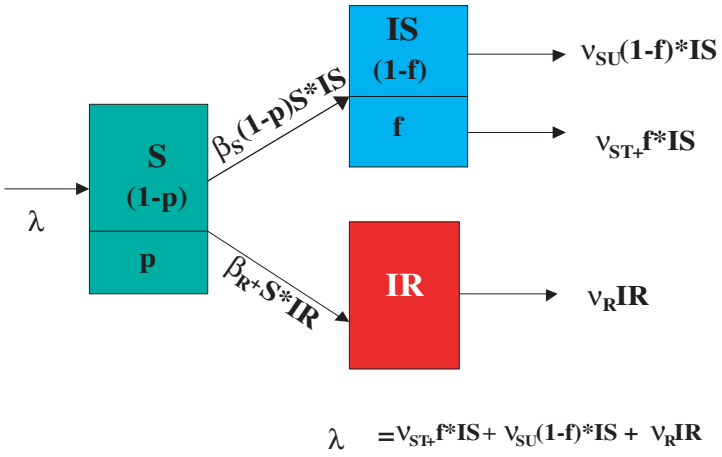

FIGURE 4-1 Compartment model for the epidemiology of the antibiotic therapy of prophylaxis. Figure courtesy of Bruce Levin.

5

From the Compartment to the Fluid

The mathematical modeling presented in Chapters 3 and 4 was based mainly on compartment models, models that lead to differential equations with unknown rate parameters. Another approach, taken by two other workshop presenters, is to investigate the fluid dynamics of the underlying biological system and use those principles to enhance our biological understanding.

George Oster, of the University of California at Berkeley, started his talk by showing a computer-generated movie representative of myxobacteria movement and then presenting a mathematical model that describes the collective behavior during the "ripple phase" of development. The ripple phase (see Figure 5-1) is characterized by a unique pattern of waves that appear to pass through one another; moreover, they can occur without any net movement of bacteria (see Igoshin et al., 2001).

Knowing that myxobacteria move with a combination of two motility mechanisms, labeled A and S, which are controlled by different physiology, Oster sought to combine mathematical models of these mechanisms with a model of the bacterial communication system into a dynamic model that could produce the ripple phase. The resulting model succeeded in emulating and explaining important characteristics of the motion of myxobacteria, such as how the bacteria in crests move *with* the wave while the bacteria in troughs move *against* the wave. When waves collide, some bacteria continue moving forward and others reverse, in accordance with experimental observations. The model also captures the ability of ripple waves to move through one another and to propagate without any net transfer of mass.

In another application of fluid dynamics, Charles Peskin, of New York University, described a mathematical model for the heart that considers the muscle tissue to be a time-dependent elastic material, which can be modeled using fluid dynamics. The geometry of his model builds on work in the 1950s by Carolyn Thomas, which described the fiber architecture of the heart as a system of spiraling muscle fibers (see Figure 5-2).

By considering the heart as a composite material of fiber and fluid, Peskin developed equations of a viscous, incompressible fluid to describe the force applied by the fibers to the fluid. Figure 5-3 illustrates the relationship between pressure and volume and how it changes as the valves open and close, while Figure 5-4 shows the force vectors, illustrating the forces on the muscle fiber of the heart.

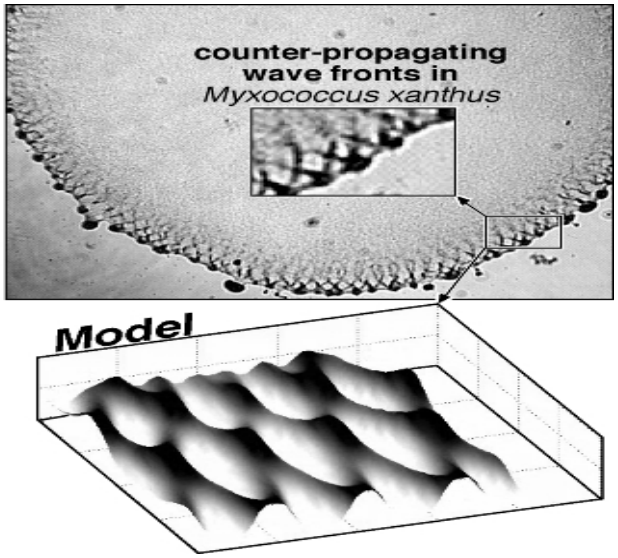

FIGURE 5-1 Micrograph and model output showing patterns in the ripple phase. Figure courtesy of George Oster.

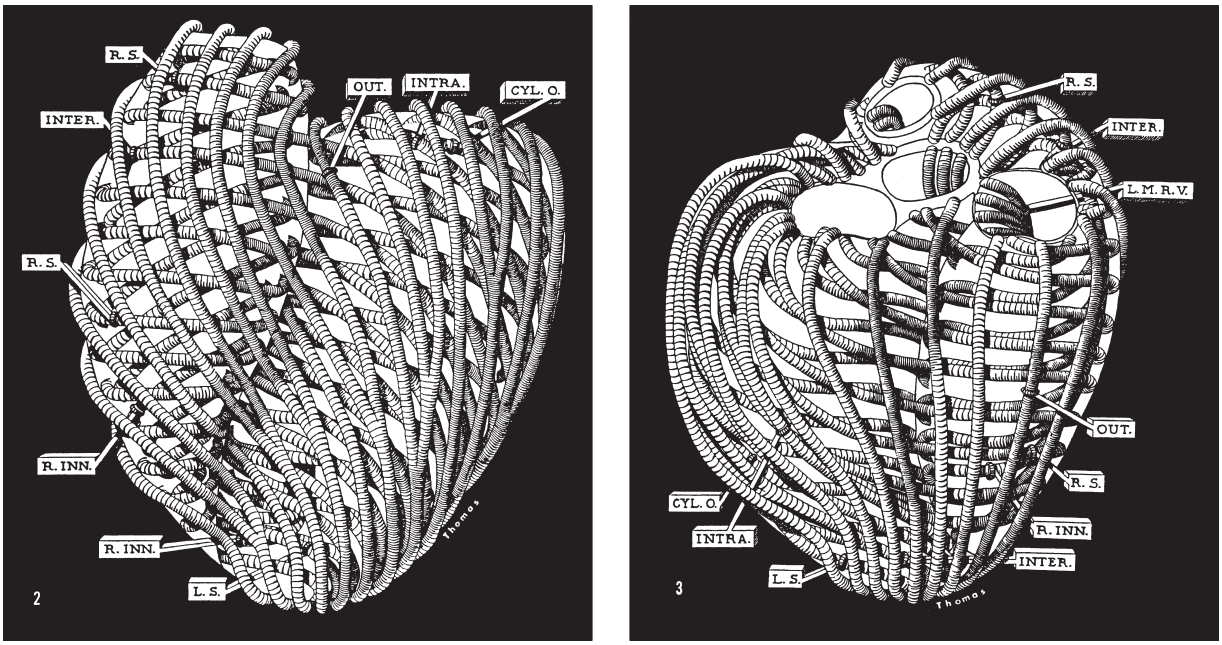

FIGURE 5-2 Muscular architecture of hog ventricle. Left, anterior view; right, posterior view. Reprinted with permission from Carolyn Eyster Thomas, Muscular architecture of the ventricles of hog and dog hearts, American Journal of Anatomy, Vol. 101 (1957), 17-57.

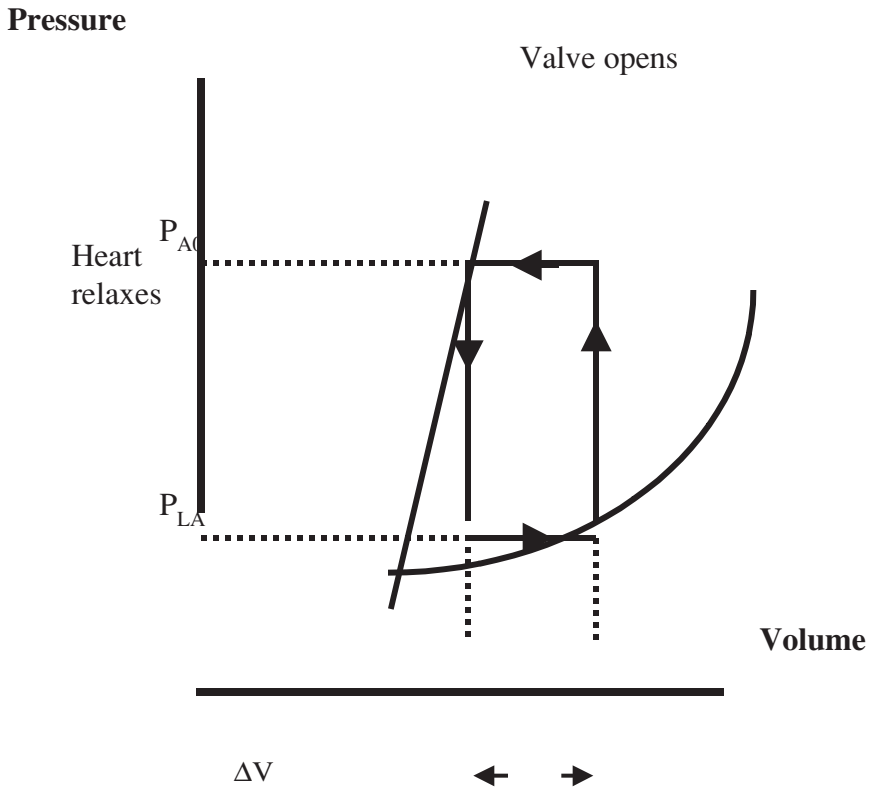

FIGURE 5-3 Pressure/volume relationship in the heart. Figure courtesy of Charles Peskin.

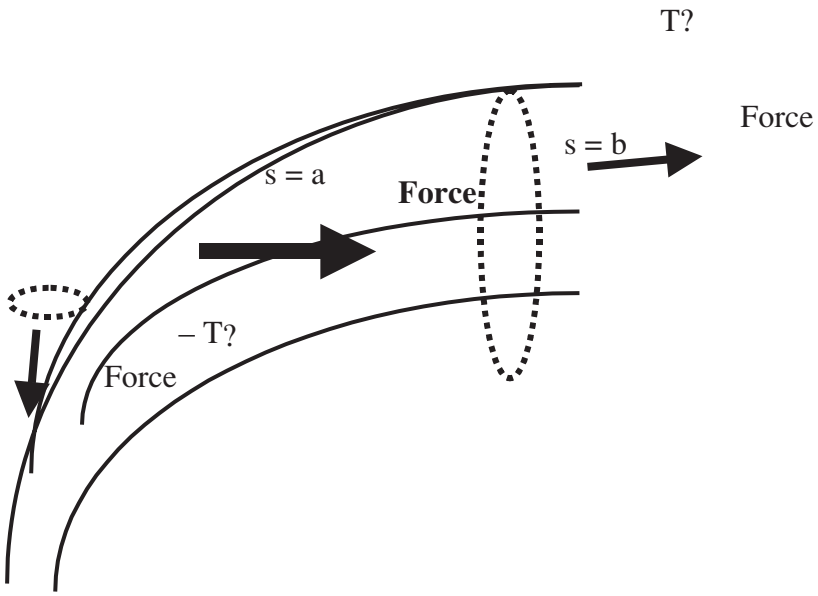

FIGURE 5-4 Illustration of forces on heart muscle fibers. Figure courtesy of Charles Peskin.

Using the following definitions,

q,r,s = material parameters (s varies along the fiber)
t = time
$x = x(q,r,s,t)$, position
$\tau = \tau(q,r,s,t)$, force
$T = T(q,r,s,t)$, stress

Peskin derived the following differential equations:

- The *fluid equation*, in Cartesian coordinates, which describes the force of the fibers on the fluid:

$$\underline{F}(\underline{x},t) = \int \frac{\partial}{\partial s}(T\underline{\tau})\delta\left[\underline{x} - \underline{X}(q,r,s,t)\right]dq\,dr\,ds$$

- The *fiber equations*, which describe the stress on the fibers:

$$T = \sigma\left(\left|\frac{\partial \underline{X}}{\partial \underline{s}}\right| - 1; q,r,s,t\right)$$

$$\underline{\tau} = \frac{\partial \underline{X} / \partial s}{|\partial \underline{X} / \partial s|}$$

- The *interaction equation*, which ties them together:

$$\frac{\partial \underline{X}}{\partial t}(q,r,s,t) = \underline{u}\left[\underline{X}(q,r,s,t),t\right] = \int \underline{u}(\underline{x},t)\delta\left[\underline{x} - \underline{X}(q,r,s,t)\right]dx$$

Note that the equations are expressed using the Dirac delta function, an alternative to computing a Jacobean. Use of the delta functions results in an algorithm that is numerically more stable than one that relies on a Jacobean.

Peskin also described a numerical method for solving the equations, using a second-order immersed boundary method derived by M.-C. Lai, which is based on an extension of the Runge-Kutta method. This method can also be used in traditional fluid mechanics problems.

The model output (example frames shown in Plates 1 and 2) compares admirably with MRI scans. Perhaps most impressive, the model captures the swirling movement of blood within the ventricles, a phenomenon of physiological importance that had not emerged in simpler models. This swirling explains why less force is required to exit the chamber than would otherwise be predicted, and it also eliminates leakage in valves from back pressure. However, some improvements are still needed in the model, including refining the representation of the up-down valve motion, getting a larger valve opening, and decreasing the movement of the base of the heart.

It is interesting to note that the Peskin work reflects two of the main themes common to other talks at the workshop. First, the evolution of his heart model over the years has been guided by a clear

understanding of what physiological features must be captured and what mathematical methods might be suitable to do so, rather than by, say, simply increasing model resolution. For instance, his previous model, although doing an adequate job of reproducing some heart actions, displayed an irregular pattern of blood flow and leakage in the valves that was significantly different from the pattern in a functioning heart. His current version has become more complex in a way that can capture this important phenomenology. Second, the Peskin model illustrates how a mathematical representation can suggest insights (subject to experimental validation) that would not be apparent from current experimental data. The Peskin heart model is an approximate surrogate for a beating heart, a surrogate that can be manipulated and inspected in ways that a living heart cannot.

6

Gene Transfer As a Biomedical Tool

Perhaps the most exciting developments in molecular biology have to do with the explosion of information and technology for dealing with biological processes at the level of the genome. A number of workshop presenters have made important progress in this area. At times during the presentations, some possibilities were suggested that would once have been in the domain of science fiction. Before delving into the data-intensive challenges that are of most relevance to mathematical scientists, we discuss two workshop presentations that set the stage.

Eduardo Marbán, of Johns Hopkins University, presented some basic background on the challenges of molecular biology, explaining, for instance, how proteins control all that we do and how DNA controls the proteins. The DNA in the human genome, if unraveled, would be about 6 feet long, and if encoded into phone books would be as high as the Washington Monument. Of this, approximately 99 percent does not code for proteins (although some of it may have other functions), while the remaining 1 percent, which we call genes, contains the blueprints for protein formation.

Genes are sometimes referred to as "wild type" (the predominant forms found in nature) or "mutant" (aberrant forms found in nature or genetically engineered). Genetic engineering or genetic therapy delivers a desired piece of genetic material by first being packaged in a modified virus (viruses being able to insert themselves into cell nuclei; the modification is whatever is necessary to make the virus benign) and injecting them into a subject. These modified viruses (common cold viruses, reengineered AIDS viruses) are specialized for gene transfer and can be very efficient. Genes can then be introduced isogenically (within the same species) or allogenically (from one species to another).

The potential use of gene transfer is unlimited. Some examples are the following:

- To test hypotheses in biological systems;
- To treat single-gene deficiency disorders such as ADA deficiency (which results in people without viable immune systems) or hemophilia;
- To treat a complex disease system, such as that promoting angiogenesis as a treatment for coronary artery disease; and

- To reengineer cells and tissue at a functional level.

Marbán described a case study in which gene therapy was used to treat cardiac arrhythmia, in particular the problem of atrial fibrillation. (There are many kinds of arrhythmias, which account for more than 2 million deaths per year in the United States and result in the implanting of approximately 255,000 pacemakers per year at about $45,000 per operation.) Current therapies for arrhythmia include drug therapy (which has many side effects) or the implant of pacemakers or defibrillators (which require a lifetime commitment to repeated procedures). Other therapies also carry complications, so it is reasonable to look to gene therapy. Gene therapy avoids the use of implantable hardware, has the potential to treat targeted tissues, and can be reversed. There are still problems to overcome, however. For example, there can be difficulty in delivering the therapy to the appropriate tissue or cell types.

The choice for an initial gene therapy trial aimed at arrhythmia is the genetic modification of the atrial-ventricular (AV) node, a site on the heart that exerts significant control over the heart's beating. The goal of the treatment is to control and slow the conduction through the AV node, thus decreasing the chance of fibrillation. The genetic approach to treatment is to insert an agent that blocks certain stimulations of the AV node. To test the effectiveness of the gene therapy, an experiment was carried out in swine. The results were successful: The gene therapy limited the pig's heart rate during atrial fibrillation to 155 beats per minute, whereas the rate during atrial fibrillation for a control subject went up to 190. The role that mathematical scientists will play is in the bioinformatics that enables such gene therapies.

Michael Phelps, of the University of California at Los Angeles, spoke about using a minuscule positron emission tomography (PET) scanner, called a micro-PET scanner, to monitor the progress of a treatment that involves gene therapy. This is a very promising technique to advance the art of gene therapy, and mathematical scientists are needed to address the reconstruction of images from the micro-PET data. The micro-PET scanner provides a tissue concentration measurement, which can be transformed into a metabolic rate by a mathematical model.

Phelps's lab has built a micro-PET scanner that is used in mice, rats, and monkeys (mice being the typical mammalian model). Their goal is in vivo gene expression measurement,[1] and to accomplish that they constructed an RNA probe that is only 13 base pairs in length to carry out in vivo hybridization. The probe carries a PET reporter gene (a gene that the PET scanner can track) and a therapeutic gene (a gene that effects some treatment). The PET reporter gene produces a signal, detectable by a PET scanner, if gene expression occurs—that is, if the therapeutic gene expresses and delivers the therapy. To test the micro-PET technique, Phelps constructed sample probes and injected them into the livers of mice. The level of expression of the therapeutic gene, as detected through the micro-PET technique, was in excellent agreement with the actual level of associated RNA measured during subsequent autopsies.

Figure 6-1 shows the PET tracking of a therapeutic gene that expresses in the liver. Gene expression first increases, as is seen after 2 days, then decreases, and after 2 weeks the immune system terminates the virus. The promoter can be turned on and off with an external switch (tetracycline), as might be needed in some studies. Among other advances, this procedure should allow cancer cells to be tracked over time as they migrate.

[1] A gene is expressed when a functional product is made from that gene. The gene is used as a template to make messenger RNA (mRNA), and the mRNA in turn serves as a template for the synthesis of a protein. The protein is what carries out a particular task in the cell.

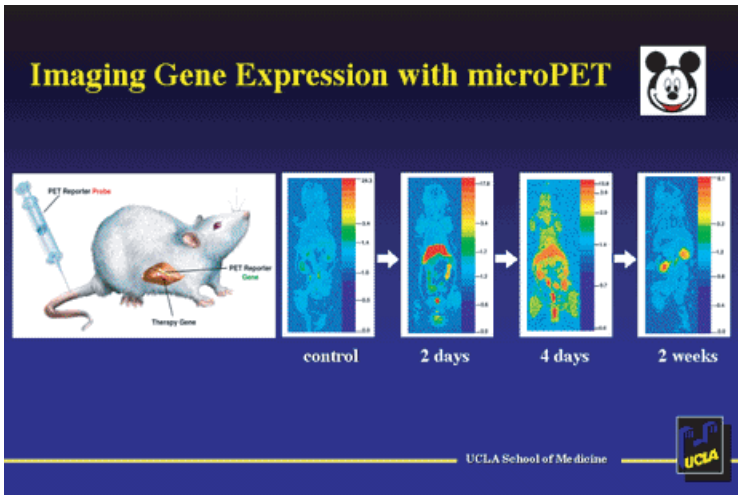

FIGURE 6-1 Imaging gene expression with micro-PET. Figure courtesy of Michael Phelps.

Phelps stressed the importance of going back and forth between humans and mice, because the science is enabled by both. The advantage of the current drug therapy/delivery system is that the drug can be given in truly tiny amounts, and the promoters can be turned on externally to increase the dose.

Several developments are needed on the mathematics/statistics front. At present, Fourier-based reconstruction is used to create images from the micro-PET data, but better methods are needed.

7

The Data Flood: Analysis of Massive and Complex Genomic Data Sets

One of the major themes brought out by the workshop was the interplay between theory and data, but the discussions in preceding chapters do not mention how much data must be dealt with. In fact, the data sets themselves are so massive that their analysis presents major challenges to statistical methodology.

As an example, Dan Roden, of Vanderbilt University, reported on research the original goal of which was to use genetics to predict individual responses to drugs. However, the research quickly evolved into the challenge of navigating through a massive data set. Pharmacologists are very interested in understanding why individuals have different responses to the same drugs, and how to predict those variations. The variability in drug response can correlate with a variety of factors, such as gender, age, disease type, concomitant drug therapies, and ethnicity.

Variability in drug response among different individuals may also be due to genetic factors. Each person has two strands of DNA in his or her genome, shown as two panels in Figure 7-1. At particular genome locations, the DNA sequences might differ between any two people. Such a difference, called a DNA polymorphism, might be associated with the occurrence of side effects in a given individual.

Mutation is one of the factors causing DNA polymorphisms, and which therefore contributes to disease onset. DNA polymorphisms may be due to the deletion, insertion, or substitution of a nucleotide, may occur at coding or noncoding regions of the DNA, and may or may not alter gene function. The occurrence of DNA polymorphism makes it possible to associate a person's response to drugs with particular DNA regions, for example, by correlating the occurrence of the polymorphism with the response. This is the basis of current phamacogenetics, which is the study of the impact of individual genetic variants on drug response.

Roden's research sought to evaluate the role of genetics in determining drug response in the case of a single nucleotide polymorphism (SNP) that is known to predispose individuals to drug-induced arrhythmias. He approached the problem with the following strategy:

- Define the drug response (phenotype) of interest.
- Test appropriate DNA samples, patients, or families.
- Identify candidate genes that might explain significant response variations.

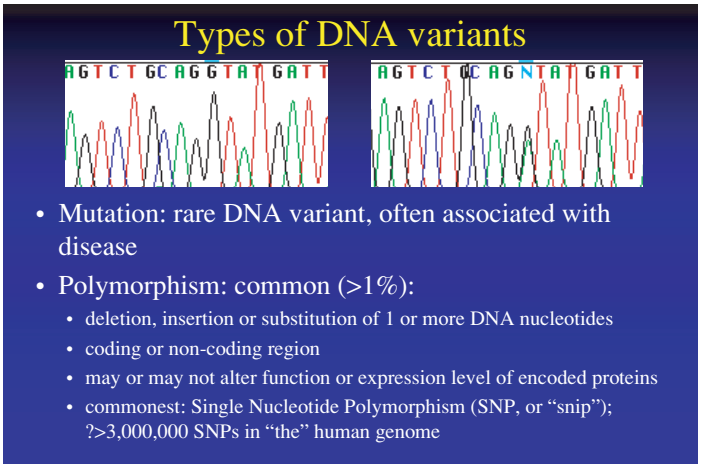

FIGURE 7-1 Types of DNA variants: mutation and polymorphisms. Figure courtesy of Dan Roden.

- Identify polymorphisms in candidate genes.
- Relate the identified polymorphism to the phenotype.

Such an analysis would produce a graph like that in Figure 7-2, where the χ^2 statistic would be calculated at each SNP. However, such an analysis would be infeasible for both statistical and economic reasons, because of the flood of data. Suppose the research has considered 100,000 SNPs in 1,000 patients (500 affected, 500 not affected). The statistical problem is that the data will result in 100,000 χ^2 statistics. With such a multiplicity of tests, there will be many false positives. How then does one set a sensible cutoff point for statistical significance?

Even if the statistical problem can be solved, basic economics makes this straightforward experiment infeasible because of the tremendous cost of recording 100,000 genotypes in each of a thousand people. (If the cost of determining a genotype were only 50 cents, the entire experiment would still cost $50 million.) Accordingly, there is a pressing need to solve the problem of handling the flood of bioinformatics data.

The data flood pointed out by Roden is only one example of the data handling challenges to be overcome. With the development of microarray experiments, the amount of data available today is enormous. At the April 2001 workshop, Terry Speed, of the University of California at Berkeley, gave an overview of microarray experiments, which provide a means of measuring expression levels of many genes in parallel.

In the so-called Stanford protocol, shown on the right side of Plate 3, genetic material from cells is apportioned into two samples, each of which is exposed to a different treatment. (One of the treatments might be the null treatment, in which case we are comparing a treated sample with a control.) The goal is to determine how the two samples differ in the way their genes are expressed—that is, how the genes cause proteins to be created in accordance with their embedded genetic information. One sample is labeled with a red dye and the other with a green dye. The two samples are distributed over a microarray slide (a "gene chip"), which typically has 5,000 to 6,000 different segments of complementary DNA (cDNA) arrayed on it. The two samples of red- and green-dye-tagged genetic material adhere to the

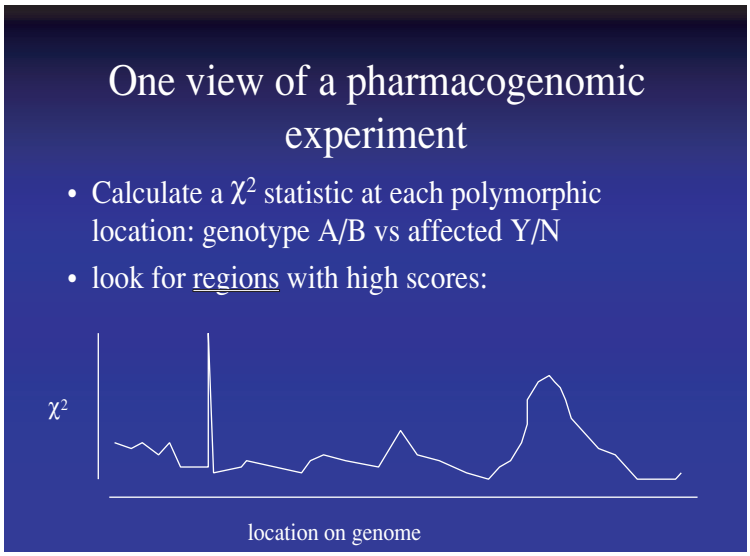

FIGURE 7-2 Data from a hypothetical pharmacogenomic experiment. Figure courtesy of Dan Roden.

slide in different patterns according to their chemical bonding to the cDNA. When the dyed genetic material is allowed to express proteins, the level of activity at each coordinate of the gene chip can be measured through the fluorescence of the dyes. From these measurements, one can develop an understanding of how the genetic material was affected by the treatment to which it was exposed. More complete background on this process, and a number of valuable other links, may be found at <http://www.stat.Berkeley.edu/users/terry/zarray/Html/index.html>.

Many statistical issues arise in the analysis of microarray data, including issues of experimental design, data preprocessing, and arriving at ultimate conclusions. For example, the typical range of expression (on a $\log_2$ scale) is about ± 5, and the amount of background noise in the data could be substantial. Thus, at present, it is usually possible to identify (with some certainty) only those genes that express at a very high or very low level.

Although there are problems with expression levels, and also with bias, a plot of M versus A, where

$$M = \log_2 (\text{red expression}) - \log_2 (\text{green expression})$$
$$A = \log_2 (\text{red expression}) + \log_2 (\text{green expression}),$$

can be extremely useful, as in the following experiment described by Speed, which identified genes with altered expression between two physiological zones (zone 1 and zone 4) of the olfactory epithelium in mice. Plate 4 shows the log ratios plotted against the average of the logs (which gives a measure of absolute expression). It illustrates the noise level in much of the data. It also shows that a number of genes have very high expression levels, and that these genes show differential expression.

Summarizing, Speed outlined some challenges to current research:

- How to address the observed bias associated with whether a sample is treated with red or green dye (which suggests the need to run the complementary experiment of interchanging the red and green labels);
- How to create better designs for microarray experiments, ones that go beyond merely comparing treatment with control;
- How to carry out the experiments' preprocessing so as to reduce the noise in the data; and
- How to deal with the fact that, because a large number of genes are tested in microarray experiments, the large number of statistical tests carried out in parallel greatly increases the chance of finding a false positive. (One attempt to address this is exemplified in Tusher et al. (2001), which uses the false discovery rate method—an approach to the multiple comparisons problem that controls for the expected proportion of false positives rather than attempting to minimize the absolute chance of false positives—to set cutoff points for these errors.)

8

Summary

Throughout the workshop there was much lively discussion among the participants. It was apparent from many of the talks and mentioned explicitly during the discussions that there has been a big cultural change in mathematics and statistics in recent years. In the past, theory and models would often be developed before data were collected; that is a viable approach when dealing with fields of study that are grounded in fundamental mathematical laws (e.g., Maxwell's equations or the Navier-Stokes equation). Now, in many areas that are attracting the attention of mathematical scientists, data drive the development of theory. This is certainly true for mathematical sciences research related to the biomedical sciences, and the resulting intellectual stimulation will likely have far-reaching effects on mathematical sciences research.

The workshop's discussions identified three general ways in which the mathematical sciences have benefited biomedical research:

- By suggesting insights that could not be observed directly (such as "viewing" the interior of the beating heart via a simulation);
- By classifying and describing generic features and processes of biomedical systems; and
- By suggesting how some biomedical systems work and what their limitations are (through tools such as dynamical analysis of mathematical models that emulate cell signaling networks).

The workshop also made clear that there is a great opportunity for many more mathematical scientists to become involved in cross-disciplinary research with biomedical scientists. A major challenge to be overcome before that interface reaches its potential is for more mathematical scientists to be exposed in depth to research in the biomedical sciences and given the opportunity to contribute. As research funding becomes increasingly available, the limiting factor becomes the availability of information about the mathematical formulations of important biomedical research. It is hoped that this workshop summary helps address that need.

References

Asthagiri, A.R., C.A. Reinhart, A.F. Horwitz, and D.A. Lauffenburger. 2000. The role of transient ERK2 signals in fibronectin- and insulin-mediated DNA synthesis, *Journal of Cell Science* 113:4499-4510.

Dayan, P., and L. Abbott. 2001. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* Cambridge, Mass.: MIT Press.

Donahue, J.K., A.W. Heldman, H. Fraser, A.D. McDonald, J.M. Miller, J.J. Rade, T. Eschenhagen, and E. Marbán. 2000. Focal modification of electrical conduction in the heart by viral gene transfer, *Nature Medicine* 6:1395-1398.

Dowling, H.F. 1977. *Fighting Infection.* Cambridge, Mass.: Harvard University Press.

Duggan, D.J., M. Bittner, Y. Chen, P. Meltzer, and J.M. Trent. 1999. Expression profiling using cDNA microarrays, *Nature Genetics* 21:10-14.

Garfinkel, A., O. Voroshilovsky, Y.-H. Kim, M. Yashima, T-J. Wu, R. Doshi, H.S. Karagueuzian, J.N. Weiss, and P.-S. Chen. 2000. Prevention of ventricular fibrillation by flattening cardiac restitution, *Proceedings of the National Academy of Sciences* 97:6061-6066.

Grakoui, A., S.K. Bromley, C. Sumen, M.M. Davis, A.S. Shaw, P.M. Allen, and M.L. Dustin. 1999. The immunological synapse: a molecular machine controlling T-cell activation, *Science* 285:221-227.

Hanahan, D., and R.A.Weinberg. 2000. The hallmarks of cancer, *Cell* 100:57.

Holt, G.R., G.W. Softky, C. Koch, and R.J. Douglas. 1996. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons, *Journal of Neurophysiology* 75:1806-1814.

Igoshin, I., A. Mogilner, R. Welch, D. Kaiser, and G. Oster. 2001. Pattern formation and traveling waves in myxobacteria: theory and modeling, *Proceedings of the National Academy of Sciences* 98(26).

Johnson, D.H., C.M. Gruner, and R.M. Glantz. 2000. Quantifying information transfer in spike generation, *Neurocomputing* 33:1047-1054.

Kohn, K. 1999. Molecular interaction map of the mammalian cell cycle control and DNA repair systems, *Molecular Biology of the Cell* 10:2703-2734.

Kovacs, S.J., D.M. McQueen, and C.S. Peskin. 2001. Modeling cardiac fluid dynamics and diastolic function, *Philosophical Transactions Royal Society A* 359:1299-1314.

Lipsitch, M., C.T. Bergstrom, and B.R. Levin. 2000. The epidemiology of antibiotic resistance in hospitals: paradoxes and prescriptions, *Proceedings of the National Academy of Sciences* 97:1938-1943.

McKeithan, T.W. 1995. Kinetic proofreading in T-cell receptor signal transduction, *Proceedings of the National Academy of Sciences* 92:5042-5046.

Nagy, L. 1998. Changing patterns of gene regulation in the evolution of arthropod morphology, *American Zoologist* 38:818-828.

Sager, B., and D. Kaiser. 1993. Two cell-density domains within the Myxococcus xanthus fruiting body, *Proceedings of the National Academy of Sciences* 90:3690-3694.

Tusher, V.G., R. Tibshirani, and G. Chu. 2001. Significance analysis of microarrays applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences* 98:5116-5121.

von Dassow, G., E. Meir, E.M. Munro, and G. Odell. 2000. The segment polarity network is a robust developmental module, *Nature* 406:188-192.

Web Sites Referenced in the Text

Emery Brown: http://neurostat.mgh.harvard.edu/brown/emeryhomepage.htm
Byron Goldstein: http://www.t10.lanl.gov/profiles/goldstein.html
Don Johnson: http://www.ece.rice.edu/~dhj/#auditory
Douglas Lauffenburger: http://web.mit.edu/cbe/dallab/Research.html
John Rinzel: http://www.cns.nyu.edu/corefaculty/Rinzel.html
Terry Speed: http://www.stat.Berkeley.EDU/users/terry/zarray/Html/index.html

Appendix

Workshop Program and Attendees

PROGRAM

Thursday, April 26

Noon	Lunch available at workshop site

Overview Session—Part 1

1:00 p.m.	Eduardo Marbán (John Hopkins University), "Gene Transfer/Gene Therapy"
1:30 p.m.	Terry Speed (University of California at Berkeley), "Statistics and Microarray Data"
2:00 p.m.	Open discussion—De Witt Sumners (Florida State University), moderator
3:00 p.m.	Break

Mathematical Sciences and Disease States—Part 1

3:30 p.m.	Charles Peskin (New York University), "A Virtual Heart for Pathophysiology and Prosthesis Design"
4:00 p.m.	James Weiss (University of California at Los Angeles), "Biological Pattern Formation: From Arrhythmias to Embryos"
4:30 p.m.	Open discussion—De Witt Sumners (Florida State University), moderator; James Keener (University of Utah), discussant
5:30 p.m.	Reception

Friday, April 27

8:00 a.m.	Continental breakfast

Overview Session—Part 2

8:30 a.m.	Michael Phelps (University of California at Los Angeles), "Genetic Engineering, Molecular Imaging, and Molecular Drug Design"
9:00 a.m.	Douglas Lauffenburger (Massachusetts Institute of Technology), "Cell Engineering: Quantitative Modeling and Experimental Studies of How Cell Functions Depend on Molecular Properties"
9:30 a.m.	Open discussion—Jim Weiss (University of California at Los Angeles), moderator
10:30 a.m.	Break

Mathematical Sciences and Disease States—Part 2

11:00 a.m.	Dan Roden (Vanderbilt University), "Using Genetics to Predict Individual Responses to Drugs—Hope or Hype?"
11:30 a.m.	Bruce Levin (Emory University), "Mathematical Models of the Population Dynamics of Antibiotic Therapy"
Noon	Open discussion—Jim Weiss (University of California at Los Angeles), moderator; James Keener (University of Utah), discussant
1:00 p.m.	Lunch

Dynamical Models of Cellular Processes

2:00 p.m.	John Tyson (Virginia Polytechnic Institute and State University), "CyberYeast: A Computational Model of Cell Cycle Regulation in Budding Yeast"
2:30 p.m.	Byron Goldstein (Los Alamos National Laboratory), "Modeling Immunoreceptor Signaling: From the Generic to the Detailed"
3:00 p.m.	Garrett Odell (University of Washington), "Modeling the Cell-signaling Network That Controls Stripes in Arthropods"
3:30 p.m.	George Oster (University of California at Berkeley), "The Mysterious Meanderings of Myxobacteria"
4:00 p.m.	Open discussion—Iain Johnstone (Stanford University), moderator; Leon Glass (McGill University), discussant
5:30 p.m.	Reception

Saturday, April 28

8:30 a.m.	Continental breakfast

Neuroscience

9:00 p.m.	John Rinzel (New York University), "Modeling the Thalamus in Sleep and Awake States"
9:30 a.m.	Don Johnson (Rice University), "Information Processing: Data Analysis and Theory"
10:00 a.m.	Larry Abbott (Brandeis University), "The Effects of Noise on Neural Response Dynamics and Gain"
10:30 a.m.	Emery Brown (Harvard Medical School), "Dynamics of Spatial Information Encoding in the Rat Hippocampus"

11:00 a.m. Open discussion—Peter Bickel (University of California at Berkeley), moderator; Keith Worsley (McGill University), discussant
12:30 p.m. Adjourn

ATTENDEES

Larry Abbott, Brandeis University
Amir Assadi, University of Wisconsin
Douglas Bauer, National Academy of Sciences
Peter J. Bickel, University of California at Berkeley
Mark Borodovsky, Georgia Institute of Technology
Emery Brown, Harvard Medical School
Jason Cantarella, University of Georgia
James Cassatt, National Institute of General Medical Sciences
Marvin Cassman, National Institute of General Medical Sciences
Andrew Conn, International Business Machines
Dawn Courtney, National Academy of Sciences
Robert Cox, National Institute of Mental Health
Keith Crank, National Science Foundation
Greg Dewey, Keck Graduate Institute
Cathryn Dippo, Bureau of Labor Statistics
William DuMouchel, AT&T Research
Joan Esnayra, National Academy of Sciences
Nina Fedoroff, Pennsylvania State University
Terry Gaasterland, Rockefeller University
David Galas, Keck Graduate Institute
Bijoy Ghosh, Washington University
Leon Glass, McGill University
Byron Goldstein, Los Alamos National Laboratory
Michael Green, Pennsylvania State University
Denny Gulick, University of Maryland
Jeff Hasty, Boston University
Ira Herskowitz, University of California at San Francisco
Colin Hill, Cornell University
David Hoel, Medical College of South Carolina
Fern Hunt, National Institute of Standards and Technology
Monica Hurdal, Florida State University
Don Johnson, Rice University
Gary Johnson, U.S. Department of Energy
Iain Johnstone, Stanford University
Jim Keener, University of Utah
Sallie Keller-McNulty, Los Alamos National Laboratory
Judy Kennedy, University of Delaware
Malgorzata Klosek, University of Wisconsin
Douglas Lauffenburger, Massachusetts Institute of Technology
Charles Lawrence, Rensselaer Polytechnic Institute

Bruce Levin, Emory University
William Levine, University of Maryland
Christine Lister, Cornell University
Jun Liu, Harvard University
John Luecke, University of Texas
Craig Malbon, SUNY, Stony Brook
Eduardo Marbán, Johns Hopkins University
Peter McCullagh, University of Chicago
Michael Mihalik, Vanderbilt University
Kathleen Morrish, National Imagery and Mapping Agency
Janet Novotny, U.S. Department of Agriculture
Garrett Odell, University of Washington
George Oster, University of California at Berkeley
Bernard Palsson, University of California at San Diego
Alan Perelson, Los Alamos National Laboratory
Charles Peskin, New York University
Don Pfatt, Rockefeller University
Michael Phelps, University of California at Los Angeles
Walt Polansky, U.S. Department of Energy
Alex Pothen, Old Dominion University
John Rinzel, New York University
Dan Roden, Vanderbilt University
Don Schwendem, Rensselaer Polytechnic Institute
David Scott, Rice University
Fran Sharples, National Academy of Sciences
Eric Siggia, Rockefeller University
Karen Skinner, National Institute on Drug Abuse
Temple Smith, Boston University
Terence Speed, University of California at Berkeley
Ken Stephenson, University of Tennessee
De Witt Sumners, Florida State University
Nancy Sung, Burroughs Wellcome Fund
Michael Tabor, University of Arizona
Michael Teitelbaum, Sloan Foundation
David Terman, Ohio State University
David Thomassen, U.S. Department of Energy
Philippe Tondeur, National Science Foundation
John Tyson, Virginia Polytechnic Institute and State University
Nora Volkow, Brookhaven National Laboratory
Scott Weidman, National Academy of Sciences
Jim Weiss, University of California at Los Angeles
Keith Worsley, McGill University
Barbara Wright, National Academy of Sciences
Rongling Wu, University of Florida
Sam Wu, University of Florida
Cheng Zhu, Georgia Institute of Technology

PLATE 1 Section of the model heart through the left ventricle in diastole. Structures at the top (from left to right in the figure) are pulmonary artery, aorta, and left atrium (with two pulmonary veins appearing in the section). At the bottom (from left to right in the figure), are the right and left ventricles. The anterior and posterior leaflets of the mitral valve appear in cross section at the top of the left ventricle. Note the prominent vortex (rotating clockwise in the figure) that has been shed from the anterior leaflet of the mitral valve, and the less prominent counter-rotating vortex that has been shed from the posterior leaflet. Together these are presumably the cross section of a ring vortex that has been shed like a smoke-ring from the mitral valve as a whole. Reprinted with permission from Kovacs, S.J., D.M. McQueen, and C.S. Peskin. Modeling cardiac fluid dynamics and diastolic function. Phil. Trans. R. Soc. Lond. A 359:1299-1314, 2001.

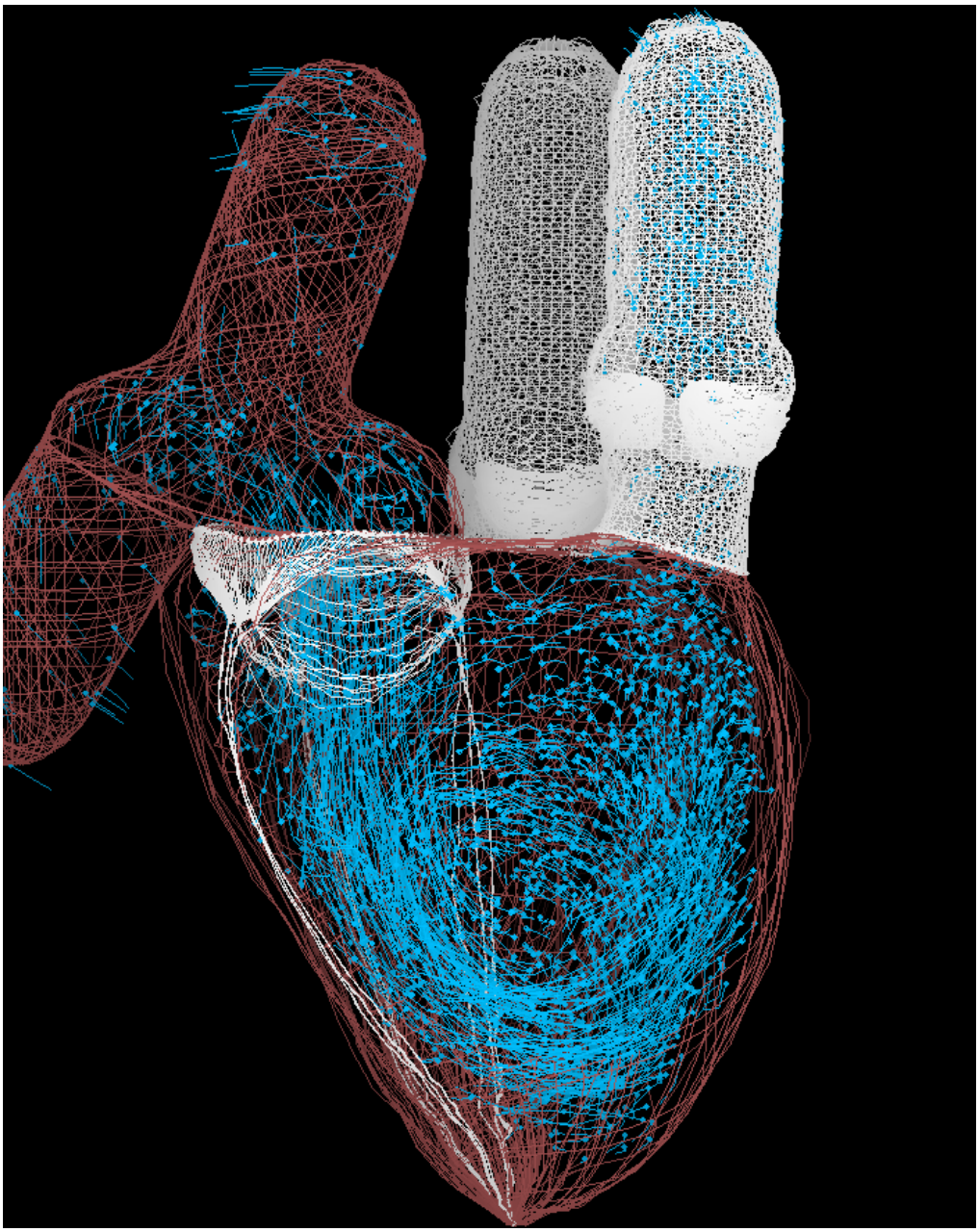

PLATE 2 Cutaway view of the model heart showing the flow pattern of blood in the right ventricle during diastole. Structures seen from left to right in the upper part of the figure are the superior vena cava connecting to the right atrium with the interior vena cava below, the aorta, and the pulmonary artery. In the lower part of the figure, a prominent vortex (rotating counterclockwise in the figure) fills the right ventricle. A jet of blood flowing through the open tricuspid valve merges with and presumably drives this vortex. Reprinted with permission from Kovacs, S.J., D.M. McQueen, and C.S. Peskin. Modeling cardiac fluid dynamics and diastolic function. Phil. Trans. R. Soc. Lond. A 359:1299-1314, 2001.

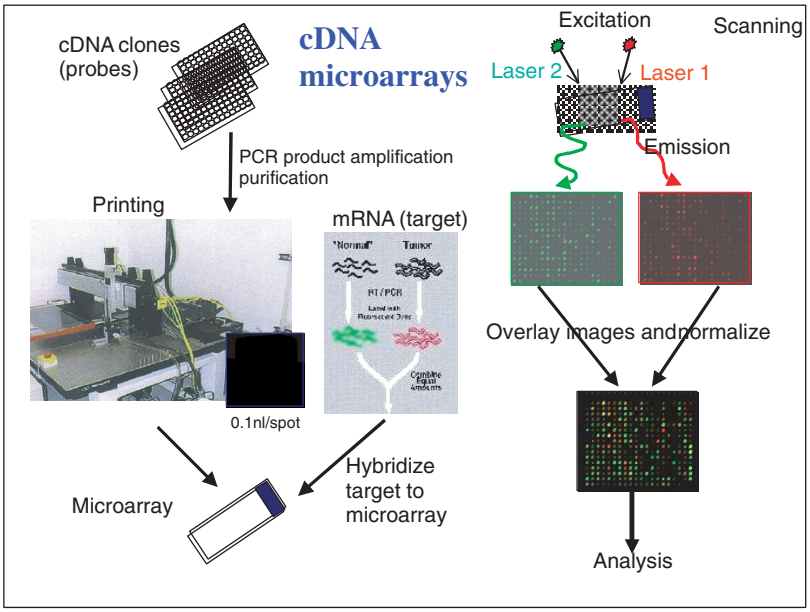

PLATE 3 Microarray construction. Figure courtesy of Terry Speed.

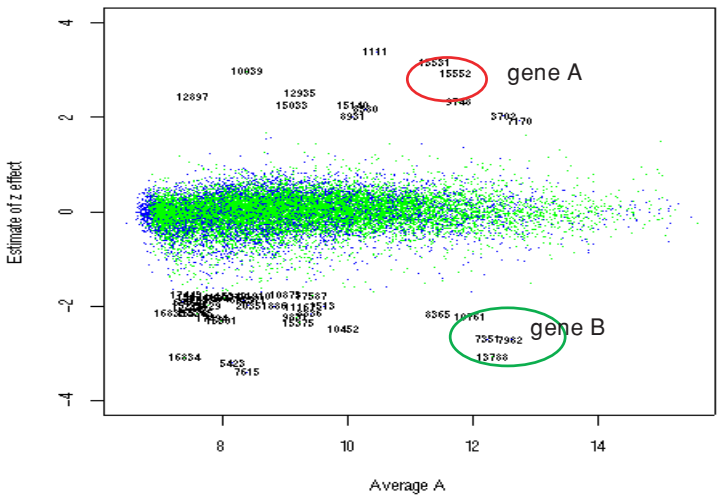

PLATE 4 Estimates of zone effects: log (zone 4/zone 1) versus average A. Figure courtesy of Terry Speed.